

НЕЙРОІНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ

NEUROINFORMATICS AND INTELLIGENT SYSTEMS

НЕЙРОИНФОРМАТИКА И ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

УДК 004.896

МЕТОД СТРУКТУРНОГО ДОНАЛАШТУВАННЯ НЕЙРОМЕРЕЖЕВИХ МОДЕЛЕЙ ДЛЯ ЗАБЕЗПЕЧЕННЯ ІНТЕРПРЕТАБЕЛЬНОСТІ

Леощенко С. Д. – аспірант кафедри програмних засобів Національного університету «Запорізька політехніка», Запоріжжя, Україна.

Олійник А. О. – д-р техн. наук, доцент кафедри програмних засобів Національного університету «Запорізька політехніка», Запоріжжя, Україна.

Субботін С. О. – д-р техн. наук, професор, завідувач кафедри програмних засобів Національного університету «Запорізька політехніка», Запоріжжя, Україна.

Гофман Є. О. – канд. техн. наук, старший науковий співробітник науково-дослідної частини Національного університету «Запорізька політехніка», Запоріжжя, Україна.

Корнієнко О. В. – аспірант кафедри програмних засобів Національного університету «Запорізька політехніка», Запоріжжя, Україна.

АНОТАЦІЯ

Актуальність. Розглянуто задачу структурного доналаштування попередньо синтезованих моделей на базі штучних нейронних мереж для забезпечення властивості інтерпретабельності при роботі із великими об'ємами даних. Об'єктом дослідження є процес структурного доналаштування штучних нейронних мереж з використанням адаптивних механізмів.

Мета роботи полягає у розробці методу структурного доналаштування нейронних мереж для підвищення швидкості їх роботи та зменшенні ресурсоемності при обробці великих даних.

Метод. Запропоновано метод структурного доналаштування нейронних мереж на основі адаптивних механізмів запозичених із нейроеволюційних методів синтезу. На початку метод використовує систему індикаторів для оцінки існуючої структури штучної нейронної мережі. Оцінка ґрунтується на структурних особливостях нейромоделі. Потім отримані індикаторні оцінки порівнюються із критеріальними значеннями для вибору типу структурного доналаштування. У якості варіантів структурного доналаштування використовується варіанти мутаційних змін із групи методів нейроеволюційної модифікації топології і ваг нейромережі. Метод дозволяє знизити ресурсоемність під час роботи нейромоделі, за рахунок пришвидшення обробки великих даних, що розширює поле практичного застосування штучних нейронних мереж.

Результати. Розроблений метод реалізовано та досліджено на прикладі використання рекурентної штучної мережі типу Довга короткочасна пам'ять при вирішенні задачі класифікації. Використання розробленого методу дозволило збільшити швидкість роботи нейромоделі з тестовою вибіркою на 25,05%, в залежності від використовуваних обчислювальних ресурсів.

Висновки. Проведені експерименти підтвердили працездатність запропонованого математичного забезпечення і дозволяють рекомендувати його для використання на практиці при структурному доналаштуванні попередньо синтезованих нейромоделей для подальшого вирішення задач діагностування, прогнозування, оцінювання та розпізнавання образів з використанням великих даних. Перспективи подальших досліджень можуть полягати в більш тонкому налаштуванні системи індикаторів для визначення зв'язків, що кодують зачумлені дані з метою додаткового підвищення точності роботи моделей на основі нейронних мереж.

КЛЮЧОВІ СЛОВА: інтерпретабельність, топологія, доналаштування структури, нейроеволюція, нейронні мережі.

АБРЕВІАТУРИ

IoT – Internet of things;

LSTM – Long short-term memory;

OBS – Optimal Brain Surgery;

TWEANN – Topology and Weight Evolving Artificial

Neural Networks;

WANN – Weight Agnostic Neural Networks;

PHM – рекурентна нейронна мережа;

ШНМ – штучна нейронна мережа.

НОМЕНКЛАТУРА

ε – незалежна змінна;

E_{Sample} – функція помилки;
 c – міжнейронні зв'язки ШНМ;
 $crit_{condcompl}$ – критерій умовної складності;
 E – відносна помилка;
 E_{Sample} – помилка моделі на тестовій вибірці;
 $error_{class}$ – помилкові екземпляри;
 $Error_{NN}$ – точність роботи ШНМ;
 $f(w, \varepsilon)$ – математична модель;
 FB_c – зворотні зв'язки у ШНМ;
 FF_c – прямі зв'язки у ШНМ;
 $ind_{activ}^{(\mu, i)}$ – індикатор логічної прозорості i -го нейрона μ -го шару мережі;
 $ind_{deg rconn}$ – індикатор середньої зв'язаності ШНМ;
 $ind_{distrat}$ – індикатор розрідженості зв'язків мережі;
 $ind_{logtransp}$ – індикатор логічної прозорості ШНМ;
 ind_{memor} – індикатор використання пам'яті;
 k – кількість міжнейронних зв'язків у ШНМ;
 L – кількість шарів у ШНМ;
 l – кількість нейронів на вході мережі;
 N_i – множина нейронів на вході мережі;
 N_o – множина нейронів на виході мережі;
 N_h – множина нейронів прихованого шару мережі;
 N_s – загальна кількість нейронів в ШНМ;
 NN – штучна нейронна мережа;
 NN_{param} – мета-параметри (інформація) про мережу;
 NN_{struct} – структура мережі;
 $NumN_{\mu}$ – число нейронів у μ -му шарі ШНМ;
 $NumW^{(\mu, i)}$ – кількість ваг i -го нейрону у μ -му шарі ШНМ;
 $NumW_{w=0}^{(\mu, i)}$ – кількість ваг i -го нейрону у μ -му шарі ШНМ, що дорівнюють 0;
 $Number_{sampler}$ – кількість екземплярів у вибірці;
 p – кількість нейронів на виході ШНМ;
 r – кількість нейронів у прихованому шарі ШНМ;
 $param$ – параметри ШНМ;
 $samp_{Inst}$ – кількість екземплярів у навчальній вибірці;
 $samp_{Feat}$ – кількість ознак у навчальній вибірці
 $struct$ – структура ШНМ;
 $Sample$ – вхідна вибірка даних;
 Val_1 – обмеження на складність ШНМ;
 Val_2 – обмеження на розмірність ШНМ;
 w – сукупність параметрів регресійної моделі;

W_{FBc} – кількість зворотних зв'язків для рекурентних нейромереж ($W_{FBc} = \{w_1, w_2, \dots, w_j\}$);

W_{FFc} – кількість прямих зв'язків нейромережі ($W_{FFc} = \{w_1, w_2, \dots, w_i\}$);

$W_{w=0}$ – ваги зв'язків, що дорівнюють нулю.

ВСТУП

Впровадження технологій, на кшталт: IoT, віддалених сенсорних систем та мереж допомагає промисловостям підвищити ефективність [1–5], проте можливості аналізувати величезні обсяги наявних даних саме тоді, коли це необхідно – в режимі онлайн, залишаються значно обмеженими. Зазвичай це обумовлено складнощами з обробкою великих даних в реальному часі [6]. Предобробка даних може іноді займати занадто багато часу і так само бути досить ресурсоємною. Тому варто звернути увагу на саму модель, яка використовується для роботи.

Завдання пошуку оптимальної структури нейронної мережі тісно пов'язана з проблемами недонавчання та перенавчання [6–8]. Занадто прості мережі не здатні адекватно моделювати цільові залежності в реальних завданнях. Занадто складні мережі мають надлишкове число вільних параметрів, які в процесі навчання налаштовуються не тільки на відновлення цільової залежності, але і на відтворення шуму [9, 10].

В роботі автори пропонують метод структурної оптимізації, що ґрунтується на адаптивних механізмах нейроevolюційних підходів для пошуку найбільш оптимальної структури ШНМ, що використовується у якості моделі при обробці даних.

Об'єкт дослідження – процес структурного доналаштування ШНМ.

Існуючі методи структурного доналаштування синтезованих нейромоделей характеризуються низкою недоліків [6–8]. Для деяких методів ці недоліки обумовлені необхідністю обрахунку додаткових метапараметрів, що значно підвищує вимоги до обчислювальних ресурсів. Для інших методів недоліки пов'язані із вузькою областю специфічних задач відповідних нейромоделей.

Предмет дослідження – метод структурного доналаштування ШНМ.

На сьогоднішній день існує група методів нейроevolюційного підходу, що одночасно виконують модифікацію топології та параметрів нейромоделі – TWEANN [11, 12]. Проте, загальний процес синтезу нової ШНМ може бути занадто ресурсоємним та недоречним. Тому, у роботі запропоновано підхід, що базується на використанні мутаційних змін та специфічних індикаторів і критеріїв для структурного доналаштування вже попередньо синтезованих ШНМ.

Мета роботи – розробка методу структурного доналаштування ШНМ для підвищення швидкості їх роботи та зменшення ресурсоємності при обробці великих даних.

1 ПОСТАНОВКА ПРОБЛЕМИ

Нехай маємо попередньо синтезовану нейромодель NN із визначеною структурою та значеннями параметрів $NN = (struct, param)$. Структура $struct = \{N, c\}$, $N = \{N_i, N_h, N_o\}$, $c = \{c\}$ – визначена сукупністю нейронів $N = \{N_i, N_o, N_h\}$, що складається із підмножин вхідних $N_i = \{N_{i_1}, N_{i_2}, \dots, N_{i_l}\}$, $l = 1, 2, \dots, |N_i|$, вихідних $N_o = \{N_{o_1}, N_{o_2}, \dots, N_{o_p}\}$, $p = 1, 2, \dots, |N_o|$ та прихованих нейронів $N_h = \{N_{h_1}, N_{h_2}, \dots, N_{h_r}\}$, $r = 1, 2, \dots, |N_h|$ та множини ваг $w = \{w_k\}$ зв'язків $c = \{c_1, c_2, \dots, c_k\}$, $k = 1, 2, \dots, |c|$ між нейронами.

Тоді задача структурного доналаштування ШНМ полягає в тому, щоб внести структурні зміни до існуючої структури мережі $struct' = \{N', c'\}$, $N' = \{N'_i, N'_h, N'_o\}$, $c' = \{c'\}$ таким чином, щоб збільшити рівень її інтерпретабельності $crit_{condcompl} \rightarrow 1$, за умови, що точність роботи не зменшується $Error_{NN'} \leq Error_{NN}$.

2 ОГЛЯД ЛІТЕРАТУРИ

Останні дослідження Бостонських науковців довели, що впродовж сну (особливо, глибокого сну) у людському мозку відбуваються процеси очищення, або «brainwashed» [13–15]. Загалом можна відмітити, що дослідження довели, що під час сну мозок проводить процеси впорядкування набутої інформації за рахунок скорочення, видалення або додавання нових зв'язків між нейронами. Роботи засвідчили, що під час сну нейрони зменшуються в розмірах і між ними виникають простору, які заповнюються мозковою рідиною. Результати спостережень вчених засновані на відкритті так званої глімфатичної системи, яка діє в мозку спеціально для видалення шкідливих речовин [14, 15]. Клітини мозку-можливо, гліальні клітини, які оточують і підтримують нейрони, зіщулюються під час сну. Це призводить до збільшення міжклітинного простору в речовині мозку, що в свою чергу підсилює приплив рідини, яка виносить з мозку токсини [14, 15]. Тож, припущення про те, що схожі методи оптимізації структури штучних ШНМ зможуть вирішити ряд існуючих проблем не є безгрунтовним (рис. 1).

OBS – метод спрощення структури регресійної моделі, наприклад, ШНМ [16, 17]. Основна ідея OBS полягає в тому, що ті елементи моделі або ШНМ, які мають малий вплив на помилку апроксимації, можна виключити з моделі без значного погіршення якості апроксимації [16, 17].

Детально представити метод OBS можна наступним чином. Нехай задана вибірка $Sample$, модель $f(w, \varepsilon)$. Для спрощення структури регресійної моделі виконуються наступні кроки [16, 17]:

1) налаштуємо модель, отримуємо параметри $w^* = \arg \min(E_{Sample}(w | f, Sample))$;

2) для прирощення $w^* + \Delta w$ вирішуємо оптимізаційну задачу, знаходимо для кожного індексу i мінімальне значення Лагранжиана L_i ;

3) обираємо серед L_i мінімальне, відсікаємо елемент моделі, відповідний i -му параметру;

4) додаємо до вектора параметрів w^* , вектор прирощень Δw , відповідний відсіченому параметру;

5) отримуємо спрощену модель. Модель переналаштовувати не потрібно;

6) процедуру можна повторювати до тих пір, поки значення помилки не перевершить заздалегідь задане.

Регуляризація в машинному навчанні – метод додавання деяких додаткових обмежень до умови з метою вирішити некоректно поставлене завдання або запобігти перенавчання [18, 19]. Ця інформація часто має вигляд штрафу за складність моделі. Наприклад, це можуть бути обмеження гладкості результуючої функції або обмеження за нормою векторного простору [18, 19]. З байєсівської точки зору багато методів регуляризації відповідають додаванню деяких апріорних розподілів на параметри моделі. Найчастіше виділяють [18, 19]:

1) L_1 -регуляризацію, або регуляризація лассо-регресією;

2) L_2 -регуляризацію, або регуляризація регресією Ріджа.

В основі L_1 -регуляризації лежить досить проста ідея. Як і у випадку L_2 -регуляризації, ми просто додаємо штраф до початкової функції витрат. Подібно до того, як в L_2 -регуляризації ми використовуємо L_2 -нормування для поправки вагових коефіцієнтів, в L_1 -регуляризації ми використовуємо спеціальне L_1 -нормування. L_2 -регуляризація допомагає вирішити проблему перенавчання моделі [18, 19]. Суть в тому, що надто важкі вагові коефіцієнти відштовхують лінію найкращої відповідності, побудовану на основі мінімізації квадрата помилок, від основної тенденції. L_1 -регуляризація сприяє розрідженості функції, коли лише деякі фактори не дорівнюють нулю. L_2 -регуляризація сприяє появі малих вагових коефіцієнтів моделі, але не сприяє їх точній рівності нулю [18, 19].

WANN – це архітектура ШНМ, представлена розробниками Google, здатна імітувати вроджені інстинкти і рефлексії живих істот, з подальшим донавчанням протягом життя [20, 21]. А також може значно зменшує кількість зв'язків всередині мережі, підвищуючи тим самим їх швидкодію.

Загальна схема роботи методу дуже схожа на нейроревольюційні методи синтезу ШНМ [20, 21]:

1) створюється популяція простих ШНМ;

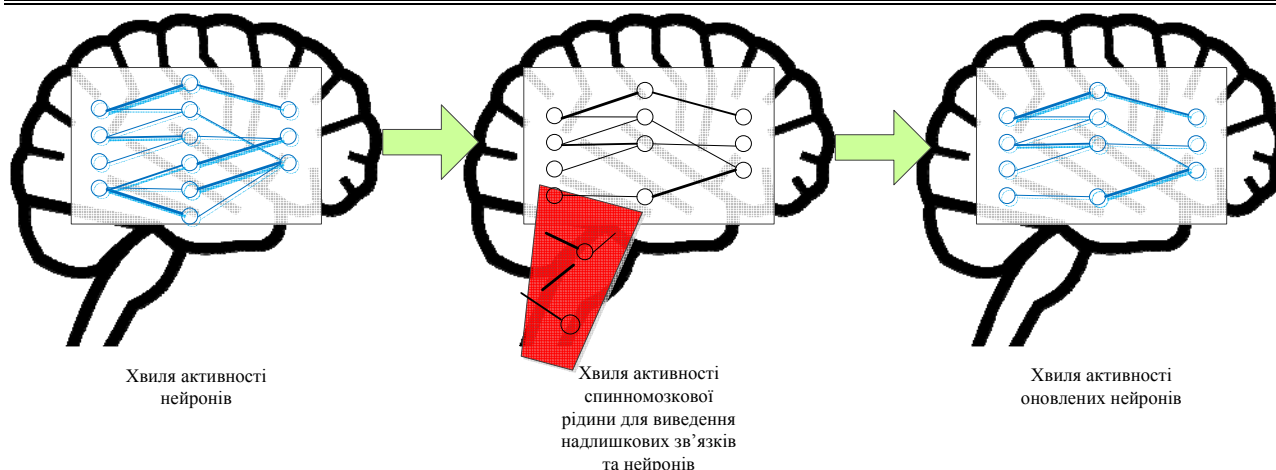


Рисунок 1 – Хвильоподібний процес впорядкування та очищення мозку [13]

2) кожна мережа ініціалізує всі свої ваги одним числом, причому для широкого діапазону чисел: $w = -2 \dots + 2$;

3) отримані мережі сортуються за якістю рішення задачі і за кількістю нейронів (у бік зменшення);

4) у частині кращих представників додається один нейрон, один зв'язок або змінюється функція активації в одному нейроні;

5) ці модифіковані мережі використовуються в якості початкових для нової ініціалізації – так метод повертається до першого кроку.

При порівнянні цих методів та підходів можна зробити наступні висновки: порівняння усі сучасні методи структурної оптимізації ШНМ мають низку недоліків. Так OBS та методи Регуляризації мають хороше теоретичне обґрунтування та використовуються вже достатньо давно, проте слід зазначити, що ці методи вимагають додаткових обрахунків мета-параметрів про ШНМ. Більш того, варто зазначити, що ці обрахунки значно підвищують ресурсоемісність: значно підвищуються вимоги до пам'яті та обчислювальних потужностей – у майбутньому це може призвести до неможливості або слабкої продуктивності паралельних модифікацій цих методів. Також важливо пам'ятати про складність застосування регуляризації для специфічних задач, адже для цих методів дійсно важлива – похідна функції. Похідна, є ключовою, оскільки методи базуються на градієнтному спуску, який в основному рухається в напрямку похідної. При квадратичному члені чим ближче величина знаходиться до нуля, тим меншою стає похідна, поки також не наблизиться до нуля. Тому в такій ситуації зміни при Регуляризації будуть неістотними. Найважливішою ж рисою є те, що OBS та методи Регуляризації майже неможливо застосовувати для структурної оптимізації ШНМ із складними структурами: рекурентні ШНМ зі зворотними зв'язками або глибокі ШНМ із декілька ми прихованими шарами нейронів. З іншого боку, WANN виграють через те, що можуть поступово змінювати топологію ШНМ та налаштовувати її відповідно до задачі. Також слід відзначити значну

перевагу WANN за рахунок того, що немає необхідності зберігати та обраховувати мета-параметри мережі. Проте через недостатнє теоретичне обґрунтування ця технологія поки що нерозповсюджена.

3 МАТЕРІАЛИ ТА МЕТОДИ

Нейроеволюційні методи передбачають використання специфічних генетичних операторів [11, 22, 23]. В ряді робіт пропонувалося використовувати адаптивні механізми мутації. Загалом визначимо основні види мутацій, що можуть використовуватися при синтезі ШНМ [22, 23]:

1) додавання прихованого нейрона. Новий нейрон додається разом з вхідним і вихідним зв'язками. При цьому вихідний зв'язок нейрона не може його зв'язувати із вхідним нейроном;

2) видалення випадково обраного прихованого нейрона разом з усіма вхідними і вихідними зв'язками. При цьому якщо утворюється пробіл в індексах нейронів, що залишилися, то проводиться корекція індексів. Вхідні та вихідні нейрони мережі не можуть бути видалені;

3) додавання зв'язку. Випадковим чином визначаються індекси початкового і кінцевого нейронів в ШНМ, представленої мутуючої особиною. При цьому зв'язок не може закінчуватися вхідним нейроном. Вага зв'язку визначається також випадково з діапазону $[-0,5; 0,5]$: $w_i = \text{Rand}[-0,5; 0,5]$. Якщо в мережі вже існує зв'язок з аналогічними вхідними і вихідними нейронами, то його вага замінюється на випадкову;

4) видалення випадково обраного зв'язку. При цьому може виникнути ситуація, коли видалається останній зв'язок у прихованого нейрона. У цьому випадку нейрон також видалається, і, якщо необхідно, проводиться корекція індексів нейронів мережі;

5) зміна функції активації випадкового прихованого нейрону.

Таким чином, за допомогою мутації можна точно змінювати параметри структури ШНМ.

Тож зазначимо, що таким чином схожі механізми можуть використовуватися для оптимізації структури попередньо синтезованих ШНМ. Саме тому пропону-

ється використовувати адаптивний механізм мутації проте з певними додатковими установками. Хаотичне додавання (видалення) нейронів і зв'язків може призвести до ситуацій, коли, наприклад, в мережі багато нейронів і мало зв'язків. Більш логічним буде застосовувати різні види мутацій в залежності від особливостей архітектури мережі, представленої мутуючої особиною. Для цього буде введено критерії та індикатори, для оцінювання та регуляції розміру та напрямку розвитку мережі.

Використання першого критерію засновано на припущенні, що чим більше елементів в сумі у вхідному та вихідному векторі навчальної вибірки (чим більшою є сумарна кількість вхідних і вихідних нейронів), то, ймовірно, більш складна мережа необхідна для вирішення поставленого завдання. Обчислюється другий коефіцієнт за такою формулою [24–28]:

$$crit_{condcompl} = \frac{N_i + N_o}{N_s}. \quad (1)$$

Тобто, чим більше буде нейронів у мережі, тим меншим буде значення критерію $crit_{condcompl}$ і з тим меншою ймовірністю буде обрана мутація, що додає новий прихований нейрон [24–28].

Перший же індикатор буде використовуватися для оцінки надмірності використаної пам'яті. Як було зазначено вище, нейромережі володіють пам'яттю, що реалізується вагами зв'язків [24–28]. Чим менше пам'ять мережі, тим менше образів вона може запам'ятати. Проте, в ситуації коли дві мережі з різним об'ємом пам'яті забезпечують необхідну точність розпізнавання (оцінювання), то мережа з меншою пам'яттю, звісно має кращі узагальнюючі властивості. Надмірність пам'яті мережі будемо характеризувати за допомогою коефіцієнта надмірності для зберігання навчальної вибірки [24–28]:

$$ind_{memor} = \frac{W_{FFc} + W_{FBc}}{samp_{Inst} \cdot samp_{Feat}}. \quad (2)$$

Якщо $ind_{memor} > 1$, то пам'ять мережі надлишкова (розмірність пам'яті мережі більше розмірності навчальної вибірки). Якщо $ind_{memor} = 1$, то мережа може запам'ятати всю навчальну вибірку (розмірність пам'яті мережі дорівнює розмірності навчальної вибірки). Якщо $ind_{memor} < 1$, то мережа не зможе в точності запам'ятати всю навчальну вибірку (розмірність пам'яті мережі менше розмірності навчальної вибірки), однак при цьому мережа буде проявляти узагальнюючі та апроксимуючі здібності [24–28].

Логічна прозорість мережі в значній мірі залежить від загальної кількості зв'язків в мережі і кількості зв'язків, що з'єднують конкретні нейрони (кількості входів нейронів прихованих шарів) [24–28]. Чим менше зв'язків, тим простіше мережа і тим зручніше вона для аналізу та інтерпретації людиною. Тож індикатор

розрідженості зв'язків мережі прямого поширення визначимо за формулою [24–28]:

$$ind_{distr} = \frac{W_{w=0}}{\sum_{\mu=1}^L NumN_{(\mu-1)} NumN_{\mu}}, \quad (3)$$

де $0 \leq W_{w=0} \leq \sum_{\mu=1}^L NumN_{(\mu-1)} NumN_{\mu}$.

Індикатор середньої зв'язаності нейронної мережі показує середню кількість входів нейронів всіх шарів, крім першого [24–28].

Індикатор логічної прозорості i -го нейрона μ -го шару мережі $ind_{activ}^{(\mu,i)}$ буде визначатися типом використовуваної функції активації [24–28]. Для лінійної та порогової функції активації приймемо: $ind_{activ}^{(\mu,i)} = 1$, для всіх інших функцій активації $ind_{activ}^{(\mu,i)} = 0$.

$$ind_{deg rconn} = \frac{1}{\sum_{\mu=2}^L NumN_{\mu}} \times \sum_{\mu=2}^L \sum_{i=1}^{NumN_{\mu}} (NumW^{(\mu,i)} - NumW_{w=0}^{(\mu,i)}). \quad (4)$$

Виходячи з цього, коефіцієнт логічної прозорості ШНМ $ind_{logtransp}$ будемо розраховуватися так [24–28]:

$$ind_{logtransp} = \frac{\sum_{\mu=1}^L \sum_{i=1}^{NumN_{\mu}} ind_{activ}^{(\mu,i)}}{ind_{deg rconn} \sum_{\mu=1}^L NumN_{\mu}}, \quad (5)$$

$ind_{deg rconn} \neq 0$.

На рис. 2 представлено у схематичному вигляді процес вибору типу мутації. На рис. 2 використовуються такі позначення: Val_1 та Val_2 – певні граничні значення, що визначаються для завдання. Умовно процес вибору типу мутації можна розділити на дві гілки по першому умовному переходу.

4 ЕКСПЕРИМЕНТИ

Для експериментального дослідження результатів використання запропонованого методу оптимізації структури ШНМ проведемо ряд тестувань. На рис. 3 наведено приклад рекурентної ШНМ. Як видно з рисунку структура ШНМ є надмірною, зазвичай в таких мережах є нульові ваги у деяких синопсисів.

Для роботи будемо використовувати вибірку даних – Набір даних про ускладнення інфаркту міокарда [29–31]. Проблеми реальної складності необхідні для

тестування і порівняння різних методів інтелектуального аналізу даних і розпізнавання образів. Пропонується вибірка може бути використана для вирішення практично важливого завдання: прогнозування ускладнень інфаркту міокарда на основі інформації про пацієнта на момент надходження і на третій день госпітального періоду [29–31].

Інфаркт міокарда – одна з найскладніших проблем сучасної медицини [29–31]. Гострий інфаркт міокарда

пов'язаний з високою смертністю в перший рік після нього. Захворюваність їм залишається високою у всіх країнах. Особливо це стосується міського населення високорозвинених країн, яке піддається впливу хронічних стресових факторів, нерегулярного і не завжди збалансованого харчування [29–31]. У Сполучених Штатах, наприклад, більше мільйона людей щорічно страждають від інфаркту, і 200–300 тисяч з них помирають від гострого ІМ до прибуття в лікарню [29–31].

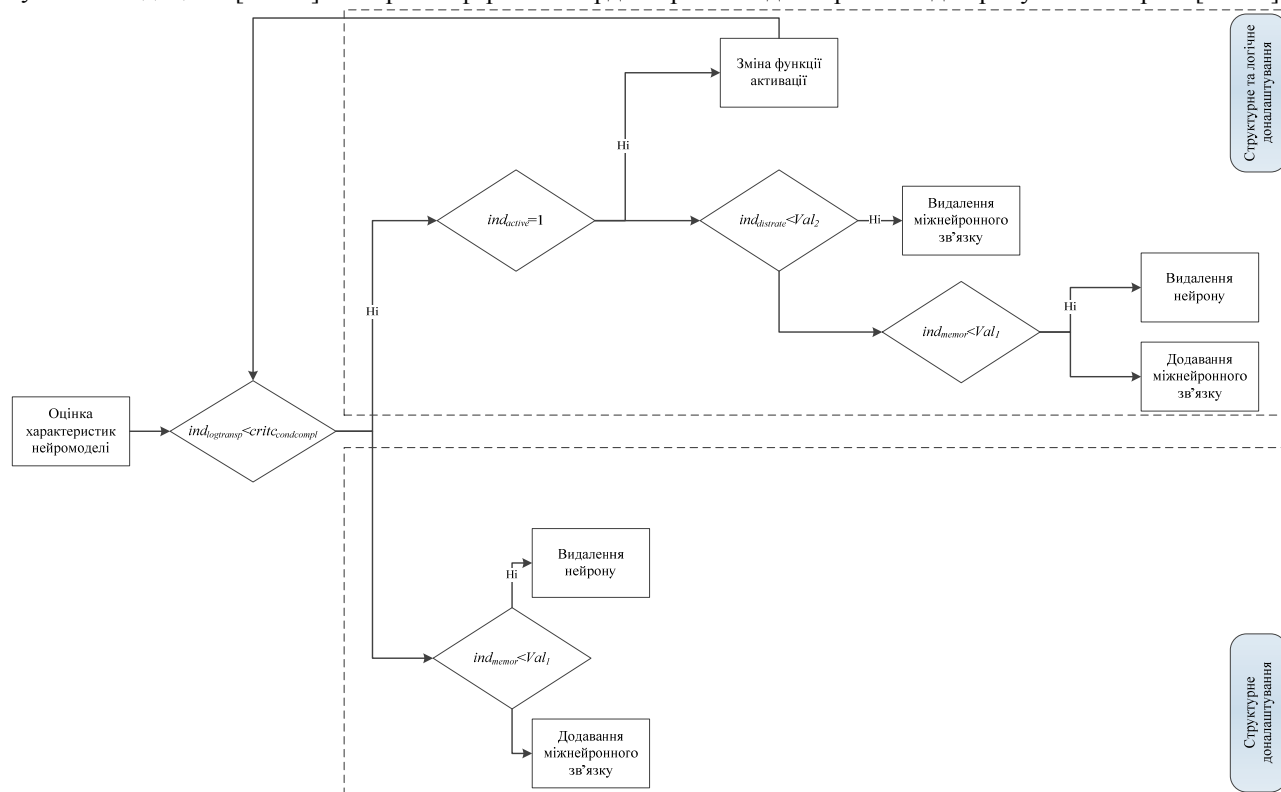


Рисунок 2 – Схематичне зображення процесу вибору типу мутації

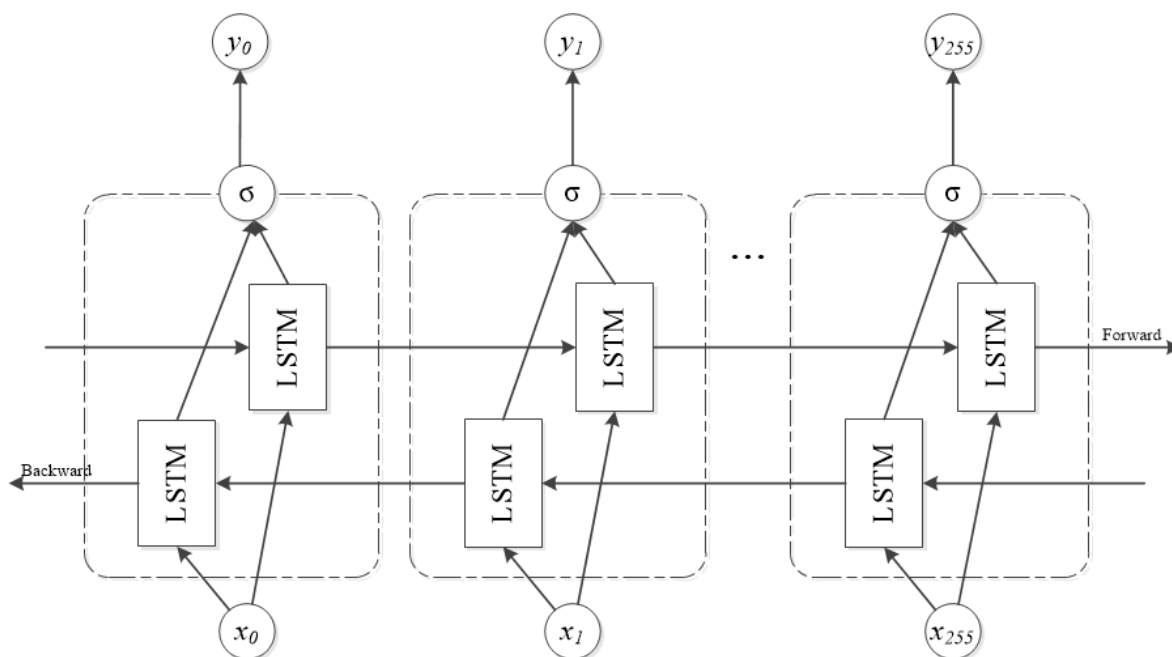


Рисунок 3 – ШНМ до оптимізації

Перебіг захворювання у пацієнтів з інфарктом відбувається по-різному. Може протікати без ускладнень або з ускладненнями, які не погіршують довгостроковий прогноз. У той же час близько половини хворих в гострому і підгострому періодах мають ускладнення, які призводять до загострення захворювання і навіть смерті. Навіть досвідчений фахівець не завжди може передбачити розвиток цих ускладнень. У зв'язку з цим прогнозування ускладнень інфаркту міокарда з метою своєчасного проведення необхідних профілактичних заходів є важливим завданням.

Характеристики набору даних у таблиці 1.

Таблиця 1 – Загальна характеристика вибірки даних Myocardial infarction complications Data Set

Характеристика вибірки	багато-вимірна	Кількість екземплярів	1700
Характеристика атрибутів	дійсні	Кількість атрибутів	124

Під час тестування будемо звертати увагу на час роботи мережі та розподіл часу за ітераціями. Додатково будемо оцінювати точність роботи [22, 23]. Для цього введемо відносну помилку, яка в цьому випадку буде обчислюватися, як відношення кількості невірно діагностованих об'єктів до загального розміру вибірки (кількості її екземплярів):

$$E = \frac{error_{class}}{Number_{sampl}} \cdot 100\% . \quad (6)$$

Точність роботи – не є цільовим параметром, проте важливо, щоб оптимізація не погіршала цього значення.

5 РЕЗУЛЬТАТИ

У таблиці 2 наведено порівняння результатів роботи для звичайної мережі та мережі після оптимізації.

Таблиця 2 – Порівняння результатів роботи

	Час роботи, с	Помилка, %
Звичайна мережа	7621	4,2
Оптимізована мережа	5712	3,4

На рис. 4 наведено графік, що демонструє різницю у розподілі часу між ітераціями під час роботи ШНМ.

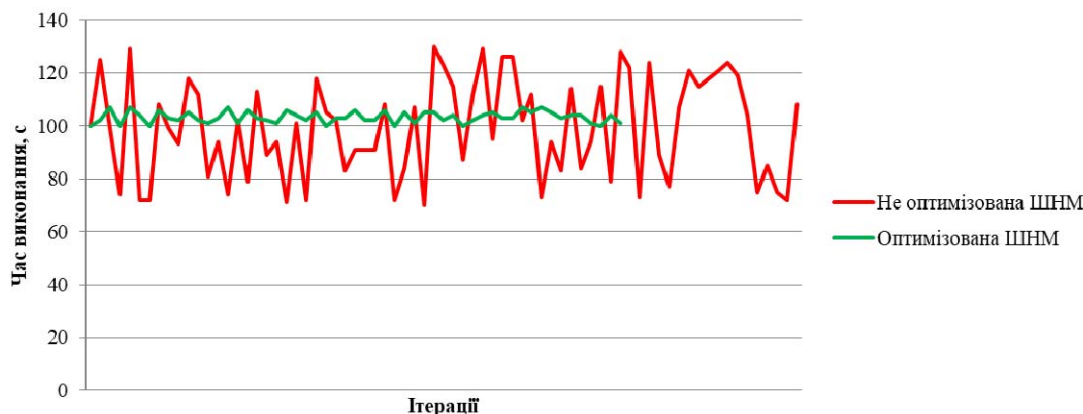


Рисунок 4 – Розподіл часу між ітераціями

6 ОБГОВОРЕННЯ

З таблиці 2 видно, що час роботи значно скоротився. Це пояснюється тим, що оптимізована мережа не задіює зайві синописи та нейрони, що значно пришвидшує роботу із нею. Також важливим є те, що точність роботи оптимізованої мережі зросла. Такий ефект можна пояснити тим, що надмірна структура мережі має надлишкове число вільних параметрів, які в процесі навчання налаштовуються не тільки на відновлення цільової залежності, але і на відтворення шуму, що може впливати на точність роботи.

Під час аналізу графіка можна зробити висновок, що як раз надмірна топологічна складність мережі до оптимізації призводить до різких стрибків у часі між ітераціями. У той же час графік розподілу для оптимізованої мережі більш плавний. І хоча певні стрибки все ж є, їх вплив значно менше та частково його можна списати на особливості складної рекурентної архітектури мережі.

Також варто зазначити, що спрощення структури ШНМ впливає і на ресурсоемістність, а саме на використання пам'яті. Зазвичай, найбільше пам'яті вимагає процес серіалізації ваг синописів. Оскільки оптимізована структура має меншу кількість нейронів та зв'язків між ними – це скорочує витрати пам'яті на збереження мережі.

ВИСНОВКИ

Вирішено актуальну науково-прикладну проблему структурного доналаштування синтезованих ШНМ.

Наукова новизна полягає в тому що розроблено метод структурного доналаштування синтезованих ШНМ на основі нейроволюційного підходу з використанням системи індикаторів та критеріїв для адаптивного визначення мутаційних змін. У запропонованому методі використовуються система індикаторів для оцінки існуючої структури ШНМ. Враховуючи отриману оцінку за структурними особливостями та спираючись на критеріальні значення, обирається тип структурного доналаштування. У якості варіантів структурного доналаштування використовується варіанти мутаційних змін із групи методів нейроволюційної модифікації топології і ваг нейромережі.

Запропонований метод дозволяє знизити ресурсоємність під час роботи нейромоделі, за рахунок пришвидшення обробки великих даних.

Практична цінність полягає в тому, що розв'язано практичні завдання структурного доналаштування та спрощення попередньо синтезованих ШНМ, які можуть використовуватися для діагностування, прогнозування, оцінювання та розпізнавання образів. Більш того, враховуючи полегшену структуру такі нейромоделі можна використовувати при роботі з великими об'ємами вхідних даних. Результати експериментів показали, що запропонований метод дозволяє більш прорідити ШНМ для її подальшого використання у якості моделі для діагностування, прогнозування, оцінювання та моделювання.

Перспективи подальших досліджень полягають у розробці, дослідженні та імплементації критеріальної системи, що дозволить більш тонко оцінювати структурні елементи, що кодуєть зашумлені дані та їх зв'язки. Це дозволить додатково зменшити обчислювальні вимоги та збільшити точність роботи ШНМ, що в свою чергу значно розширить коло практичного використання нейромоделей.

ПОДЯКИ

Робота була проведена за підтримки держбюджетних науково-дослідницьких проектів державного бюджету Національного університету «Запорізька політехніка» «Інтелектуальні методи та програмне забезпечення для діагностики та неруйнівного контролю якості військової та цивільної техніки» (номер державної реєстрації 0119U100360) та «Розроблення методів та засобів для аналізу та прогнозування динамічної поведінки нелінійних об'єктів» номер державної реєстрації 0121U107499).

ЛІТЕРАТУРА / LITERATURA

1. Veneri G. Hands-On Industrial Internet of Things: Create a powerful Industrial IoT infrastructure using Industry 4.0 / G. Veneri, A. Capasso. – Birmingham : Packt Publishing, 2018. – 556 p.
2. Roshak M. Artificial Intelligence for IoT Cookbook: Over 70 recipes for building AI solutions for smart homes, industrial IoT, and smart cities / M. Roshak. – Birmingham : Packt Publishing, 2021. – 260 p.
3. Practical IoT Hacking: The Definitive Guide to Attacking the Internet of Things / [F. Chantzis, I. Stais, P. Calderon et al.]. – San Francisco : No Starch Press, 2021. – 464 p.
4. Nam C. S. Neuroergonomics: Principles and Practice (Cognitive Science and Technology) / C. S. Nam. – Berlin : Springer, 2020. – 492 p.
5. Intelligent Systems and Learning Data Analytics in Online Education (Intelligent Data-Centric Systems: Sensor Collected Intelligence) / [S. Caballé, S. N. Demetriadis, E. G.-Sánchez, P.M. Papadopoulos, A. Weinberger]. – Cambridge : Academic Press, 2021. – 424 p.
6. Aggarwal C. C. Neural Networks and Deep Learning: A Textbook / C. C. Aggarwal. – Berlin : Springer, 2018. – 520 p.

7. Shikhman V. Mathematical Foundations of Big Data Analytics / V. Shikhman, D. Müller. – Berlin : Springer, 2021. – 288 p.
8. Deitel P. Intro to Python for Computer Science and Data Science: Learning to Program with AI, Big Data and The Cloud / P. Deitel, H. Deitel. – London : Pearson, 2019. – 880 p.
9. D'souza R. N. Structural Analysis and Optimization of Convolutional Neural Networks with a Small Sample Size / R. N. D'souza, P.Y. Huang, F.C. Yeh // Scientific Reports. – 2020. – № 10. – P. 1–13. DOI: 10.1038/s41598-020-57866-2.
10. Nowakowski G. Neural Network Structure Optimization Algorithm / G. Nowakowski, Y. Dorogyy, O. Dorogayevaniuk // Journal of Automation, Mobile Robotics and Intelligent Systems. – 2018 – № 12. – P. 5–13. DOI: 10.14313/JAMRIS_1-2018/1.
11. Iba H. Evolutionary Approach to Machine Learning and Deep Neural Networks: Neuro-Evolution and Gene Regulatory Networks / H. Iba. – New York : Springer, 2018. – 258 p.
12. Committee of SGTm Neural-Like Structures with RBF kernel for Insurance Cost Prediction Task / [I. Izonin, R. Tkachenko, N. Kryvinska et al.] // 2nd Ukraine Conference on Electrical and Computer Engineering (UKRCON), 2–6 July 2019 : proceedings. – Lviv : IEEE, 2019. – P. 1037–1040. DOI: 10.1109/UKRCON.2019.8879905.
13. Coupled electrophysiological, hemodynamic, and cerebrospinal fluid oscillations in human sleep / [N. E. Fultz, G. Bonmassar, K. Setsompop et al.] // Science. – 2019. – Vol. 366, Issue 6465. – P. 628–631. DOI: 628–631. 10.1126/science.aax5440.
14. Are We “Brain Washed” during Sleep? [Electronic resource]. – Access mode: <https://www.bu.edu/articles/2019/cerebrospinal-fluid-washing-in-brain-during-sleep/>
15. Deep Sleep Gives Your Brain a Deep Clean [Electronic resource]. – Access mode: <https://www.scientificamerican.com/article/deep-sleep-gives-your-brain-a-deep-clean/>
16. Optimal Brain Surgeon on Artificial Neural Networks in Nonlinear Structural Dynamics / [N. H. Christiansen, J. Hultmann Job, K. Klyver, J. Høgsberg] // 25th Nordic Seminar on Computational Mechanics (NSCM), 25–26 October 2012 : proceedings. – Lund : Lund University, 2012. – P.319–324.
17. Endisch C. Optimal Brain Surgeon for General Dynamic Neural Networks / C. Endisch, C. Hackl, D. Schröder // Progress in Artificial Intelligence. – 2007. – Vol. 4874. – P. 15–28. DOI: 10.1007/978-3-540-77002-2_2.
18. Oppermann, A. Regularization in Deep Learning – L1, L2, and Dropout [Electronic resource]. Access mode: <https://www.deeplearning-academy.com/p/ai-wiki-regularization>
19. Classic Regularization Techniques in Neural Networks [Electronic resource]. Access mode: <https://medium.com/@ODSC/classic-regularization-techniques-in-neural-networks-68bccee03764>
20. Weight Agnostic Neural Networks [Electronic resource]. Access mode: <https://arxiv.org/pdf/1906.04358.pdf>
21. Weight Agnostic Neural Networks [Electronic resource]. Access mode: <https://weightagnostic.github.io/>
22. Modification and parallelization of genetic algorithm for synthesis of artificial neural networks / [S. D. Leoshchenko, A. O. Oliynyk, S.A. Subbotin et al.] // Radio Electronics,

- Computer Science, Control. – 2019. – № 4. – P. 68–82. DOI: 10.15588/1607-3274-2018-3-12.
23. Implementation of the indicator system in modeling complex technical systems / [S. D. Leoshchenko, S. A. Subbotin, A. O. Oliinyk, O. E. Narivs'kiy] // Radio electronics, computer science, control. – 2021. – № 1. – P. 117–127. DOI: 10.15588/1607-3274-2021-1-12.
24. Sattari M. T. Performance evaluation of artificial neural network approaches in forecasting reservoir inflow / M. T. Sattari, K. Yurekli, M. Pal // Applied Mathematical Modelling. – 2012. – Vol. 36, Issue 6. – P. 2649–2657. DOI: 10.1016/j.apm.2011.09.048.
25. Hassan M. Evaluating the performance of a neural network-based multi-criteria recommender system / M. Hassan, M. Hamada // International Journal of Spatio-Temporal Data Science. – 2019. – Vol. 1 (54) – P. 54–66. DOI: 10.1504/IJSTDS.2019.10018848
26. Jamróz D. The examination of the effect of the criterion for neural network's learning on the effectiveness of the qualitative analysis of multidimensional data / D. Jamróz // Knowledge and Information Systems. – 2020. – Vol. 62. – P. 3263–3289. DOI: 10.1007/s10115-020-01441-8.
27. Субботін С. А. Критерии сравнения распознающих моделей на основе нейронных сетей и анализ их взаимосвязей / С.А. Субботин // Искусственный интеллект. – 2014. – Vol. 1. – P. 142–152.
28. Leoshchenko S. Adaptive Mechanisms for Parallelization of the Genetic Method of Neural Network Synthesis / S. Leoshchenko, A. Oliinyk, S. Subbotin // 10th International Conference on Advanced Computer Information Technologies (ACIT 2020), Deggendorf, 16–18 November : proceedings. – Ternopil: IEEE, 2020. – P. 446–450. DOI: 10.1109/ACIT49673.2020.9208905.
29. Myocardial infarction complications Data Set [Electronic resource]. – Access mode: <https://archive.ics.uci.edu/ml/datasets/Myocardial+infarction+complications>
30. Medical, psychological and physiological applications of MultiNeuron neural simulator / [A. N. Gorban, D. A. Rossiev, E. V. Butakova et al.] // The Second International Symposium on Neuroinformatics and Neurocomputers, Rostov-on-Don, 20–23 September : proceedings. – Rostov-on-Don: IEEE, 1995. – P. 7–14. DOI: 10.1109/ISNINC.1995.480831.
31. Trajectories, bifurcations, and pseudo-time in large clinical datasets: applications to myocardial infarction and diabetes data / [S. E. Golovenkin, J. Bac, A. Chervov et al.] // GigaScience. – 2020. – Vol. 9(11). – P. 1–20. DOI: 10.1093/gigascience/giaa128.

Стаття надійшла до редакції 21.06.2021.
Після доробки 12.07.2021.

УДК 004.896

МЕТОД СТРУКТУРНОЙ ДОНАСТРОЙКИ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ ДЛЯ ОБЕСПЕЧЕНИЯ ИНТЕРПРЕТАБЕЛЬНОСТИ

Леошенко С. Д. – аспирант кафедры программных средств Национального университета «Запорожская политехника», Запорожье Украина.

Олейник А. А. – д-р техн. наук, доцент, доцент кафедры программных средств Национального университета «Запорожская политехника», Запорожье Украина.

Субботин С. А. – д-р техн. наук, профессор, заведующий кафедрой программных средств Национального университета «Запорожская политехника», Запорожье Украина.

Гофман Е. А. – старший научный сотрудник научно-исследовательской части Национального университета «Запорожская политехника», Запорожье Украина.

Корниенко А. В. – аспирант кафедры программных средств Национального университета «Запорожская политехника», Запорожье Украина.

АННОТАЦИЯ

Актуальность. Рассмотрена задача структурного изменений предварительно синтезированных моделей на базе искусственных нейронных сетей для обеспечения свойства интерпретабельности при работе с большими объемами данных. Объектом исследования является процесс структурного изменений искусственных нейронных сетей с использованием адаптивных механизмов.

Цель работы заключается в разработке метода структурного изменения нейронных сетей для повышения скорости их работы и уменьшении ресурсоемкости при обработке больших данных.

Метод. Предложен метод структурного донастройки нейронных сетей на основе адаптивных механизмов заимствованных из нейроэволюционных методов синтеза. В начале метод использует систему индикаторов для оценки существующей структуры искусственной нейронной сети. Оценка основывается на структурных особенностях нейромоделей. Затем полученные индикаторные оценки сравниваются с критериальными значениями для выбора типа структурного изменений. В качестве вариантов структурного изменения используются варианты мутационных изменений из группы методов нейроэволюционной модификации топологии и весов нейросети. Метод позволяет снизить ресурсоемкость во время работы нейромоделей, за счет ускорения обработки больших данных, что расширяет поле практического применения искусственных нейронных сетей.

Результаты. Разработанный метод реализован и исследован на примере использования рекуррентной искусственной сети типа долгая кратковременная память при решении задачи классификации. Использование разработанного метода позволило увеличить скорость работы нейромоделей с тестовой выборкой на 25,05 %, в зависимости от используемых вычислительных ресурсов.

Выводы. Проведенные эксперименты подтвердили работоспособность предложенного математического обеспечения и позволяют рекомендовать его для использования на практике при структурной донастройке предварительно синтезированных нейромоделей для дальнейшего решения задач диагностирования, прогнозирования, оценивания и распознавания образов с использованием больших данных. Перспективы дальнейших исследований могут заключаться в более тонкой настройке параметров. © Леошенко С. Д., Олейник А. А., Субботин С. А., Гофман Е. А., Корниенко А. В., 2021
DOI 10.15588/1607-3274-2021-3-8

стройке системы индикаторов для определения связей, кодирующих зашумленные данные с целью дополнительного повышения точности работы моделей на основе нейронных сетей.

КЛЮЧЕВЫЕ СЛОВА: интерпретабельность, топология, донастройка структуры, нейроэволюция, нейронные сети.

UDC 004.896

THE METHOD OF STRUCTURAL ADJUSTMENT OF NEURAL NETWORK MODELS TO ENSURE INTERPRETATION

Leoshchenko S. D. – Post-graduate student of the Department of Software Tools, National University “Zaporizhzhia Polytechnic”, Zaporizhzhia, Ukraine.

Oliinyk A. O. – Dr. Sc., Professor, Associate Professor, Associate Professor of the Department of Software Tools, National University “Zaporizhzhia Polytechnic”, Zaporizhzhia, Ukraine.

Subbotin S. A. – Dr. Sc., Professor, Head of the Department of Software Tools, National University “Zaporizhzhia Polytechnic”, Zaporizhzhia, Ukraine.

Gofman Ye. O. – PhD, Senior Researcher of the Research Unit, National University “Zaporizhzhia Polytechnic”, Zaporizhzhia, Ukraine

Korniienko O. V. – Post-graduate student of the Department of Software Tools, National University “Zaporizhzhia Polytechnic”, Zaporizhzhia, Ukraine.

ABSTRACT

Context. The problem of structural modification of pre-synthesized models based on artificial neural networks to ensure the property of interpretation when working with big data is considered. The object of the study is the process of structural modification of artificial neural networks using adaptive mechanisms.

Objective of the work is to develop a method for structural modification of neural networks to increase their speed and reduce resource consumption when processing big data.

Method. A method of structural adjustment of neural networks based on adaptive mechanisms borrowed from neuroevolutionary synthesis methods is proposed. At the beginning, the method uses a system of indicators to evaluate the existing structure of an artificial neural network. The assessment is based on the structural features of neuromodels. Then the obtained indicator estimates are compared with the criteria values for choosing the type of structural changes. Variants of mutational changes from the group of methods of neuroevolutionary modification of the topology and weights of the neural network are used as variants of structural change. The method allows to reduce the resource intensity during the operation of neuromodels, by accelerating the processing of big data, which expands the field of practical application of artificial neural networks.

Results. The developed method is implemented and investigated by the example of using a recurrent artificial network of the long short-term memory type when solving the classification problem. The use of the developed method allowed speed up of the neuromodel with a test sample by 25.05%, depending on the computing resources used.

Conclusions. The conducted experiments confirmed the operability of the proposed mathematical software and allow us to recommend it for use in practice in the structural adjustment of pre-synthesized neuromodels for further solving problems of diagnosis, forecasting, evaluation and pattern recognition using big data. The prospects for further research may consist in a more fine-tuning of the indicator system to determine the connections encoding noisy data in order to further improve the accuracy of models based on neural networks.

KEYWORDS: interpretation, topology, structural adjustment, neuroevolution, neural networks.

REFERENCES

1. Veneri G., Capasso A. Hands-On Industrial Internet of Things: Create a powerful Industrial IoT infrastructure using Industry 4.0. Birmingham, Packt Publishing, 2018, 556 p.
2. Roshak M. Artificial Intelligence for IoT Cookbook: Over 70 recipes for building AI solutions for smart homes, industrial IoT, and smart cities. Birmingham, Packt Publishing, 2021, 260 p.
3. Chantzis F., Stais I., Calderon P., Deirmentzoglou E., Woods B. Practical IoT Hacking: The Definitive Guide to Attacking the Internet of Things. San Francisco, No Starch Press, 2021, 464 p.
4. Nam C. S. Neuroergonomics: Principles and Practice (Cognitive Science and Technology). Berlin, Springer, 2020, 492 p.
5. Caballé S., Demetriadis S. N., Sánchez E. G., Papadopoulos P. M., Weinberger A. Intelligent Systems and Learning Data Analytics in Online Education (Intelligent Data-Centric Systems: Sensor Collected Intelligence). Cambridge, Academic Press, 2021, 424 p.
6. Aggarwal C. C. Neural Networks and Deep Learning: A Textbook, Berlin, Springer, 2018, 520 p.
7. Shikhman V., Müller D. Mathematical Foundations of Big Data Analytics. Berlin, Springer, 2021, 288 p.
8. Deitel P., Deitel H. Intro to Python for Computer Science and Data Science: Learning to Program with AI, Big Data and The Cloud. London, Pearson, 2019, 880 p.
9. D'souza R. N., Huang P.Y., Yeh F.C. Structural Analysis and Optimization of Convolutional Neural Networks with a Small Sample Size, *Scientific Reports*, 2020, No. 10, pp. 1–13. DOI: 10.1038/s41598-020-57866-2.
10. Nowakowski G., Dorogyy Y., Doroga-Ivaniuk O. Neural Network Structure Optimization Algorithm, *Journal of Automation, Mobile Robotics and Intelligent Systems*, 2018, No. 12, pp. 5–13. DOI: 10.14313/JAMRIS_1-2018/1.
11. Iba H. Evolutionary Approach to Machine Learning and Deep Neural Networks, *Neuro-Evolution and Gene Regulatory Networks*. New York, Springer, 2018, 258 p.
12. Izonin I., Tkachenko R., Kryvinska N., Gregus M., Tkachenko P., Vitynskiy P. Committee of SGTN Neural-Like Structures with RBF kernel for Insurance Cost Prediction Task, *2nd Ukraine Conference on Electrical and Computer Engineering (UKRCON), 2–6 July 2019, proceedings*. Lviv, IEEE, 2019, pp. 1037–1040. DOI: 10.1109/UKRCON.2019.8879905.

13. Fultz N. E., Bonmassar G., Setsompop K., Stickgold R. A., Rosen B. R., Polimeni J. R., Lewis L. D. Coupled electrophysiological, hemodynamic, and cerebrospinal fluid oscillations in human sleep, *Science*, 2019, Vol. 366, Issue 6465, pp. 628–631. DOI: 628–631. 10.1126/science.aax5440.
14. Are We “Brain Washed” during Sleep? [Electronic resource], Access mode: <https://www.bu.edu/articles/2019/cerebrospinal-fluid-washing-in-brain-during-sleep/>
15. Deep Sleep Gives Your Brain a Deep Clean [Electronic resource], Access mode: <https://www.scientificamerican.com/article/deep-sleep-gives-your-brain-a-deep-clean/>
16. Christiansen N. H., Hultmann Job J., Klyver K., Høgsbrg J. Optimal Brain Surgeon on Artificial Neural Networks in Nonlinear Structural Dynamics, *25th Nordic Seminar on Computational Mechanics (NSCM), 25–26 October 2012, proceedings*. Lund, Lund University, 2012, pp. 319–324.
17. Endisch C., Hackl C., Schröder D. Optimal Brain Surgeon for General Dynamic Neural Networks, *Progress in Artificial Intelligence*, 2007, Vol. 4874, pp. 15–28. DOI: 10.1007/978-3-540-77002-2_2.
18. Oppermann A. Regularization in Deep Learning – L1, L2, and Dropout [Electronic resource]. Access mode: <https://www.deeplearning-academy.com/p/ai-wiki-regularization>
19. Classic Regularization Techniques in Neural Networks [Electronic resource]. Access mode: <https://medium.com/@ODSC/classic-regularization-techniques-in-neural-networks-68bccee03764>
20. Weight Agnostic Neural Networks [Electronic resource]. Access mode: <https://arxiv.org/pdf/1906.04358.pdf>
21. Weight Agnostic Neural Networks [Electronic resource]. Access mode: <https://weightagnostic.github.io/>
22. Leoshchenko S. D., Oliinyk A. O., Subbotin S. A., Lytvyn V. A., Shkarupilo V. V. Modification and parallelization of genetic algorithm for synthesis of artificial neural networks, *Radio Electronics, Computer Science, Control*, 2019, № 4, P. 68–82. DOI: 10.15588/1607-3274-2018-3-12.
23. Leoshchenko S. D., Oliinyk A. O., Subbotin S. A., Nariv'skiy O. E. Implementation of the indicator system in modeling complex technical systems, *Radio electronics, Computer science, Control*, 2021, No. 1, pp. 117–127. DOI: 10.15588/1607-3274-2021-1-12.
24. Sattari M. T., Yurekli K., Pal M. Performance evaluation of artificial neural network approaches in forecasting reservoir inflow, *Applied Mathematical Modelling*, 2012, Vol. 36, Issue 6, pp. 2649–2657. DOI: 10.1016/j.apm.2011.09.048.
25. Hassan M., Hamada M. Evaluating the performance of a neural network-based multi-criteria recommender system, *International Journal of Spatio-Temporal Data Science*, 2019, Vol. 1 (54), pp. 54–66. DOI: 10.1504/IJSTDS.2019.10018848
26. Jamróz D. The examination of the effect of the criterion for neural network's learning on the effectiveness of the qualitative analysis of multidimensional data, *Knowledge and Information Systems*, 2020, Vol. 62, pp. 3263–3289. DOI: 10.1007/s10115-020-01441-8.
27. Subbotin S.A. Criteria for Comparison of Recognition Models Based on Neural Networks and Analysis of Their Mutual Relations, *Artificial Intelligent*, 2014, Vol. 1, pp. 142–152.
28. Leoshchenko S., Oliinyk A., Subbotin S. Adaptive Mechanisms for Parallelization of the Genetic Method of Neural Network Synthesis. *10th International Conference on Advanced Computer Information Technologies (ACIT 2020), Deggendorf, 16–18 November, proceedings*. Ternopil, IEEE, 2020, pp. 446–450. DOI: 10.1109/ACIT49673.2020.9208905.
29. Myocardial infarction complications Data Set [Electronic resource]. Access mode: <https://archive.ics.uci.edu/ml/datasets/Myocardial+infarction+complications>
30. Gorban A. N., Rossiev D. A., Butakova E. V., Gilev S. E. et al.] Medical, psychological and physiological applications of MultiNeuron neural simulator, *The Second International Symposium on Neuroinformatics and Neurocomputers, Rostov-on-Don, 20–23 September : proceedings*. Rostov-on-Don, IEEE, 1995, pp. 7–14. DOI: 10.1109/ISNINC.1995.480831.
31. Golovenkin S. E., Bac J., Chervov A., Mirkes E. M., Orlova Y. V., Barillot E., Gorban A. N., Zinovyev A. Trajectories, bifurcations, and pseudo-time in large clinical datasets: applications to myocardial infarction and diabetes data, *GigaScience*, 2020, Vol. 9(11), pp. 1–20. DOI: 10.1093/gigascience/giaa128

A MODEL AND TRAINING METHOD FOR CONTEXT CLASSIFICATION IN CCTV SEWER INSPECTION VIDEO FRAMES

Moskalenko V. V. – PhD, Associate Professor of Computer Science department, Sumy State University, Sumy, Ukraine.

Zaretsky M. O. – Postgraduate Student of Computer Science department, Sumy State University, Sumy, Ukraine.

Moskalenko A. S. – PhD, Senior Lecturer of Computer Science department, Sumy State University, Sumy, Ukraine.

Panych A. O. – M. Eng., Lecturer of Computer Science department, Sumy State University, Sumy, Ukraine.

Lysyuk V. V. – M. Eng., Co-Founder of Molfar.AI, Molfar.AI sp. z o.o., Gdansk, Poland.

ABSTRACT

Context. A model and training method for observational context classification in CCTV sewer inspection video frames was developed and researched. The object of research is the process of detection of temporal-spatial context during CCTV sewer inspections. The subjects of the research are machine learning model and training method for classification analysis of CCTV video sequences under the limited and imbalanced training dataset constraint.

Objective. Stated research goal is to develop an efficient context classifier model and training algorithm for CCTV sewer inspection video frames under the constraint of the limited and imbalanced labeled training set.

Methods. The four-stage training algorithm of the classifier is proposed. The first stage involves training with soft triplet loss and regularisation component which penalises the network's binary output code rounding error. The next stage is needed to determine the binary code for each class according to the principles of error-correcting output codes with accounting for intra- and interclass relationship. The resulting reference vector for each class is then used as a sample label for the future training with Joint Binary Cross Entropy Loss. The last machine learning stage is related to decision rule parameter optimization according to the information criteria to determine the boundaries of deviation of binary representation of observations for each class from the corresponding reference vector. A 2D convolutional frame feature extractor combined with the temporal network for inter-frame dependency analysis is considered. Variants with 1D Dilated Regular Convolutional Network, 1D Dilated Causal Convolutional Network, LSTM Network, GRU Network are considered. Model efficiency comparison is made on the basis of micro averaged F1 score calculated on the test dataset.

Results. Results obtained on the dataset provided by Ace Pipe Cleaning, Inc confirm the suitability of the model and method for practical use, the resulting accuracy equals 92%. Comparison of the training outcome with the proposed method against the conventional methods indicated a 4% advantage in micro averaged F1 score. Further analysis of the confusion matrix had shown that the most significant increase in accuracy in comparison with the conventional methods is achieved for complex classes which combine both camera orientation and the sewer pipe construction features.

Conclusions. The scientific novelty of the work lies in the new models and methods of classification analysis of the temporal-spatial context when automating CCTV sewer inspections under imbalanced and limited training dataset conditions. Training results obtained with the proposed method were compared with the results obtained with the conventional method. The proposed method showed 4% advantage in micro averaged F1 score.

It had been empirically proven that the use of the regular convolutional temporal network architecture is the most efficient in utilizing inter-frame dependencies. Resulting accuracy is suitable for practical use, as the additional error correction can be made by using the odometer data.

KEYWORDS: Sewer pipe inspection, convolutional neural network, error-correction output codes, Siamese network, Information-Extreme Learning, information criterion, LSTM, GRU.

ABBREVIATIONS

BB is a Building Block;
CNN is a Convolutional Neural Network;
GRU is a Gated Recurrent Unit;
LSTM is a Long Short-Term Memory;
MSCC is a Manual of Sewer Condition Classification;
PACP is a Pipeline Assessment Certification Program;
SLAM is a Simultaneous Localization and Mapping;
TCN is a Temporal Convolutional Network;
1D is a One-Dimensional space;
2D is a Two-Dimensional space.

NOMENCLATURE

D_v^{train} is the labeled frame sequences for training;

D_w^{test} is the labeled frame sequences for testing;

$I_{v,k}$ is a v -th set of ordered video frame sequences for training;

$L_{v,k}$ is a v -th set of ordered labels of video frame sequences for training;

V is a number of labeled frame sequences for training;

W is a number of labeled frame sequences for testing;

K_v is a size of v -th set for training;

K_w is a size of w -th set for testing;

N is a size of high-level feature set;

Z is a size of set of classes;

e_{ξ_1} is a ξ_1 -th parameter which impacts on feature representation, $\xi_1 = \overline{1, \Xi_1}$;

f_{ξ_2} is a ξ_2 -th parameter which impacts on efficiency of decision rules, $\xi_2 = \overline{1, \Xi_2}$;

TP_z is a numbers of true positives for decision rule of z -th class;

FP_z is a numbers of false positives for decision rule of z -th class;

FN_z is a numbers of false negatives for decision rule of z -th class;

b_z^* is a binary reference vector (center of optimal container) for class X_z^o ;

d_z^* is a radius of optimal container for class X_z^o in Hamming distance units;

$f(x)$ is a function describing the feature extractor;

x_a is a image randomly selected from the mini-batch;

x_{ep} is a nearest neighbour in the minibatch belonging to the same class;

$C(x)$ is a function returning the image class;

x_{shn} is a sample image from the mini-batch which is the closest among the samples of opposite classes, but located further than hard negative sample;

e is a single column matrix, $e = [1, 1, \dots, 1]^T$;

λ is a regularization coefficient;

$f_i(x)$ is a the value of sigmoid layer output i for input image;

$b_{z,i}$ is a value of i -th bit of the reference vector of the z -th class to which the image x belongs;

E_z is a Z -class information criteria, a function of the accuracy characteristics;

$\alpha_z^{(k)}$ is a false postive rate on k -th traning step for z -th class;

$\beta_z^{(k)}$ is a false negative rate on k -th traning step for z -th class;

$D_{1,z}^{(k)}$ is a true positive rate or sensitivity on k -th traning step for z -th class;

$D_{2,z}^{(k)}$ is a true negative rate or specificity on k -th traning step for z -th class;

d is a distance measure defining the hyperspherical container radii in the radial basis of Hamming space;

$\{d\}$ is a set of concentric radii of data distribution in class z with the centre b_z .

INTRODUCTION

Sewer pipes are critical infrastructure items which require frequent monitoring. The most widely used method for analysis of sewer pipe conditions involves the CCTV inspection of the pipes view to identify the defects

and faults inside the pipe. Each of the detected defects and faults is assigned a standardised code in accordance with the applicable local standards, among which the most common are the British MSCC5 and American PACP6 or PACP7 [1].

The preparation of a report on the condition of inspected sewer pipes in accordance with the standards requires careful examination and detailed analysis of the collected CCTV inspection videos. The use of computer vision and machine learning techniques for CCTV inspection footage analysis can increase productivity and reduce costs [2].

To achieve the correct defect coding it is necessary to have information about the location, orientation, shape, severity and proximity of the defect to the upstream and downstream manholes and sewer line branches (laterals/service connections/taps). In turn, contextual data on the orientation and relative location of the inspection camera in the pipe is needed to extract such information. Such data, however, as a rule is not available in the explicit form. This makes observation context recognition a relevant task.

Orientation and relative position of the camera can be determined with the help of visual odometry or simultaneous localization and mapping (SLAM) methods [3]. However, CCTV inspection videos generally already contain superimposed distance readings measured by a mechanical odometer. Correct defect coding also does not require high degree of precision in respect of the camera optical axis angle and relative position in relation to the center of the pipe. Employing a computationally efficient frame sequence classifier to estimate the camera orientation instead of computationally complex SLAM algorithms is therefore more appropriate. In this case alphabet of classes can be expanded to include various non-standard situations which need to be processed correctly.

A modern approach to the classification analysis of the sequence of video frames involves the use of deep neural networks. Important steps in the classification analysis of individual video frames or their sequences are the feature descriptions of both individual frames and the relationships between them, as it has a direct bearing on the effectiveness of the resulting decision rules.

Convolutional neural networks are still the most effective approach to image feature description at present. Where the analysis of time series is concerned, the undisputed leaders are recurrent and temporal convolutional networks, dilated versions of which provide a speed advantage without information loss [4]. The end results, however, depend not only on the architecture of the model, but also on the methods of machine learning and regularization employed. This is especially true where the labelled training dataset is small and variability of observations is high. The study of the effectiveness of different model architectures and training methods for specific applications thus remains a relevant task, as no single universal approach to data analysis exists and steady emergence of new research continually changes

the paradigm, forcing a rethink of the current body of knowledge and highlight new directions for further research.

The research goal is development of an effective deep learning model and its training method for recognizing the context of observations during CCTV sewer inspection. **The object of research** is the process of detection of temporal-spatial context during CCTV sewer inspections. **The subjects of the research** are machine learning model and training method for classification analysis of CCTV video sequences under the limited and imbalanced training dataset constraint.

1 PROBLEM STATEMENT

Let V sequences $D_v^{train} = \{I_{v,k}, L_{v,k} \mid v = \overline{1, V}; k = \overline{1, K_v}\}$ and W sequences $D_w^{test} = \{I_{w,k}, L_{w,k} \mid w = \overline{1, W}; k = \overline{1, K_w}\}$ are collected of labeled video frames for training and testing, respectively. Let the set $\{X_z^o \mid z = \overline{1, Z}\}$ is characterized by observation context in pipe, be given. In this case, the dataset is unbalanced, the minority class can contain twice as many samples as the majority.

Moreover, the structure of the vector of model parameters is known

$$g = \langle e_1, \dots, e_{\xi_1}, \dots, e_{\Xi_1}, f_1, \dots, f_{\xi_2}, \dots, f_{\Xi_2} \rangle, \quad (1)$$

$$\Xi_1 + \Xi_2 = \Xi.$$

In this case, the constraints $R_{\xi_1}(e_1, \dots, e_{\xi_1}, \dots, e_{\Xi_1}) \leq 0$, $R_{\xi_2}(f_1, \dots, f_{\xi_2}, \dots, f_{\Xi_2}) \leq 0$ are impose on parameters.

It is necessary to find by machine learning an optimal values of parameters g (1) which provide to achieve the maximum value of micro averaged F1 score for context classifier

$$F1 = \frac{2 \sum_z TP_z}{2 \sum_z TP_z + \sum_z FP_z + \sum_z FN_z}. \quad (2)$$

$$g^* = \arg \max_G \{F1(g)\}. \quad (3)$$

When the model functions in its inference mode, it is necessary to provide high confidence of classification of frame context on test images.

2 REVIEW OF THE LITERATURE

Early algorithms for CCTV sewer inspection video frames classification analysis employed edge and contour detection methods for feature description [5]. Such an approach, however, ignores a large amount of contextual information and necessitates particular attention to the design of a post-processing algorithm. An algorithm of this kind would require a large number of handcrafted parameters and conditions, which can lead to

incompleteness of decision rules or contradictions between them due to the human error. Gabor filters offer a more flexible and theoretically sound approach to visual feature extraction [6]. However, models of this type are characterized by insufficient information capacity for computationally efficient description of contexts under conditions of complex defect and/or design features combinations.

Much progress in the field of visual data analysis had been achieved within the framework of the deep machine learning, based on hierarchical feature description. A distinguishing feature of hierarchical feature extractors is their higher information capacity in comparison to models with one hidden layer [7, 8]. At present deep convolutional neural networks are still considered the most effective for image feature description [8]. In the field of time-series analysis, the leading positions are occupied by recurrent and temporal convolutional networks. Their dilated versions provide a speed advantage without loss of information [4] in both customary, centered, and causal model output variants. However, some observational contexts or their parts are rare and can have significant intra-class variability, leading to imbalances and a scarcity of labeled samples corresponding to complex and irregular situations. This imposes limitations on the use of deep models sensitive to the volume and balance of labelled training data.

One of the ways to increase the efficiency of models with a limited amount of marked data is to use the ideas and methods of information theory and synthesis of decision rules within a geometric approach. An example of geometric approach methods are Siamese neural networks, where the fitness function makes use of constraints and relationships between the distances between samples of the same and different classes [9]. Siamese networks have shown the greatest efficiency in few-shot learning and meta-learning algorithms, but they are most commonly used for feature embedding.

Information theory and coding methods are a natural choice for increasing the resistance to noise, such as artifacts and limited visibility. For example, error-correcting output codes implement end-to-end pseudo-ensemble, increase the multi-class classification accuracy and Improve Probability Estimation for Adversarial Robustness of Deep Neural Networks [10].

However, the existing methods of binary class code selection does not take into account the internal structure of classes. Information-extreme machine learning methods provide optimization in the information sense of the decision rule parameters based on error-correcting output codes [11]. However, information-extreme learning does not provide end to end deep model learning mechanisms. Thus, the combination of ideas and methods of Siamese neural networks and information-extreme learning offers considerable promise for further improvement of data analysis models, in particular for context analysis in CCTV sewer inspections.

3 MATERIALS AND METHODS

Classification analysis of CCTV video frames in the simplest case can be performed by a single convolutional network. Such network can be trained both in the traditional way and as a part of Siamese or generative models. However, in the situations of loss of visibility or significantly close proximity of the camera to the pipe walls, images lose the large part of their useful context information. This necessitates the analysis of each frame in conjunction with the neighbouring frames to restore the context information. In general case context detection model will have the 2D convolutional neural network at the lower level for spatial feature extraction and 1D temporal network for analysis of cross-frame dependencies (Fig. 1).

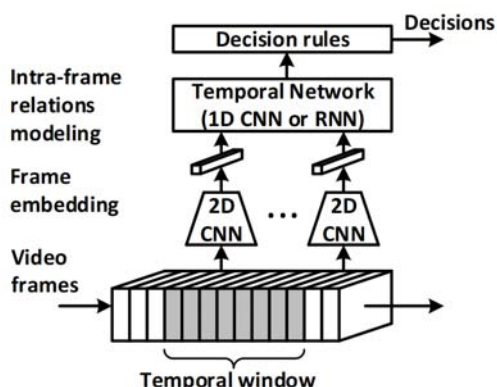


Figure 1 – Generalized architecture of context classifier model

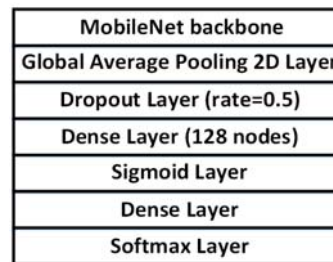
For separate frame analysis the use of MobileNet general purpose convolutional network is proposed [9]. Only a convolutional backbone without fully connected layers is used [9]. Fig 2 a depicts the classic convolutional network variant and Fig 2b its modification used for research of the proposed training method.

Global Average Pooling is used for dimensionality reduction and a Dropout pseudo-ensemble with 50% of the input features dropping is used for regularization [5, 9]. Fully connected and sigmoid layers form the output feature set.

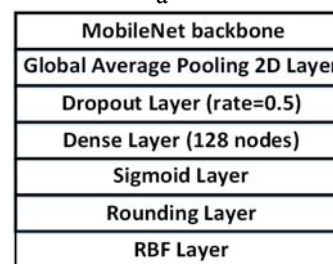
Image classifier model’s decision rules contain the rounding layer which produces the binary coded representation and radial-basis function defining the object’s belonging to a certain class, with classes separated by hyper-spherical containers in binary Hamming space. Each hyper-spherical container is defined by the binary reference vector (container center) and container radius in Hamming distance units. In this case radial-basis membership function $\mu_z(b)$ for N -dimensional binary vector b is

$$\mu_z(b) = 1 - \sum_{i=1}^N b_i \oplus b_{z,i}^* / d_z^*, \quad (4)$$

where b_z^* – binary reference vector (center of optimal container) for class X_z^o ; d_z^* – radius of optimal container for class X_z^o in Hamming distance units.



a



b

Figure 2 – Architecture of classic and modified variants of the convolutional network for frame-by-frame classification : a – baseline image classifier model structure; b – proposed image classifier model structure

Temporal network can be implemented using one of the popular model architectures, such as 1D dilated non-causal convolutional network, 1D dilated causal convolutional network, Recurrent neural networks with Long Short-Term Memory (LSTM), Recurrent neural networks with Gated Recurrent Units (GPU).

The Base Block (BB) convolutional temporal networks is depicted in Fig. 3 The first layer of the BB is a 1D dilated convolution with kernel size $k = 3$ where the dilation factor is doubled for each subsequent BB, i.e. 1, 2, 4, 8.

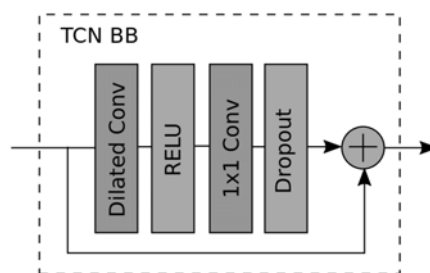


Figure 3 – Basic block of Temporal Convolutional Network

Fig. 3a gives an illustration of the receptive field (black arrows) of an output activation from a single stage with three stacked BBs with regular convolutions. Regular convolutions have a receptive field that expands equally wide to the right as it does to the left. This means that it looks as far into the future (right) as it looks into the past (left). Thus the current frame context will be clarified using the information in both past and future frames.

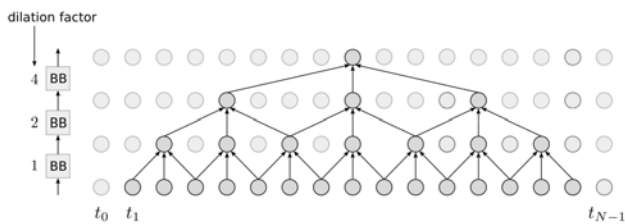


Figure 4 – Temporal Network with Non-causal (regular) convolutions

Fig. 5 illustrates temporal neural network with casual convolutions which amplify the forecast productivity nearer to the right edge. In this case, the curt frame context will be completely defined on the basis of the characteristics and interrelationships of the preceding frames.

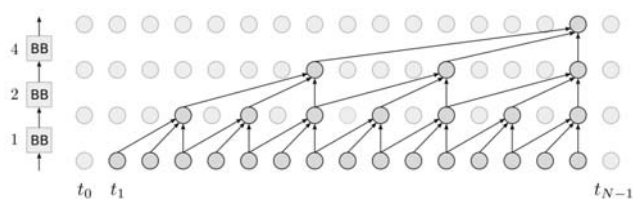
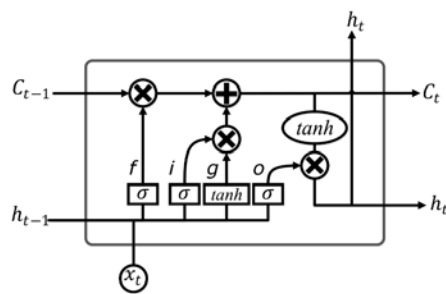


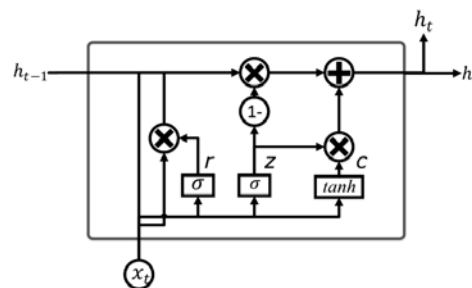
Figure 5 – Temporal Network with casual convolutions

LSTM has memory blocks connected by successive layers, and it enables the network to selectively memorize the input training data through a unique three-gate structure (Fig. 6a). The line across the top of the diagram is the cell state c , and represents the internal memory of the unit. The line across the bottom is the hidden state h , and the i , f , o , and g gates are the mechanism by which the LSTM works around the vanishing gradient problem. During training, the LSTM learns the parameters for these gates. Instead of the input, forget, and output gates in the LSTM cell, the GRU cell has two gates, an update gate z , and a reset gate r (Fig. 6b). The update gate defines how much previous memory to keep around and the reset gate defines how to combine the new input with the previous memory. There is no persistent cell state distinct from the hidden state as in LSTM.

To compare and trace the changes in productivity as a function of the proposed solutions, training will be performed in stages. First the single-frame detection model will be trained in a traditional manner and with the proposed training method without taking the neighbouring frames into account. Then the best trained model is chosen and its feature extractor, the layers located up to and including the sigmoid layer, is used for frame embedding into the temporal detection model. Every neural network type will be trained with both traditional and the proposed training method. The capacity hyperparameter, responsible for the network size will be grid-optimized for each model.



a



b

Figure 6 – Cell of Recurrent Neural Network: a – LSTM cell; b – GRU cell

Traditional training method involves an addition of Dense layer with Softmax output normalization, error backpropagation and cross-entropy loss function, such as Adam, to the feature extractor.

The modified method consists of 4 stages necessary to create a binary feature description used to form the information-extreme decision rules (Fig. 7)

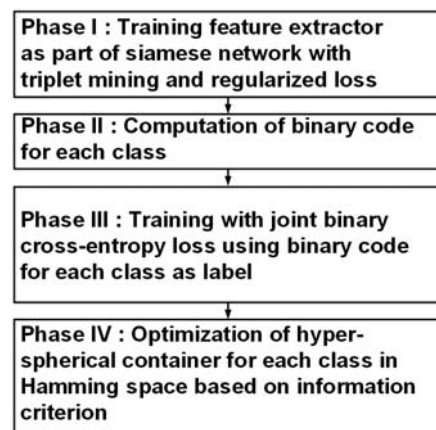


Figure 7 – Proposed training method stages

The first stage involves training with soft triplet loss and regularization component for penalising the binary code output rounding error. The mode input receives a mini-batch with M images of every class. The loss function is calculated as

$$L = -\log \frac{\exp(\|f(x_a) - f(x_{ep})\|)}{\exp(\|f(x_a) - f(x_{ep})\|) + \exp(\|f(x_a) - f(x_{shn})\|)} + \lambda(f(x_a)^T(e - f(x_a)) + f(x_{ep})^T(e - f(x_{ep})) + f(x_{shn})^T(e - f(x_{shn}))), \quad (5)$$

$$x_{ep} = \arg \min_{x:C(x)=C(x_a)} \|f(x_a) - f(x)\|, \quad (6)$$

$$x_{shn} = \arg \min_{\substack{C(x) \neq C(x_a) \\ x: \|f(x_a) - f(x)\| > \|f(x_a) - f(x_p)\|}} \|f(x_a) - f(x)\|, \quad (7)$$

The next phase is needed to determine the binary class code in accordance with the error-correction output codes principles but also accounting for intra-class and inter-class relationships. The training set of Z classes $\{x_{z,s} \mid z = \overline{1, Z}, s = \overline{1, n_z}\}$ containing n_z samples of z -class, is encoded with a binary representation $\{b_{z,s,i} \mid z = \overline{1, Z}, s = \overline{1, n_z}, i = \overline{1, N}\}$ with dimensionality N . The binary coding of input image $x_{z,s}$ is achieved by rounding the i sigmoid layer output to the whole number.

$$b_{z,s,i} = \begin{cases} 1, & \text{if } f_i(x_{z,s}) > 0.5; \\ 0, & \text{otherwise.} \end{cases}$$

Binary reference vector b_z for class z can be calculated by [rank-wise] comparison of frequency of binary ones in z class with the background binary ones frequency in the training set

$$b_{z,i} = \begin{cases} 1, & \text{if } \frac{1}{n_z} \sum_{s=1}^{n_z} b_{z,s,i} > \frac{1}{Z} \sum_{c=1}^Z \frac{1}{n_c} \sum_{s=1}^{n_c} b_{c,s,i}; \\ 0, & \text{otherwise.} \end{cases}$$

Reference vector b_z for class z is used as a sample label in further training with Joint Binary Cross Entropy Loss, which is calculated for each training sample x as

$$L = - \sum_{i=1}^N (b_{z,i} \log f_i(x) + (1 - b_{z,i}) \log(1 - f_i(x))).$$

The last stage of machine learning is related to the optimization of container radius by information criterion to account for the boundaries of deviation of binary representation of observations in each class from the corresponding reference vectors.

$$E_z^* = \max_{\{d\}} E_z(d). \quad (8)$$

Entropy criteria for a bi-alternative evaluation system ($Z = 2$) and equally probable hypothesis, representing the most statistically difficult case, can be expressed as a function of accuracy characteristics as follows [11]

$$E_z^{(k)} = 1 + \frac{1}{2} \left(\frac{\alpha_z^{(k)}(d)}{\alpha_z^{(k)}(d) + D_{2,z}^{(k)}(d)} \log_2 \frac{\alpha_z^{(k)}(d)}{\alpha_z^{(k)}(d) + D_{2,z}^{(k)}(d)} + \frac{\beta_z^{(k)}(d)}{D_{1,z}^{(k)}(d) + \beta_z^{(k)}(d)} \log_2 \frac{\beta_z^{(k)}(d)}{D_{1,z}^{(k)}(d) + \beta_z^{(k)}(d)} + \frac{D_{1,z}(d)}{D_{1,z}^{(k)}(d) + \beta_z^{(k)}(d)} \log_2 \frac{D_{1,z}(d)}{D_{1,z}^{(k)}(d) + \beta_z^{(k)}(d)} + \frac{D_{2,z}^{(k)}(d)}{\alpha_z^{(k)}(d) + D_{2,z}^{(k)}(d)} \log_2 \frac{D_{2,z}^{(k)}(d)}{\alpha_z^{(k)}(d) + D_{2,z}^{(k)}(d)} \right). \quad (9)$$

4 EXPERIMENTS

The current article considers all the training stages and their corresponding results obtained on the dataset provided by Ace Pipe Cleaning, Inc.

Class alphabet used to detect the observation context without taking into account the content of the neighbouring frames contains 10 main context classes (Table 1). This alphabet contains ignore, side and connection classes. Samples for these classes are easy to collect and label, however they do not provide complete certainty as to the camera orientation, as it is not clear which pipe wall camera is facing (Fig. 8).

Class alphabet used to detect the observation context with taking into account the content of the neighbouring frames contains 11 main context classes (Table 2). Temporal features have to provide complete certainty as to the camera orientation, hence there is no ignore class and side and connection classes have been replaced by Right, Left, Top, Bottom, Right connection, Left connection, Top connection and Bottom connection. Semi Left, Semi Right, Semi top, Semi bottom classes are replaced by the corresponding Right, Left, Top, Bottom. Temporal window for consideration of the neighbouring frames is set to 128 frames. This window was selected as a multiple of 2 and was experimentally selected as being close to optimal for various models. Parsing labelled video files $\{D_v^{train}\}$ and $\{D_w^{test}\}$ ensures the formation of variable quantity of samples for each class.

Prior unsupervised learning of the upper convolutional layers on unlabelled samples from the intended usage domain is aimed at increasing the subsequent supervised machine learning efficiency. It is worthwhile considering the influence the parameters of the growing sparse coding neural gas algorithm used in unsupervised learning have on the results of supervised learning. Table 1 presents the machine learning results and quantity N_c of generated convolutional filters (neurons) as a function of the parameter v , which characterises the accuracy of coverage of the training set by the convolutional filters.

Before training, the entire dataset is balanced by applying augmentation to minor classes (0–5% change in scale, $\pm 5\%$ rotation, $\pm 5\%$ change in brightness).

Each model has a number of hyperparameters which define its configuration and capacity. Optimal hyperparameters are first selected for each architecture

and a comparison of results obtained from various architectures with these parameters. Models are trained during 60 epochs. The hyperparameters are selected with a view to avoid the noticeable overtraining effect.

Table 1 – Class set for the single-frame context classifier model

Designation of class	Number of examples	Names of context	Description
X_1^0	5000	Forward	Camera pointing forward along the pipe
X_2^0	3000	Side	Camera is pointing at the pipe wall facing left, right, down or up, when it is hard to understand which part of the pipe (top, bottom, left or right) the camera is pointing at
X_3^0	1000	Semi right	Incomplete right turn, orientation of the camera can be clearly determined from a single frame
X_4^0	1000	Semi left	Incomplete left turn, orientation of the camera can be clearly determined from a single frame
X_5^0	320	Semi top	Incomplete turn upwards, orientation of the camera can be clearly determined from a single frame
X_6^0	180	Semi bottom	Incomplete turn downwards, orientation of the camera can be clearly determined from a single frame
X_7^0	2100	Connection	Another pipe connecting to the main
X_8^0	500	Man-hole	Point of entry into the pipe for inspection, from the manhole to the beginning of the pipe proper
X_9^0	152	Collapse	Collapsed pipe, further movement forward is impossible
X_{10}^0	100	Ignore	Situations to be ignored for processing purposes

Backbone of Mobilenet with the capacity coefficient set to 0.25 and input resolution set to 160x160 pixels was used as a single-frame feature extractor. Temporal convolutional network has 1 stage configuration with 7 BBs. Quantity of feature channels in convolutional filters of the first layer is set as $C=128$, to match the dimensionality of image embedding. The last layer of the temporal network is connected to the Dense layer, which contains 128 nodes with sigmoid activation function. GRU and LSTM recurrent networks contain one layer with 128 units. States of each unit are submitted to the Dense layer which also contains 128 nodes with sigmoid activation function. After the sigmoid layer of either model two additional output layers analogous to those depicted in Fig. 2 are used, dependent on the chosen training method.

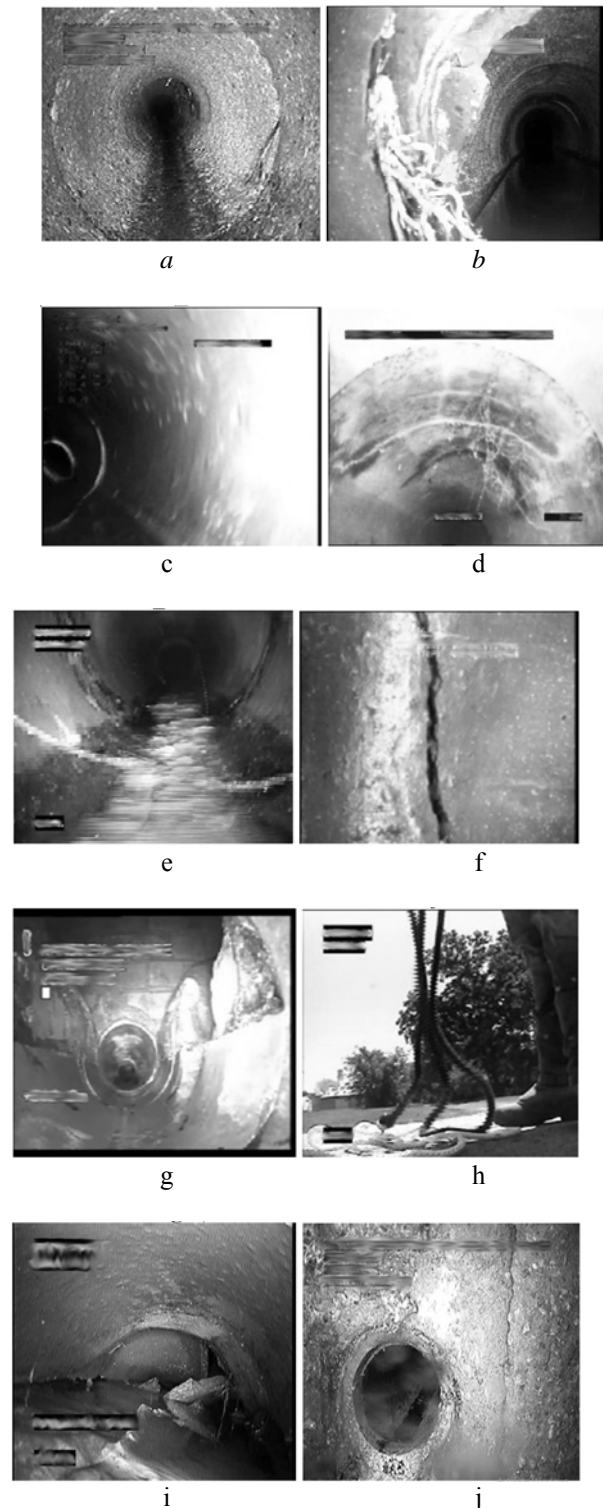


Figure 8 – Sample images of each class :
 a – class X_1^0 ; b – class X_2^0 ; c – class X_3^0 ; d – class X_4^0 ; e – class X_5^0 ; f – class X_6^0 ; g – class X_7^0 ; h – class X_8^0 ; i – class X_9^0 ; j – class X_{10}^0

Table 2 – Class set for context classifier model accounting for neighbouring frames content

Designation of class designation	Number of examples	Names of context	Description
X_1^0	4500	Forward	Camera pointing forward along the pipe
X_2^0	1500	Right	Incomplete or complete right camera turn
X_3^0	1420	Left	Incomplete or complete left camera turn
X_4^0	400	Up	Incomplete or complete camera turn upwards
X_5^0	180	Down	Incomplete or complete camera turn downwards
X_6^0	500	Right connection	Connecting pipe on the right
X_7^0	500	Left connection	Connecting pipe on the left
X_8^0	150	Top connection	Connecting pipe at the top
X_9^0	75	Down connection	Connecting pipe at the bottom
X_{10}^0	500	Manhole	Point of entry into the pipe for inspection, from the manhole to the beginning of the pipe proper
X_{11}^0	170	Collapse	Collapsed pipe, further movement forward is impossible

Fig. 9 depicts the change in F1 score on the test dataset during the classifier training with the baseline and proposed (Fig. 7) methods. The test dataset was created by selecting 15% of the samples from each class represented in Table 1.

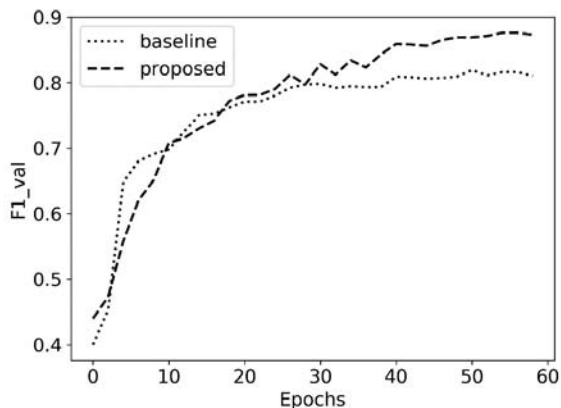


Figure 9 – Dependency of F1 score on test dataset from training epochs for single frame context classifier without accounting of neighbouring frames

Analysis of Fig. 9 shows that up to 30 epochs both training methods are performing with similar efficiency. However, the third phase (Fig. 7) of the proposed method, employed after 30th epoch, increases accuracy by 6%. Hence the feature extractor trained with the proposed training method is used for embedding into the high-level neural network.

Fig. 10 depicts the change in the F1 score on the test dataset when training the frame classifier which accounts for the neighbouring frames. Training is done with the baseline method but with varying architectures of the temporal network. Test dataset is created in the same way, but selecting 15% of the samples of each class represented in Table 2.

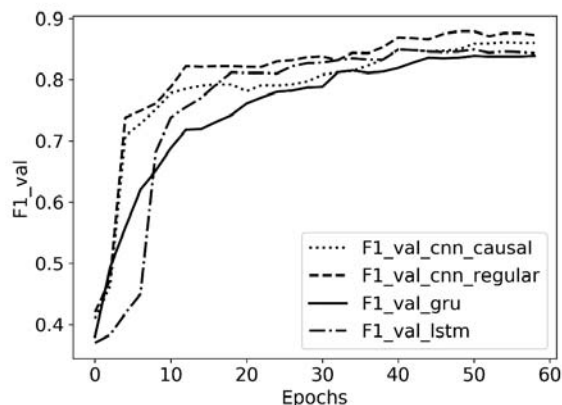


Figure 10 – Dependency of F1 score on test dataset from quantity of training epochs when training with the baseline method for context classifier with accounting of neighbouring frames

Analysis of Fig. 10 shows that the accuracy of recurrent networks during the initial training epochs increased slower than that of the convolutional networks. However, during the final stages the accuracy effectively reached a plateau with the results for different networks converging to virtually indistinguishable values. Amongst the convolutional network regular rather than causal architecture shown the best result. F1 score for the regular temporal network reached 0.87, whereas F1 for LSTM and GRU model stands at 0.843 and 0.839 respectively.

Fig. 11 shows the change in test dataset F1 score when training the classifier which accounts for the neighbouring frames trained with the proposed (Fig. 7) method using different temporal network architectures. Test network was formed in the same way by selecting 15% of the samples of each class illustrated in Table 2.

Analysis of Fig. 11 shows that accuracy of recurrent networks in the case is also inferior to those of the convolutional networks. Regular structure model had likewise shown to be the leader among the convolutional networks. However, the maximal F1 score for the model with the regular structure trained with the proposed method is 92%, which exceeds the baseline training method results on the same network by 4%.

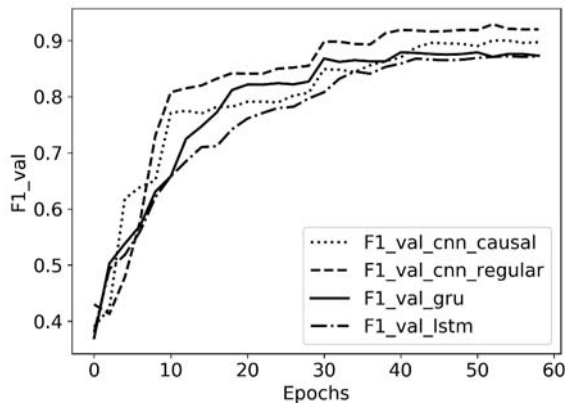


Figure 11 – Dependency of F1 score on test dataset from quantity of training epochs when training with the proposed method for context classifier with accounting of neighbouring frames

Thus the advantage of using a convolutional model for observation context analysis and the proposed multi-phase training method had been established. It was also empirically proven that the regular convolutional temporal network architecture is the most efficient in utilising interframe dependencies. Resulting accuracy is suitable for practical use, as the additional error correction can be made by using the odometer / distance counter data.

6 DISCUSSION

It is worth considering not just the aggregated metrics but also the confusion matrix to properly evaluate the efficiency of the proposed approach. A normalised confusion matrix derived from the model with regular TCN trained with the conventional approach is presented in Fig. 12.

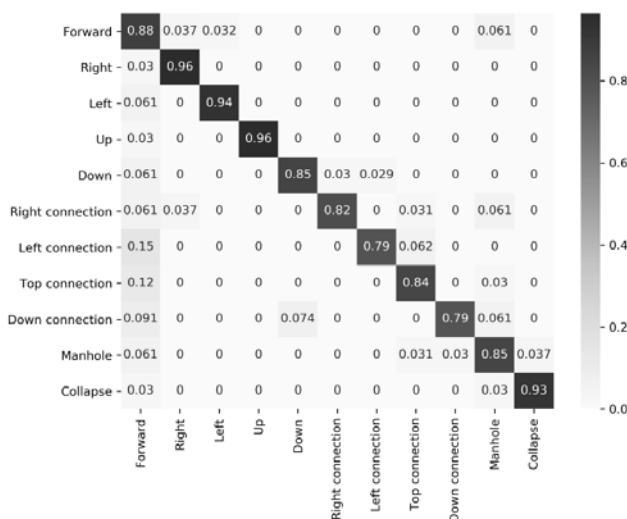


Figure 12 – Normalized confusion matrix for the optimal context classifier based on regular TCN trained with the conventional approach

Analysis of the confusion matrix presented in Fig. 12 shows that the lowest accuracy corresponds to the complex classes : Right Connection, Left Connection, Top Connection and Manhole. These classes combine both camera orientation and distinctive structural elements of the sewer system.

Fig. 13 depicts the normalized confusion matrix for the context classifier based on the regular TCN trained with the proposed approach.

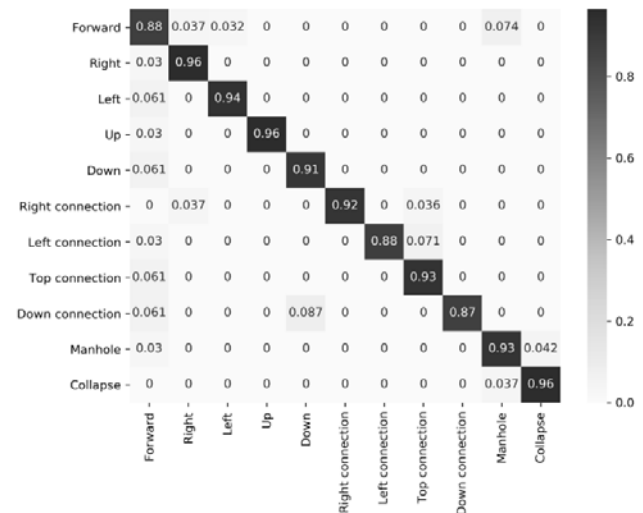


Figure 13 – Normalized confusion matrix for the optimal context classifier based on regular TCN trained with the proposed multi-phased approach

Analysis of Fig. 13 shows that the *forward* class is intersecting with almost all the other classes. In comparison with the conventional training methods, the proposed solution ensures increase in detection accuracy for complex context classes, such as Right Connection, Left Connection, Top Connection and Manhole.

Despite the improvement in accuracy, training on the same data is still not error-free. Some connections can be missed and for connections in 10:00–02:00 sector the camera orientation can be confused with the connections in the 02:00–05:00 and 07:00–10:00 sectors. This can be due to the insignificant difference between the specific degrees of partial camera turn. However, the impact of the erroneous context classification at the intersection of the context classes on the effectiveness of sewer pipe inspection is outside the scope of this study.

CONCLUSIONS

The scientific novelty of the work lies in the new models and methods of classification analysis of spatial and temporal context for automation of CCTV sewer inspections under the limited training dataset and data imbalance constraints.

Model contains a 2D convolutional network which produces feature embedding as an output of the sigmoid layer and 1D temporal convolutional regular network with the sigmoid and [rounding] output layers with radial-basis

decision rules created according to the error-correcting output codes and information-extreme learning.

The proposed method includes 4 stages : training with soft triplet loss and regularization component for penalising the binary code output rounding error; determination of the binary class code in accordance with the error-correcting output codes principles but also accounting for intraclass and interclass relationships; optimization of container radius by information criterion to account for the boundaries of deviation of binary representation of observations in each class from the corresponding reference vectors.

Training results obtained with the proposed method were compared with the results obtained by conventional training method, with a resulting 4% improvement in micro averaged F1 score metric. Confusion matrix analysis had shown that the biggest improvement in accuracy is observed for the classes which combine both the camera orientation and the distinctive structural features of the sewer system.

The practical significance of the achieved outcomes is an increase in accuracy of the classification analysis of temporal and spatial context during the CCTV sewer inspections under conditions of limited labelled training data availability and uncertainty related to the arbitrary observation conditions.

REFERENCES

1. Moradi S., Zayed T., Golkhoo F. Review on Computer Aided Sewer Pipeline Defect Detection and Condition Assessment, *Infrastructures*, 2019, Vol. 4, No. 1: 10. DOI: 10.3390/infrastructures4010010.
2. Myrans J., Everson R., Kapelan Z. Automated detection of fault types in CCTV sewer surveys, *Journal of Hydroinformatics*, 2018, Vol. 21, No. 1, pp. 153–163. DOI: 10.2166/hydro.2018.073.
3. He M., Zhu Ch., Huang Q. et al. A review of monocular visual odometry, *The Visual Computer*, 2020, Vol. 36, No. 2, pp. 1053–1065. DOI: 10.1007/s00371-019-01714-6.
4. Lim B., Zohren S. Time-series forecasting with deep learning: a survey, *Philosophical Transactions of the Royal Society A*, 2021, Vol. 379, Issue 2194, P. 14. DOI: 10.1098/rsta.2020.0209.
5. Syahrian N. M., Risma P., Dewi T. Vision-Based Pipe Monitoring Robot for Crack Detection Using Canny Edge Detection Method as an Image Processing Technique, *Kinetik*, 2017, Vol. 2, No. 4, pp. 243–250. DOI: 10.22219/kinetik.v2i4.243.
6. Czimmermann T., Ciuti G., Milazzo M. et al. Visual-Based Defect Detection and Classification Approaches for Industrial Applications – A SURVEY, *Sensors*, 2020, Vol. 20, No. 5: 1459. DOI: 10.3390/s20051459.
7. Cheng J. C. P., Wang M. Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques, *Automation in Construction*, 2018, Vol. 95, pp. 155–171. DOI: 10.1016/j.autcon.2018.08.006.
8. Panella F., Boehm J., Loo Y. et al. Deep learning and image processing for automated crack detection and defect measurement in underground structures, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2018, Vol. XLII-2, pp. 829–835. DOI: 10.5194/isprs-archives-xlii-2-829-2018.
9. Zhan H., Shi B., Duan L.-Y. et al. DeepShoe: An improved Multi-Task View-invariant CNN for street-to-shop shoe retrieval, *Computer Vision and Image Understanding*, 2019, Vol. 180, pp. 23–33. DOI: 10.1016/j.cviu.2019.01.001.
10. Zhang B., Tondi B., Lv X. et al. Challenging the Adversarial Robustness of DNNs Based on Error-Correcting Output Codes, *Security and Communication Networks*, 2020, Vol. 2020: 8882494. DOI: 10.1155/2020/8882494.
11. Moskalenko V., Moskalenko A., Korobov A. et al. The Model and Training Algorithm of Compact Drone Autonomous Visual Navigation System, *Data*, 2019, Vol. 4, No. 1: 4. DOI: 10.3390/data4010004.

Received 23.04.2021.

Accepted 24.08.2021.

УДК 004.891.032.26:629.7.01.066

МОДЕЛЬ І МЕТОД НАВЧАННЯ ДЛЯ КЛАСИФІКАЦІЙНОГО АНАЛІЗУ КОНТЕКСТУ КАДРІВ ВІДЕОІНСПЕКЦІЇ СТІЧНИХ ТРУБ

Москаленко В. В. – канд. техн. наук, доцент кафедри комп'ютерних наук, Сумський державний університет, Суми, Україна.

Зарєцький М. О. – аспірант кафедри комп'ютерних наук, Сумський державний університет, Суми, Україна.

Москаленко А. С. – канд. техн. наук, старший викладач кафедри комп'ютерних наук, Сумський державний університет, Суми, Україна.

Панич А. О. – магістр інженерії, асистент кафедри комп'ютерних наук, Сумський державний університет, Суми, Україна.

Лисюк В. В. – магістр інженерії, співзасновник компанії Molfar.AI sp. z o.o., Гданськ, Польща.

АНОТАЦІЯ

Актуальність. Розроблено та досліджено модель та метод навчання для класифікації контекстів спостереження на кадрах відеоінспекції стічних труб. Об'єктом дослідження є процес виявлення просторово-часового контексту під час інспекцій стічних труб. Предметом дослідження є модель та метод машинного навчання для класифікаційного аналізу кадрів відеоінспекції в умовах обмеженого та незбалансованого набору розмічених навчальних даних.

Мета дослідження – розроблення ефективних моделі і методу машинного навчання для класифікаційного аналізу контексту відеокадрів інспекції стічних труб в умовах обмеженого обсягу та незбалансованості розміченого навчального набору даних.

Методи дослідження. Запропоновано чотирьох етапний алгоритм навчання класифікатора. Перший етап полягає у навчанні з нормалізованою триплетною функцією втрат і регуляризуючою складовою, яка штрафує за помилку округлення вихідного сигналу до двійкового подання. Наступний етап полягає у визначенні двійкового коду для кожного класу для

реалізації кодів, що виправляють помилки, але з урахуванням внутрішньокласових та міжкласових відношень. Отриманий еталонний двійковий вектор для кожного класу потім використовується як цільова мітка під час наступного етапу навчання з бінарною крос-ентропійною функцією втрат. Останній етап машинного навчання пов'язаний з оптимізацією параметрів правила прийняття рішень за інформаційним критерієм для визначення допустимих меж відхилення двійкового подання спостережень кожного класу від відповідного еталонного вектора. Розглядається 2D згортковий екстрактор ознак у поєднанні з темпоральною мережею для аналізу міжкадрових залежностей. Розглядаються варіанти з 1D згортковою мережею з дірними регулярними згортками, 1D згорткова мережа з дірними причинно-наслідковими згортками, рекурентна мережа LSTM та рекурентна мережа GRU. Порівняння ефективності моделей проводиться на основі мікро усередненої F1-міри, обчисленої на тестовому наборі даних.

Результати. Результати, отримані за набором даних, наданим Ace Pipe Cleaning, Inc, підтверджують придатність моделі та методу для практичного використання, оскільки отримана точність дорівнює 92%. Порівняння результатів навчання із запропонованим методом та традиційним методом показало перевагу на 4% за мікро-усередненим значенням F1-міри. Подальший аналіз матриці помилок показав, що найбільш суттєве підвищення точності порівняно зі традиційними методами досягається для складних класів, які поєднують як орієнтацію камери, так і особливості конструкції стічної труби.

Висновки. Наукова новизна роботи полягає у нових моделях та методах класифікаційного аналізу просторово-часового контексту для автоматизації відеоінспекції стічних труб в умовах обмеженого обсягу та незбалансованості розмічених навчальних даних. Результати навчання, отримані за запропонованим методом, порівнюються з результатами, отриманими за допомогою традиційного методу класифікаційного аналізу зображень. Запропонований метод продемонстрував перевагу на 4% за мікро-усередненим значенням F1-міри.

Емпірично було доведено, що темпоральна мережа на основі 1D згорткової мережі з дірними регулярними згортками є найбільш ефективною для аналізу міжкадрових залежностей. Отримана точність забезпечує придатність отриманих моделей для практичного використання, оскільки додаткове виправлення помилок можна реалізувати на основі даних одометра.

КЛЮЧОВІ СЛОВА: інспекція стічних труб, згорткова нейронна мережа, коди з самокорекцією помилок, сіамська нейронна мережа, інформаційно-екстремальне навчання, інформаційний критерій, LSTM, GRU.

УДК 004.891.032.26:629.7.01.066

МОДЕЛЬ И МЕТОД ОБУЧЕНИЯ ДЛЯ КЛАССИФИКАЦИОННОГО АНАЛИЗА КОНТЕКСТА КАДРОВ ВИДЕОИНСПЕКЦИИ СТОЧНЫХ ТРУБ

Москаленко В. В. – канд. техн. наук, доцент кафедры компьютерных наук, Сумской государственной университет, Сумы, Украина.

Зарецкий Н. А. – аспирант кафедры компьютерных наук, Сумской государственной университет, Сумы, Украина.

Москаленко А. С. – канд. техн. наук, старший преподаватель кафедры компьютерных наук, Сумской государственной университет, Сумы, Украина.

Паныч А. А. – магистр инженерии, ассистент кафедры компьютерных наук, Сумской государственной университет, Сумы, Украина.

Лысюк В. В. – магистр инженерии, сооснователь компании Molfar.AI sp. z o.o., Гданск, Польша.

АННОТАЦИЯ

Актуальность. Разработаны и исследованы модель и метод обучения для классификации контекстов наблюдения на кадрах видеоинспекции сточных труб. Объектом исследования является процесс распознавания пространственно-временного контекста во время инспекций сточных труб. Предметом исследования является модель и метод машинного обучения для классификационного анализа кадров видеоинспекции в условиях ограниченного и несбалансированного набора размеченных обучающих данных.

Цель исследования – разработка эффективных модели и метода машинного обучения для классификационного анализа контекста видеокладов инспекции сточных труб в условиях ограниченного объема и несбалансированности размеченного обучающего набора данных.

Методы исследования. Предложено четырехэтапный алгоритм обучения классификатора. Первый этап заключается в обучении с нормализованной триплетной функцией потерь и регуляризирующей составляющей, которая штрафует за ошибку округления выходного сигнала к двоичному представлению. Следующий этап заключается в определении двоичного кода для каждого класса для реализации кодов, исправляющих ошибки, но с учетом внутрикласовых и межкласовых отношений. Полученный эталонный двоичный вектор для каждого класса затем используется как целевая метка во время следующего этапа обучения с бинарной кросс-энтропийной функцией потерь. Последний этап машинного обучения связан с оптимизацией параметров правила принятия решений за информационным критерием для определения допустимых пределов отклонения двоичного представления наблюдений каждого класса от соответствующего эталонного вектора. Рассматривается 2D сверточный экстрактор признаков в сочетании с темпоральной сетью для анализа межкадровых зависимостей. Рассматриваются варианты 1D сверточной сети с дырявыми регулярными свертками, 1D сверточной сети с дырявыми причинно-следственными свертками, рекуррентная сеть LSTM и рекуррентная сеть GRU. Сравнение эффективности моделей производится на основе микро усредненной F1-меры, которая вычисляется на тестовом наборе данных.

Результаты. Результаты, полученные на наборе данных, предоставленным Ace Pipe Cleaning, Inc, подтверждают пригодность модели и метода для практического использования, так как полученная точность равна 92%. Сравнение результатов обучения за предложенным методом с результатами за традиционным методом показало преимущество на 4% за микро-усредненным значением F1-меры. Дальнейший анализ матрицы ошибок показал, что наиболее существенное

повышение точности по сравнению с традиционными методами достигается для сложных классов, которые объединяют как ориентацию камеры, так и особенности конструкции сточной трубы.

Выводы. Научная новизна работы заключается в новых модели и методе классификационного анализа пространственно-временного контекста для автоматизации видеоинспекции сточных труб в условиях ограниченного объема и несбалансированности размеченных обучающих данных. Результаты обучения, полученные по предлагаемому методу, сравниваются с результатами, полученными с помощью традиционного метода классификационного анализа изображений. Предложенный метод продемонстрировал преимущество на 4% за микро-усредненным значением F1-меры.

Эмпирически было доказано, что темпоральная сеть на основе 1D сверточной сети с дырявыми регулярными свертками является наиболее эффективной для анализа межкадровых зависимостей. Полученная точность обеспечивает пригодность полученных моделей для практического использования, поскольку дополнительное исправление ошибок можно реализовать на основе данных одометра.

КЛЮЧЕВЫЕ СЛОВА: инспекция сточных труб, сверточная нейронная сеть, коды с самокоррекцией ошибок, сиамская нейронная сеть, информационно-экстремальное обучение, информационный критерий, LSTM, GRU.

ЛІТЕРАТУРА / LITERATURA

1. Defect Detection and Condition Assessment / S. Moradi, T. Zayed, F. Golkhoo // *Infrastructures*. – 2019. – Vol. 4, No. 1: 10. DOI: 10.3390/infrastructures4010010.
2. Myrans J. Automated detection of fault types in CCTV sewer surveys / J. Myrans, R. Everson, Z. Kapelan // *Journal of Hydroinformatics*. – 2018. – Vol. 21, No. 1. – P. 153–163. DOI: 10.2166/hydro.2018.073.
3. A review of monocular visual odometry / [M. He, Ch. Zhu, Q. Huang et al.] // *The Visual Computer*. – 2020. – Vol. 36, No. 2. – P. 1053–1065. DOI: 10.1007/s00371-019-01714-6.
4. Lim B. Time-series forecasting with deep learning: a survey / B. Lim, S. Zohren // *Philosophical Transactions of the Royal Society A*. – 2021. – Vol. 379, Issue 2194. – P. 14. DOI: 10.1098/rsta.2020.0209.
5. Syahrian N. M. Vision-Based Pipe Monitoring Robot for Crack Detection Using Canny Edge Detection Method as an Image Processing Technique / N. M. Syahrian, P. Risma, T. Dewi // *Kinetik*. – 2017. – Vol. 2, No. 4. – P. 243–250. DOI: 10.22219/kinetik.v2i4.243.
6. Visual-Based Defect Detection and Classification Approaches for Industrial Applications – A SURVEY / [T. Czimmermann, G. Ciuti, M. Milazzo et al.] // *Sensors*. – 2020. – Vol. 20, No. 5: 1459. DOI: 10.3390/s20051459.
7. Cheng J. C. P. Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques / J. C. P. Cheng, M. Wang // *Automation in Construction*. – 2018. – Vol. 95. – P. 155–171. DOI: 10.1016/j.autcon.2018.08.006.
8. Deep learning and image processing for automated crack detection and defect measurement in underground structures / [F. Panella, J. Boehm, Y. Loo et al.] // *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. – 2018. – Vol. XLII-2. – P. 829–835. DOI: 10.5194/isprs-archives-xlii-2-829-2018.
9. DeepShoe: An improved Multi-Task View-invariant CNN for street-to-shop shoe retrieval / [H. Zhan, B. Shi, L.-Y. Duan et al.] // *Computer Vision and Image Understanding*. – 2019. – Vol. 180. – P. 23–33. DOI: 10.1016/j.cviu.2019.01.001.
10. Challenging the Adversarial Robustness of DNNs Based on Error-Correcting Output Codes / [B. Zhang, B. Tondi, X. Lv et al.] // *Security and Communication Networks*. – 2020. – Vol. 2020: 8882494. DOI: 10.1155/2020/8882494.
11. The Model and Training Algorithm of Compact Drone Autonomous Visual Navigation System / [V. Moskalenko, A. Moskalenko, A. Korobov et al.] // *Data*. – 2019. – Vol. 4, No. 1: 4. DOI: 10.3390/data4010004.

PROBLEM OF A DISCRETE DATA ARRAY APPROXIMATION BY A SET OF ELEMENTARY GEOMETRIC ALGORITHMS

Povkhan I. F. – Dr. Sc., Assistant Professor, Assistant Professor at the System Software Department, State Higher Education Institution Uzhhorod National University, Uzhhorod, Ukraine.

Mitsa O. V. – Dr. Sc., Assistant Professor, Head of the Information Control Systems and Technologies Department, State Higher Education Institution Uzhhorod National University, Uzhhorod, Ukraine.

Mulesa O. Y. – Dr. Sc., Assistant Professor, Assistant Professor at the Cybernetics and Applied Mathematics, State Higher Education Institution Uzhhorod National University, Uzhhorod, Ukraine.

Melnyk O. O. – PhD, Assistant Professor, Assistant Professor at the System Software Department, State Higher Education Institution Uzhhorod National University, Uzhhorod, Ukraine.

ABSTRACT

Context. In this paper, a problem of a discrete data array approximation by a set of elementary geometric algorithms and a recognition model representation in a form of algorithmic classification tree has been solved. The object of the present study is a concept of a classification tree in a form of an algorithm trees. The subject of this study are the relevant models, methods, algorithms and schemes of different classification tree construction.

Objective. The goal of this work is to create a simple and efficient method and algorithmic scheme of building the tree-like recognition and classification models on the basis of the algorithm trees for training selections of large-volume discrete information characterized by a modular structure of independent recognition algorithms assessed in accordance with the initial training selection data for a wide class of applied tasks.

Method. A scheme of classification tree (algorithm tree) synthesis has been suggested being based on the data array approximation by a set of elementary geometric algorithms that constructs a tree-like structure (the ACT model) for a preset initial training selection of arbitrary size. The latter consists of a set of autonomous classification/recognition algorithms assessed at each step of the ACT construction according to the initial selection. A method of the algorithmic classification tree construction has been developed with the basic idea of step-by-step arbitrary-volume and structure initial selection approximation by a set of elementary geometric classification algorithms. When forming a current algorithm tree vertex, node and generalized attribute, this method provides alignment of the most effective and high-quality elementary classification algorithms from the initial set and complete construction of only those paths in the ACT structure, where the most of classification errors occur. The scheme of synthesizing the resulting classification tree and the ACT model developed allows one to reduce considerably the tree size and complexity. The ACT construction structural complexity is being assessed on the basis of a number of transitions, vertices and tiers of the ACT structure that allows the quality of its further analysis to be increased, the efficient decomposition mechanism to be provided and the ACT structure to be built in conditions of fixed limitation sets. The algorithm tree synthesis method allows one to construct different-type tree-like recognition models with various sets of elementary classifiers at the preset accuracy for a wide class of artificial intelligence theory problems.

Results. The method of discrete training selection approximation by a set of elementary geometric algorithms developed and presented in this work has received program realization and was studied and compared with those of logical tree classification on the basis of elementary attribute selection for solving the real geological data recognition problem.

Conclusions. Both general analysis and experiments carried out in this work confirmed capability of developed mechanism of constructing the algorithm tree structures and demonstrate possibility of its promising use for solving a wide spectrum of applied recognition and classification problems. The outlooks of the further studies and approbations might be related to creating the other-type algorithmic classification tree methods with other initial sets of elementary classifiers, optimizing its program realizations, as well experimental studying this method for a wider circle of applied problems.

KEYWORDS: algorithmic classification tree, image recognition, classification, classification algorithm, branching criterion, geometric algorithm.

ABBREVIATIONS

TS is a training selection;
ST is a test selection;
RS is a recognition system;
IR is an image recognition;
GA is a generalized attribute;
RF is a recognition function;
LCT is a logical classification tree;
ACT is an algorithmic classification tree;
GAT is a generalized attribute tree;
BAS is a branched attribute selection.

NOMENCLATURE

M^n is a manifold of real vectors of dimensionality n ;
 n is a total number of the problem attributes (the attribute space dimensionality);
 w_i is a discrete object of the initial TS;
 u is a value of a class of discrete object w belonging;
 m is a total number of training pairs (known classification objects) of the initial TS;
 l is a value of a class of discrete object x belonging;
 f_R is a recognition function;

x_S^i is a vector (point) of a space set by the problem condition;

H_i is a set of classes set by the initial TS;

G_i is a set of manifolds of relevant objects w_j ;

P_i is a constructed generalized attribute;

A_i, A_j is a pair of arbitrary vectors with M^n ;

a_j, b is a pair of arbitrary real numbers;

c_0, c_1 is a center of mass of the classes H_0 and H_1 ;

n_0, n_1 is a capacities (masses) of the classes H_0 and H_1 ;

z is an arbitrary point of the classes H_0 and H_1 partition plane;

r is a fixed hemisphere radius in the attribute space of the problem;

S is a vector of the classes H_0 and H_1 partition;

ξ is a vicinity of a fixed point with M^n ;

ρ is a constructed RS efficiency;

K_0, U is a manifold of all the class H_0 hemispheres;

K_1, Y is a manifold of all the class H_1 hemispheres;

k_i, l_i is a hemisphere sequence;

$d_{\min}(w_i)$ is a radius of a hemisphere with a center at the point w_i ;

$d(w_0, w_1)$ is a distance between the points w_0 and w_1 ;

d_0 is a minimal distance to a neighboring class;

p_i is a capacity of the relevant GA in a form of a hyperellipse;

E is a GA manifold in a form of elementary hyperellipses;

S_i^j is a GA set in a form of hyperplanes;

P_i^j is a GA set in a form of hyperparallelepipeds;

K_i^j is a GA set in form of hyperspheres;

q is a total number of GA in the algorithm tree.

INTRODUCTION

An important problem that often faces the engineer is the task of automatic building the systems for processing large arrays of information and decision-making systems. Effective solving these problems will allow one to pass to computer a difficult work of designing a complex recognition system and to release the creative engineer's potential to solve other, more important, problems. In addition, automation of algorithmic and software design of specific recognition systems is the key to their high efficiency for each real task, and, therefore, it will ensure the rapid development of various branches of science and technology [1–5].

As of today, more than 3.500 recognition algorithms (based on different approaches and concepts) are known
© Povkhan I. F., Mitsa O. V., Mulesa O. Y., Melnyk O. O., 2021
DOI 10.15588/1607-3274-2021-3-10

having certain limitations when being used (i.e. accuracy, speed, versatility, reliability etc). In addition, each of the algorithms is limited to a specific application, and this is certainly the weakest point not only of these algorithms, but also of recognition systems based on the relevant concepts. Information technologies based on the mathematical models of image recognition are widely used in socio-economic and environmental information processing systems. This is due to the fact that this approach allows one to eliminate the shortcomings of classical methods and achieve fundamentally new results, rationally using the capacity of computer systems [3].

The most of available methods of training selection processing in the recognition function construction do not allow one to achieve the required level of the recognition system accuracy and adjust their complexity in the process of designing these systems. The methods of constructing recognition systems based on the classification tree methods [2] are free of such shortcoming. The peculiarity of the logical tree method is the possibility of complex use of many known recognition algorithms (methods) to solve each specific problem of constructing a recognition scheme. It is based on a single methodology – the optimal approximation of the training selection by a set of generalized attributes (autonomous algorithms) included in some scheme (operator), built in the course of training process.

The **object of study** is a general concept of the decision tree, namely, the algorithmic classification tree constructions built on the basis of a scheme of approximating the training selection array by a set of elementary classification algorithms.

The decision tree concept [6–8] and its relevant library realizations (LightGBM, XGBoost), though being close in idea (the logic tree scheme), do not allow realization of the concept of algorithmic classification tree consisting of a set of vertices – the different-type autonomous classification algorithms. One should take into account that the generalized attribute set tree is, in fact, the algorithmic tree reflection.

The **subject of studies** includes the methods, algorithms and schemes of constructing the classification algorithmic trees in the image recognition tasks.

Presented study allows overcoming these system limitations of constructing the recognition systems by synthesizing the algorithmic trees of classification. Their main specific feature is the modular tree structure and the possibility of applying an arbitrary recognition algorithm or method in the process of synthesizing the classification scheme [9–11].

The **objective of the work** is to elaborate the method of the RS classification model construction on the basis of the scheme of training selection array approximation by a set of elementary classification algorithms. Note that the classification system schemes obtained are characterized by a tree-like structure and the presence of autonomous classification algorithms as their own structural elements.

1 PROBLEM STATEMENT

In this paper, we shall consider algorithmic implementations of a generalized attribute (algorithm) formation in the algorithmic tree method, which is used in the finished software system and is based on the geometric separation of images using geometric objects. It should be noted that for an arbitrary geometric algorithm or recognition method the main role is played by the distance between the objects in the space for which the recognition problem is solved.

Let M^n be a manifold n of dimensional real vectors (i.e. we deal with the n -dimensional space). Under images we mean the system of subsets (classes) H_1, \dots, H_k in the manifold M^n . The training selection, in turn, is set by a sequence:

$$(w_1, f_R(w_1), \dots, w_m, f_R(w_m)). \quad (1)$$

Here $w_i \in M^n, f_R(w_j) \in \{1, 2, \dots, k\} (i=1, \dots, m)$, and if $f_R(w_i) = u, u \in \{1, \dots, k\}$, then the object w_i belongs to the image H_u . The objects w_i are, in fact, the vectors (points) $x_1^i, x_2^i, \dots, x_S^i$ set in this space.

To simplify explanation of the principal idea, let us assume that $k \in \{0, 1\}$, i.e. the objects may belong to the two classes H_0 or H_1 (the binary classification case) only. Furthermore, we shall assume that a certain metrics is set at the space M^n , i.e. this space is metrical. We shall denote some distance between the vectors A_i and A_j in this space as $\|A_i - A_j\|$. It is evident that the way of choosing the distance between them has several solutions, for instance:

$$\begin{aligned} \text{a) } \|A_i - A_j\| &= \sqrt{\sum_{m=1}^S (x_i^m - x_j^m)^2}; \\ \text{b) } \|A_i - A_j\| &= \sum_{m=1}^S |x_i^m - x_j^m|; \\ \text{c) } \|A_i - A_j\| &= \max_m |x_i^m - x_j^m|. \end{aligned} \quad (2)$$

Note that here $A_i = x_1^i, \dots, x_S^i, A_j = x_1^j, \dots, x_S^j$. For example, the hyperplane algorithm operation results in one or more generalized attributes P_j , which, in fact, are the parameters of some hyperplanes that allow all the classes H_i of the manifold M^n to be partitioned. A quite large number of such partitioning (hyperplane constructing) ways are available here.

Thus, this work faces a problem of constructing the classification tree model with parameters p and structure L that should be optimal $F(L(p, w_i), f_R(w_i)) \rightarrow opt$ with respect to the initial TS data.

2 REVIEW OF THE LITERATURE

This study continues a cycle of works devoted to the problems of discrete object recognition (classification) tree-like schemes [9–15]. They raise issues of constructing, using and optimizing logical trees. Thus, it is known from [16] that the resulting classification rule (scheme) constructed by an arbitrary method or algorithm of branched attribute selection has a tree-like logical structure. The logical tree consists of vertices (attributes) grouped by tiers and obtained at a certain step (stage) of the recognition tree construction [17]. An important problem that arises from [10] is that of synthesizing recognition trees, which will be actually represented by the algorithm tree (graph).

In contrast to existing methods, the main feature of tree recognition systems is that the importance of individual attributes (groups of attributes or algorithms) is determined with respect to the function that defines the objects partition into classes [18, 19]. Please keep in mind that the numerical value of the specified importance characterizes the error of the objects partition into classes. Suppose that at the first step of recognition tree constructing, an arbitrary recognition algorithm is used, and, as a result, we obtain some formula (generalized attribute). This formula implements a certain level of recognition. The function takes several values depending on the attribute values. These values characterize the paths (classes), and there are ways in which this formula “works well”, there are those in which it “works bad”, and there is no improvement in the level of recognition. It is clear that for these attribute values (paths) one has to take other algorithm that will create other formula (generalized attribute), etc. Thus, in recognition methods based on (algorithmic) trees, it is necessary to repeat such algorithm selection until we obtain the required level of recognition quality [11].

A basic issue of selecting branching criterion in the classification tree scheme structure, whose selection may be affected by the problem specificity, was raised in [20]. The issues of generation and interpretation of classification rules in the LCT structures are raised in [14]. The problem of assessing the attribute informativity [15] in constructing the classification tree vertices remains topical and requires further studying the direction of pre-processing and analyzing the initial data structures [21]. Moreover, the presence of limitations in terms of generating the LCT structure is the disadvantage of the functionalities of assessing the attribute quality in the above works. Thus, the work [8] opens fundamental questions regarding decision tree generation for the case of uninformative attributes. A possible way to improve this work might be the use of attribute combinations and sets to generate informative vertices of the LCT structures. The ability of the LCT/ACT (i.e. logical and algorithmic classification trees) structures to perform one-dimensional branching and analyze individual variable impact, importance and quality allows one to work with different-type variables in the form of predicates.

For the case of the ACT models, the question of assessing the quality of the corresponding structure branches, i.e. of the autonomous classification algorithms, remains relevant [10]. Here the search for effective criteria for branching tree structures is the way of improving the ACT structure methods. This concept of logical trees is actively used in the intellectual data analysis, where the ultimate goal is to synthesize a model that predicts the value of the target variable based on a set of initial data at the system input. In [22], an important issue of analysis of the classification quality of the decision tree sets was raised. A possible way to improve the overall quality of classification is the use of decision tree ensembles, bug-ging and boosting mechanisms [23–25]. Note that these schemes will provide the classification model with necessary accuracy only if there is an effective branching criterion [24]. In this case, as a test of the constructed LCT models, it is appropriate to use the cross-validation scheme taken from [26].

The disadvantage of this approach is the final complexity of the classification models and the need in the procedure of final model cutting-off. In regards to application, there are a large number of methods and algorithms that implement the decision tree concept, but the most widely used and widespread are their two representatives (i.e. the C4.5/C5.0 Ross Quinlan scheme and the CART (Classification and Regression Tree) scheme). The C4.5/C5.0 scheme uses the so-called theoretical-information criterion as a criterion for selecting a node, while the CART algorithm is based on calculating the Gini index (a statistical indicator of the attribute difference), which takes into account the relative distances between class distributions within metrics. The main examples, parameters and mechanisms of this classification tree scheme can be developed from open resources [27–31]. The shortcoming here is the relatively weak efficiency in terms of vertex selection in comparison with other modern methods and schemes of the LCT structures. The main idea of methods and algorithms of branched selection of ACT algorithm attributes and vertices, in contrary to the neural network approach [32–34], can be defined as the optimal approximation of some initial TS by a set of ranked classification algorithms. Then the central question from [11] comes to the fore, i.e. a problem of choosing an effective branching criterion, selecting vertices, attributes and features of discrete objects for the LCT schemes and choosing algorithms for ACT.

3 MATERIALS AND METHODS

At the next stage of the study, for the case of binary classification, we shall construct such a generalized attribute (a particular hyperplane) that would separate these classes in the most efficient way. One of the easiest ways of such construction would be:

$$P(x_1^i, x_2^i, \dots, x_s^i) = \begin{cases} 1, & \text{if } \sum_{j=1}^S a_j x_j^i + b \geq 0; \\ 0, & \text{if } \sum_{j=1}^S a_j x_j^i + b < 0. \end{cases}$$

$$P(x_1^i, x_2^i, \dots, x_s^i) = \begin{cases} 1, & \text{if } \sum_{j=1}^S (x_j^i - a_j)^2 - b^2 \leq 0; \\ 0, & \text{if } \sum_{j=1}^S (x_j^i - a_j)^2 - b^2 > 0. \end{cases}$$

Note that here a_j, b are the arbitrary real numbers.

Also, it is clear that a simple and effective division of the classes H_0, H_1 is possible only if the hypothesis of compactness in the space of these classes is valid. The latter means that all points of each of these classes are at a relatively small distance from each other, i.e. there are such points (centers of mass of these classes), around which they are grouped. Obviously, in this case it is appropriate to choose the arithmetic mean of the submanifolds H_0, H_1 as the above centers. Thus, when constructing the GA, it is sometimes important to take into account the capacity of each of the submanifolds. We shall mean by the capacity (mass) of a manifold the number of pairs of objects from the training selection of type (1), for which relation $f_R(w_i) = j$ holds true:

$$c_0 = \frac{\sum_{w_i \in H_0} w_i}{n_0}, c_1 = \frac{\sum_{w_i \in H_1} w_i}{n_1}. \quad (3)$$

Just the points $c_0 = (c_1^0, \dots, c_s^0)$ and $c_1 = (c_1^1, \dots, c_s^1)$ shall be the centers of mass of the classes H_0 and H_1 , while the numbers n_0 and n_1 will be their capacities (masses). Since n_0, n_1 have to influence directly P (i.e. the hyperplane location in space), it seems reasonable to assume as follows:

$$\frac{s - c_0}{c_1 - s} = \frac{n_0}{n_1}. \quad (4)$$

It follows directly from (4) that:

$$s = \frac{n_1 c_0 + n_0 c_1}{n_0 + n_1}. \quad (5)$$

In accordance with the training selection data, the centers of mass of the classes H_0, H_1 and their capacities

n_0, n_1 , as well as the point S , are calculated. This point S will be a result of selection of the most successful division of the classes H_0, H_1 . Such selection will be the better, the wider are the relevant classes separated from each other in space. After constructing the attribute P , all the points w not exceeding S belong to the class H_0 , whereas all the remaining points belong to the class H_1 .

In view of all mentioned above, we will step-by-step describe the scheme of division of classes by means of a hyperplane. Note only that in our case the training process is carried out at once for all objects of the initial selection, but this assumption is not a limitation.

At the initial stage, based on the training selection data, calculation of the centers of mass and capacity of the classes H_0, H_1 is carried out. Note that these two operations can be done with a single processing of the training selection. The next step is to find the point of division between the centers of mass of the classes, through which the desired plane will pass separating them. Therefore, first the distance between the centers of mass of the classes is calculated depending on the type of the distance determination in space, which was fixed at the beginning of the algorithm. At the second stage, depending on the capacity of the classes H_0, H_1 , location of the point S that lies on the line connecting them is to be found. We draw a plane through a certain point S normally to the line connecting the centers of mass of two opposite classes. The equation of this plane will look like:

$$(z - s) \cdot (c_1 - c_0) = 0. \quad (6)$$

Note that z denotes here an arbitrary point of the plane (6), and we deal with a scalar product of the vectors $x \cdot y$. Then the following attribute could be taken as the generalized attribute P :

$$P(z) = \begin{cases} 1, & \text{if } (z - s) \cdot (c_1 - c_0) \geq 0; \\ 0, & \text{if } (z - s) \cdot (c_1 - c_0) < 0. \end{cases} \quad (7)$$

If $P(z) = 1$, then z should be related to the class H_1 , while if $P(z) = 0$, z should be related to the class H_0 . Thus, in this case the classes H_0 and H_1 will be separated by a plane (6) that crosses a straight line, which connects the centers of mass of the above classes at the point S and, in addition, is orthogonal to this line.

After the generalized attribute P is built, it should be checked. Such check of the generalized attribute P shall include repeated checking all the objects of the training selection for their recognition correctness. One essential peculiarity should be noted here, namely, the errors of two categories are possible in testing the attribute P during the recognition tree method operation, when the attribute P will be only one of many constructed generalized attributes. The first category of errors is not critical, when the dividing plane due to not fully correct

construction will fail to cover a part of objects belonging to one of classes (it is not critical, because 'not covered' part of objects will be passed for processing to another recognition algorithm). The second error category, a critical one, will lead to impossibility to build the dividing plane and occurs in case the above plane covers objects of other class.

The algorithmic implementation of both training and recognition suggested above can be applied to various class regions, the full separation of which is achieved especially for a set of objects of class selection from non-intersecting space regions. Such sets of objects (points) can be completely separated from those of points from any other class, if their regions in space do not intersect. Note that the main disadvantage of this implementation is that, in general, separation of the class regions is, as a rule, not optimal and depends directly on the efficiency of the choice of the attributes. The training selection informativity and adequacy are also important [9, 15].

Consider the local algorithm of generating attributes in a form of hyperspheres. The most efficient and universal geometric algorithm is the hypersphere one. The scheme of this algorithm, which will be described below, in order to form generalized attributes (the GA sets), builds at each step a hypersphere of the following form:

$$\sqrt{\sum_{i=1}^n (x_i - a_i)^2} \leq r. \quad \text{The algorithm of training selection}$$

approximation by hyperspheres is very similar, in general, to the hyperellipse algorithm (and is its further conceptual development), where the attribute is described by

$$\sum_{i=1}^n c_i (x - x_i)^2 \leq r. \quad \text{Its high versatility, in contrary}$$

to other algorithms, is based on the fact that even if the compactness hypothesis is not fulfilled, it is possible to construct a sequence of attributes that would separate objects of one class from those of the other one. Note that the only condition for this is prohibition of overlapping objects of different classes. The principle of operation of this algorithm is to approximate the region of the corresponding class by a set of hyperspheres that would cover all its objects (points). Moreover, the main attention should be paid not so much to the description of the class region by hyperspheres, as to the efficiency and cost-effectiveness of this description. That is, the process of approximation of an arbitrary class should take place with the construction of a minimum number of hyperspheres (generalized attributes).

Note that in order to store in the computer memory some hyperspheres that approximate a certain region, it is enough for each of them to memorize the coordinates of the hypersphere center and radius. If we give a certain parameter (radius), each point in the space (training selection element) will become a hypersphere. Let us describe a simplified scheme of the hypersphere algorithm and consider the following possible assumptions:

a) there are no object with M^n in the ξ -vicinity of each point w_i of the initial selection;

b) there are objects with M^n in the ξ -vicinity of point w_i , but all of them belong to the same class together with point w_i ;

c) there are objects with M^n in the ξ -vicinity of point w_i , but they may belong to different classes.

Assumption (a) is not entirely preferable, as it is unlikely that some approximation by a certain sequence of geometric figures will be better than the training selection itself. Assumption (b) will be the closest to reality, however, in this case of the ξ -vicinity of the point w_i , we know information about the point w_i itself only. From the previous statements we can draw the following conclusion: each point (object) w_i belongs to a certain class H_j together with the ξ -vicinity.

The initial stage of the hypersphere algorithm deals with processing the data of the training selection in order to calculate the center of mass of the class H_0 . When processing the data of the training selection of the form (1), for each point (vector) w_i , for which the relation $f_R(w_i) = 0$ is valid (it is assumed that the object belongs to the neighboring class H_0), the distance to the neighboring class H_1 is calculated in parallel, being determined by the ratio:

$$d_{\min}(w_i) = \min_{w_j \in H_1} (w_i - w_j). \quad (8)$$

The quantity $d_{\min}(w_i)$ is used to determine the radius of a hyper sphere centered at the point w_i . The radius $r(w_i)$ will be equal to $d_{\min}(w_i)/2$ and results from the following considerations: dividing the distance to the neighboring class in half is necessary to ensure that the hyperspheres with centers at the points $w_i (w_i \in H_0)$ and $w_j (w_j \in H_1)$ do not intersect. This, in turn, may lead to uncertainty in the classification process.

After actual determining the radii of the hyperspheres describing the class H_0 , one may construct a certain manifold K_0 and write to it all the hyperspheres of the class H_0 . Note that their number will be equal to that of objects in the training selection for which equality $f_R(w_i) = 0, (i = 1, \dots, m)$ will hold true. Note that we have fulfilled the first task (i.e. description of the class H_0 region with the help of hyperspheres), although it is clear that this approximation will not be optimal.

This stage of the algorithm operation relates to finding the most efficient and cost-effective description of the hyperspheres of the class H_0 region. To do this, for each

of the elements of the manifold K_0 (hyperspheres w_i with radius $d_{\min}(w_i)/2$) we calculate the quantity $m(w_i)$ equal to the number of all w_j for which $f_R(w_i) = 0, (i = 1, \dots, m)$ and the following relation holds true:

$$r(w_i) < d(w_i, w_j). \quad (9)$$

Here $d(w_i, w_j)$ determines the distance in space between the objects w_i and w_j . Therefore, it is clear that the quantity $m(w_i)$ characterizes the number of points in the space of the class H_0 , which will be described by a hypersphere centered at w_i . Let's call this quantity the hypersphere capacity. After calculating the hypersphere capacities from the manifold K_0 , one has to select (and remove) that with the largest capacity, i.e. k_1 , and place it into the manifold U , which will contain the hyperspheres of the optimal H_0 class approximation. The manifold K_0 is also subject to additional processing, i.e. removing from it all the elements (hyperspheres), the centers of which fall under the hypersphere k_1 region, i.e. meet the condition $r(k_1) < d(w_i, w_j)$.

After constructing the first element of the sequence of approximating hyperspheres of the manifold U , this construction scheme must be repeated, returning to the second stage. The process will continue until the desired sequence of hyperspheres is constructed that fully describes the class H_0 . Note that even in the worst case of the location of the classes H_0, H_1 in the space the generalized attribute, nevertheless, will be constructed, and only the number of hyperspheres will be equal to the number of objects w_i from the training selection that belong to H_0 . After describing the region of the class H_0 (hyperspheres belong to the manifold $U(k_1, k_2, \dots)$), a sequence of approximating hyperspheres for the next class H_1 is constructed according to the similar scheme (result is $Y(l_1, l_2, \dots)$).

The process of recognizing new objects w_i that enter the input of the classification system will proceed via the following scheme: first the hyperspheres of the manifold U will be checked, i.e., if $|k_j - w_i| \leq r(k_j)$, the object w_i belongs to the class H_0 . Then the hyperspheres of the manifold Y will be checked, and if $|l_j - w_i| \leq r(l_j)$, then the object w_i belongs to the class H_1 . However, there is a third option, when for a certain w_i none of the previous conditions will be met, then this object will be denied classification. The reason is not getting into any of the

class H_0, H_1 regions. The restriction we imposed at the beginning of the description of the hypersphere algorithm scheme (i.e. that the training selection specifies partitioning into two classes H_0, H_1) is not mandatory. An arbitrary selection of the objects w_i that define partitioning into an arbitrary number of classes H_i can be reduced to a training selection of two classes, for example, partitioning into H_0 and into all other classes, i.e. H_1, H_2, \dots, H_i . After finding the approximating sequence for H_0 , the training selection is again partitioned into H_1 and H_2, H_3, \dots, H_i , and so on.

The main disadvantage of the above scheme of approximation by the sampling hyperspheres is the significant amount of calculations that appear with a dramatic increase of the training selection volume. First of all, this is due again to the constant cyclic processing of the selection to build each subsequent hypersphere, that is, if the desired sequence of hyperspheres consists of ten elements, the selection will be processed the same number of times. There is only one way to overcome this problem, i.e. to find a sequence of hyperspheres that approximates the working class with a less number of elements. This can be achieved after making structural changes to the scheme of the hypersphere algorithm.

To do this, at the very beginning of the algorithm operation the center of mass of the class c_0 according to scheme (3) and the value $d_{\min}(c_0)$ of the distance to the neighboring class (8) are calculated for the class H_0 . After performing the above operations, we obtain the first desired hypersphere with a center at the point c_0 of the radius $r(c_0) = d_{\min}(c_0)/2$. Of course, having calculated its capacity and corrected the training selection, one can see that the selection size is significantly reduced, i.e. most of working class elements will be excluded from it. Next, the scheme of operation of the algorithm for constructing a sequence of hyperspheres will coincide with that previously described, but the number of steps will be significantly reduced due to the first, successfully constructed, element of the sequence sought.

Consider the local algorithm for generating attributes in the form of hyperellipses. In the previous local algorithm, some hypersphere was chosen at each step to form attributes, but it is clear that other spatial geometric objects, such as hyperellipses or hyperparallelepipeds, can also be used as attributes (see discussion below). Let us describe the essence of the algorithmic implementation of training selection approximation by hyperellipse, which is a further development of the concept of space approximation by hyperspheres and in some cases allows a simpler and more cost-saving description of the class region to be achieved.

Let us assume again that a known classification training selection with the objects w_i of type (1) is set in the n -dimensional space. To simplify the algorithm step

explanation, we assume that $k = \{0,1\}$, i.e. the objects may belong to the two classes, i.e. H_0 or H_1 only. In addition, we shall assume that a certain metrics (2) is set in this space.

The initial stage will be the primary processing of the training sample and the construction of two manifolds G_0 and G_1 , which will contain the objects of selection of the corresponding classes H_0, H_1 , i.e. $w_j \in G_i$, if $w_j \in H_i$.

Such distribution of the training selection objects is necessary for the further construction of hyperellipses and calculation of the minimum distance to the next class. Having fixed, for example, a manifold H_0 , we shall calculate for each object w_i the following number d_{\min} that belongs to it:

$$d_{\min}(w_i) = \min_{\substack{w_0 \in H_0 \\ w_1 \in H_1}} d(w_0, w_1). \quad (10)$$

Here $d(w_0, w_1)$ is a distance between the points w_0 and w_1 . In fact, the quantity $d_{\min}(w_i)$ shall fix for each object w_i with H_0 a minimal distance to the neighboring class H_1 (i.e. to the nearest object of this class).

Recall that, by definition, an ellipse is the geometric location of points, the sum of the distances from which to two fixed points (foci) is a constant value. It is clear that it is natural to take pairs of objects (vectors, points) from the set H_0 as the above foci.

After calculating the measure of distance between each of the objects of the class H_0 and the neighboring class H_1 , one has to sort and select all the pairs of points (w_i, w_j) from this class, for which the following inequality will be valid:

$$d(w_i, a) + d(w_j, a) < d_0(w_i, w_j). \quad (11)$$

Here $(a \in H_1)$, and $d_0(w_i, w_j) = \min_{a \in H_1} ((w_i, w_j), a)$.

In fact, this inequality expresses the condition of the ellipse. The quantity d_0 characterizes the minimum distance to the next class. All the pairs of points (w_i, w_j) found will be selected in a special manifold E , having additionally introduced the quantity $r(w_i, w_j)$, i.e. the radius of the hyperellipse, which will be equal to $d_0(w_i, w_j)/2$. Its division in half is carried out for the reasons similar to the reasoning of the division of the hypersphere radius in the previous algorithm. Note only that all pairs of points (w_i, w_j) found must satisfy the condition of the hyperellipse (11), in the right part of which should be the value of its radius.

After performing the second stage, we obtain a manifold E that will contain all possible hyperellipses

actually describing the class H_0 . An important point here is that none of them will contain points (objects) of another class H_1 , but this approximation will not be complete. That is, in contrast to the hypersphere algorithm, in our case there may be points (objects) that will not lie in the region of any of the hyperellipses of the manifold E . It is clear that to a large extent this will be the result of the complex arrangement of classes (images) in space.

After the previous actions, we obtain an approximation by the sequence of hyperellipses of the manifold E from a given training selection. After that, one has only to select the most effective description of the class, i.e. to find among the elements of the manifold E those that would provide the most optimal description of the class H_0 . To ensure this condition fulfillment, it is necessary to calculate for each hyperellipse its capacity p_i . This value for each hyperellipse will show the number of objects with H_0 , the distance to which is less than or equal to its radius.

It is clear that, knowing the capacities of each of the hyperellipse of the manifold E , we can already select that one for which this value will be the largest, and therefore, carry out a partial approximation. Note only that, in the terminology of the logical tree methods, hyperellipse will already be some generalized attribute, but covering (unambiguously classifying) a certain part of the class H_0 only. That is, the process of construction (selection) of further hyperellipses must be continued. To do this, one has to remove from the class H_0 those objects w_i that fall within the classification region of the first generalized attribute found (it is clear that their number is equal to the hyperellipse capacity). After that, one may proceed to the next step, i.e. to select the next best hyperellipse (generalized attribute), for which it is necessary to return to the first stage of the algorithm.

This sequence of actions will be carried out until such a sequence of hyperellipses is generated that completely approximates the class H_0 (the number of steps of the algorithm will coincide with their number). However, a second option is possible, when such elements will belong to the class H_0 from which it is impossible to construct such hyperellipses that will not cover the objects of the neighboring class H_1 . The appearance of such objects is explained by the complex arrangement of classes (images) in space; and will result in classification failures that require some other recognition algorithm to recognize.

It is clear that after finding the sequence E of hyperellipses that will cover the class H_0 , the hyperellipses are constructed in a similar manner to approximate all other classes of training selection (H_1 in our case). The process of recognizing new objects in the hyperellipse algorithm corresponds by its structure to the same stage in the previous hypersphere algorithm. That is, © Povkhan I. F., Mitsa O. V., Mulesa O. Y., Melnyk O. O., 2021
DOI 10.15588/1607-3274-2021-3-10

each w_i of the unknown classification that enters the input of the recognition system should be checked sequentially for compliance with the conditions of hyperellipse. The condition of the test will be to verify the fact that the distance between w_i and the centers of the hyperellipse should not exceed its radius. If w_i is absorbed by one of the hyperellipses, it will be assigned to a class that approximates this hyperellipse. If none of the conditions are met, an object of unknown classification (failure) will be generated for the object w_i .

The weakest point in such an algorithm scheme in terms of time consumption is the construction of all possible hyperellipses, which will simply be reduced to the usual search. It is clear that in many cases a significant part of them will be rejected at the stage of verification either due to overlapping of the class area (incorrectness) or due to relatively insignificant capacity values (duplication of hyperellipses). A cardinal solution of this problem is to find and select the so-called "boundary" points and build a manifold E of hyperellipses based on them. Note only that the scheme of this process will be considered in the following geometric algorithm of hyperparallelepipeds, and can proceed in several stages.

It is clear that the relatively high cost of computing capacity with constant each-step (cyclic) processing of the training selection is also an important disadvantage of such an algorithm scheme for constructing hyperellipses. One of the possible solutions of this problem may be to work only with those hyperellipses that are obtained at the first step of the algorithm. The downside in this case is that there is a high probability of non-optimal or incomplete approximation of the training selection, first of all, this is due to the fact that the elements of the manifold E do not vary with each step. It should be emphasized that in the software implementation of real recognition problems, it is important to find the middle ground between the volume of calculations to elaborate the resulting classification rule. That is, the choice of, for example, the scheme of initial processing of the training selection should depend, first of all, on its volume.

Also note that since the hypersphere algorithm is a partial case of the hyperellipse algorithm, that is, if the hyperellipses turns into a hypersphere, then the number of hyperellipses will not exceed that of hyperspheres, which approximate this class based on the hypersphere algorithm. It does not follow from this that such an algorithm scheme will always lead us to the desired coverage of the class region (at least, by hyperspheres).

4 EXPERIMENTS

Note that the tree algorithm method uses one of the four simple geometric recognition algorithms as the attributes, i.e. the hypersphere algorithm, the hyperplane algorithm, the hyperellipse algorithm and the hyperparallelepiped algorithm. Their operation principle is to approximate the training selection by appropriate geometric objects [11]. The result of each of these algorithms is one or more generalized attribute(s)

(corresponding geometric objects), which describe a certain part of the training selection. Moreover, there may be the cases when the algorithm fails to construct a generalized attribute due to the complex arrangement of classes in the n -dimensional space. It is also possible that the constructed attributes do not approximate fully the selection (objects that do not fall under this approximated region are called classification failures).

Let a training selection of 2.000 elements be the only information about the nature of the manifold M partitioned into the classes H_i . Note that here we deal with the objects w_i described by three attributes and grouped into the four classes. So, the problem is to distribute objects of unknown classification over one of four classes. The presented selection contains the data of chemical analysis of the diesel fuel content (the task of assessing the quality of fuel) in a simplified version (the number of attributes is reduced from fifteen to three) to demonstrate the very concept of the algorithmic tree. Note that in this case there is no test sample, i.e. we will not assess the efficiency of the constructed system, but will only approximate the data with a set of generalized attributes (algorithms).

At the first stage, we shall assess the efficiency of each of the algorithms, on the basis of which the general classification scheme will be built, with respect to the initial training selection (Table 1).

Table 1 – Assessment of the efficiency of elementary geometric algorithms for classifying discrete objects with respect to the initial selection

Class number Algorithm type	Class 1	Class 2	Class 3	Class 4
Hypersphere algorithm	0/32	0/16	0/18	0/11
Hyperellipse algorithm	0/12	2/4	0/10	15/3
Hyperparallelepiped algorithm	0/6	1/5	0/7	9/6
Hyperplane algorithm	21/9	14/6	0/2	12/6

The cells present the efficiency of each of the algorithms with respect to the classes of the initial training selection. The first number is responsible for the quantity of objects that are denied in classification by the appropriate algorithm, and the second one is responsible for that of generalized attributes (geometric objects), which approximate the corresponding selection class. Depending on the initial choice of the algorithm as the recognition tree vertex, the process of constructing the resulting classification scheme can be completed with a different number of steps. Possible classification scheme is presented in (Fig. 1).

It can be seen from Table 1 that the efficiency of all the algorithms, except for the hyperplane one, with respect to the *Class 1* is 100%, so it can be applied to any algorithm (of course, except for the hyperplane algorithm). At any further stages of recognition scheme

construction, this algorithm is selected again, and it has been proven to be the most efficient and economical one with respect to all other classes of the initial selection. In addition, each of the generalized attributes generated by it represents the coordinates of the center of the hypersphere and its radius and requires a minimum amount of memory for its storage.

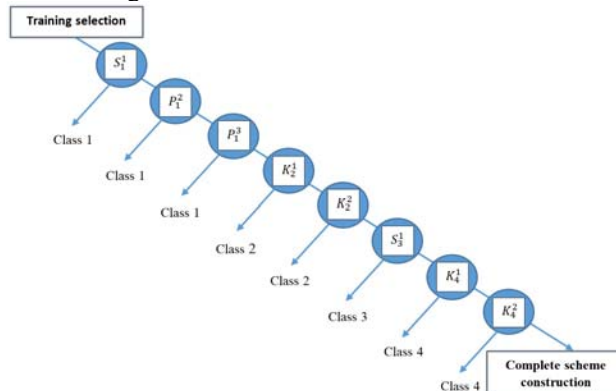


Figure 1 – Example of RS constructed

The recognition scheme (Fig. 1) is built of algorithms, the efficiency of which was evaluated with respect to the number of generalized attributes, which they use to describe the training selection. Thus, two algorithms were used to approximate the *Class 1*: first, the hyperplane algorithm had constructed the GA S_1^1 , which described it only partially. At the second stage, the hyperparallelepiped algorithm is applied, i.e. the attributes P_1^2 and P_1^3 that finally completed the recognition of this class. At the following stages of recognition the hypersphere (attributes $K_2^1, K_2^2, K_4^1, K_4^2$) and the hyperplane algorithms (attribute S_3^1) are again applied. Note that in order to build this scheme three different recognition algorithms are used not directly affecting each other's operation. That is, completely different in principle and ideology algorithms could be applied instead of them, allowing one to construct a recognition scheme with arbitrary complexity and efficiency. Only the efficiency of each of them for a fixed selection and the information capacity of the generalized attributes generated by them are important. That is, the tree method operates with ready-made (constructed) attributes only, and it may not be interested in what algorithm or method they were obtained.

This recognition scheme is constructed on the basis of a logical tree method and can be represented as a certain algorithmic scheme (operator) built by some algorithm to minimize or maximize the corresponding functionality, based on which the importance of the attribute, group of attributes or the efficiency of the recognition algorithm uniquely related to classification errors are assessed (it crosses the method of attribute branched selection).

Note that the tree method based on the input (training selection) data and the range of algorithms for generalized attributes stored in its library constructs (generates) a

certain scheme optimal by memory costs (complexity) and recognition efficiency (system). Under the scheme in this case we mean a set of numerical parameters for elementary attributes that best approximate the initial data array [14] (meaning the definition of the decisive scheme and the I. Vitenko's generalized attribute). Thus, in our case, the arguments of the constructed recognition scheme are the class attributes (hypercoils, hyperellips, etc.) or interclass attributes (hyperplanes). The parameters of the specified attributes and the general structure of the system (scheme) are stored in the computer memory.

Each of the schemes constructed by the tree method will be a general recognition system that can be used for practical work (processing large arrays of experimental data in the form of data arrays). Note also that the resulting scheme will be to some extent a new recognition algorithm (of course, synthesized from known algorithms and methods). In addition, for these classification systems it is not necessary to store in the computer memory the objects of selection on which it was constructed, i.e. large information arrays. The latter, in turn, leads to the fact that the process of constructing recognition system based on the tree method is largely similar to the process of information compression (meaning methods of information compression with losses) or encoding.

5 RESULTS

Thus, based on the classification tree method and modularity principle, Uzhhorod National University has developed a software package Orion III to generate autonomous recognition systems. The algorithmic library of this system has 11 recognition algorithms, among which are the geometric algorithmic implementations suggested above. This system allows other autonomous algorithms to be connected given the provision of the data exchange interface with the module for generating recognition schemes (the open architecture principle). Due to the use of external algorithms, which can be based on arbitrary concepts, a high versatility of the software for a wide range of recognition tasks is provided. Note also that the system allows generation of autonomous recognition systems in two modes – the automatic (with a step-by-step assessment of the efficiency of a set of algorithms with respect to the training selection) and interactive (choice of vertices and algorithms of the algorithmic tree depend on the operator) ones. This approach provides to a great extent high versatility in solving application tasks (it shifts responsibility to the implementation of an autonomous recognition algorithm), requires less attention to the task specifics (interpretation of attributes), and, on the other hand, imposes high requirements on the completeness and adequacy (quality) of training selections.

The principal task here was to construct an autonomous recognition system based on geological data (the problem of oil-bearing bed separation). The mathematical model of recognition objects in this case is presented in the form of attributes of the x_1, x_2, \dots, x_n -

sets. Their following main properties were used to recognize objects (12 basic elementary attributes and 10 additional ones):

- bed thickness;
- clay solution resistance;
- resistance ρ_k on the standard potential probe;
- resistance on the gradient probes at the lateral logging sounding with the $A_0 = 0.5$ m size;
- resistance on the gradient probes at the lateral logging sounding with the $A_0 = 1$ m size;
- resistance on the gradient probes at the lateral logging sounding with the $A_0 = 2$ m size;
- resistance on the gradient probes at the lateral logging sounding with the $A_0 = 4$ m size;
- resistance on the gradient probes at the lateral logging sounding with the $A_0 = 8$ m size;
- inverse probe;
- rock resistance;
- well diameter;
- clay cake thickness.

The training sample provides information about the objects of two classes (the oil-bearing bed class and the water-bearing bed class). At the stage of examination, the constructed classification system should ensure effective recognition of objects of unknown classification with respect to these two classes. Before starting work, the training selection was automatically checked for correctness (searching and deleting the same objects of different belonging, i.e. the first-kind errors), although the system has implemented a scheme for additional training and error correction in the classification tree (the ATEC algorithm), since generation took place automatically, then this algorithm was not used.

Note that the training selection consisted of 1.250 objects (of which 756 were the oil-bearing ones), and the efficiency of the constructed recognition system was assessed on a test sample of 240 objects. The data from training and test selections were obtained on the basis of geological exploration in the Transcarpathian region during the period from 2001 to 2013. The methods of training selection approximation on the basis of hyperspheres and hyperellipses were used as the fixed algorithms (only geometric algorithms were selected from the library, and their algorithmic schemes were described above). The test use of a set of other library algorithms provided generation of an algorithmic tree of much greater complexity. It should also be noted that increasing of the set of algorithms negatively affects the total generation time of the system (in automatic mode) due to their step-by-step assessment with respect to the training selection.

Also, the resulting number of generalized features was 18 per 756 objects of the oil-bearing bed class and 22 per 494 objects of the water-bearing bed class. The constructed autonomous classification system is based on the recognition scheme presented in Fig. 2.

Note that if one assesses the efficiency (with respect to compression – the description of the training selection data) of the constructed scheme of the system of discrete object classification by the formula $\rho = 100 - (100 \cdot \frac{q}{m})$ and amounted to 96.8%.

Note that this recognition system was constructed for two different configurations:

- 1) (Conf No. 1 Intel I5 8500 / Ram 8GB);
- 2) (Conf No. 2 AMD FX8370 / Ram 16GB).

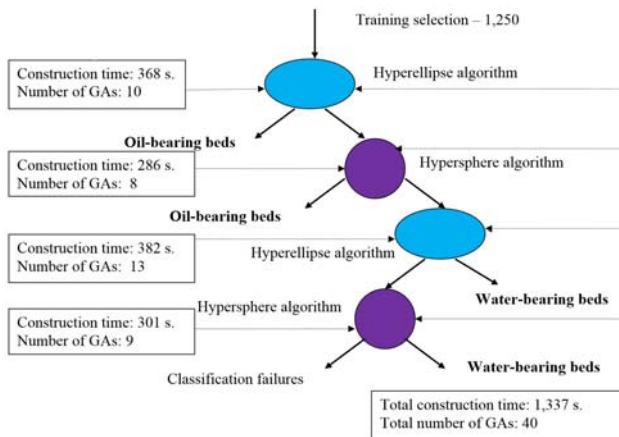


Figure 2 – The layout of the RS constructed on the basis of geological data

The whole process took 1.586 seconds and 1.337 seconds, respectively, what is largely due to the increase in the processor frequencies, disk subsystem speed and assembler optimization.

Note also that the hyperplane and hyperparallelepiped algorithms appeared not very suitable for this recognition system, since they 'refused' to work on this array of geological data (training selection). It should also be noted that the hyperellipse algorithm 'failed' to perform a complete approximation of the class regions, and was used in conjunction with another algorithm to simplify the resulting recognition scheme. At the stage of evaluating the effectiveness of the system constructed, 221 of 240 objects of the test sample were successfully classified (5 objects were denied classification). This pretty good result (certainly, 1250 objects are not enough to build a normal highly efficient classification system volume) is mainly due to the use of the hypersphere algorithm to approximate the training selection. This allowed one to ensure effective separation of the class region even at an incomplete (inadequate) selection due to increasing the number of generalized features.

6 DISCUSSION

Note that the structure of the algorithm tree constructed within this work (the ACT model) operates only with ready-made (constructed) sets of generalized attributes (elementary geometric classification algorithms), and it may not be interested in any general algorithm or

method (scheme, rule, method) they are obtained, and each of the schemes constructed by the algorithmic tree method will be a general recognition system that can be used for practical work (processing large arrays of experimental data in the form of discrete sets of arbitrary nature). An important point is that the resulting classification scheme (tree of algorithms) will be to some extent a new recognition algorithm (of course, synthesized from known algorithms and methods), and the resulting ACT structure (new classification scheme) is characterized by high versatility with respect to the application and relatively compact structure of the model itself (within the scope of the problem presented in the work, the GA sets of only two types were used). However, the ACT structure requires relatively high hardware costs for storing the generalized attributes (or their sets) and the initial assessment of the quality of classification algorithms according to TS. Moreover, the ACT models, in comparison with the LCT structures, have a high-speed classification rules, comparable hardware costs for storage and operation of the tree structure and high quality of classification.

CONCLUSIONS

New simple algorithmic implementations for approximating an array of geological data by a set of generalized features (elementary geometric algorithms of algorithms) have been suggested in this paper. The ACT structure is a graph-schematic structure with a dimension of 40 generalized attributes and generation time of 1.337 seconds, and for the tree synthesis two elementary geometric algorithms have been used.

An approach to the synthesis of new recognition algorithms based on a library (set) of already known algorithms and methods has been developed. That is, an effective scheme for recognizing discrete objects based on the tree method is presented with a step-by-step assessment and selection of generalized features at each step of the scheme synthesis. The efficiency of the developed discrete object classification scheme (geological data) was 96%.

The **scientific novelty** of the obtained results is based on the fact that the proposed method of the ACT structures on the basis of a set of autonomous recognition and classification algorithms assessment and ranking for generating the classification tree structure (the ACT model). Moreover, at each step of the classification tree branching a certain part of the TS (or its submanifold) is approximated.

The **practical value** of the obtained results is that the method of constructing the ACT models (the LCT/ACT structure) was implemented in the library of algorithms of the universal software system ORION III to solve various practical classification (recognition) problems for different arrays of discrete objects. The practical tests confirmed the efficiency of mathematical software, proposed ACT models and developed software that allows one to make recommendations on the use of this approach (the algorithm tree models) and its software implementation

for a wide range of applications for discrete object classification and recognition.

Prospects for further research may be directed towards the development of algorithmic classification trees methods, i.e. the methods of cutting-off and minimizing the ACT structures, optimizing software implementations of the ACT construction method, as well as its practical testing on the manifold of real classification and recognition problems.

ACKNOWLEDGEMENTS

The work was carried out within the framework of the scientific topic of the Department of software systems of the state higher educational institution “Uzhgorod National University” – methods and tools of software engineering, implementation of large data analysis processes based on information platforms (State Registration Number 0119U100703).

REFERENCES

1. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Stanford, 2008, 768 p.
2. Quinlan J. R. Induction of Decision Trees, *Machine Learning*, 1986, No. 1, pp. 81–106.
3. Mitchell T. Machine learning. New York, McGraw-Hill, 1997, 432 p.
4. Dietterich T. G., Kong E. B. Machine learning bias, statistical bias, and statistical variance of decision tree algorithms [Electronic resource]. Corvallis, Oregon State University, 1995, 14 p. Access mode : <http://www.cems.uwe.ac.uk/~irjohnso/coursenotes/uqc832/tr bias.pdf>
5. Breiman L. L., Friedman J. H., Olshen R. A., Stone C. J. Classification and regression trees. Boca Raton, Chapman and Hall/CRC, 1984, 368 p.
6. Vtoghoff P.E. Incremental Induction of Decision Trees, *Machine Learning*, 1989, No. 4, pp. 161–186.
7. Vasilenko Y. A., Vasilenko E. Y., Kuhayivsky A. I., Papp I. O. Construction and optimization of recongnizing systems, *Scientific and technical journal "Information technologies and systems"*, 1999, No. 1, pp. 122–125.
8. Subbotin S.A. Construction of decision trees for the case of low-information features, *Radio Electronics, Computer Science, Control*, 2019, No. 1, pp. 121–130.
9. Povhan I.F. Logical recognition tree construction on the basis a step-to-step elementary attribute selection, *Radio Electronics, Computer Science, Control*, 2020, No. 2, pp. 95–106.
10. Povkhan I. F. The general concept of the methods of algorithmic classification trees, *Radio Electronics, Computer Science, Control*, 2020, No. 3, pp. 108–121.
11. Povhan I. F. Limited method for the case of algorithmic classification tree, *Radio Electronics, Computer Science, Control*, 2020, No. 4, pp. 106–118.
12. Povhan I. Designing of recognition system of discrete objects, *2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP), 2016, Lviv, Ukraine*. Lviv, 2016, pp. 226–231.
13. Vasilenko Y. A., Vashuk F. G., Povkhan I. F. Automating the construction of classification systems based on agent – schemes, *Mathematical modeling, optimization and information technologies : International Joint Conference MDIF-2012, Kisheneu, Moldova, 2012*. Kisheneu, 2012, pp. 444–446.
14. Povkhan I.F. Features of synthesis of generalized features in the construction of recognition systems using the logical tree method, *Information technologies and computer modeling ITKM-2019 : materials of the international scientific and practical conference, Ivano-Frankivsk, May 20–25, 2019*. Ivano-Frankivsk, 2019, pp. 169–174.
15. Vasilenko Y. A., Vashuk F. G., Povkhan I. F. The importance of discrete signs, *XX International Conference Promising ways and directions of improving the educational system, Uzhgorod, November 16–19, 2010*. Uzhgorod, 2010, Vol. 21, No. 1, pp. 217–222.
16. Alpaydin E. Introduction to Machine Learning. London, The MIT Press. 2010, 400 p.
17. De Mántaras R. L. A distance-based attribute selection measure for decision tree induction, *Machine learning*, 1991, Vol. 6, No. 1, pp. 81–92.
18. Painsky A., Rosset S. Cross-validated variable selection in tree-based methods improves predictive performance, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, Vol. 39, No. 11, pp. 2142–2153. DOI:10.1109/tpami.2016.2636831.
19. Miyakawa M. Criteria for selecting a variable in the construction of efficient decision trees, *IEEE Transactions on Computers*, 1989, Vol. 38, No. 1, pp. 130–141.
20. Kotsiantis S.B. Supervised Machine Learning: A Review of Classification Techniques, *Informatica*, 2007, No. 31, pp. 249–268.
21. Deng H., Runger G., Tuv E. Bias of importance measures for multi-valued attributes and solutions, *Proceedings of the 21st International Conference on Artificial Neural Networks (ICANN), Espoo, Finland, Jun 14–Jun 17, 2011*. Espoo, 2011, pp. 293–300.
22. Dietterich T. G. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization, *Machine learning*, 2000, Vol. 40, No. 2, pp. 139–157.
23. Subbotin S., Oliinyk A. eds. R. Szewczyk, M. Kaliczynska. The dimensionality reduction methods based on computational intelligence in problems of object classification and diagnosis, *Recent Advances in Systems, Control and Information Technology*. Cham, Springer, 2017, pp. 11–19. (Advances in Intelligent Systems and Computing, vol. 543).
24. Subbotin S. A. Methods and characteristics of localitypreserving transformations in the problems of computational intelligence, *Radio Electronics, Computer Science, Control*, 2014, No. 1, pp. 120–128.
25. Subbotin S.A. Methods of sampling based on exhaustive and evolutionary search, *Automatic Control and Computer Sciences*, 2013, Vol. 47, No. 3, pp. 113–121. DOI: 10.3103/s0146411613030073
26. Koskimaki H., Juutilainen I., Laurinen P., Roning J. Two-level clustering approach to training data instance selection: a case study for the steel industry, *Neural Networks : International Joint Conference (IJCNN-2008), Hong Kong, 1–8 June 2008, proceedings*. Los Alamitos, IEEE, 2008, pp. 3044–3049. DOI: 10.1109/ijcnn.2008.4634228
27. Srikant R., Agrawal R. Mining generalized association rules *Future Generation Computer Systems*, 1997, Vol. 13, No. 2, pp. 161–180.
28. Amit Y., Geman D., Wilder K. Joint induction of shape features and tree classifiers, *IEEE Transactions on Pattern*

- Analysis and Machine Intelligence*, 1997, Vol. 19, No. 11, pp. 1300–1305.
29. Mingers J. An empirical comparison of pruning methods for decision tree induction, *Machine learning*, 1989, Vol. 4, No. 2, pp. 227–243.
30. Karimi K., Hamilton H. J. Generation and Interpretation of Temporal Decision Rules, *International Journal of Computer Information Systems and Industrial Management Applications*, 2011, Vol. 3, pp. 314–323.
31. Kamiński B., Jakubczyk M., Szufel P. A framework for sensitivity analysis of decision trees, *Central European Journal of Operations Research*, 2017, Vol. 26 (1), pp. 135–159.
32. Lupei M., Mitsa A., Repariuk V., Sharkan V. Identification of authorship of Ukrainian-language texts of journalistic style using neural networks, *Eastern-European Journal of Enterprise Technologies*, 2020, Vol. 1 (2 (103)), pp. 30–36. DOI: <https://doi.org/10.15587/1729-4061.2020.195041>
33. Bodyanskiy Y., Vynokurova O., Setlak G. and Pliss I. Hybrid neuro-neo-fuzzy system and its adaptive learning algorithm, *Computer Sciences and Information Technologies (CSIT), Xth Scien. and Tech. Conf., Lviv, 2015*. Lviv, 2015, pp. 111–114.
34. Subbotin S. The neuro-fuzzy network synthesis and simplification on precedents in problems of diagnosis and pattern recognition, *Optical Memory and Neural Networks (Information Optics)*, 2013, Vol. 22, No. 2, pp. 97–103. DOI: 10.3103/s1060992x13020082

Received 15.06.2021.
Accepted 13.08.2021.

УДК 001.891:65.011.56

ЗАДАЧА АПРОКСИМАЦІЇ МАСИВУ ДИСКРЕТНИХ ДАНИХ НАБОРОМ ЕЛЕМЕНТАРНИХ ГЕОМЕТРИЧНИХ АЛГОРИТМІВ

Повхан І. Ф. – д-р техн. наук, доцент, доцент кафедри програмного забезпечення систем ДВНЗ Ужгородський національний університет, м. Ужгород, Україна.

Міца О. В. – д-р техн. наук, доцент, зав. кафедри інформаційних управляючих систем і технологій. ДВНЗ Ужгородський національний університет, м. Ужгород, Україна.

Мулеса О. Ю. – д-р техн. наук, доцент, доцент кафедри кібернетики та прикладної математики ДВНЗ Ужгородський національний університет, м. Ужгород, Україна.

Мельник О. О. – канд. техн. наук, доцент, доцент кафедри програмного забезпечення систем ДВНЗ Ужгородський національний університет, м. Ужгород, Україна.

АНОТАЦІЯ

Актуальність. В роботі розв’язана задача апроксимації масиву дискретних даних набором елементарних геометричних алгоритмів і представлення побудованої моделі розпізнавання у вигляді алгоритмічного дерева класифікації. Об’єктом даного дослідження є концепція дерева класифікації у вигляді дерева алгоритмів. Предметом дослідження є актуальні моделі, методи, алгоритми та схеми побудови різнотипних дерев класифікації.

Мета. Метою даної роботи є створення простого та ефективного методу та алгоритмічної схеми побудови деревоподібних моделей розпізнавання та класифікації на основі дерев алгоритмів для навчальних вибірок дискретної інформації великого об’єму, який характеризується модульною структурою з незалежних алгоритмів розпізнавання оцінених на основі даних початкової початкової вибірки для широкого спектру прикладних задач.

Метод. Пропонується схема синтезу дерев класифікації (дерев алгоритмів) на основі апроксимації масиву даних набором елементарних геометричних алгоритмів, яка для заданої початкової навчальної вибірки довільного розміру буде деревоподібну структуру (модель АДК), яка складається з набору автономних алгоритмів класифікації та розпізнавання оцінених на кожному кроці, етапі побудови АДК за даною початковою вибіркою. Розроблений метод побудови алгоритмічного дерева класифікації основна ідея якого полягає в по кроковій апроксимації початкової вибірки довільного об’єму та структури набором елементарних геометричних алгоритмів класифікації. Даний метод при формуванні поточної вершини дерева алгоритмів, вузла, узагальненої ознаки, забезпечує виділення найбільш ефективних, якісних елементарних алгоритмів класифікації з початкового набору та побудову лише тих шляхів в структурі АДК де відбувається найбільша кількість помилок класифікації. Розроблена схема синтезу результуючого дерева класифікації, моделі АДК дозволяє значно скоротити розмір та складність дерева. Структурна складність конструкції АДК оцінюється на основі кількості переходів, вершин та ярусів структури АДК, що дозволяє підвищити якість його наступного аналізу, забезпечити ефективний механізм декомпозиції, та будувати структури АДК в умовах фіксованих наборів обмежень. Метод синтезу дерев алгоритмів дозволяє будувати різнотипні деревоподібні моделі розпізнавання з різними початковими наборами елементарних класифікаторів з наперед заданою точністю для широкого класу задач теорії штучного інтелекту.

Результати. Розроблений та представлений в даній роботі метод апроксимації дискретних навчальних вибірок набором елементарних геометричних алгоритмів отримав програмну реалізацію та був досліджений і порівняний з методами логічних дерев класифікації на основі селекції елементарних ознак при розв’язку задачі розпізнавання реальних даних геологічного типу.

Висновки. Проведені в даній роботі загальний аналіз та експерименти підтвердили працездатність розробленого механізму побудови структур дерев алгоритмів та показують можливість його перспективного використання для розв’язку широкого спектру практичних задач розпізнавання та класифікації. Перспективи подальших досліджень та апробацій можуть полягати в створенні методів алгоритмічного дерева класифікації інших типів з іншими початковими наборами елементарних класифікаторів, оптимізації його програмних реалізацій, а також експериментальних дослідженнях даного методу на більш широкому колі практичних задач.

КЛЮЧОВІ СЛОВА: алгоритмічне дерево класифікації, розпізнавання образів, класифікація, алгоритм класифікації, критерій розгалуження, геометричний алгоритм.

УДК 001.891:65.011.56

ЗАДАЧА АППРОКСИМАЦИИ МАССИВА ДИСКРЕТНЫХ ДАННЫХ НАБОРОМ ЭЛЕМЕНТАРНЫХ ГЕОМЕТРИЧЕСКИХ АЛГОРИТМОВ

Повхан И. Ф. – д-р техн. наук, доцент, доцент кафедры программного обеспечения систем ГВУЗ Ужгородский национальный университет, г. Ужгород, Украина.

Мица А. В. – д-р техн. наук, доцент, зав. кафедры информационных управляющих систем и технологий ГВУЗ Ужгородский национальный университет, г. Ужгород, Украина.

Мулеса О. Ю. – д-р техн. наук, доцент, доцент кафедры кибернетики и прикладной математики ГВУЗ Ужгородский национальный университет, г. Ужгород, Украина.

Мельник Е. А. – канд. техн. наук, доцент, доцент кафедры программного обеспечения систем ГВУЗ Ужгородский национальный университет, г. Ужгород, Украина.

АННОТАЦИЯ

Актуальность. В работе решена задача аппроксимации массива дискретных данных набором элементарных геометрических алгоритмов и представления построенной модели распознавания в виде алгоритмического дерева классификации. Объектом данного исследования является концепция деревьев классификации в виде дерева алгоритмов. Предметом исследования являются актуальные модели, методы, алгоритмы и схемы построения разнотипных деревьев классификации.

Цель. Целью данной работы является создание простого и эффективного метода и алгоритмической схемы построения древовидных моделей распознавания и классификации на основе деревьев алгоритмов для обучающих выборок дискретной информации большого объема. Причем они характеризуется модульной структурой из независимых алгоритмов распознавания, оцененных на основе данных начальной обучающей выборки для широкого спектра прикладных задач.

Метод. Предлагается схема синтеза деревьев классификации (деревьев алгоритмов) на основе аппроксимации массива данных набором элементарных геометрических алгоритмов, которая для заданной исходной обучающей выборки произвольного размера строит древовидную структура (модель АДК), которая состоит из набора автономных алгоритмов классификации и распознавания, оцененных на каждом этапе построения АДК по данной исходной выборке. Разработан метод построения алгоритмического дерева классификации основная идея которого заключается в по шаговой аппроксимации начальной выборки произвольного объема и структуры набором элементарных геометрических алгоритмов классификации. Данный метод при формировании текущей вершины дерева алгоритмов, узла, обобщенного признака, обеспечивает выделение наиболее эффективных, качественных элементарных алгоритмов классификации из начального набора и доработку только тех путей в структуре АДК где происходит наибольшее количество ошибок классификации. Разработана схема синтеза результирующего дерева классификации, модели АДК позволяет значительно сократить размер и сложность дерева. Структурная сложность конструкции АДК оценивается на основе количества переходов, вершин и ярусов структуры АДК, что позволяет повысить качество его последующего анализа, обеспечить эффективный механизм декомпозиции, и строить структуры АДК в условиях фиксированных наборов ограничений. Метод синтеза деревьев алгоритмов позволяет строить разнотипные древовидные модели распознавания с различными начальными наборами элементарных классификаторов с заранее заданной точностью для широкого класса задач теории искусственного интеллекта.

Результаты. Разработанный и представленный в данной работе метод аппроксимации дискретных обучающих выборок набором элементарных геометрических алгоритмов получил программную реализацию и был исследован и сравнен с методами логических деревьев классификации на основе селекции элементарных признаков при решении задачи распознавания реальных данных геологического типа.

Выводы. Проведенные в данной работе общий анализ и эксперименты подтвердили работоспособность разработанного механизма построения структур деревьев алгоритмов и показывают возможность его перспективного использования для решения широкого спектра практических задач распознавания и классификации. Перспективы дальнейших исследований и апробаций могут заключаться в построении методов алгоритмического дерева классификации других типов с другими начальными наборами элементарных классификаторов, оптимизации его программных реализаций, а также экспериментальных исследованиях данного метода в более широком спектре практических задач.

КЛЮЧЕВЫЕ СЛОВА: алгоритмическое дерево классификации, распознавания образов, классификация, алгоритм классификации, критерий ветвления, геометрический алгоритм.

ЛИТЕРАТУРА / LITERATURA

1. Hastie T. The Elements of Statistical Learning / T. Hastie, R. Tibshirani, J. Friedman. – Stanford, 2008. – 768 p.
2. Quinlan J.R. Induction of Decision Trees / J. R. Quinlan // Machine Learning. – 1986. – №1. – P. 81–106.
3. Mitchell T. Machine learning / T. Mitchell. – New York : McGrawHill, 1997. – 432 p.
4. Dietterich T. G. Machine learning bias, statistical bias and statistical variance of decision tree algorithms [Electronic resource] / T. G. Dietterich, E. B. Kong. – Corvallis : Oregon State University, 1995. – 14 p. Access mode : <http://www.cems.uwe.ac.uk/~irjohnso/coursenotes/uqc832/trbias.pdf>
5. Classification and regression trees / L. L. Breiman, J. H. Friedman, R. A. Olshen et al.]. – Boca Raton : Chapman and Hall/CRC, 1984. – 368 p.
6. Vtogofov P. E. Incremental Induction of Decision Trees / P. E. Vtogofov // Machine Learning. – 1989. – № 4. – P. 161–186.

7. Construction and optimization of recognizing systems / [Y. A. Vasilenko, E. Y. Vasilenko, A. I. Kuhayivsky, I. O. Papp] // Scientific and technical journal "Information technologies and systems". – 1999. – № 1. – P. 122–125.
8. Subbotin S. A. Construction of decision trees for the case of low-information features / S. A. Subbotin // Radio Electronics, Computer Science, Control. – 2019. – № 1. – P. 121–130.
9. Povhan I. F. Logical recognition tree construction on the basis a step-to-step elementary attribute selection / I. F. Povhan // Radio Electronics, Computer Science, Control. – 2020. – № 2. – P. 95–106.
10. Povkhan I. F. The general concept of the methods of algorithmic classification trees / I. F. Povkhan // Radio Electronics, Computer Science, Control. – 2020. – № 3. – P. 108–121.
11. Povhan I. F. Limited method for the case of algorithmic classification tree / I. F. Povhan // Radio Electronics, Computer Science, Control. – 2020. – № 4. – P. 106–118.
12. Povhan I. Designing of recognition system of discrete objects / I. Povhan // 2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP), Lviv, 2016, Ukraine. – Lviv, 2016. – P. 226–231.
13. Vasilenko Y. A. Automating the construction of classification systems based on agent-schemes / Y. A. Vasilenko, F. G. Vashuk, I. F. Povkhan // Mathematical modeling, optimization and information technologies : International Joint Conference MDIF-2012, Kishineu, Moldova, 2012. – Kishineu, 2012. – P. 444–446.
14. Povkhan I. F. Features of synthesis of generalized features in the construction of recognition systems using the logical tree method / I. F. Povkhan // Information technologies and computer modeling ITKM-2019 : materials of the international scientific and practical conference, Ivano-Frankivsk, May 20–25, 2019. – Ivano-Frankivsk, 2019. – P. 169–174.
15. Vasilenko Y. A. The importance of discrete signs / Y. A. Vasilenko, F. G. Vashuk, I. F. Povkhan // XX International Conference Promising ways and directions of improving the educational system, Uzhgorod, November 16–19, 2010. – Uzhgorod, 2010. – Vol. 21, № 1. – P. 217–222.
16. Alpaydin E. Introduction to Machine Learning / E. Alpaydin. – London : The MIT Press, 2010. – 400 p.
17. De Mántaras R. L. A distance-based attribute selection measure for decision tree induction / De Mántaras R. L. // Machine learning. – 1991. – Vol. 6, № 1. – P. 81–92.
18. Painsky A. Cross-validated variable selection in tree-based methods improves predictive performance / A. Painsky, S. Rosset // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2017. – Vol. 39, № 11. – P. 2142–2153. DOI:10.1109/tpami.2016.2636831
19. Miyakawa M. Criteria for selecting a variable in the construction of efficient decision trees / M. Miyakawa // IEEE Transactions on Computers. – 1989. – Vol. 38, № 1. – P. 130–141.
20. Kotsiantis S. B. Supervised Machine Learning: A Review of Classification Techniques / S. B. Kotsiantis // Informatica. – 2007. – № 31. – P. 249–268.
21. Deng H. Bias of importance measures for multi-valued attributes and solutions / H. Deng, G. Runger, E. Tuv // Proceedings of the 21st International Conference on Artificial Neural Networks (ICANN), Espoo, Finland, Jun 14–Jun 17, 2011. – Espoo, 2011. – P. 293–300.
22. Dietterich T. G. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization / T. G. Dietterich // Machine learning. – 2000. – Vol. 40, № 2. – P. 139–157.
23. Subbotin S. The dimensionality reduction methods based on computational intelligence in problems of object classification and diagnosis / S. Subbotin, A. Oliinyk // Recent Advances in Systems, Control and Information Technology / eds.: R. Szewczyk, M. Kaliczyńska. – Cham : Springer, 2017. – P. 11–19. – (Advances in Intelligent Systems and Computing, vol. 543).
24. Subbotin S. A. Methods and characteristics of locality-preserving transformations in the problems of computational intelligence / S. A. Subbotin // Radio Electronics, Computer Science, Control. – 2014. – № 1. – P. 120–128.
25. Subbotin S. A. Methods of sampling based on exhaustive and evolutionary search / S. A. Subbotin // Automatic Control and Computer Sciences. – 2013. – Vol. 47, № 3. – P. 113–121. DOI: 10.3103/s0146411613030073
26. Two-level clustering approach to training data instance selection: a case study for the steel industry / [H. Koskimaki, I. Juutilainen, P. Laurinen, J. Roning] // Neural Networks : International Joint Conference (IJCNN-2008), Hong Kong, 1–8 June 2008 : proceedings. – Los Alamitos : IEEE, 2008. – P. 3044–3049. DOI: 10.1109/ijcnn.2008.4634228
27. Srikant R. Mining generalized association rules / R. Srikant, R. Agrawal // Future Generation Computer Systems. – 1997. – Vol. 13, №2. – P. 161–180.
28. Amit Y. Joint induction of shape features and tree classifiers / Y. Amit, D. Geman, K. Wilder // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1997. – Vol. 19, № 11. – P. 1300–1305.
29. Mingers J. An empirical comparison of pruning methods for decision tree induction / J. Mingers // Machine learning. – 1989. – Vol. 4, № 2. – P. 227–243.
30. Karimi K. Generation and Interpretation of Temporal Decision Rules / K. Karimi, H. J. Hamilton // International Journal of Computer Information Systems and Industrial Management Applications. – 2011. – Vol. 3. – P. 314–323.
31. Kamiński B. A framework for sensitivity analysis of decision trees / B. Kamiński, M. Jakubczyk, P. Szufel // Central European Journal of Operations Research. – 2017. – Vol. 26 (1). – P. 135–159.
32. Identification of authorship of Ukrainian-language texts of journalistic style using neural networks / [M. Lupei, A. Mitsa, V. Repariuk, V. Sharkan] // Eastern-European Journal of Enterprise Technologies. – 2020. – Vol. 1 (2 (103)). – P. 30–36. DOI: <https://doi.org/10.15587/1729-4061.2020.195041>
33. Hybrid neuro-neo-fuzzy system and its adaptive learning algorithm / [Y. Bodyanskiy, O. Vynokurova, G. Setlak and I. Pliss] // Computer Sciences and Information Technologies (CSIT) : Xth Scien. and Tech. Conf., Lviv, 2015. – Lviv, 2015. – P. 111–114.
34. Subbotin S. The neuro-fuzzy network synthesis and simplification on precedents in problems of diagnosis and pattern recognition / S. Subbotin // Optical Memory and Neural Networks (Information Optics). – 2013. – Vol. 22, № 2. – P. 97–103. DOI: 10.3103/s1060992x13020082

AN INTELLIGENT MODEL BASED ON DEEP TRANSFER LEARNING FOR DETECTING ANOMALIES IN CYBER-PHYSICAL SYSTEMS

Sukhostat L. V. – PhD, Associate Professor, Institute of Information Technology, Azerbaijan National Academy of Sciences, Baku, Azerbaijan.

ABSTRACT

Context. The problem of detecting anomalies from signals of cyber-physical systems based on spectrogram and scalogram images is considered. The object of the research is complex industrial equipment with heterogeneous sensory systems of different nature.

Objective. The goal of the work is the development of a method for signal anomalies detection based on transfer learning with the extreme gradient boosting algorithm.

Method. An approach based on transfer learning and the extreme gradient boosting algorithm, developed for detecting anomalies in acoustic signals of cyber-physical systems, is proposed. Little research has been done in this area, and therefore various pre-trained deep neural model architectures have been studied to improve anomaly detection. Transfer learning uses weights from a deep neural model, pre-trained on a large dataset, and can be applied to a small dataset to provide convergence without overfitting. The classic approach to this problem usually involves signal processing techniques that extract valuable information from sensor data. This paper performs an anomaly detection task using a deep learning architecture to work with acoustic signals that are preprocessed to produce a spectrogram and scalogram. The SPOCU activation function was considered to improve the accuracy of the proposed approach. The extreme gradient boosting algorithm was used because it has high performance and requires little computational resources during the training phase. This algorithm can significantly improve the detection of anomalies in industrial equipment signals.

Results. The developed approach is implemented in software and evaluated for the anomaly detection task in acoustic signals of cyber-physical systems on the MIMII dataset.

Conclusions. The conducted experiments have confirmed the efficiency of the proposed approach and allow recommending it for practical use in diagnosing the state of industrial equipment. Prospects for further research may lie in the application of ensemble approaches based on transfer learning to various real datasets to improve the performance and fault-tolerance of cyber-physical systems.

KEYWORDS: anomaly detection, acoustic signal, transfer learning, spectrogram, scalogram, cyber-physical system.

ABBREVIATIONS

XGBoost is an extreme gradient boosting;
AUC is the area under the receiver operating characteristic curve;
CPS is a cyber-physical system;
SPOCU is a scaled polynomial constant unit;
PCA is a principal component analysis;
MIMII is a malfunctioning industrial machine investigation and inspection;
LOF is a local factor outlier;
GMM is a Gaussian mixture model;
OC-SVM is a one-class support vector machine;
STFT is a short-time Fourier transform;
CNN is a convolutional neural network.

NOMENCLATURE

γ is the minimum loss reduction needed for splitting;
 λ is a regularization term;
 X is a time-frequency signal representation;
 F is the number of frequency bins;
 T is the time dimension;
 x_i is a signal block;
 l is the length of the feature vector;
 σ is a window function;
 $\Theta(\omega, \tau)$ is the Fourier transform;
 v is the loss value of the XGBoost algorithm;
 $G(x)$ is an activation function;

\tilde{y}_i is an objective optimization function;
 K is the number of decision trees;
 ϕ_k is an independent tree with leaf scores,
 Φ is the space of the regression tree;
 Ω is a regularization term.

INTRODUCTION

An abnormal state of a cyber-physical system (CPS) can be caused by faulty components, temporary failures, misconfiguration, cyberattacks, or their combination [1, 2]. An adversary intervenes in CPS to manipulate the readings of sensors or actuators, leading to abnormal system operation.

Anomaly detection in an industrial scenario is essential because undetected failures can lead to critical damage. Early detection of anomalies can improve the reliability of fault-prone industrial equipment and reduce operating and maintenance costs.

The development of Industry 4.0 has led to new technologies for efficient and reliable monitoring of such systems. Thus, modern CPSs include devices that form a multi-sensor configuration. These systems simplify the data collection process, resulting in the availability of large datasets. Consequently, there has been an increase in the development of data mining methods for detecting anomalies [3].

The classical approach to such problems usually involves signal processing techniques that extract useful information from sensor data.

The object of study is complex industrial equipment with heterogeneous sensory systems of various nature. For this purpose, preliminary data processing is required to extract the most informative features [4]. It is usually a very time-consuming task that requires expert knowledge.

The subject of study is methods for detecting anomalies in industrial equipment signals based on transfer learning. Images of signal spectrogram and scalogram are reviewed for a more accurate classification of equipment failures. The SPOCU (scaled polynomial constant unit) activation function [5] is considered to improve the accuracy of the proposed approach. The XGBoost algorithm is applied because it has high performance and requires little computational resources at the training stage.

The purpose of the work is to use transfer learning in combination with the XGBoost algorithm to improve the accuracy of detecting an abnormal state from acoustic signals of CPS.

1 PROBLEM STATEMENT

Suppose we are given an acoustic signal that has a time-frequency representation $X \in R^{F \times T}$, where T is the time dimension, and F is the number of frequency bins. For a given signal dataset, it is necessary to find the function $F: X \rightarrow R$ such that $F(X)$ is higher for abnormal samples than for normal operation recordings. The acoustic signal is split into fragments using a sliding window $x_i \in R^{\tau \times T}$ ($\tau < T$). Here it is proposed to extract the l -dimensional feature vector using a feature extractor for each x_i . A pre-trained deep neural network is considered a feature extractor. Then some anomaly detection algorithm F is trained on all features from the fragments of the dataset $\Lambda = \{X_j \in R^{F \times T}\}_{j=1}^N$.

2 REVIEW OF THE LITERATURE

The detection of anomalies in industrial equipment is becoming an important area of research. The difficulty here is to obtain information from several sensors that differ in their specific acoustic properties [6]. Researchers propose new methods and expand existing algorithms for detecting industrial equipment faults [6–13].

Morita et al. [7] proposed principal component analysis (PCA) with local factor outlier (LOF) and Gaussian mixture model (GMM) to detect abnormal sounds in the presence of limited computing resources and a small dataset.

Paper [8] described an approach that combines pre-trained OpenL3 embeddings with the reconstruction error of an interpolation autoencoder using GMM as the final predictor. The parameters were set individually for each machine using the results from the development set.

Michau and Fink [9] developed an architecture for learning a meaningful and sparse representation of high-frequency signals. They combined both the wavelets the-

ory and deep learning for classification and anomaly detection tasks.

The application of autoencoder deep learning architectures for unsupervised acoustic anomaly detection based on Dense and convolutional neural networks (CNN) was considered in [10]. The energy features of the mel-spectrogram were extracted from the raw sounds. Several preliminary experiments were conducted to tune the autoencoder hyperparameters.

Tiwari et al. [11] proposed an ensemble of two systems capable of recording anomalous system behavior. In the first system, an outlier detection method based on the nearest neighbor search was proposed. In the second system, i -vectors and GMM are applied for anomaly detection. The negative log-likelihood is used as its anomaly scores.

OutlierNets, a family of very compact deep convolutional autoencoder architectures adapted for real-time acoustic anomaly detection, were proposed [12]. It has extremely low complexity and matches or exceeds large convolutional autoencoder architecture by AUC (area under the receiver operating characteristic curve) exhibiting microsecond scale latency on embedded hardware.

The efficiency of acoustic anomaly detection based on image transfer learning was studied [13]. The authors considered various deep neural models. Results showed that features extracted with ResNet18 and ResNet34 with GMM and OC-SVM (one-class support vector machine) achieved the best average AUC. It confirmed that the image-based features with transfer learning models might achieve competitive results in acoustic anomaly detection.

The following conclusions can be drawn summarising the analysis of the current research state in detecting anomalies from industrial facilities acoustic signals:

1) A small amount of work was focused on transfer learning for feature extraction and failure detection in industrial machines.

2) All the functionality of deep neural networks is not taken into account.

All this confirms the relevance of this study.

This paper proposes a new method for the automatic detection of acoustic signal anomalies based on transfer learning. The signal spectral information is considered as input data for the proposed model. The addition of the XGBoost algorithm improves the accuracy of CPS fault detection. Experiments on a real MIMII (malfunctioning industrial machine investigation and inspection) dataset have shown the effectiveness of the proposed approach and can help experts diagnose equipment malfunctions.

3 MATERIALS AND METHODS

The paper proposes an approach to detecting machine signal anomalies using transfer learning. Transfer learning uses weights from a deep neural model, pre-trained on a large dataset. It can be applied to a small dataset, providing convergence without overfitting.

The proposed approach to detecting machine signal anomalies from images using transfer learning consists of the following steps: pre-processing, feature extraction

using a deep neural network, feature fusion, and classification based on the XGBoost algorithm (Fig. 1).

The considered signals are pre-divided into cells of 128 samples with 64 samples overlap. A scalogram based on a wavelet transform and a spectrogram based on a short-term Fourier transform are extracted from the signals (STFT). STFT splits the signal into several overlapping blocks, multiplying them by the Hanning window function:

$$\Theta(\omega, \tau) = \int_{-\infty}^{+\infty} \theta(t) \sigma(t - \tau) e^{-i2\pi\omega t} dt, \quad (1)$$

where σ is a window function.

And the Morlet wavelet is considered in the wavelet transformation to obtain more informative images [14]:

$$\psi(t) = e^{-\beta^2 t^2 / 2} \cos(\pi t), \quad (2)$$

where β is a parameter that controls the shape of the mother wavelet.

Acquisition of the scalogram and spectrogram images is performed in parallel. Visual representations (RGB) of 128x128x3 size are then sent to a deep neural network.

Since deep learning models are trained on large datasets of various images, they can be applied to anomalies detection in signals from industrial facilities. Each of the model layers is responsible for different image features.

In this paper, in order to extract features from spectrogram and scalogram images, the following pre-trained deep neural networks are considered: Xception [15], MobileNet [16], DenseNet-121 [17] and InceptionV3 [18].

The MobileNet model is a small network that contains depth-separable convolutions and improves recognition performance [16]. The InceptionV3 network includes parallel convolutional layers that are then combined to

produce a result [18, 19]. The Xception network is a linear set of residual convolutional layers [15, 20]. The summation operation speeds up the transition from one model layer to another [21]. The considered DenseNet model is used to collect information from all levels of the network and transfer it to subsequent levels when there is not enough data for training [17].

The fully connected layer was treated as a feature vector using a pre-trained model. The results are combined to extract information about various characteristics and reduce recognition errors. Thus, the total size of the feature vector is 1x1024.

In this work, SPOCU [5] is considered in the proposed model as an activation function in hidden layers to improve the accuracy of anomaly detection in image-based signals and is calculated as follows:

$$G(x) = \rho H\left(\frac{x}{\zeta} + \xi\right) - \rho H(\xi), \quad (3)$$

where $\xi \in (0,1)$, $\rho, \zeta > 0$ and

$$H(x) = \begin{cases} r(c), & x \geq c \\ r(x), & x \in [0, c) \\ 0, & x < 0 \end{cases} \quad (4)$$

with $r(x) = x^3(x^5 - 2x^4 + 2)$ and $1 \leq c < \infty$. According to (3), here $c = \infty$, $\rho = 3.0937$, $\xi = 0.6653$ and $\zeta = 4.437$.

Then the resulting feature vector is fed to the XGBoost classifier, which was proposed by Chen et al. in 2016 [22]. XGBoost is a regression tree that supports the classification task. The basis of the algorithm is to optimize the value of the objective function. In this case, the objective optimization function is defined as follows:

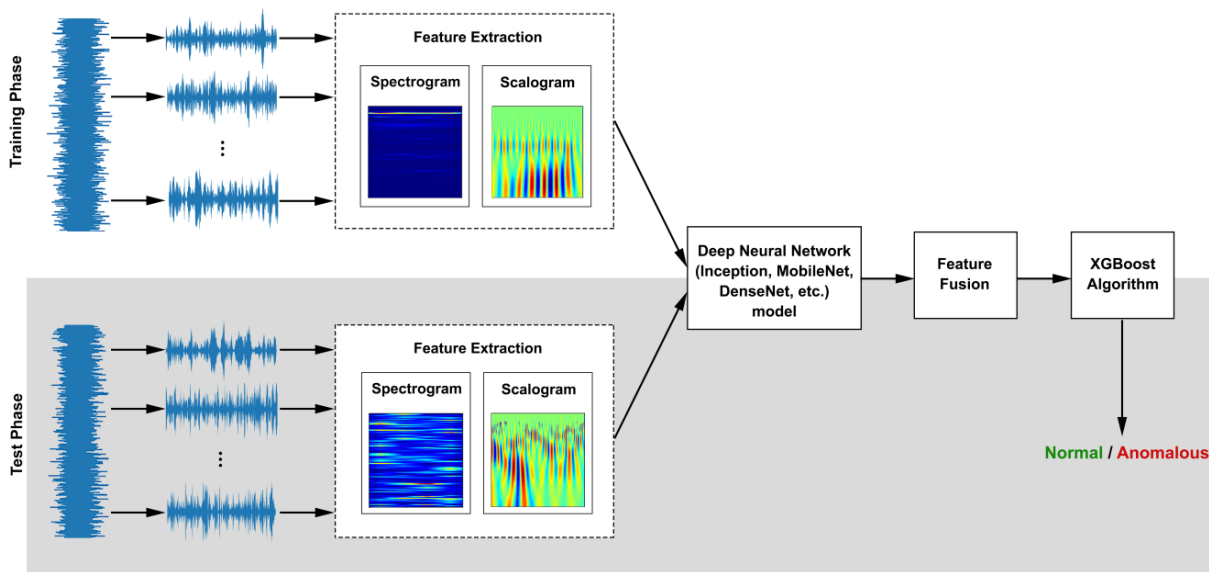


Figure 1 – Flowchart of the proposed approach

$$\tilde{y}_i = \sum_{k=1}^K \phi_k(x_i), \quad \phi_k \in \Phi, \quad (5)$$

where K is the number of decision trees, ϕ_k is an independent tree with leaf scores, Φ is the space of the regression tree. In this case, the loss function is given by the following equation:

$$L(\phi_t) = \sum v(\tilde{y}_i, y_i) + \sum \Omega(\phi_t), \quad (6)$$

where v is the loss value of the XGBoost algorithm, \tilde{y}_i is the predicted output, Ω is a regularization term that prevents overfitting (7).

$$\Omega(\phi) = \gamma K + \frac{1}{2} \lambda \|w\|^2, \quad (7)$$

where K is the number of leaf nodes, w is the score on each leaf, γ and λ are constants to control the degree of regularization.

Thus, we get the following:

$$L(\phi_t) \approx \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma K. \quad (8)$$

where g_i is the first derivative, and h_i is the second derivative of loss function, respectively.

For the XGBoost method, the learning rate is 0.001, number of trees to fit 100, maximum tree depth 6, $\gamma=0$ and $\lambda=1$.

4 EXPERIMENTS

This section provides the experimental dataset description, the evaluation metrics, and the experimental results to evaluate the proposed approach based on transfer learning.

The dataset of CPS under normal and abnormal operating conditions is considered to evaluate the proposed approach [23]. The audio dataset was collected using a circular microphone array consisting of eight separate microphones as 16-bit audio signals with a sampling rate

of 16 kHz [23]. It contains eight separate channels for each segment. The MIMII dataset contains the sound of four different machine types: valves, pumps, fans, and sliders. For each type of machine, different real anomalous scenarios were considered: pollution, leakage, rotating imbalance, rail damage, etc. MIMII also contains data for four machine IDs (00, 02, 04, and 06). Different signal-to-noise ratio levels (6 dB, 0 dB, and -6 dB) were considered in the dataset. It consists of 26,092 “normal” sound segments and 6,065 abnormal sound segments.

The “normal” and abnormal signatures for all machine types in the time domain of the MIMII dataset are shown in Fig. 2 ((a) – (d)) and Fig. 2 ((e) – (h)), respectively.

STFT spectrogram and scalogram based on wavelet transform for fans, pumps, sliders, and valves are shown in Fig. 3 and 4, respectively.

Performance evaluation of the proposed model is based on the following metrics: precision, recall, and F-measure.

The precision measure determines as the number of objects classified as positive that are truly positive:

$$Precision = \frac{TP}{TP + FP}. \quad (9)$$

where TP defines true positive values, TN are true negative values, FP are false positive values, and FN are false negative values.

The recall measure is used to determine the part of the positive samples selected by the classifier:

$$Recall = \frac{TP}{TP + FN}. \quad (10)$$

F-measure combines the recall and precision metrics:

$$F - measure = \frac{2 \times Recall \times Precision}{Recall + Precision}. \quad (11)$$

All considered metrics are widely used performance indicators in machine learning [24].

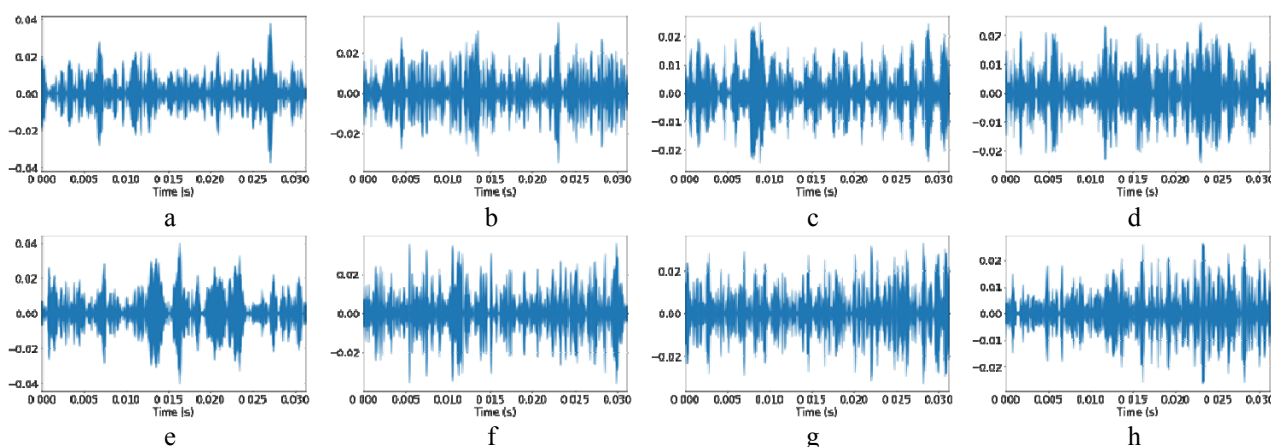


Figure 2 – Waveforms of normal ((a)–(d)) and abnormal operation ((e)–(h)) for four considered machines: a, e – Fan; b, f – Pump; c, g – Slider; d, h – Valve

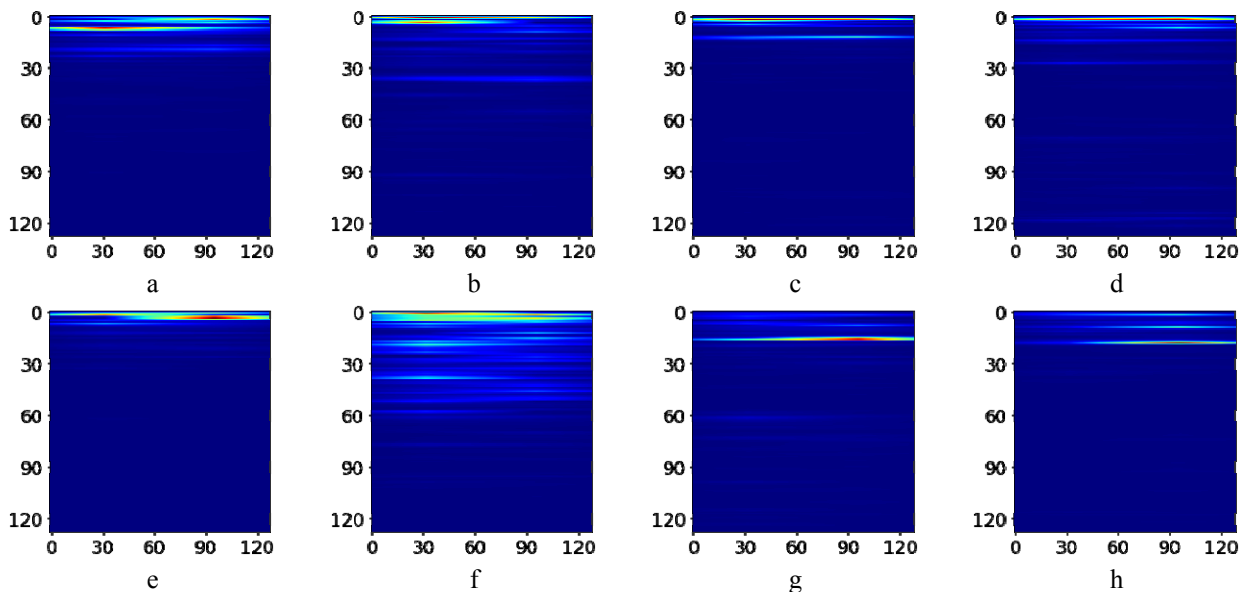


Figure 3 – Short-time Fourier transform spectrogram of the four considered machines (ID: 00) under normal ((a)–(d)) and anomalous ((e)–(h)) conditions at -6dB SNR:

a, e – Fan; b, f – Pump; c, g – Slider; d, h – Valve

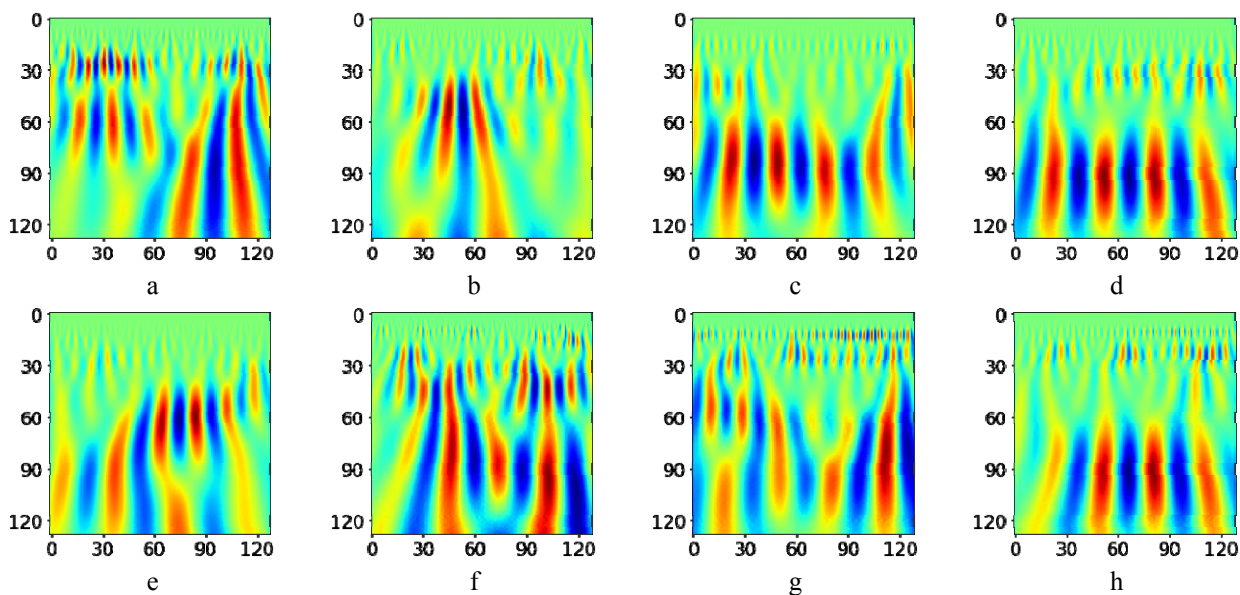


Figure 4 – Continuous wavelet transform scalogram of the four considered machines (ID: 00) under normal ((a)–(d)) and anomalous ((e)–(h)) conditions at -6dB SNR:

a, e – Fan; b, f – Pump; c, g – Slider; d, h – Valve

5 RESULTS

In this paper, the experiments are conducted in Python 2.7.13 using various libraries, including Tensorflow, Librosa, and Keras. Intel Xeon (R) CPU X5670 @ 2.93GHz * 24 with 24 GB of RAM machine was used.

This study analyses and compares various deep learning models (such as Xception, Inception, DenseNet, and MobileNet). They are trained on the MIMII dataset and applied to feature extraction from spectrogram and scalogram. Recall, Precision, and F-measure were considered as evaluation metrics.

Table 1 – Performance evaluation of the proposed approach

Model	Metric	Machine			
		Fan	Pump	Slider	Valve
Inception+ XGBoost	Recall	87.0	95.0	98.0	100
	Precision	89.1	87.9	100	94.3
	F-measure	92.3	90.6	98.1	96.3
Xception+ XGBoost	Recall	100	88.0	98.0	100
	Precision	84.2	100	100	92.6
	F-measure	96.8	92.8	98.1	95.3
Mobilenet+ XGBoost	Recall	96.0	88.0	94.0	100
	Precision	78.7	95.7	100	96.2
	F-measure	92.0	90.9	96.1	97.2
Densenet+ XGBoost	Recall	99.9	96.0	98.0	100
	Precision	87.0	100	100	97.1
	F-measure	98.2	97.1	98.1	97.7

Hyperparameters optimization was performed using cross-validation. The combination of parameters was chosen based on the lowest training loss and the highest accuracy. The best performance of the model was observed with a batch size equal to 32. Also, the evaluation of the learning rate was performed on different values. The accuracy of the model decreased with the increasing learning rate. It is important to note that all four considered models achieved high accuracy by decreasing the learning rate to 0.001.

The results of the experiments are shown in Table 1. Comparison of various deep neural models showed that Densenet+XGBoost outperformed the other considered models in detecting anomalies from machine signals according to the F-measure metric. Models Inception+XGBoost, Xception+XGBoost, and Densenet per-

formed well for a Slider machine using recall and precision metrics. Even though all models showed the best results in terms of recall for a Valve-type machine, they were inferior to Densenet+XGBoost according to precision and F-measure.

Thus, the Densenet+XGBoost model turned out to be the best. In order to evaluate its performance, it was compared with such models as Alexnet [13], ResNet18 [13], ResNet34 [13], SqueezeNet [13], IAEO3_opt [8] and PCA+LOF+GMM [7] for all types machines and machine IDs (00, 02, 04 and 06) according to the AUC metric (Table 2). Densenet+XGBoost showed the best performance of all machine types, resulting in average AUCs of 93.1%, 97.3%, 98.4%, and 93% for fan, pump, slider, and valve, respectively.

Table 2 – Comparison of mean AUC of different transfer learning models on the MIMII dataset

Machine Model	Fan				Pump				Slider				Valve			
	00	02	04	06	00	02	04	06	00	02	04	06	00	02	04	06
AlexNet+GMM [13]	57.7	61.7	53.9	94.5	84.1	70.8	81.6	66.0	98.3	80.9	61.4	57.5	60.2	69.2	59.9	53.5
AlexNet+OC-SVM [13]	51.0	73.1	59.7	93.2	77.5	56.4	81.1	60.1	96.2	81.4	53.6	56.5	61.6	73.6	48.3	48.9
ResNet18+GMM [13]	62.6	64.1	59.3	94.4	84.5	71.3	84.0	68.3	99.1	85.8	68.8	65.6	58.3	73.3	60.2	56.9
ResNet34+GMM [13]	58.7	65.6	57.0	90.9	78.4	66.8	87.9	63.2	99.6	90.4	82.5	69.1	73.0	79.1	60.1	61.9
ResNet34+OC-SVM [13]	50.1	67.4	57.5	83.0	64.9	51.5	81.2	60.2	96.8	85.0	71.4	64.3	75.6	77.8	64.3	53.1
SqueezeNet+OC-SVM [13]	55.6	64.8	46.2	78.8	86.7	49.4	88.4	62.3	99.2	81.5	59.4	71.6	69.0	71.3	53.1	58.2
Morita et al. (2020) [7]	67.4	87.1	79.3	96.2	72.5	70.4	94.2	87.1	97.7	75.9	96.9	94.2	99.4	91.8	94.2	80.7
IAEO3_opt [8]	65.5	83.3	71.4	98.1	84.4	77.8	98.0	78.9	95.9	84.0	97.9	85.9	100	99.7	99.8	98.8
Densenet+XGBoost	93.6	95.1	87.9	95.9	97.7	96.9	97.8	97.1	99.6	99.5	97.2	97.3	93.8	95.7	93.9	88.6

6 DISCUSSION

The proposed model made it possible to achieve a significant improvement in anomaly detection according to the data of machine sensors according to AUC 95.45%, compared to the previously proposed models [7, 8, 10, 13]. Densenet+XGBoost improved by about 8% over the PCA model [7] applied to the log spectrogram of the audio signal combined with LOF and GMM on the MIMII dataset. Grollmisch et al. [8] proposed a method combining OPENL3 embeddings and interpolation autoencoder (IAEO3_opt) for acoustic signals anomaly detection. Compared to the IAEO3_opt model, the Densenet+XGBoost model has improved by approximately 7% [8]. Densenet+XGBoost gave comparable results (AUC 95.5%) on the reviewed MIMII dataset. Coelho et al. [10] used CNN and Dense network in combination with an autoencoder for the task of unsupervised acoustic anomaly detection, where the results averaged 72.0%, 73.1%, 91.8%, and 78.9% for the fan, pump, slider, and valve, respectively. The accuracy of the proposed method is 93.1% for the fan, 97.3% for the pump, 98.4% for the slider, and 93.0% for the valve, which is significantly higher than the above result. The results on the MIMII dataset showed that the Densenet + XGBoost model outperformed other approaches. The 10-fold-cv results dem-

onstrated the reliability and robustness of the proposed model.

CONCLUSIONS

The urgent problem of detecting anomalies is solved based on acoustic signals from industrial equipment.

The scientific novelty of obtained results is that a transfer learning approach with the XGBoost classifier is proposed. There has been little research done in this area, and therefore studies are underway on various pre-trained deep neural model architectures to detect anomalies. The spectrogram and scalogram of the acoustic signal were considered as input data for the proposed architecture. The SPOCU activation function [5] was used to improve the accuracy of the proposed approach.

The practical significance of the obtained results is that experiments on a real MIMII dataset showed the effectiveness of the proposed approach and can help experts in diagnosing equipment malfunctions. Comparison with other known methods proves the superiority of Densenet+XGBoost in terms of anomaly detection accuracy.

Prospects for further research are in the development of ensemble approaches based on transfer learning using other real datasets to improve the performance and fault-tolerance of CPS.

REFERENCES

1. Langner R. Stuxnet: dissecting a cyberwarfare weapon, *IEEE Security & Privacy*, 2011, Vol. 9, No. 3, pp. 49–51. DOI: 10.1109/MSP.2011.67
2. Ahmed C. M., Zhou J. Challenges and opportunities in CPS security: a physics-based perspective, *IEEE Security & Privacy*, 2020, Vol. 18, No. 6, pp. 14–22. DOI: 10.1109/MSEC.2020.3002851
3. Ahmed C. M., Murguia C., Ruths J. Model-based attack detection scheme for smart water distribution networks, *Computer and Communications Security: 17th ACM ASIA Conference, Abu Dhabi, 2–6 April 2017, proceedings*. 2017, pp. 101–113. DOI: 10.1145/3052973.3053011
4. Canizo M., Triguero I., Conde A., Onieva E. Multi-head CNN–RNN for multi-time series anomaly detection: an industrial case study, *Neurocomputing*, 2019, Vol. 363, pp. 246–260. DOI: 10.1016/j.neucom.2019.07.034
5. Kiseřák J., Lu Y., Švihra J., Szépe P., Stehlík M. “SPOCU”: scaled polynomial constant unit activation function, *Neural Computing and Applications*, 2021, No. 33, pp. 3385–3401. DOI: 10.1007/s00521-020-05182-1
6. M. R. G. R., Ahmed C. M., Mathur A. Machine learning for intrusion detection in industrial control systems: challenges and lessons from experimental evaluation, *Cybersecurity*, 2021, Vol. 4, No. 27, pp. 1–12. DOI: 10.1186/s42400-021-00095-5
7. Morita K., Yano T., Tran K. Q. Anomalous sound detection by using local outlier factor and Gaussian mixture model, 2020 [Electronic resource]. Access mode: http://dcase.community/documents/challenge2020/technical_reports/DCASE2020_Morita_51_t2.pdf
8. Grollmisch S., Johnson D., Abeßer J., Lukashevich H. IAEO3 – combining OpenL3 embeddings and interpolation autoencoder for anomalous sound detection, 2020 [Electronic resource]. Access mode: http://dcase.community/documents/challenge2020/technical_reports/DCASE2020_Grollmisch_15_t2.pdf
9. Michau G., Fink O. Fully learnable deep wavelet transform for unsupervised monitoring of high-frequency time series, 2021 [Electronic resource]. Access mode: <https://arxiv.org/pdf/2105.00899.pdf>
10. Coelho G., Pereira P., Matos L., Ribeiro A., Nunes E., Ferreira A., Cortez P., Pilastrri A. Deep dense and convolutional autoencoders for machine acoustic anomaly detection, *Artificial Intelligence Applications and Innovations: 17th IFIP International Conference, Hersonissos, 25–27 June 2021: proceedings*. Cham, Springer, 2021, pp. 337–348. DOI: 10.1007/978-3-030-79150-6_27
11. Tiwari P., Jain Y., Avila A., Monteiro J., Kshirsagar S., Gaballah A., Falk T. H. Modulation spectral signal representation and i-vectors for anomalous sound detection, 2020 [Electronic resource]. Access mode: http://dcase.community/documents/challenge2020/technical_reports/DCASE2020_Tiwari_84_t2.pdf
12. Abbasi S., Famouri M., Shafiee M. J., Wong A. OutlierNets: highly compact deep autoencoder network architectures for on-device acoustic anomaly detection, *Sensors*, 2021, Vol. 21, No. 14, pp. 1–12. DOI: 10.3390/s21144805
13. Müller R., Ritz F., Illium S., Linnhoff-Popien C. Acoustic anomaly detection for machine sounds based on image transfer learning, *Agents and Artificial Intelligence, 13th International Conference, 4–6 February 2021, proceedings*. pp. 49–56. DOI: 10.5220/0010185800490056
14. Lin J., Qu L. Feature Extraction based on Morlet wavelet and its application for mechanical fault diagnosis, *Journal of Sound and Vibration*, 2000, Vol. 234, No. 1, pp. 48–135. DOI: 10.1006/jsvi.2000.2864
15. Chollet F. Xception: deep learning with depthwise separable convolutions, *Computer Vision and Pattern Recognition: 30th IEEE Conference. Honolulu, 21–26 July 2017, proceedings*, pp. 1800–1807. DOI: 10.1109/CVPR.2017.195
16. Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L. C. MobileNetV2: inverted residuals and linear bottlenecks, *Computer Vision and Pattern Recognition: IEEE/CVF Conference, Salt Lake City, 18–23 June 2018: proceedings*. pp. 4510–4520. DOI: 10.1109/CVPR.2018.00474
17. Huang G., Liu Z., van der Maaten L., Weinberger K. Q., Densely connected convolutional networks, *Computer Vision and Pattern Recognition, 30th IEEE Conference. Honolulu, 21–26 July 2017, proceedings*, pp. 2261–2269. DOI: 10.1109/CVPR.2017.243
18. Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. Rethinking the inception architecture for computer vision, *Computer Vision and Pattern Recognition, IEEE Conference. Las Vegas, 27–30 June 2016, proceedings*. pp. 2818–2826. DOI: 10.1109/CVPR.2016.308
19. Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. Going deeper with convolutions, *Computer Vision and Pattern Recognition, IEEE Conference. Boston, 7–12 June 2015, proceedings*, pp. 1–9. DOI: 10.1109/CVPR.2015.7298594
20. Kassani S. H., Kassani P. H., Khazaeinezhad R., Wesolowski M. J., Schneider K. A., Deters R. Diabetic retinopathy classification using a modified Xception architecture, *Signal Processing and Information Technology, IEEE International Symposium, Ajman, 10–12 December 2019, proceedings*, pp. 1–6. DOI: 10.1109/ISSPIT47144.2019.9001846
21. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition, *Computer Vision and Pattern Recognition: IEEE Conference, Las Vegas, 27–30 June 2016: proceedings*. pp. 770–778. DOI: 10.1109/CVPR.2016.90
22. Chen T., Guestrin C. XGBoost: a scalable tree boosting system, *Knowledge Discovery and Data Mining, 22nd ACM SIGKDD International Conference. San Francisco, 13–16 August 2016, proceedings*, pp. 785–794. DOI: 10.1145/2939672.2939785
23. Purohit H., Tanabe R., Ichige K., Endo T., Nikaido Y., Suefusa K., Kawaguchi Y. MIMII dataset: sound dataset for malfunctioning industrial machine investigation and inspection, *Detection and Classification of Acoustic Scenes and Events, 4th Workshop. New York, 25–26 October 2019, proceedings*, pp. 209–213. DOI: 10.33682/m76f-d618
24. Ferri C., Hernández-Orallo J., Modroui R. An experimental comparison of performance measures for classification, *Pattern Recognition Letters*, 2009, Vol. 30, No. 1, pp. 27–38. DOI: 10.1016/j.patrec.2008.08.010

Received 30.06.2021.
Accepted 13.08.2021.

УДК 004.056

ІНТЕЛЕКТУАЛЬНА МОДЕЛЬ ВИЯВЛЕННЯ АНОМАЛІЙ У КІБЕРФІЗИЧНИХ СИСТЕМАХ НА ОСНОВІ ГЛИБОКОГО ТРАНСФЕРНОГО НАВЧАННЯ

Сухостат Л. В. – канд. техн. наук, доцент, Інститут інформаційних технологій, Національна академія наук Азербайджану, Баку, Азербайджан.

АНОТАЦІЯ

Актуальність. Розглянуто задачу виявлення аномалій сигналів кіберфізичних систем на основі зображень спектрограм і скалограм. Об'єктом дослідження є складне промислове устаткування, яке має неоднорідні сенсорні системи різної природи.

Мета роботи. Розробка методу виявлення аномалій сигналів на основі трансферного навчання у поєднанні з алгоритмом екстремального градієнтного бустінгу.

Метод. Запропоновано підхід на основі трансферного навчання і екстремального градієнтного бустінгу, розроблений для виявлення аномалій в акустичних сигналах кіберфізичних систем. У цій області було проведено мало досліджень, і тому вивчалися різні архітектури заздалегідь навчених глибоких нейронних моделей, щоб поліпшити виявлення аномалій. Трансферне навчання використовує ваги з глибокої нейронної моделі, попередньо навченої на великому наборі даних, і може бути застосоване до невеликого набору навчальних даних, що забезпечує збіжність без перенавчання. Класичний підхід до такого роду проблем зазвичай включає в себе методи обробки сигналів, які дозволяють отримувати корисну інформацію з даних сенсорів. У цій статті виконується завдання виявлення аномалій з використанням архітектури глибокого навчання для роботи з акустичними сигналами, з яких попередньо витягуються спектрограми і скалограми. Функція активації SPOCU була розглянута для поліпшення точності запропонованого підходу. Алгоритм екстремального градієнтного бустінгу був використаний, оскільки він має високу продуктивність і вимагає мало обчислювальних ресурсів на етапі навчання. Застосування даного алгоритму дозволяє домогтися значного поліпшення виявлення аномалій в сигналах промислового обладнання.

Результати. Розроблений підхід реалізований програмно і досліджений під час вирішення завдання виявлення аномалій в акустичних сигналах кіберфізичних систем на наборі даних МІМІІ.

Висновки. Проведені експерименти підтвердили працездатність запропонованого підходу і дозволяють рекомендувати його для використання на практиці при вирішенні завдань діагностування стану промислового устаткування. Перспективи подальших досліджень можуть полягати в застосуванні ансамблевих підходів на основі трансферного навчання до різних реальних наборів даних для підвищення продуктивності та відмовостійкості кіберфізичних систем.

КЛЮЧОВІ СЛОВА: виявлення аномалій, акустичний сигнал, трансферне навчання, спектрограма, скалограма, кіберфізична система.

УДК 004.056

ИНТЕЛЛЕКТУАЛЬНАЯ МОДЕЛЬ ОБНАРУЖЕНИЯ АНОМАЛИЙ В КИБЕРФИЗИЧЕСКИХ СИСТЕМАХ НА ОСНОВЕ ГЛУБОКОГО ТРАНСФЕРНОГО ОБУЧЕНИЯ

Сухостат Л. В. – канд. техн. наук, доцент, Институт Информационных Технологий, Национальная Академия Наук Азербайджана, Баку, Азербайджан.

АННОТАЦИЯ

Актуальность. Рассмотрена задача обнаружения аномалий сигналов киберфизических систем на основе изображений спектрограмм и скалограмм. Объектом исследования является сложное промышленное оборудование, которое имеет неоднородные сенсорные системы различной природы.

Цель работы. Разработка метода обнаружения аномалий сигналов на основе трансферного обучения в сочетании с алгоритмом экстремального градиентного бустинга.

Метод. Предложен подход на основе трансферного обучения и экстремального градиентного бустинга, разработанный для обнаружения аномалий в акустических сигналах киберфизических систем. В этой области было проведено мало исследований, и поэтому изучались различные архитектуры заранее обученных глубоких нейронных моделей, чтобы улучшить обнаружение аномалий. Трансферное обучение использует веса из глубокой нейронной модели, предварительно обученной на большом наборе данных, и может быть применено к небольшому набору обучающих данных, что обеспечивает сходимость без переобучения. Классический подход к такого рода проблемам обычно включает в себя методы обработки сигналов, которые позволяют извлекать полезную информацию из данных сенсоров. В этой статье выполняется задача обнаружения аномалий с использованием архитектуры глубокого обучения для работы с акустическими сигналами, из которых предварительно извлекаются спектрограммы и скалограммы. Функция активации SPOCU была рассмотрена для улучшения точности предложенного подхода. Алгоритм экстремального градиентного бустинга был использован, потому что он обладает высокой производительностью и требует мало вычислительных ресурсов на этапе обучения. Применение данного алгоритма позволяет добиться значительного улучшения обнаружения аномалий в сигналах промышленного оборудования.

Результаты. Разработанный подход реализован программно и исследован при решении задачи обнаружения аномалий в акустических сигналах киберфизических систем на наборе данных МІМІІ.

Выводы. Проведенные эксперименты подтвердили работоспособность предложенного подхода и позволяют рекомендовать его для использования на практике при решении задач диагностирования состояния промышленного оборудования. Перспективы дальнейших исследований могут заключаться в применении ансамблевых подходов на основе трансферного обучения к различным реальным наборам данных для повышения производительности и отказоустойчивости киберфизических систем.

КЛЮЧЕВЫЕ СЛОВА: обнаружение аномалий, акустический сигнал, трансферное обучение, спектрограмма, скалограмма, киберфизическая система.

ЛІТЕРАТУРА / LITERATURE

1. Langner R. Stuxnet: dissecting a cyberwarfare weapon / R. Langner // *IEEE Security & Privacy*. – 2011. – Vol. 9, № 3. – P. 49–51. DOI: 10.1109/MSP.2011.67
2. Ahmed C. M. Challenges and opportunities in CPS security: a physics-based perspective / C. M. Ahmed, J. Zhou // *IEEE Security & Privacy*. – 2020. – Vol. 18, № 6. – P. 14–22. DOI: 10.1109/MSEC.2020.3002851
3. Ahmed C. M. Model-based attack detection scheme for smart water distribution networks / C. M. Ahmed, C. Murguia, J. Ruths // *Computer and Communications Security: 17th ACM ASIA Conference, Abu Dhabi, 2–6 April 2017: proceedings*. – P. 101–113. DOI: 10.1145/3052973.3053011
4. Multi-head CNN–RNN for multi-time series anomaly detection: an industrial case study / [M. Canizo, I. Triguero, A. Conde, E. Onieva] // *Neurocomputing*. – 2019. – Vol. 363. – P. 246–260. DOI: 10.1016/j.neucom.2019.07.034
5. Kiseľák J. “SPOCU”: scaled polynomial constant unit activation function / [J. Kiseľák, Y. Lu, J. Švihra et al.] // *Neural Computing and Applications*. – 2021. – № 33. – P. 3385–3401. DOI: 10.1007/s00521-020-05182-1
6. M. R. G. R. Machine learning for intrusion detection in industrial control systems: challenges and lessons from experimental evaluation / G. R. M. R., C. M. Ahmed, A. Mathur // *Cybersecurity*. – 2021. – Vol. 4, № 27. – P. 1–12. DOI: 10.1186/s42400-021-00095-5
7. Morita K. Anomalous sound detection by using local outlier factor and Gaussian mixture model / K. Morita, T. Yano, K. Q. Tran. – 2020 [Electronic resource]. Access mode: http://dcase.community/documents/challenge2020/technical_reports/DCASE2020_Morita_51_t2.pdf
8. IAEO3 – combining OpenL3 embeddings and interpolation autoencoder for anomalous sound detection / [S. Grollmisch, D. Johnson, J. Abeßer, H. Lukashevich]. – 2020 [Electronic resource]. Access mode: http://dcase.community/documents/challenge2020/technical_reports/DCASE2020_Grollmisch_15_t2.pdf
9. Michau G. Fully learnable deep wavelet transform for unsupervised monitoring of high-frequency time series / G. Michau, O. Fink. – 2021 [Electronic resource]. Access mode: <https://arxiv.org/pdf/2105.00899.pdf>
10. Deep dense and convolutional autoencoders for machine acoustic anomaly detection / G. Coelho, P. Pereira, L. Matos et al.] // *Artificial Intelligence Applications and Innovations: 17th IFIP International Conference, Hersonissos, 25–27 June 2021: proceedings*. – Cham : Springer, 2021. – P. 337–348. DOI: 10.1007/978-3-030-79150-6_27
11. Modulation spectral signal representation and i-vectors for anomalous sound detection / [P. Tiwari, Y. Jain, A. Avila et al.] – 2020 [Electronic resource]. Access mode: http://dcase.community/documents/challenge2020/technical_reports/DCASE2020_Tiwari_84_t2.pdf
12. OutlierNets: highly compact deep autoencoder network architectures for on-device acoustic anomaly detection / [S. Abbasi, M. Famouri, M. J. Shafiee, A. Wong] // *Sensors*. – 2021. – Vol. 21, № 14. – P. 1–12. DOI: 10.3390/s21144805
13. Acoustic anomaly detection for machine sounds based on image transfer learning / [R. Müller, F. Ritz, S. Illium, C. Linnhoff-Popien] // *Agents and Artificial Intelligence: 13th International Conference, 4–6 February 2021 : proceedings*. – P. 49–56. DOI: 10.5220/0010185800490056
14. Lin J. Feature Extraction based on Morlet wavelet and its application for mechanical fault diagnosis / J. Lin, L. Qu // *Journal of Sound and Vibration*. – 2000. – Vol. 234, № 1. – P. 48–135. DOI: 10.1006/jsvi.2000.2864
15. Chollet F. Xception: deep learning with depthwise separable convolutions / F. Chollet // *Computer Vision and Pattern Recognition: 30th IEEE Conference, Honolulu, 21–26 July 2017: proceedings*. – P. 1800–1807. DOI: 10.1109/CVPR.2017.195
16. Sandler M. MobileNetV2: inverted residuals and linear bottlenecks / [M. Sandler, A. Howard, M. Zhu et al.] // *Computer Vision and Pattern Recognition: IEEE/CVF Conference, Salt Lake City, 18–23 June 2018: proceedings*. – P. 4510–4520. DOI: 10.1109/CVPR.2018.00474
17. Densely connected convolutional networks / [G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger] // *Computer Vision and Pattern Recognition: 30th IEEE Conference, Honolulu, 21–26 July 2017: proceedings*. – P. 2261–2269. DOI: 10.1109/CVPR.2017.243
18. Rethinking the inception architecture for computer vision / [C. Szegedy, V. Vanhoucke, S. Ioffe et al.] // *Computer Vision and Pattern Recognition: IEEE Conference, Las Vegas, 27–30 June 2016: proceedings*. – P. 2818–2826. DOI: 10.1109/CVPR.2016.308
19. Going deeper with convolutions / [C. Szegedy, W. Liu, Y. Jia et al.] // *Computer Vision and Pattern Recognition: IEEE Conference, Boston, 7–12 June 2015: proceedings*. – P. 1–9. DOI: 10.1109/CVPR.2015.7298594
20. Kassani S. H. Diabetic retinopathy classification using a modified Xception architecture / [S. H. Kassani, P. H. Kassani, R. Khazaiezhad et al.] // *Signal Processing and Information Technology: IEEE International Symposium, Ajman, 10–12 December 2019: proceedings*. – P. 1–6. DOI: 10.1109/ISSPIT47144.2019.9001846
21. Deep residual learning for image recognition / [K. He, X. Zhang, S. Ren, J. Sun] // *Computer Vision and Pattern Recognition: IEEE Conference, Las Vegas, 27–30 June 2016: proceedings*. – P. 770–778. DOI: 10.1109/CVPR.2016.90
22. Chen T. XGBoost: a scalable tree boosting system / T. Chen, C. Guestrin // *Knowledge Discovery and Data Mining: 22nd ACM SIGKDD International Conference, San Francisco, 13–16 August 2016: proceedings*. – P. 785–794. DOI: 10.1145/2939672.2939785
23. MIMII dataset: sound dataset for malfunctioning industrial machine investigation and inspection / [H. Purohit, R. Tanabe, K. Ichige et al.] // *Detection and Classification of Acoustic Scenes and Events: 4th Workshop, New York, 25–26 October 2019: proceedings*. – P. 209–213. DOI: 10.33682/m76f-d618
24. Ferri C. An experimental comparison of performance measures for classification / C. Ferri, J. Hernández-Orallo, R. Modroiu // *Pattern Recognition Letters*. – 2009. – Vol. 30, № 1. – P. 27–38. DOI: 10.1016/j.patrec.2008.08.010

ПРОГРЕСИВНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

PROGRESSIVE INFORMATION TECHNOLOGIES

ПРОГРЕССИВНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 004.9

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПРОСУВАННЯ ІНТЕРНЕТ-РЕСУРСІВ В ПОШУКОВИХ СИСТЕМАХ НА ОСНОВІ КОНТЕНТ-АНАЛІЗУ КЛЮЧОВИХ СЛІВ WEB-СТОРІНОК

Висоцька В. А. – канд. техн. наук, доцент, доцент кафедри «Інформаційні системи та мережі», Національний університет «Львівська політехніка», Львів, Україна.

АНОТАЦІЯ

Актуальність. Своєчасний та коректний аналіз процесу відвідувань Інтернет-ресурсів, які призвели до успішної конверсії електронного бізнесу, досить важливий та актуальний для успішного управління Web-сайтом. Відповідний коректний аналіз трафіку, який як приносить успішні конверсії, так і не успішні, дасть можливість виявити причинену впливу на показники та критерії конверсії та дозволить виміряти ефективність змін, які здійснюються на сайті для збільшення конверсії трафіку. Для вирішення даних проблем та досягнення відповідних цілей електронної комерції необхідно не лише збирати інформацію щодо діяльності користувачів системи на Web-сайті, а і визначити конкретні показники ефективності Web-сайту, щоб в подальшому впливати на їхні зміни в бік покращення стратегії ведення електронного бізнесу. Таким чином треба розробити та впровадити аналітичний метод супроводу текстового контенту Інтернет-ресурсів електронної комерції на основі аналізу ключових показників ефективності Web-сайту, особливо звернувши увагу на визначення множини релевантних та актуальних ключових слів, за якими роблять переходи з пошукових систем постійні користувачі, дії яких призвели до зростання конверсії електронного бізнесу.

Метою дослідження є розроблення технології просування Інтернет-ресурсів електронної комерції на основі результатів Web-аналітики ключових показників сторінок як KPI та KSI через формування релевантної множини ключових слів як feedback активності постійної аудиторії.

Метод. Розроблено аналітичний метод просування Інтернет-ресурсів на основі аналізу ключових показників ефективності Web-сайту, який основний на трьох основних алгоритмах алгоритмі визначення проблемних місць структури сайту для подальшої оптимізації, алгоритмі оптимізації заходів з маркетингу пошукових систем, алгоритмі просування сайту та розрахунку його ефективності.

Розроблено загальні рекомендації з проектування систем опрацювання інформаційних ресурсів, відмінні від існуючих наявністю додаткових модулів, які суттєво впливають на результати просування Web-сайту в просторі Інтернет для подальшого досягнення успішності конверсії електронної комерції або збільшення значень цих показників. Серед них модуль Інтернет-магазину, маркетингу, модуль-копірайтер та Web-майстера. Для кожного з них розраховують власний KPI. Це дасть змогу ефективно просто реалізувати опрацювання інформаційних ресурсів на рівні розробника систем (зменшення ресурсів та часу на розроблення, покращення якості роботи систем опрацювання інформаційних ресурсів).

Результати. В роботі розроблено та детально описано на основі результатів процесу Web-аналітики параметри та критерії оцінювання рівня успішності ведення електронного бізнесу. Також розроблено програмні засоби супроводу текстового контенту Інтернет-ресурсів на основі аналізу ключових показників ефективності Web-сайту. Для детального аналізу функціонування та просування Інтернет-систем електронної комерції типу Інтернет-газета та Інтернет-журнал розроблено та впроваджено 12 різних систем відповідно з підтриманням для кожної з них різною кількістю етапів життєвого циклу контенту. Проведено комп'ютерний експеримент аналізу ключових показників ефективності Web-сайту. Сервіс ведення статистики відвідувань Web ресурсу дозволяє оцінити збільшення обсягів продажу текстового контенту від прямо пропорційної залежності збільшення кількості відвідувань Web ресурсу, кількості постійних користувачів, перспективності маркетингових заходів.

Висновки. Виявлено, що наявність відповідних модулів в системах опрацювання інформаційних ресурсів сприяє збільшенню обсягів реалізації текстового контенту постійному користувачу на 9 %, активного залучення унікальних відвідувачів, потенційних користувачів та розширення меж цільової та регіональної аудиторії на 11 %, переглянутих сторінок на 12 %, часу відвідування інформаційних ресурсів на 7 %.

КЛЮЧОВІ СЛОВА: контент, текстовий контент, Інтернет-ресурс, бізнес-процес, система управління контентом, життєвий цикл контенту, Інтернет-газета, Web-сайт, відвідувачі, Web-сторінка, кількість відвідань, інформаційний пошук, відсоток відвідувань, конверсія відвідувань, показник конверсії, KPI.

АБРЕВІАТУРА

OKR – Objectives and Key Results;
KPI – Key Performance Indicators;
KSI – Key Success Indicators;
BCS – Balanced Score Cards;
CPM – Cost Per Mille;
RIA – Rich Internet application;
IIA – Installable Internet application.

НОМЕНКЛАТУРА

K_{cv} – коефіцієнт конверсії;
 N_{cv} – кількість конверсій;
 N_{vt} – загальна кількість відвідувань Web-сайту;
 N_{vr} – загальна кількість відвідувачів Web-сайту;
 N_{vrb} – загальна кількість відвідувань Web-сайту, під час яких товар додавався у кошик;
 N_{vrb} – загальна кількість відвідувачів Web-сайту, під час яких товар додавався у кошик;
 S_{tm} – Web-ресурс;
 M_{is} – модуль Інтернет-магазину;
 M_{mr} – модуль маркетингу;
 M_{cp} – модуль-копірайтер;
 M_{vm} – модуль для Web-майстера;
 S_{cc} – середній коефіцієнт конверсії;
 S_{co} – середня вартість замовлень;
 S_{cv} – середня вартість на відвідування або середня корисність відвідування;
 S_{ro} – середня ROI або середнє повернення на інвестиції;
 P_{iv} – відсоток прибутку від нових відвідувачів;
 I_{nv} – індекс нових покупців при першому відвідуванні – новий визначений KPI;
ROI – розрахунок повернення на інвестиції в Google Analytics;
Income – прибуток;
Expenses – витрати;
RR – розрахунок норми прибутку;
 P_{iv} – відсоток транзакцій від нових відвідувачів;
 P_{nv} – відсоток нових відвідувачів сайту;
 P_{vz} – відсоток відвідувань за типом засобу через AdWords;
 K_{cz} – коефіцієнт конверсії цілей (у відсотках) за типом засобу через AdWords;
 P_{vk} – відсоток відвідувань за типом компанії через AdWords;
 P_{ck} – відсоток конверсії цілей за типом компанії через AdWords;

I_{ck} – індекс конверсії цілей за типом компанії;
 S_{rk} – середній ROI за типом компанії;
 P_{vk} – відсоток нових та тих, що повернулися, відвідувачів;
 P_{ap} – відсоток нових та тих, що повернулися, покупців;
 K_{vb} – коефіцієнт впізнання бренду;
 I_{yk} – індекс якості компанії;
 $P_{cv}(x)$ – функція визначення відсотку конверсії цілей для відвідувань від компанії x ;
 N_{bq} – кількість пошукових запитів із назвою бренду;
 N_{tv} – кількість прямих відвідувань Web-сайту;
 N_{aq} – загальна кількість пошукових запитів;
 S_{tp} – середня тривалість перебування на сайті через AdWords;
 S_{pv} – середня кількість переглядів сторінок за відвідування через Google Analytics;
 P_{vv} – показник відмов (у відсотках), наприклад для однієї сторінки як P_{vvp} ;
 S_{nc} – середня кількість кліків на рекламі за N_{vr} відвідувань;
 P_{zv} – відсоток зацікавленості відвідувачів;
 P_{av} – відсоток нових відвідувачів та тих, що повернулися через Google Analytics;
 P_{ht} – відсоток повторних відвідувань користувача, які відбулися з попереднього відвідування менше ніж за t_1 днів (розраховують за даними Google Analytics);
 P_{st} – відсоток повторних відвідувань користувача, які відбулися з попереднього відвідування між t_1 та t_2 днів при $t_1 < t_2$ (розраховують за даними Google Analytics);
 P_h – відсоток повторних відвідувань користувача, які відбулися з попереднього відвідування понад t_2 днів (розраховують за даними Google Analytics);
 N_{vnp} – кількість одно-сторінкових відвідувань для цієї сторінки через Google Analytics;
 N_{inp} – кількість відвідування користувачами цієї сторінки напряму через Google Analytics;
 N_{cr} – середня кількість кліків на рекламі через AdWords;
 N_{ov} – загальна кількість відвідувань через Google Analytics;

N_{vr} – кількість відвідувань для аналізу (зазвичай $N_{vr}=1000$, як для CPM);

N_{ad} – загальна кількість дій на сайті через AdWords;

N_{zv} – загальна кількість зацікавлених відвідувачів через AdWords;

N_{vk} – загальна кількість відвідувачів через Google Analytics;

K_{du} – показник кількості відвідувачів, відвідувань та переглядів сторінок;

P_{ul} – відсоток відвідувачів, у яких підтримується англійська/українська мова;

P_{ub} – відсоток відвідувачів, які використовують конкретний браузер через Google Analytics;

P_{us} – відсоток відвідувачів, які використовують конкретну операційну систему;

P_{ur} – відсоток відвідувачів, які мають високу/середню/низьку роздільну здатність екрану;

P_{up} – відсоток відвідувачів, у яких високошвидкісне підключення до Інтернет;

P_{um} – відсоток мобільних відвідувачів через Google Analytics;

P_{ep} – відсоток видання сторінок з помилкою;

K_{is} – показник внутрішнього пошуку;

S_{nvt} – середня кількість відвідувань за визначений період часу;

S_{nut} – середня кількість унікальних відвідувачів за визначений період часу – ;

S_{npt} – середня кількість переглядів сторінок за визначений період часу;

S_{nrv} – середня кількість переглядів сторінок за відвідування;

N_{ep} – загальна кількість виданих сторінок з помилкою;

N_{pp} – загальна кількість переглянутих сторінок;

P_{ius} – відсоток відвідувань, в яких використовують пошук по сайту;

S_{vrs} – середня кількість результатів пошуку, переглянутих за пошук;

P_{uos} – відсоток відвідувачів, які покинули сайт після перегляду результатів пошуку;

P_{uns} – відсоток відвідувачів, які проводять декілька пошуків по сайту на протязі відвідування (враховуючи декілька пошуків для одного і того ж ключового слова);

T_{svs} – середній час перебування на сайті для відвідування після пошуку;

S_{nup} – середня кількість сторінок, переглянутих відвідувачами після пошуку;

P_{uur} – відсоток відвідувачів, які використовують пошук по сайту;

P_{uur} – відсоток відвідувачів, які не використовують пошук по сайту;

P_{ucs} – відсоток конверсій від відвідувачів, що використовують пошук по сайту;

P_{pop} – відсоток відмов після відвідування однієї сторінки як результату пошуку;

P_{cus} – відсоток покупців серед відвідувачів, що використовують пошук по сайту;

P_{bus} – відсоток здійснених покупок серед відвідувачів, що використовують пошук по сайту;

P_{ksp} – відсоток відвідувачів, що переглянули більше k сторінок після здійсненого пошуку;

P_{uts} – відсоток відвідувачів, що провели більше t часу на сайті після здійсненого пошуку;

N_{nms} – кількість нульових результатів пошуку по сайту;

P_{nrv} – відсоток нульових результатів пошуку по сайту сторінок;

N_{nps} – загальна кількість нульових результатів пошуку сторінок;

N_{nps} – загальна кількість переглянутих сторінок пошуку;

K_{ps} – показник використання пошуку по сайту;

N_{sv} – кількість відвідувань із пошуком по сайту;

N_{ns} – кількість відвідувань без пошуку по сайту;

P_{nv} – відсоток нових відвідувачів сайту;

P_{uv} – відсоток унікальних відвідувачів;

S_{nv} – середня кількість переглядів за відвідування;

S_{rv} – середня тривалість відвідування;

S_{cc} – середній коефіцієнт конверсії;

P_{vcu} – відсоток відвідувачів, взаємодіючих з різними видами подання контенту, такими як масштабування, панорамування, перегляд наступного спілкування;

P_{sau} – відсоток відвідувачів, які запускають різні виконання події, наприклад, програти, пауза, наступний, рейтинг, натискання на рекламі;

P_{siu} – відсоток взаємодії з сайтом, тобто виконання певних дій, наприклад, підписатися, зареєструватися, зробити коментар, визначити рейтинг, додати в папку обране;

S_{Idx} – значення міри корисності сторінки;

R_{cv} – корисність мети відвідування (на основі корисності цілей) та корисність відвідування (на основі даних транзакцій електронної комерції);

R_{ec} – дохід (корисність) від електронної комерції;

N_{upv} – унікальні перегляди сторінок;

N_{vt} – кількість відвідувань;

N_{tr} – кількість транзакцій;

N_{cv} – кількість конверсій;

R_{ssv} – корисність відвідування з пошуком по сайту;

R_{snv} – корисність відвідування без пошуку по сайту;

N_{ssv} – кількість відвідувань з пошуком по сайту;

A_p – розмір прибутку.

ВСТУП

Аналіз службового контенту про відвідувачів дуже важливий, але мало хто використовує його в якості керівництва до дії в управлінні Web-ресурсом [1]. Але більшість починають розуміти, що можна суттєво збільшити прибуток – вдвічі і навіть втричі – якщо просто взнати, який трафік з більшою ймовірністю приносить конверсії, що люди роблять (і чого не роблять) на конкретному Web-сайті і як виміряти ефективність змін, які здійснюють на сайті для збільшення конверсії трафіку [2].

Цілі та ключові результати OKR пов'язані із розумінням цілей бізнесу. Необхідно спочатку усвідомити ці цілі, а вже потім займатися конкретними показниками ефективності Web-сайту [3]. Це повинно бути узгодженим, та починати треба з постановки OKR [1–5].

1. Скласти список зацікавлених осіб.

2. Провести «мозковий штурм» із зацікавленими особами.

3. Визначити список OKR (включити все, що можна оцінити як успіх для Web-сайта).

– Генерувати більше звернень, які приводять до продажів.

– Завантажувати більше каталогів у форматі PDF.

– Спонукаати клієнта до одночасної покупки декількох товарів/послуг (тим самим збільшуючи середню вартість замовлення).

– Сприяти створенню більш впізнаваного бренда або продукту.

– Збільшити трафік.

– Забезпечувати обслуговування клієнтів (скоротити кількість дзвінків в кулцентр).

– Будувати взаємовідношення із відвідувачами (наприклад, досягнути збільшення коментарів в блозі, повідомлень на форумі тощо).

4. Виділити та чітко сформулювати OKR (зосередитися на самих важливих 5–10 OKR).

Для вирішення даних проблем та досягнення відповідних цілей електронної комерції необхідно не лише збирати інформацію щодо діяльності користувачів системи на Web-сайті, а і визначити конкретні показники ефективності Web-сайту, щоб в подальшому впливати на їхні зміни в бік покращення стратегії ведення електронного бізнесу [6–8]. Таким чином треба розробити та впровадити аналітичний метод супроводу текстового контенту Інтернет-ресурсів електронної комерції на основі аналізу ключових показників ефективності Web-сайту [9–12].

Метою дослідження є розроблення інформаційної технології просування Інтернет-ресурсів електронної комерції на основі результатів Web-аналітики ключових показників сторінок.

Для досягнення мети були поставлені такі завдання:

– розробити аналітичний метод просування Інтернет-ресурсів на основі аналізу ключових показників ефективності Web-сайту;

– розробити на основі результатів процесу Web-аналітики параметри та критерії оцінювання рівня успішності ведення електронного бізнесу;

– розробити програмний засіб супроводу текстового контенту Інтернет-ресурсів на основі аналізу ключових показників ефективності Web-сайту;

– провести комп'ютерний експеримент аналізу ключових показників ефективності Web-сайту;

– отримати та проаналізувати результати експериментальної апробації запропонованої технології просування Інтернет-ресурсів електронної комерції на основі результатів Web-аналітики ключових показників сторінок.

1 ПОСТАНОВКА ПРОБЛЕМИ

KPI суттєво змінюються в залежності від сектора е-бізнесу, наприклад, роздрібна торгівля, туризм, технології, B2B, фінанси тощо [1]. Навіть всередині під секторів існують великі відмінності, наприклад, продаж путівок, турів чи авіа білетів, роздрібна торгівля сувенірами або одягом [2]. Навіть при порівнянні з конкурентами з точно такими ж цілями оцінки можуть бути тільки дуже приблизними [3]. Точний шлях, який повинні пройти відвідувачі для досягнення мети, та їх враження від цього процесу будуть відмінними для кожного Web-сайту [4]. Найбільші зміни в цих напрямках можуть зробити значний вплив на коефіцієнти конверсії [5]. Наприклад, менеджери з роздрібною торгівлі хочуть розрізнити відвідування існуючих покупців від відвідувачів не-покупців [1]. Тому використання стандартного коефіцієнта конверсії за галуззю може вести у оману. Коефіцієнти конверсії для е-комерції можна розраховувати різними способами [1].

$$K_{cv} = \frac{N_{cv}}{N_{vt}}; K_{cv} = \frac{N_{cv}}{N_{vr}}; K_{cv} = \frac{N_{cv}}{N_{vib}}; K_{cv} = \frac{N_{cv}}{N_{vrb}}.$$

В наведеному списку можна також слово конверсії замінити на слово транзакції. Іншими словами, відвідувач може здійснити покупку і, якщо йому дуже сподобався сайт та організація процесу покупки, то він вертається, щоби зробити додаткову покупку на протязі одного і того ж сеансу відвідування [6]. В залежності від того, який інструмент Web-аналітики використовують та які згоди прийняті в організації, це можна визначити як одну конверсію з двома транзакціями або дві конверсії з двома транзакціями [7–8]. Наприклад, Google Analytics покаже одну конверсію та дві транзакції, так як відвідувач став покупцем, а це може відбутися тільки один раз на протязі сеансу відвідування [1]. До інших внутрішньо сайтових факторів, які суттєво впливають на K_{cv} , та тому можуть ускладнити можливості для порівняння, відносяться наступні [9–12]:

– видимість Web-сайту в пошукових системах (природні та оплачу вальні результати пошуку);

- юзабіліті (зручність та простота використання) сайту (легкість переміщення по сайту, інтуїтивно зрозуміла система навігації);

- адекватність відображення Web-сайту у всіх основних браузерях;

- необхідність попередньої реєстрації/авторизації для покупки;

- час відповіді та завантаження сторінки;

- якість текстового та графічного контенту сторінки;

- використання факторів збільшення довіри, яких як логотипи систем безпеки покупок, політика конфіденційності, гарантія, використання шифрування для сторінок, де виконується оплата, рекомендації клієнтів тощо;

- наявність непрацюючих посилань чи відсутність деяких зображень;

- швидкий та точний пошук на сайті.

2 АНАЛІЗ ЛІТЕРАТУРНИХ ДЖЕРЕЛ

Існують відомі програмні засоби для аналізу ведення електронного бізнесу. Одним із самих відомих система Google Analytics [1]. Це потужний інструмент збору даних та складання звітів. Але він не здатний поліпшити роботу сайту [2–4]. Для того, щоб проаналізувати та інтерпретувати дані звітів, а потім приступити до дій, необхідний чіткий алгоритм дій та злагоджена команда [13]. Крім того необхідний додатковий інструмент, який був налаштований на конкретний електронний бізнес з врахуванням його особливостей [14]. В більшості бізнес-організаціях по всьому світу для оцінки результатів роботи використовують ключові показники ефективності КРІ [3]. Їх інколи називають ключовими показниками успішності KSI або збалансованою системою показників BCS. КРІ використовують в аналітичних відділах компаній для оцінки стану справ в бізнесі [7–8]. Як тільки організація встановить ОКР, їй потрібний спосіб оцінки успішності своєї діяльності. Таку оцінку дозволяють отримати ключові показники ефективності. Аналогічно, у Web-аналітиці КРІ – це Web-показник, який суттєво важливий для досягнення успіху організації в он-лайн. Вимоги до КРІ [1].

1. В більшості випадків КРІ – це коефіцієнт, відсоткове відношення або середнє, а не опрацьоване число. Це дає можливість подавати дані в контексті. Приклад неопрацьованих даних:

- Web-сайт вчора втратив 15 замовлень, так як сервер, який опрацьовує замовлення Інтернет-магазину, не відповідав на протязі M хвилини;

- За минулий тиждень втрачено N можливого прибутку, так як система замовлень не працює для відвідувачів, які використовують Firefox;

- Минулого місяця витрачено L грошових одиниць на ключові слова PPC-системи, які не принесли ніякої конверсії.

2. КРІ повинен бути пов'язаний із часом. Це підкреслює зміни та їх швидкість.

3. КРІ слугує стимулом для дій, важливих для бізнесу. Більшість параметрів можна виміряти та оцінити, але це не робить їх ключовими для успішності організації.

Більша частина складної роботи по підготовці КРІ складається із визначення ОКР. Ключові результати, які використовують для визначення цілей, – це і є КРІ. Треба лише перетворити їх в реальні Web-показники, які є в наявності [5].

Алгоритм підготовки КРІ [1–3].

1. Визначити ОКР. КРІ має відповідати бізнес-цілям організації.

2. Перетворити ОКР в КРІ, тобто визначити конкретні Web-показники, які відповідають бізнес-цілям ОКР (табл. 1), наприклад, час неперервної роботи сервера, швидкість відповіді сервера, нотатки про будь-які оффланові компанії, які можуть вливати на цифри, змін, внесені на сайті, інформацію про запуск нових товарів або відгуках користувачів тощо. Все це допоможе краще зрозуміти дані звітів та, відповідно, збільшити їх цінність [1–5].

3. Перевірити, що КРІ – це показники, які можна виміряти та які спонукають до дій.

4. Створити ієрархічні КРІ-звіти. Треба впевнитися, що до кожного отримувача КРІ-звіту приходять необхідні йому дані. Чим більш доречно запропонована інформація, з тим більшою увагою та зацікавленістю до неї віднесуться.

5. Визначити часткові КРІ. Одна із самих популярних – збільшення коефіцієнта конверсії сайту. Зазвичай цей показник легко оцінити, але в той же час він занадто суперечливий – відвідувач або здійснює конверсію, або ні. Наприклад, якщо конверсія складається в тому, щоб завантажити файл, то перехід на сторінку може бути частковим КРІ. До аналогічних часткових КРІ відносяться:

- перехід на сторінку контакти;

- для багатосторінкової форми запиту – заповнення першої сторінки;

- досягнення визначального моменту в процесі заповнення форми;

- перехід на сторінку акції;

- заповнення поля запиту пошуку по сайту.

6. Поеднати КРІ. Після формування списку необхідних КРІ для кожного представника відділу поєднайте їх та виключіть повторення. Мета КРІ – зосередитися на важливих показниках для бізнесу. Якщо в КРІ-звіті подані всі ключові фактори, які необхідні для оцінки успішності, то кожний КРІ повинен складати щонайменше 10% від цілого. Якщо важливість одного КРІ складає набагато менше 10%, то його треба видалити або поєднати з іншим, щоби отримати більш значний КРІ.

Таблиця 1 – Приклад перетворення ОКР в КРІ [1]

ОКР представників відділів	Запропоновані КРІ
Збільшити кількість відвідувачів, які приходять на сайт із пошукових систем	Відсоток відвідувачів з пошукових систем. Відсоток конверсій, які здійснюють відвідувачі, що прийшли на сайт з пошукових систем.
Продавати більше товарів/послуг	Відсоток відвідувань, під час яких відвідувачі додають товари в кошик. Відсоток відвідувань, під час яких відвідувачі додають товари у кошик та оформляють замовлення. Відсоток відвідувань, під час яких відвідувачі після додавання товарів у кошик переривають оформлення замовлення
Збільшити кількість відвідувачів, які беруть участь в роботі сайту	Відсоток відвідувань, під час яких відвідувачі залишають коментарій в блоці або завантажують документ Відсоток відвідувань під час яких відвідувачі заповнюють форму зворотного зв'язку або натискають на посилання mailto. Середній час знаходження на сайті за відвідування Середня кількість переглядів сторінок за відвідування
Продавати клієнтам більше супутніх товарів	Середня вартість замовлень Середня кількість товарів за транзакцію
Збільшити позитивне враження клієнтів від сайту	Відсоток відвідувань, під час яких була переглянута тільки одна сторінка *показник відмов) Відсоток внутрішніх пошуків по сайту, які привели до нульових результатів Відсоток відвідувань, які привели до відправлення заявки в службу підтримки.

КРІ-звіти не є щось не мінне у часі. Вони можуть та повинні змінюватися та еволюціонувати, по мірі того як представники відділів навчаються розуміти показники Web-сайту та робити дії для внесення змін [6–7]. Рекомендовано переглядати список КРІ хоча б раз в квартал. Наприклад, представнику відділу онлайн-маркетингу явно буде цікава різниця між відвідувачами з пошукових систем (SE) та не з пошуковим трафіком [14–18], та яка ймовірність конвертації цього трафіку (наприклад, конверсією може бути замовлення туру) [4–7]. Приклад КРІ-звіту.

1. В місяці $i+1$ прибуток від замовлень в он-лайн знизився на $x\%$ порівняно з місяцем i .

2. Приблизно $y\%$ всі відвідувачів Web-сайту приходять з пошукових систем.

3. Для відвідувачів з пошукових систем ймовірність входження в систему оформлення замовлення майже в k разів вища, ніж для відвідувачів, які прийшли не з пошукувачів.

4. Для відвідувачів з PPC-систем ймовірність входження в систему оформлення замовлень на $a_1-a_2\%$ вище, ніж для відвідувачів, які знайшли сайт в природних результатах пошукової видачі.

5. Так як механізм оформлення замовлення Web-сайту не підтримує браузер, відмінні від Name, Web-сайт втрачає z_1-z_2 грошових одиниць в місяць.

Дії, які повинні зробити представники відділів на основі цього КРІ-звіту:

– Перевірити, чи є причиною падіння онлайн-прибутків сезонний фактор, який характерний для всієї галузі чи тільки для онлайн-каналу.

– На перший погляд, $y\%$ відвідувачів, які приходять на сайт з пошукових систем – може бути дуже високим показником. Але чи це результат ефективної стратегії маркетингу пошукових систем чи просто інші канали працюють не дуже ефективно?

– При високих $a_1-a_2\%$ збільшить бюджет PPC-компаній – вони відмінно працюють. Але можливо гарні результати тут обумовлені недопрацюванням в області пошукової оптимізації, тому дане питання необхідно дослідити детальніше. Хоча в короткотерміновій перспективі має сенс збільшити бюджет PPC-компаній.

– Створить кращий механізм оформлення замовлень, який буде підтримувати і інші браузер.

В результаті сегментації КРІ-звітів в зазвичай отримуємо занадто багато детальної інформації, яка зовсім не потрібна для того, щоб давати вказівки робітникам [7–8]. А розробнику маркетингової стратегії така інформація, навпаки, просто необхідна, наприклад, КРІ-звіти по принципу ієрархії [1–5].

1. Директору по маркетингу компанії роздрібною торгівлі необхідні ієрархічні КРІ-звіти, наприклад,

– Середній коефіцієнт конверсії.

– Середня вартість замовлення.

– Витрати на залучення відвідувачів.

2. Розробнику маркетингової стратегії необхідна та ж інформація, сегментована за засобами (оплачені або природні результати пошуку, маркетинг по е-пошти або показ банерів тощо).

Сегментація за допомогою методу поглиблення в данні – це відмінний спосіб швидко зрозуміти поведінку різних сегментів відвідувачів. Визначивши ключові сегменти відвідувачів, які приходять на сайт, можна створити для них спеціальні профілі, щоби робити окремі звіти. Такі окремі сегментовані звіти дозволяють більш детально, швидко та ефективно досліджувати поведінку відвідувачів.

Сегментація в більшості випадків включає тип відвідувача, джерело переходу або географічне місцезнаходження відвідувача [19–21], наприклад:

– приклади сегментів за типом відвідувача;

– нові та ті що повернулися відвідувачі;

– клієнти та не клієнти;

– нові відвідувачі, які є клієнтами (або ті, що повернулися відвідувачі, які є клієнтами);

– приклади сегментів за джерелом відвідувачів;

– відвідувачі, які прийшли в результаті пошуку (або не в результаті пошуку);

– відвідувачі, які прийшли за партнерськими програмами (або не за партнерськими програмами);

- відвідувачі, які прийшли тільки за оплати вальними результатами пошуку;
- відвідувачі, які прийшли тільки за природними результатами пошуку;
- відвідувачі, які прийшли тільки завдяки розсилці по електронній пошті;
- приклади сегментів за географічним місцезнаходженням;
- тільки відвідувачі зі Львова, тільки з України, Європи тощо;
- тільки регіональні відвідувачі (Європа, Азія, Африка, Океанія тощо);
- україномовні відвідувачі (або відвідувачі, які використовують всі інші мови світу).

При виконанні сегментації необхідно знайти баланс між ясністю інформації про поведінку відвідувачів та великими обсягами даних [19–21].

Для ілюстрації прикладу за ролями в організації обрано Web-сайт Інтернет-туризму. Бізнес-цілі його двоякі: продавати тури та намагатися отримати заявки на професійні послуги. Для цього на Web-сайті є декілька основних розділів:

- Розділ Інтернет-магазин. Мета: продати туристичні тури, ціна яких відносно висока порівняно з цінами в більшості Інтернет-магазинів такого напрямку.
- Розділ генерації звернень за послугами. Мета: намагатися, щоби відвідувачі робили заявки на професійні послуги (екскурсії, тренінги, поради на проведення ексклюзивних мандрівок із гідями, консультації на складання індивідуальних турів). Це також дорогі послуги
- Розділ просування бренду. Написання статей для блогів, в яких даються позитивні практичні поради по організації відпочинку на основі послуг та турів, поданих на Web-сайті компанії.

3 МАТЕРІАЛИ ТА МЕТОДИ

З точки зору призначень на Web-сайті S_{tm} існує модуль Інтернет-магазину M_{is} , модуль маркетингу M_{mr} , модуль-копірайтер M_{cp} (допомога для написанням якісного та ефективного унікального контенту, тобто для авторів, журналістів, копірайтерів контентного Web-сайту) та M_{vm} модуль для Web-майстера. Для кожного розраховують власний KPI:

$$S_{tm} = \langle M_{is}, M_{mr}, M_{cp}, M_{vm} \rangle.$$

У сайту з модулем Інтернет-магазином, ймовірно, більше всього KPI, із яких можна обрати, так як основну мету (покупку) достатньо просто оцінити. І мета сайту (спонукати відвідувачів додати товар в кошик) визначена достатньо ясно. В Google Analytics є цілий розділ, який присвячений звітам з електронної комерції. Але більшість KPI краще взяти із інших

модулів. Окрім кількості відвідувачів N_{vr} , пропонувані KPI для модуля Інтернет-магазину:

$$M_{is} = \langle N_{vr}, S_{cc}, S_{co}, S_{cv}, S_{ro}, P_{nv}, I_{nv} \rangle.$$

S_{cc} та S_{co} розраховує Google Analytics в розділі електронна комерція. За замовченням він також розраховує два типи значень S_{cv} – корисність мети відвідування (на основі корисності цілей) та корисність відвідування (на основі даних транзакцій електронної комерції). Формула для розрахунку повернення на інвестиції в Google Analytics

$$ROI = \frac{Income - Expenses}{Expenses}.$$

Від'ємне значення ROI говорить про те, що втрачає гроші: витрати на залучення відвідувачів більше за витрати. Звичайно, при запуску нової компанії ROI, скоріше всього, буде від'ємним до тих пір, поки не збільшиться кількість відвідувачів, що повернулися, та бренд не стане впізнаним, що приведе до збільшення конверсії. Звичайно, ROI – це показник ефективності для загального валового прибутку. Він не враховує, який прибуток отримає від продаж. Він не враховує також кількість транзакцій чи відвідувачів. Наприклад, ROI для деякої компанії (занадто централізованої) може бути високим, а прибуток при цьому – невеликим. А при більш низькому ROI менш спеціалізованої компанії прибуток може бути достатньо великим завдяки великій кількості відвідувачів. Формула для норми прибутку виглядає наступним чином:

$$RR = \frac{Income - Expenses}{Income}.$$

Звіти по нормі прибутку та ROI присутні в AdWords. Для покупки дорогих товарів, в тому числі турів, зазвичай необхідно більший час на роздуми, ніж для покупки дешевих товарів, наприклад, сувенірів. Зазвичай це еквівалентно кількості відвідувань, які необхідні для переконання здійснити покупку. Таке KPI, як P_{iv} , дозволить взяти, характерно чи це для сайту з Інтернет-туризму. Значення ймовірності того, що нові відвідувачі стануть новими клієнтами при першому відвідуванні допоможе розрахувати I_{nv} :

$$I_{nv} = \frac{P_{iv}}{P_{nv}}.$$

Значення $I_{nv}=1$ говорить про те, що новий та той, що повернувся відвідувачі стануть клієнтами з однаковою ймовірністю. Значення менше 1 означає,

що новий відвідувач стане клієнтом з меншою ймовірністю, ніж той, що повернувся. А значення більше за 1 означає, що новий відвідувач стане клієнтом з більшою ймовірністю, ніж той, що повернувся. Залучення на Web-сайт гарних відвідувачів (тих, що генерують продажі або звернення за послугами) – це одна із головних задач маркетингу. Онлайн-маркетинг включає наступні джерела: пошукова оптимізація (безкоштовне ранжування в пошукових системах), PPC-реклама (платні результати пошуку), банерна реклама, партнерські мережі, маркетинг блогів, посилання з інших сайтів та маркетинг по електронній пошті.

Для визначення найкращого трафіку необхідно аналізувати коефіцієнт конверсії K_{cv} , витрати на компанію, отриманий прибуток та ROI. Тому KPI для модуля-маркетолога суттєво перетинаються з KPI для модуля-менеджера он-лайн туризму. Важлива відмінність полягає в тому, що маркетологи звертають увагу не лише на коефіцієнт конверсії для покупок, але і на конверсію цілей, так як це говорить про побудову відношень з відвідувачами, які, ймовірно, згодом перейдуть до покупки. Якщо пропустити такі показники як загальна кількість відвідувачів сайту, то KPI для модуля-маркетолога є такими:

$$M_{mr} = < P_{vz}, K_{cz}, P_{vk}, P_{ck}, I_{ck}, S_{rk}, P_{vk}, P_{ob}, K_{vb}, I_{yk} > .$$

Індекс якості компанії I_{yk} – новий KPI – пов'язаний з оцінкою того, наскільки добре побудований таргетинг компанії, тобто наскільки вони ефективні в залученні цільового трафіку на сайт он-лайн туризму.

$$I_{yk}(x) = \frac{P_{cv}(x)}{P_{vk}(x)} .$$

Припустимо, наприклад, що $P_{vk}=50\%$ відвідувачів приходять з AdWords, але цьому джерелу компанії відповідає лише $P_{cv}=20\%$ конверсій. Ця компанія працює неефективно, так як якщо дві компанії однаково таргетовані і кожна генерує 50% трафіку, то дві повинні давати 50% конверсій. Якщо одна працює ефективніше за іншу, генеруючи більше власної частини конверсій, то, за визначенням, ця компанія найкраще таргетована. Значення індексу $I_{yk}=1.0$ означає, що відвідувач з даної компанії здійснить конверсію з такою ж ймовірністю, як і відвідувач з будь-якої іншої компанії. Значення $I_{yk}<1.0$ означає відповідно, що відвідувач з даної компанії здійснить конверсію з меншою ймовірністю, ніж відвідувач з будь-якої іншої компанії. А значення $I_{yk}>1.0$ відповідно – відвідувач здійснить конверсію з більшою ймовірністю, ніж відвідувач з будь-якої іншої компанії (табл. 1).

Таблиця 2 – Приклади аналізу конверсій [1]

Компанія	P_{vk} , % відвідувань	P_{cv} , % конверсій	P_{ak} , % всіх конверсій (P_{cv} / N_{cv})	I_{yk} (P_{ak} / P_{vk})
Google organic	40,84	83,00	31,44	0,77
Referral	29,59	73,00	27,65	0,93
Direct	16,22	39,00	14,77	0,91
Google cpc	4,90	11,00	4,17	0,85
YSM ppc	3,62	3,00	1,14	0,31
Forum	2,02	51,00	19,32	9,56
Yahoo! Organic	0,56	3,00	1,14	2,04
Інше	2,25	1,00	0,38	0,17
Разом N	$N_{vk}=100$	$N_{cv}=264$		

Компанія для форуму – дуже добре таргетована компанія. Компанія для Yahoo! Organic також непогано таргетована, але кількість конверсій досить низка, тому на нього не треба звертати увагу, поки не буде зібрано більше даних. Можна втратити багато часу та зусиль на пошук причини, того, чому відвідування з Yahoo! Organic майже в три рази ефективніша, ніж з Google organic. Хоча насправді статистична вибірка занадто мала.

Коефіцієнт впізнання бренду:

$$K_{vb} = \frac{N_{bq} + N_{tv}}{N_{aq} + N_{tv}} .$$

Зверніть увагу, що в формулі пошукові запити – це ключові слова, які вводяться в пошукових системах. Прямі відвідування включені тому, що їх здійснюють люди, які знають адресу Web-сайту, а значить, цей бренд – при умові, що виключено із звітів відвідування Web-сайту співробітниками.

Для розробників контенту сайту головна мета – максимальне залучення аудиторії. Скільки часу люди витрачають на читання контенту сайту та який його осяг вони читають – ось ключові показники оцінка залучення аудиторії.

Існує три категорії контентних Web-сайтів:

– інформація про продукт та організації – сайти з корпоративною інформацією, сайти з обзорами товарів, блоги, сайти служби технічної підтримки, сайти онлайн-тренінгів тощо.

– Рекламний контент – сайти з безкоштовним контентом, які отримують прибуток від продажу реклами (банерів чи текстових об'яв), яку розміщують поряд з іншими контентом сайту.

– Контент за підпискою – поряд з отриманням прибутку від реклами, можуть пропонувати контент за підпискою, тобто користувач платить за підписку на отримання матеріалів (можливо, більш повні версії статей).

Незалежно від бізнес-моделі сайту, збільшення зацікавленості відвідувачів є ключовим фактором успіху. Тому м по контенту завжди шукають способи включити в кожну статтю чи сторінку додаткові теми, щоби збільшити цю зацікавленість. Відповідно, для сайтів кількість відвідувань за день, тиждень чи місяць – це важливий KPI.

$$M_{cp} = \langle S_{tp}, S_{pv}, P_{vv}, S_{nc}, P_{zv}, P_{av}, K_{vb}, P_{ht}, P_{st}, P_{lt} \rangle.$$

Середній показник відмов (у відсотках) P_{vv} можна дізнатися через Google Analytics. З точки зору автора контенту високе значення P_{vv} означає слабе зацікавлення відвідувачів, тобто слабкий інтерес до сайту. Сегментація в цьому випадку – найважливіша умова прийняття інформаційних рішень. Показник відмов для однієї сторінки:

$$P_{vvp} = \frac{N_{vnp}}{N_{inp}}.$$

Середня кількість кліків на рекламі за N_{vr} відвідувань:

$$S_{nc} = \frac{N_{cr}}{N_{av}} \cdot N_{vr}.$$

Показник зацікавленості відвідувачів: $K_{zv} = \frac{N_{ad}}{N_{av}}.$

Відсоток зацікавленості відвідувачів: $P_{zv} = \frac{N_{zv}}{N_{vk}},$

Для аналізу часових показників, пов'язаних з повторним відвідуванням, необхідно обрати ідеальні для конкретної моделі е-бізнесу часові проміжки $t_1 < t_2$. При успішному е-бізнесі: $P_{ht} \gg P_{st} \gg P_{lt}.$

Зазвичай цього досягти неможливо. Але при періодичному дослідженні цих показників можна вивести закономірності для корегування контенту, що в свою чергу покращує співвідношення хоча б як

$$P_{ht} \geq P_{st} \geq P_{lt}.$$

Веб-майстри відповідають за ефективну та безперервну роботу веб-сайту. Тому вони повинні знати, яке буде навантаження на сервери, тобто яку кількість звернень відвідувачів до сервера можна очікувати. Також необхідно знати, які браузерери та установки мови частіше всього використовують користувачі.

$$M_{vm} = \langle K_{du}, P_{ul}, P_{ub}, P_{us}, P_{ur}, P_{up}, P_{um}, P_{ep}, K_{is} \rangle.$$

Показник N_{du} є базовим для Веб-майстрів та визначається через Google Analytics:

$$N_{du} = \langle S_{nvt}, S_{nut}, S_{npt}, S_{npv} \rangle.$$

Відсоток видання сторінок з помилкою, який необхідно звести до мінімуму:

$$P_{ep} = \frac{N_{ep}}{N_{pp}}.$$

Показник внутрішнього пошуку також визначається через Google Analytics:

$$K_{is} = \langle P_{ius}, S_{vrs}, P_{uos}, P_{uns}, T_{svs}, S_{nup}, P_{uur}, P_{unr}, P_{ucs}, P_{pop}, P_{cus}, P_{bus}, P_{ksp}, P_{uis}, N_{nns}, P_{nnp} \rangle,$$

зокрема, $P_{nnp} = N_{nps}/N_{vps}$. Також можна розрахувати для подальшого аналізу:

$$K_{ps} = \frac{N_{sv}}{N_{ns}}.$$

У міру збільшення кількість сайтів електронної комерції, що підтримують RIA-технології, зростає необхідність визначення для них KPI. RIA або ІА – це Web-додаток, що має багато характеристик традиційного програмного забезпечення. Концепція тісно пов'язана з односторінковою програмою і може дозволити користувачеві інтерактивні функції, такі як перетягування, фонове меню, редагування WYSIWYG тощо. HTML5 є сучасним стандартом для розроблення розширених веб-програм, що підтримуються усіма основними браузерерами.

Як правило, RIA-система

- запускається локально в середовищі безпеки – «пісочниці» (sandbox) – механізм для безпечного виконання програм;

- запускається в браузері та не потребує додаткового встановлення програмного забезпечення;

- передає Web-клієнту необхідну частину користувацького інтерфейсу, залишаючи більшу частину (ресурси програми, дані, тощо) на сервері.

Аналітики повинні мислити ні з точки зору переглядів сторінок K_{du} , а дій і подій K_{as} , які вказують на взаємодії користувачів/ відвідувачів/ покупців з сайтом. Тобто необхідно згідно результатів аналізу показників перевизначити K_{as} множини дій, які мають здійснити відвідувачі, щоб це вважалося взаємодією з сайтом, зокрема,

$$K_{as} = \langle P_{nv}, P_{uv}, S_{nv}, S_{iv} \rangle.$$

Поєднуючи KPI взаємодії з сайтом K_{du} та відстеження подій K_{as} , можна визначити такі KPI:

$$K_{usa} = \alpha(K_{du}, K_{as}) = \langle P_{vcu}, P_{sau}, P_{siu}, \alpha \rangle.$$

Визначення різних KPI дозволяє зосередитися на тих елементах онлайн-стратегії, які є

найефективнішими для залучення відвідувачів, генерації звернень, конверсій і прибутків електронної комерції. Також це допоможе визначити оптимальну структуру веб-сайту для покращення ефективності його використання та збільшення обсягів постійних відвідувачів та покупців. Зокрема, таким чином можна визначити множину неефективних веб-сторінок. Займаючись аналізом даних про відвідувачів, необхідно оптимізувати сторінки сайту для ефективності перебування користувачів на ньому. У багатьох випадках покращити сайт можна, наприклад, через виправлення непрацюючих посилань, зміни URL-адрес сторінок входу для ефективного відвідування відвідувачів необхідних сторінок, або відкоригувати контент сторінки для подання необхідного рекламного повідомлення.

Алгоритм визначення проблемних місць структури сайту для подальшої оптимізації.

1. Визначення множини неефективних веб-сторінок через аналіз їх міри корисності.

2. Визначення множини популярних сторінок входу через аналіз показників відмов.

3. Аналіз джерел входу (пошукові системи, платна реклама, посилання в листах е-пошти, посилання на інших сайтах, прямі входження за адресою, наприклад, з історії попередніх відвідувань користувача або перше пряме відвідування).

4. Аналіз ключових слів входу.

5. Візуалізація переходів по сайту з боку користувача для досягнення мети.

6. Оцінка успішності пошуку по сайту.

Визначити множину неефективних сторінок інструментами веб-аналітики можна через аналіз списку показників:

– значення міри корисності сторінки $\$I_{dx}$;

– множини найпопулярніших сторінок входу/виходу (Top Landing and Exit Pages);

– дерево візуалізації послідовностей (Funnel Visualization).

Міру корисності сторінки розраховують як

$$\$I_{dx} = (R_{cv} + R_{ec}) / N_{upv}.$$

Тобто, якщо сторінку x_i переглядають відвідувачі, які досягають мету, то корисність цієї мети збільшує корисність сторінки x_i . Чим частіше переглядають сторінку x_i відвідувачі, які досягають цілі, і чим вища корисність мети, тим більшим стає значення $\$I_{dx}$. Цей спосіб оцінки корисності сторінок не має відношення до цілей і конверсії. Ранжування сторінок за значенням $\$I_{dx}$ встановлює черговість їх оптимізації. Неочікувані сторінки в цій множині (які не мають відношення до цілей) вказують на проблему структури та контенту веб-сайту. Паралельно необхідно проаналізувати самі популярні сторінки. Основне значення при аналізі множини популярних сторінок – це показник відмов; якщо відвідувачі потрапляють на сторінку входу x_i і зразу залишають сайт, то це ознака низької залученості користувачів в

роботу сайту. Якщо у сторінки входу x_i високий показник відмов, то контент сторінки x_i не задовольняє очікування користувачів. Необхідно дослідити джерело переходів як в середині сайту, так і переходів з інших джерел. Статистика низьких показників останніх спричиняє активізувати роботу в залежності від самих переходів в таких напрямках:

– пошукова оптимізація (SEO);

– кампанії з оплачуваними результатами пошуку;

– офлайн/Онлайн маркетингові заходи;

– реклама та ведення сторінок в соцмережах.

Аналіз ключових слів входу – це справжнє дослідження ринку, тобто відвідувачі повідомляють, який зміст очікували отримати при відвідуванні сайту. Візуалізація переходів по сайту з боку користувача для досягнення мети дозволить оцінити проблемні місця структури сайту, де потенційний відвідувач/покупець зіштовхується з проблемами, наприклад, некоректні або незрозумілі або складні етапи оплати/оформлення замовлення.

Пошук по сайту є внутрішнім пошуковим механізмом, який відвідувачі часто замінюють системою навігації по сайту або меню. Для крупних веб-сайтів з сотнями або тисячами сторінок контенту механізм внутрішнього пошуку є важливим компонентом для відвідувачів, який дозволяє швидко знайти необхідний контент. Для внутрішніх механізмів пошуку зазвичай використовують таку ж архітектуру та механізми, як і для зовнішніх пошукових систем як Google. Оцінка успішності пошуку по сайту полягає в аналізі показника відмов, а також ще ряду інших показників, зокрема:

– показник досягнутої конверсії $P_{cv} = N_{cv} / N_{vt} \cdot 100\%$.

– показник доходу $P_{ip} = R_{cv} + R_{ec}$.

– показник середньої корисності:

$$S_{cv} = (R_{cv} + R_{ec}) / (N_{cv} + N_{tr}).$$

– рейтинг конверсії в електронній торгівлі:

$$R_{ecc} = N_{tr} / N_{vt} \cdot 100\%.$$

– показник корисності відвідування:

$$P_{uv} = (R_{cv} + R_{ec}) / N_{vt}.$$

Відвідувач, який використовує пошук по сайту, в декілька разів цінніший, ніж відвідувач, який пошук по сайту не використовує. Тому розроблення та розвиток сервісу пошуку по сайту ефективно впливає на показники відвідування сайту за збільшення обсягів постійної аудиторії. Для цього використовують обчислення впливу на дохід функції пошуку по сайту $I_{ssp} = (R_{ssv} - R_{smv}) \cdot N_{ssv}$.

Цей показник регулює плани та стратегії щодо подальших інвестицій в розвиток сервісу пошуку по сайту. Цей показник має становити 80% місячного доходу для веб-сайту.

Алгоритм оптимізації заходів з маркетингу пошукових систем (SEM).

1. Дослідження ключових слів (для оплачуваних/неоплачуваних результатів пошуку).

– Відвідувачі, які прийшли відповідно до природних пошукових результатів.

– Відвідувачі, які використовують внутрішній пошук по сайту.

2. Оптимізація кампанії (оплачувані результати пошуку).

3. Оптимізація сторінки входу і SEO (пошукова оптимізація) (для оплачуваних/неоплачуваних результатів пошуку).

4. Оптимізація позицій оголошення для кампанії AdWords (оплачувані результати пошуку).

– Оптимізація позицій за відвідуванням (сторінки/відвідування, середня тривалість перебування на сайті)

– Оптимізація позицій за відсотком нових відвідувань (показник відмов, показник досягнутих переходів для мети 1 [для цілей 2–4], показник досягнутої конверсії, [прибуток, транзакції, середня корисність, рейтинг конверсії в електронній торгівлі, корисність відвідування])

– Оптимізація позицій за корисністю відвідування

– Оптимізація за часом дня в AdWords

5. Оптимізація версій оголошень для кампанії AdWords (оплачувані результати пошуку).

Для ключових слів, які приносять конверсії, необхідно оптимізувати інвестиції, встановлюючи максимальну плату за клік (CPC) в AdWords. Величина повернення на інвестиції (ROI) повинна бути позитивною, тобто отриманий дохід повинен перевищувати затрати, тобто:

$$ROI = (Income - Expenses) / Expenses \cdot 100\% > 0.$$

ROI для валового прибутку складає

$$ROI_{vp} = ((Income \cdot A_p) / 100 - Expenses) / Expenses \cdot 100\%.$$

Тобто можна розрахувати на скільки відсотків ($q\%$) більше грошей дозволено витратити на конкретне ключове слово в AdWords, без ризику отримати негативний ROI. На початку запуску кампанії ROI може бути негативним, поки бренд і веб-сайт невідомі. Відвідувачам зазвичай потрібно кілька відвідувань нового для них веб-сайту, перш ніж вони здійснять конверсію. Але така ситуація (з негативним ROI) може бути прийнятною тільки протягом короткого періоду часу – приблизно кілька тижнів, в залежності від ситуації. Щоб обчислити максимальну суму, яку можна витратити на залучення відвідувачів – максимальні витрати на залучення C_{max} , необхідно скористатися формулою:

$$C_{max} = (Income \cdot A_p / 100) / (ROI_{vp} / 100 + 1).$$

Знаючи коефіцієнт конверсії для кожного ключового слова, тепер можна вирахувати

максимальну плату за клік C_{max} , допустиму для даного ключового слова:

$$C_{max} = C_{amax} \cdot R_{ecc} / 100.$$

Результатом описаної системи є те, що не треба переплачувати за ключові слова AdWords.

Тематика ключових слів – це термін, який використовується в маркетингу пошукових систем для опису набору ключових слів, які точно описують контент сторінки. Коректно визначена тематика ключових слів для пошукових систем значно покращує ефективність відвідування користувачами системи як результат пошуку.

Як правило, теми зазвичай містять 5–10 фраз на сторінку, ключові слова в яких перетинаються. Більше десяти пересічних фраз послаблюють вплив і ефективність сторінки – з точки зору і вражень користувача, і ранжирування в пошукових системах. Якщо вже є сторінка, зміст якої орієнтоване на більш ніж 10 ключових фраз, то краще створити окрему сторінку, присвячену додатковим ключовим словам. Основні поради:

1. Завжди треба ставити на перше місце інтереси відвідувачів і клієнтів.

2. Для кампаній треба використовувати спеціальні сторінки входу – для відвідувачів, які прийшли як по оплаченим, так і неоплаченим результатам пошуку.

3. Сторінки входу мають знаходитися поряд заклику до дії.

4. Контент веб-сайту треба будувати навколо тематики ключових слів із 5–10 ключових слів і фраз, що перетинаються.

5. Треба розмістити контент, насичений ключовими словами, ближче до верху сторінки.

6. Треба використовувати в HTML-тегах `<title>` ключові слова.

7. Треба використовувати ключові слова в анкерах, тобто в HTML-тегах `<a>`.

8. Треба уникати розміщувати текст в зображеннях, Flash або в іншому упровадженому контенті.

9. Треба використовувати файл robots.txt для контролю над тим, які сторінки повинні індексуватися пошуковими системами.

10. Не треба зловживати ключовими словами і не займатися спамом пошукових систем.

1. Присвоєння корисності цілям.

2. Активація звітів електронної торгівлі.

– Визначити необмежену кількість цілей (стандартна кількість – 4 цілі на кожний профіль).

– Визначити кількість часу та відвідувань, які необхідні користувачу, щоб здійснити конверсію.

– Дослідити обсяги вкладу кожної цілі (товару) в загальний прибуток веб-сайту.

– Групувати цілі за категоріями.

– Генерувати списки окремих транзакцій як окремих цілей.

3. Відслідковування некомерційного контенту сайту як елементів електронної комерції (завантаження pdf-файлів, зображень тощо).

4. Відслідковування оффлайнних маркетингових заходів або офлайнних відвідувачів.

– Престижні URL-адреси – в разі відомого бренду весь веб-контент розміщати на одному центральному домені.

– Кодовані URL-адреси – в разі відомого бренду або якщо у продуктів вже є окремі веб-сайти.

– Поєднання з пошуком – відомість бренду менша, ніж відомість продуктів або послуг, або цільова аудиторія більше орієнтована на ціну, ніж на бренд.

4 ЕКСПЕРИМЕНТИ

Для детального аналізу функціонування та просування Інтернет-систем електронної комерції типу Інтернет-газета та Інтернет-журнал розроблено та впроваджено 12 різних систем відповідно з підтриманням для кожної з них різною кількістю етапів життєвого циклу контенту. Тобто для різних реалізованих систем були розроблені не всі модулі або взагалі не були розроблені підсистеми опрацювання інформаційних ресурсів як модуль Інтернет-магазину M_{is} , модуль маркетингу M_{mr} , модуль-копірайтер M_{cp} та M_{vm} модуль для Web-майстера. В табл. 3 подано список реалізованих Інтернет-систем електронної комерції із вказанням наявності реалізованих підсистем опрацювання інформаційних ресурсів з підтриманням життєвого циклу текстового контенту.

Таблиця 3 – Реалізація модулів опрацювання інформаційних ресурсів в розроблених системах

№	Адреса	M_{is}	M_{mr}	M_{cp}	M_{vm}
1	fotoghalereja-vysocjkykh.com	–	+/-	+	+/-
2	vgholos.com.ua	+	+	+	+
3	tatjana.in.ua	–	–	+/-	+/-
4	presstime.com.ua	+	+/-	+	+
5	www.autochip.vn.ua	+/-	–	+	+/-
6	kursyvalyut.com	–	+	+	–
7	dobryjranok.com	+/-	+/-	–	+/-
8	goodmornin-gua.com	+/-	+/-	+/-	–
9	zsh3lviv.in.ua	–	–	–	–
10	victana.lviv.ua	–	–	+/-	+
11	Colins.in.ua	–	–	+	+/-
12	Momlet.in.ua	–	–	+/-	+/-

5 РЕЗУЛЬТАТИ

В табл. 4 подані результати роботи розроблених систем згідно Google Analytics за період часу 2018–2021 р. Аналіз результатів супроводу текстового контенту дозволяє визначити причин формування цільової аудиторії за набором характеристик функціонування Web сайту. Регулюючи тематичний набір текстового контенту, його унікальність, оперативність його формування та адекватне управління ним згідно індивідуальних потреб постійного користувача, можна моделювати межі цільової соціальної аудиторії та кількість унікальних відвідувачів з пошукових систем.

На рис. 1–3 подані результати роботи розроблених систем у вигляді графіків, з яких випливає, що при наявності всіх етапів життєвого циклу контенту суттєво збільшується обсяг відвідувань та унікальних користувачів. Спадання на рис. 1 обумовлено відключенням модулів на ресурсі.

Таблиця 4 – Результати роботи систем за період часу з 10.2018р. по 03.2021 р

Показник	1	2	3	4	5	6	7	8	9	10	11	12
N_{vt}	4865	5997052	1381	3654456	9606	20132	8724	25	7	3138	7583	1578
T_{sys}	4:41	2:14	3:56	2:04	1:51	1:02	2:27	8:12	0:46	4:15	3:24	1:18
P_{pop}	56,14	71,90	53,15	83,08	55,67	82,92	68,15	48,0	97,02	32,92	75,87	96,25
K_{cv}	7,83	3,3	0,06	0,31	12,51	0,04	0,12	0,07	0,01	4,5	2,3	1,2
$P_{nv} \cdot N_{vt}/100$	3215	2501402	728	1501202	7105	16586	4996	734	562	1345	4824	548
N_{pp}	22071	11588861	5464	769923	24908	31982	18892	81	12	18132	9387	2876
K_{du}	4,54	1,93	3,96	1,67	2,59	1,59	2,17	3,24	1,67	5,78	3,17	2,02
P_{nv}	65,45	41,68	52,57	39,88	73,88	82,39	57,23	28,0	97,32	42,86	35,32	12,31
$P_{av} - P_{nv}$	34,55	58,32	47,43	60,12	26,12	17,61	42,77	72,0	2,68	57,14	24,41	7,65
P_{zv}	91,14	52,48	62,49	42,46	88,13	58,22	26,984	23	6,72	31,90	21,45	12,43
P_{uur}	45,67	36,10	22,23	31,22	8,91	34,83	24,34	0	0,13	26,04	4,35	2,54
N_{tv} / N_{vt}	2,14	11,20	7,53	26,12	2,35	5,34	48,73	77	93,12	27,88	2,31	1,25
P_{us}	1,99	0,20	3,19	0	0,58	1,55	0,04	0	0,02	6,37	0,32	0,12
P_{tv}	0,06	0,02	4,56	0	0,02	0,06	0,01	0	0,01	7,81	1,49	0,36

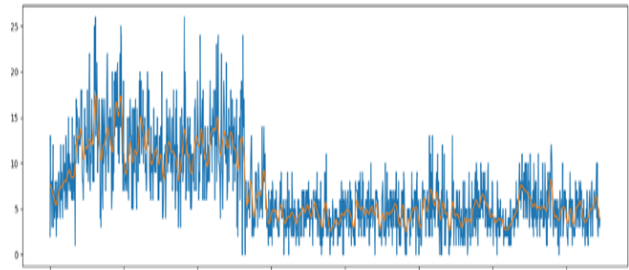
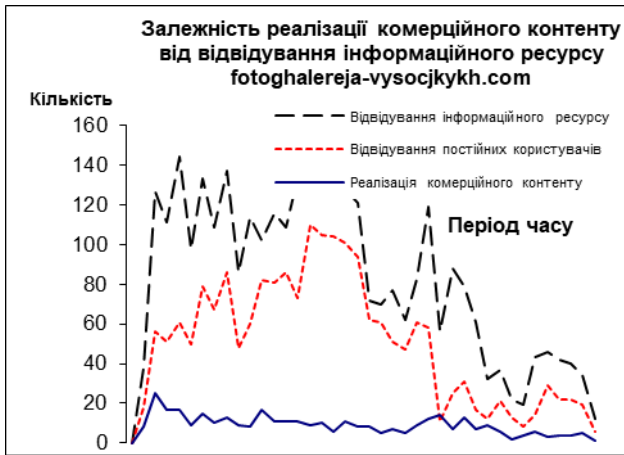


Рисунок 1 – Аналіз функціонування ресурсу fotoghalereja-vysocjkykh.com

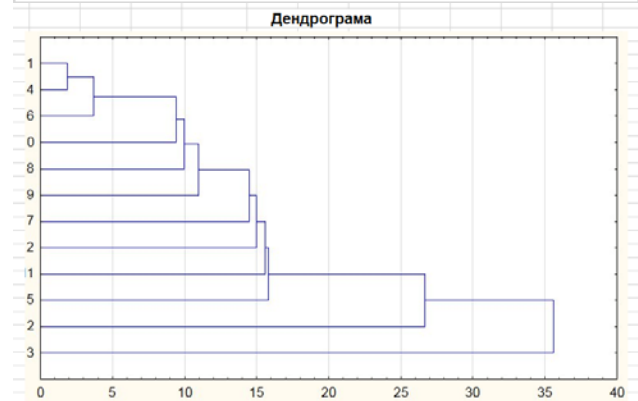
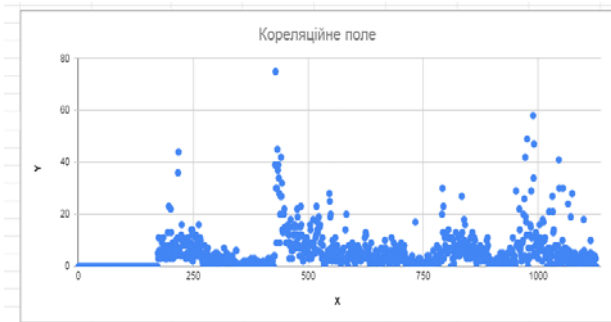
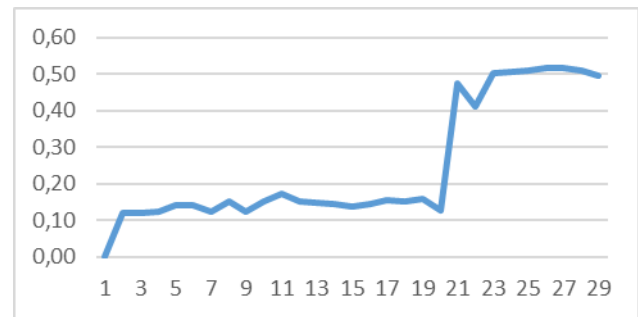
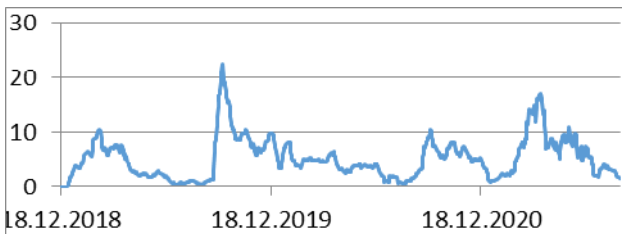


Рисунок 3 – Графік кореляції для victana.lviv.ua

Рисунок 2 – Аналіз функціонування victana.lviv.ua

Пік зростання на рис. 2–3 припадає на період підключення відповідних модулів в структуру сайту. Аналогічно на ресурсі kursyvalyut.com також в певний момент були відключені розроблені модулі, що сприяло до суттєвого зниження кількості відвідувань, відповідно значно знизило значення коефіцієнта конверсії (рис. 4).

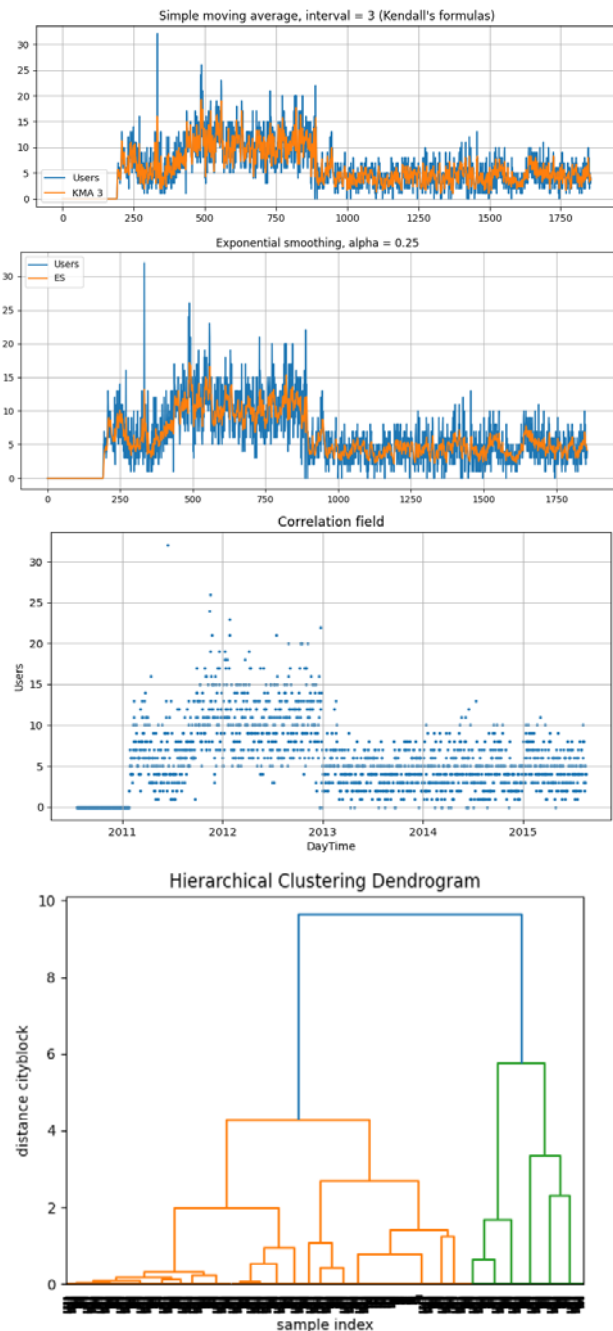


Рисунок 4 – Аналіз функціонування kursyvalyut.com

Також повисився такий трафік переходів для нових сесій з пошукових систем для victana.lviv.ua (рис. 5) та знизився трафік для kursyvalyut.com (рис. 6).

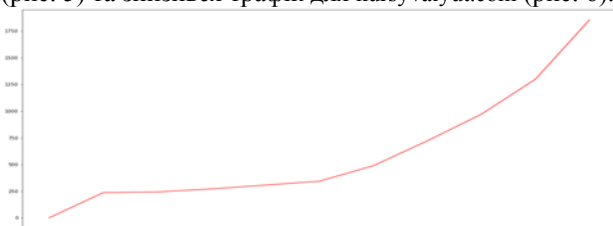


Рисунок 5 – Аналіз трафіку переходів victana.lviv.ua

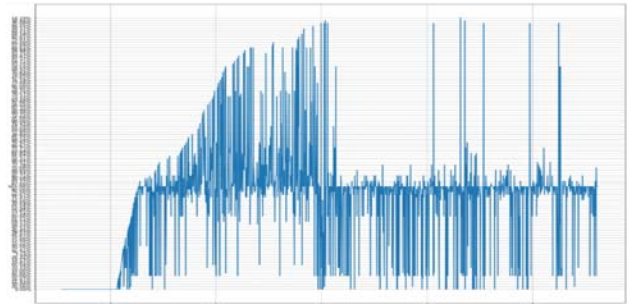


Рисунок 6 – Аналіз трафіку переходів kursyvalyut.com

Для ресурсу tatjana.in.ua він виріс на початках запуску та залишається на постійному майже рівні (модулі розроблені частково і не всі, ніколи не відключалися). Суттєвих зростань піків не спостерігається (рис. 7), а ось зниження відвідування систематично припадає на відпуски та свята.

Суттєвих зростань піків не спостерігається (рис. 7), а ось зниження відвідування систематично припадає на відпуски та свята.

6 ОБГОВОРЕННЯ

Сервіс ведення статистики відвідувань Web ресурсу дозволяє оцінити збільшення обсягів продажу текстового контенту від прямо пропорційної залежності збільшення кількості відвідувань Web ресурсу, кількості постійних користувачів, перспективності маркетингових заходів (рис. 8).

Наявність $S_{tm} = \langle M_{is}, M_{mr}, M_{cp}, M_{vm} \rangle$ відповідних модулів в системах опрацювання інформаційних ресурсів сприяє збільшенню обсягів реалізації текстового контенту постійному користувачу на 9%, активного залучення унікальних відвідувачів, потенційних користувачів та розширення меж цільової та регіональної аудиторії на 11%, переглянутих сторінок на 12%, часу відвідування інформаційних ресурсів на 7%.

ВИСНОВКИ

Стаття присвячена технології просування Інтернет-ресурсів електронної комерції на основі результатів Web-аналітики ключових показників сторінок як KPI та KSI через формування релевантної множини ключових слів як feedback активності постійної аудиторії. В роботі розроблений аналітичний метод просування Інтернет-ресурсів на основі аналізу ключових показників ефективності Web-сайту, який основний на трьох основних алгоритмах як:

1. Алгоритм визначення проблемних місць структури сайту для подальшої оптимізації.
2. Алгоритм оптимізації заходів з маркетингу пошукових систем (SEM).
3. Алгоритм просування сайту та розрахунку його ефективності.

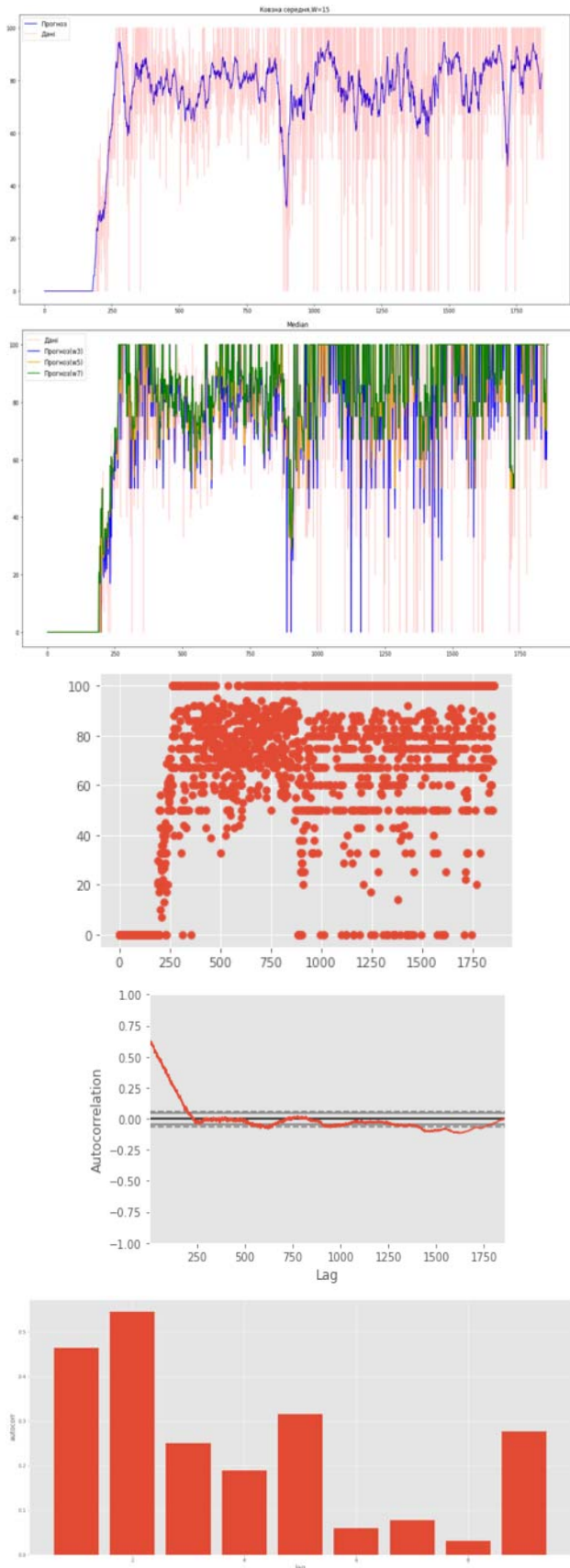


Рисунок 7 – Аналіз трафіку переходів tatjana.in.ua

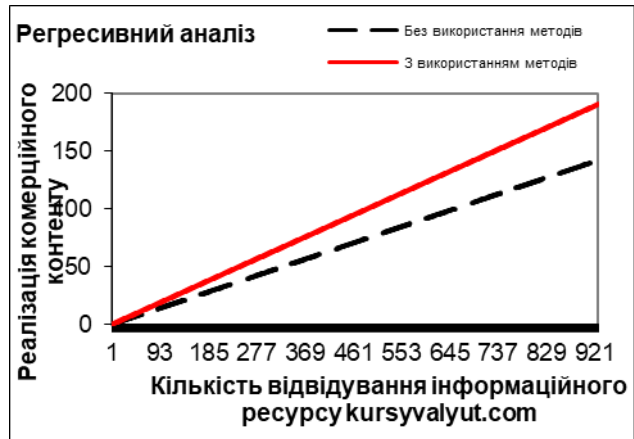


Рисунок 8 – Регресивний аналіз збільшення продажів текстового контенту

Розроблено загальні рекомендації з проектування систем опрацювання інформаційних ресурсів, відмінні від існуючих наявністю додаткових модулів, які суттєво впливають на результати просування Web-сайту в просторі Інтернет для подальшого досягнення успішності конверсії електронної-комерції або збільшення значень цих показників. Серед них модуль Інтернет-магазину M_{is} , модуль маркетингу M_{mr} , модуль-копірайтер M_{cp} (допомога для написанням якісного та ефективного унікального контенту, тобто для авторів, журналістів, копірайтерів контентного Web-сайту) та M_{vm} модуль для Web-майстера. Розроблено структури модулів системи опрацювання інформаційних ресурсів для реалізації етапів життєвого циклу текстового контенту.

Для кожного з них розраховують власний KPI. Це дасть змогу ефективно просто реалізувати опрацювання інформаційних ресурсів на рівні розробника систем (зменшення ресурсів та часу на розроблення, покращення якості роботи систем опрацювання інформаційних ресурсів). У сайта з модулем Інтернет-магазином, ймовірно, більше всього KPI, із яких можна обрати, так як основну мету (покупку) достатньо просто оцінити. І мета сайту (спонукати відвідувачів додати товар в кошик) визначена достатньо ясно. В Google Analytics є цілий розділ, який присвячений звітам з електронної комерції. Але більшість KPI краще взяти із інших модулів. Окрім кількості відвідувачів N_{vr} , в роботі запропоновані додаткові KPI для модуля Інтернет-магазину. Для визначення найкращого трафіку необхідно аналізувати коефіцієнт конверсії K_{cv} , витрати на компанію, отриманий прибуток та ROI. Тому KPI для модуля-маркетолога суттєво перетинаються з KPI для модуля-менеджера он-лайн туризму. Важлива відмінність полягає в тому, що маркетологи звертають увагу не лише на коефіцієнт конверсії для покупок, але і на конверсію цілей, так як це говорить про побудову відношень з відвідувачами, які, ймовірно, згодом перейдуть до покупки.

Незалежно від бізнес-моделі сайту, збільшення зацікавленості відвідувачів є ключовим фактором успіху. Тому по контенту завжди шукають способи включити в кожну статтю чи сторінку додаткові теми, щоби збільшити цю зацікавленість. Відповідно, для сайтів кількість відвідувань за день, тиждень чи місяць – це важливий KPI. Веб-майстри відповідають за ефективну та безперервну роботу веб-сайту. Тому вони повинні знати, яке буде навантаження на сервери, тобто яку кількість звернень відвідувачів до сервера можна очікувати. Також необхідно знати, які браузерери та установки мови частіше всього використовують користувачі.

В роботі розроблено та детально описано на основі результатів процесу Web-аналітики параметри та критерії оцінювання рівня успішності ведення електронного бізнесу. Також розроблено програмні засоби супроводу текстового контенту Інтернет-ресурсів на основі аналізу ключових показників ефективності Web-сайту. Для детального аналізу функціонування та просування Інтернет-систем електронної комерції типу Інтернет-газета та Інтернет-журнал розроблено та впроваджено 12 різних систем відповідно з підтриманням для кожної з них різною кількістю етапів життєвого циклу контенту. Проведено комп'ютерний експеримент аналізу ключових показників ефективності Web-сайту. Розроблено та впроваджено модулі Інтернет-магазину, маркетингу, модуль-копірайтер та Web-майстера для досягнення ефекту від роботи на рівні власника (підвищення рентабельності, зростання інтересу користувачів) та користувача (зрозумілість, спрощення інтерфейсу, уніфікація, розширення вибору) систем опрацювання інформаційних ресурсів. Розроблено метод супроводу текстового контенту на основі аналізу статистики функціонування системи опрацювання інформаційних ресурсів для зміни значень параметрів управління та вимог формування текстового контенту, що дало змогу збільшити обсяги реалізації текстового контенту постійному користувачу на 9%. Вдосконалено структуру системи опрацювання інформаційних ресурсів на основі аналізу процесів опрацювання інформаційних ресурсів, відмінну від існуючих наявністю підсистем Інтернет-магазину, маркетингу, модуль-копірайтер та Web-майстера, що дало можливість реалізувати етапи життєвого циклу текстового контенту та розробити рекомендації з проектування типових систем. Розроблено рекомендації з проектування структури системи опрацювання інформаційних ресурсів, відмінної від існуючих деталізацією етапів та наявністю підсистем опрацювання інформаційних ресурсів, що дають змогу підтримувати життєвий цикл текстового контенту на рівні розробника систем (зменшення ресурсів та часу на розроблення, покращення якості роботи систем).

Розроблено та впроваджено програмні засоби для аналізу результатів просування Інтернет-ресурсів електронної комерції на основі результатів Web-

© Висоцька В. А., 2021
DOI 10.15588/1607-3274-2021-3-12

аналітики ключових показників сторінок як KPI та KSI через формування релевантної множини ключових слів як feedback активності постійної аудиторії. Виявлено, що наявність відповідних модулів в системах опрацювання інформаційних ресурсів сприяє збільшенню обсягів реалізації текстового контенту постійному користувачу на 9%, активного залучення унікальних відвідувачів, потенційних користувачів та розширення меж цільової та регіональної аудиторії на 11%, переглянутих сторінок на 12%, часу відвідування інформаційних ресурсів на 7%. Отримано та проаналізовано результати експериментальної апробації запропонованої технології просування Інтернет-ресурсів електронної комерції на основі результатів Web-аналітики ключових показників сторінок. Використані відповідні модулі сприяє збільшенню активного залучення потенційних користувачів та розширення меж цільової аудиторії на 11% для покращення функціонування системи опрацювання інформаційних ресурсів на рівні власника (підвищення рентабельності, зростання інтересу користувачів) та користувача (зрозумілість, спрощення інтерфейсу, автоматизація процесів опрацювання інформаційних ресурсів та розширення вибору функціональних можливостей).

ПОДЯКИ

Роботу виконано в рамках держбюджетної теми «Методи та засоби функціонування систем підтримки прийняття рішень на основі онтологій» (ID:839 2017-05-15 09:20:01 (2459–315)). Дослідження провадилось в межах спільних наукових досліджень кафедри інформаційних систем та мереж НУ «Львівська політехніка» на тему «Дослідження, розроблення і впровадження інтелектуальних розподілених інформаційних технологій та систем на основі ресурсів баз даних, сховищ даних, просторів даних та знань з метою прискорення процесів формування сучасного інформаційного суспільства». Наукові дослідження провадилися також в рамках ініціативної тематики досліджень кафедри ICM НУ «Львівська політехніка» на тему «Розроблення інтелектуальних розподілених систем на основі онтологічного підходу з метою інтеграції інформаційних ресурсів».

ЛІТЕРАТУРА / LITERATURA

1. Clifton B. Advanced web metrics with Google Analytics / B. Clifton. – Indianapolis : John Wiley & Sons, 2012. – 589 p.
2. Sulova S. A system for e-commerce website evaluation / S. Sulova. – Access mode: https://www.researchgate.net/profile/Snezhana-Sulova/publication/334734832_A_System_for_E-Commerce_Website_Evaluation/links/5d4549a0299bf1995b60d51f/A-System-for-E-Commerce-Website-Evaluation.pdf
3. Saura J. R. Understanding the digital marketing environment with KPIs and web analytics / J. R. Saura, P. Palos-Sánchez, L. M. Cerdá Suárez // Future Internet. – 2019. – 9(4). – P. 76. DOI: <https://doi.org/10.3390/fi9040076>
4. García M. D. M. R. An ontology-based data integration approach for web analytics in e-commerce /

- M. D. M. R. García, J. García-Nieto, J. F. Aldana-Montes, // *Expert Systems with Applications*. – 2016. – Vol. 63. – P. 20–34. DOI: <https://doi.org/10.1016/j.eswa.2016.06.034>
5. Heller D. Web analytics: functions, KPIs and reports in SMEs. – Access mode: <https://kola.opus.hbz-nrw.de/opus45-kola/frontdoor/deliver/index/docId/1295/file/BachelorThesisDominatorHeller.pdf>
 6. Golyash I. The performance audit of a corporate website as a tool for its internet marketing strategy / I. Golyash, V. Panasiuk, S. Sachenko // *EUREKA: Social and Humanities*. – 2017. – Vol. 5. – P. 57–66.
 7. Rodello I. A. Evaluation of the impact of promotional campaign through a social networks on the key performance indicators of website for online of group-buying in Brazil / I. A. Rodello, V. Dandolo, M. M. Grande // *European Journal of Management Issues*. – 2016. – Vol. 7. – P. 244–249.
 8. Performance evaluation of the electronic commerce systems / [S. V. Shaytura, Y. P. Kozhayev, K. V. Ordov et al.] // *Performance evaluation*. – 2017. – Vol. 38. – P. 1–11.
 9. Design of the architecture of an intelligent system for distributing commercial content in the internet space based on SEO-technologies, neural networks, and Machine Learning / [V. Lytvyn, V. Vysotska, A. Demchuk et al.] // *Eastern-European Journal of Enterprise Technologies*. – 2019. – Vol. 2 (2–98). – P. 15–34. DOI: <https://doi.org/10.15587/1729-4061.2019.164441>
 10. Batiuk T. Intelligent system for socialization by personal interests on the basis of SEO technologies and methods of machine learning / T. Batiuk, V. Vysotska, V. Lytvyn // *Computational Linguistics and Intelligent Systems (COLINS 2020)* : 4th International Conference, Lviv, 23–24 April 2020 : CEUR workshop proceedings. – Aachen: CEUR-WS.org, 2020. – Vol. 2604. – P. 1237–1250.
 11. Online Tourism System Development for Searching and Planning Trips with User's Requirements / [N. Antonyuk, M. Medykovskyy, L. Chyrun et al.] // *Advances in Intelligent Systems and Computing*. – 2020. – Vol. 1080. – P. 831–863. doi: https://doi.org/10.1007/978-3-030-33695-0_55
 12. Application of Online Marketing Methods and SEO Technologies for Web Resources Analysis within the Region / [V. Kuchkovskiy, V. Andrunyk, M. Krylyshyn et al.] // *Computational Linguistics and Intelligent Systems (COLINS 2021)* : 5th International Conference, Lviv, 22–23 April 2021 : CEUR workshop proceedings. – Aachen: CEUR-WS.org, 2021. – Vol. 2870. – P. 1652–1693.
 13. An Intelligent System for Commercial of Information Products Distribution Based SEO and Sitecore CMS / [B. Rusyn, L. Pohreliuk, O. Kapshii et al.] // *Computational Linguistics and Intelligent Systems (COLINS 2020)* : 4th International Conference, Lviv, 23–24 April 2020 : CEUR workshop proceedings. – Aachen: CEUR-WS.org, 2020. – Vol. 2604. – P. 760–777.
 14. Using Internet News Flows as Marketing Data Component / [S. Orekhov, H. Malyhon, I. Liutenko, T. Goncharenko] // *Computational Linguistics and Intelligent Systems (COLINS 2020)* : 4th International Conference, Lviv, 23–24 April 2020 : CEUR workshop proceedings. – Aachen: CEUR-WS.org, 2020. – Vol. 2604. – P. 358–373.
 15. Pavlenko O. Search Query Data Analysis: Challenges and Opportunities / O. Pavlenko, I. Tymofieieva // *Computational Linguistics and Intelligent Systems (COLINS 2020)* : 4th International Conference, Lviv, 23–24 April 2020 : CEUR workshop proceedings. – Aachen: CEUR-WS.org, 2020. – Vol. 2604. – P. 452–461.
 16. Estimation of the Level of Interest and Modeling of the Topic of Innovation Through Search in Google / [O. Kliuiev, N. Vnukova, S. Hlibko et al.] // *Computational Linguistics and Intelligent Systems (COLINS 2020)* : 4th International Conference, Lviv, 23–24 April 2020 : CEUR workshop proceedings. – Aachen: CEUR-WS.org, 2020. – Vol. 2604. – P. 523–535.
 17. Radiuk P. A Framework for Exploring and Modelling Neural Architecture Search Methods / P. Radiuk, N. Hryppynska // *Computational Linguistics and Intelligent Systems (COLINS 2020)* : 4th International Conference, Lviv, 23–24 April 2020 : CEUR workshop proceedings. – Aachen: CEUR-WS.org, 2020. – Vol. 2604. – P. 1060–1074.
 18. Choosing the Method of Finding Similar Images in the Reverse Search System / [O. Veres, B. Rusyn, A. Sachenko, I. Rishnyak] // *Computational Linguistics and Intelligent Systems (COLINS 2018)* : 2nd International Conference, Lviv, 25–27 June 2018 : CEUR workshop proceedings. – Aachen: CEUR-WS.org, 2018. – Vol. 2136. – P. 99–107.
 19. Basyuk T. Mathematical Model of Semantic Search and Search Optimization / T. Basyuk, A. Vasyliuk, V. Lytvyn // *Computational Linguistics and Intelligent Systems (COLINS 2019)* : 3rd International Conference, Kharkiv, 18–19 April 2019 : CEUR workshop proceedings. – Aachen: CEUR-WS.org, 2019. – Vol. 2362. – P. 96–105.
 20. Adamuthe A. Adaptive harmony search for optimizing constrained resource allocation problem / A. Adamuthe, T. Nitave // *International Journal of Computing*. – 2018. – Vol. 17(4). – P. 260–269.
 21. Search engine marketing is not all gold: Insights from Twitter and SEO Clerks / [R. Aswani, A. K. Kar, P. V. Ilavarasan, Y. K. Dwivedi] // *International Journal of Information Management*. – 2018. – Vol. 38(1). – P. 107–116.

Стаття надійшла до редакції 25.05.2021
Після доробки 22.07.2021

УДК 004.9

ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ ПРОДВИЖЕНИЯ ИНТЕРНЕТ-РЕСУРСОВ В ПОИСКОВЫХ СИСТЕМАХ НА ОСНОВЕ КОНТЕНТ-АНАЛИЗА КЛЮЧЕВЫХ СЛОВ WEB-СТРАНИЦ

Высоцкая В. А. – канд. техн. наук, доцент, доцент кафедры «Информационные системы и сети», Национальный университет «Львовская политехника», Украина.

АННОТАЦИЯ

Актуальность. Своевременный и корректный анализ процесса посещений интернет-ресурсов, которые привели к успешной конверсии электронного бизнеса, достаточно важный и актуальный для успешного управления Web-сайтом. Соответствующий корректный анализ трафика, который не только приносит успешные конверсии, так и не успешные, даст возможность выявить причине не влияния на показатели и критерии конверсии и позволит измерить эффективность изменений, которые осуществляют на сайте для увеличения конверсии трафика. Для решения данных проблем и достижения соответствующих целей электронной коммерции необходимо не только собирать информацию о деятельности пользователей системы на Web-сайте, а и определить конкретные показатели эффективности Web-сайта, чтобы в дальнейшем влиять на их изменения в сторону улучшения стратегии ведения электронного бизнеса. Таким образом надо разработать и внедрить аналитический метод сопровождения текстового контента интернет-ресурсов электронной коммерции на основе анализа ключевых показателей эффективности Web-сайта,

особенно обратив внимание на определение множества релевантных и актуальных ключевых слов, по которым делают переходы с поисковых систем постоянные пользователи, действия которых привели к росту конверсии электронного бизнеса.

Целью исследования является разработка технологии продвижения интернет-ресурсов электронной коммерции на основе результатов Web-аналитики ключевых показателей страниц как KPI и KSI через формирование релевантного множества ключевых слов как feedback активности постоянной аудитории.

Метод. Разработан аналитический метод продвижения Интернет-ресурсов на основе анализа ключевых показателей эффективности Web-сайта, который основан на трех основных алгоритмах алгоритме определения проблемных мест структуры сайта для дальнейшей оптимизации, алгоритме оптимизации мероприятий по маркетингу поисковых систем, алгоритме продвижения сайта и расчета его эффективности.

Разработаны общие рекомендации по проектированию систем обработки информационных ресурсов, отличные от существующих наличием дополнительных модулей, существенно влияют на результаты продвижения Web-сайта в пространстве Интернет для дальнейшего достижения успешности конверсии электронной коммерции или увеличения значений этих показателей. Среди них модуль Интернет-магазина, маркетинга, модуль-копирайтер и Web-мастера. Для каждого из них рассчитывают собственный KPI. Это позволит эффективно просто реализовывать обработки информационных ресурсов на уровне разработчика систем (уменьшение ресурсов и времени на разработку, улучшение качества работы систем обработки информационных ресурсов).

Результаты. В работе разработано и подробно описано на основе результатов процесса Web-аналитики параметры и критерии оценки уровня успешности ведения электронного бизнеса. Также разработаны программные средства сопровождения текстового контента интернет-ресурсов на основе анализа ключевых показателей эффективности Web-сайта. Для детального анализа функционирования и продвижения Интернет-систем электронной коммерции типа Интернет-газета и интернет-журнал разработаны и внедрены 12 различных систем в соответствии с поддержанием для каждой из них разным количеством этапов жизненного цикла контента. Проведен компьютерный эксперимент анализа ключевых показателей эффективности Web-сайта. Сервис ведения статистики посещений Web ресурса позволяет оценить увеличение объемов продаж текстового контента от прямо пропорциональной зависимости увеличения количества посещений Web ресурса, количества постоянных пользователей, перспективности маркетинговых мероприятий.

Выводы. Выявлено, что наличие соответствующих модулей в системах обработки информационных ресурсов способствует увеличению объемов реализации текстового контента постоянному пользователю на 9%, активного привлечения уникальных посетителей, потенциальных пользователей и расширение границ целевой и региональной аудитории на 11%, просмотренных страниц на 12%, времени посещения информационных ресурсов на 7%.

КЛЮЧЕВЫЕ СЛОВА: контент, текстовый контент, Интернет-ресурс, бизнес-процесс, система управления контентом, жизненный цикл контента, Интернет-газета, Web-сайт, посетители, Web-страница, количество посещений, информационный поиск, процент посещений, конверсия посещений, показатель конверсии, KPI.

UDC 004.9

INFORMATION TECHNOLOGY FOR INTERNET RESOURCES PROMOTION IN SEARCH SYSTEMS BASED ON CONTENT ANALYSIS OF WEB-PAGE KEYWORDS

Vysotska V. – PhD, Associate Professor of Information Systems and Networks Department, Lviv Polytechnic National University, Lviv, Ukraine.

ABSTRACT

Context. Timely and correct analysis of the process of visiting Internet resources, which led to the overall conversion of e-business, is fundamental and relevant for successfully managing the website. Appropriate, accurate traffic analysis, which brings both successful and unsuccessful conversions, will identify the cause of the impact on conversion metrics and criteria and will measure the effectiveness of changes made to the site to increase traffic conversions. It is necessary to collect information on the activities of system users on the website and determine specific performance indicators of the website to improve e-business strategy further to solve these problems and achieve the relevant goals of e-commerce. Thus, it is necessary to develop and implement an analytical method of text content support for e-commerce Internet resources based on the analysis of key performance indicators of the website, paying particular attention to determining the set of relevant and relevant keywords used by regular users and led to an increase in e-business conversions.

Objective of the study is to develop a technology for promoting Internet resources of e-commerce based on the results of Web-analytics of critical indicators of pages as KPI and KSI through forming a relevant set of keywords as feedback activity of a regular audience.

Method. An analytical method for promoting Internet resources based on the analysis of key performance indicators of the website, which is based on three main algorithms algorithm for identifying problem areas of the site structure for further optimization, algorithm for optimizing search engine marketing activities (SEM), algorithm for site promotion and calculation of its efficiency.

General recommendations for the design of information resources processing systems have been developed, different from the existing ones, by the presence of additional modules that significantly affect promoting the website on the Internet to further the success of e-commerce or improve the values of these indicators. Among them is the module of online shopping, marketing, module-copywriter and Web-master. For each of them, calculate their own KRI. It will allow you to effectively implement the processing of information resources at the level of system developers (reducing resources and time for development, improving the quality of information processing systems).

Results. The paper develops and describes in detail, based on the results of Web-analytics, the parameters and criteria for assessing the level of success of e-business. Software tools for monitoring the textual content of Internet resources based on the analysis of key performance indicators of the website have also been developed. For a detailed analysis of the functioning and promotion of Internet e-commerce systems such as Internet newspaper and Internet magazine, 12 different methods have been developed and implemented, respectively, with support for each of them with a different number of stages of the content life cycle. A computer experiment of analysis of key performance indicators of the website was conducted. The service of keeping statistics of visits to the Web resource allows you to estimate the increase in sales of textual content in direct proportion to the rise in the number of visits to the Web resource, the number of regular users, the prospects of marketing activities.

Conclusions. It was found that the presence of appropriate modules in the systems of information resources processing increases the sales of textual content to the regular user by 9%, active involvement of unique visitors, potential users and expanding the target and regional audience by 11%, viewed pages by 12%, resources by 7%.

KEYWORDS: content, textual content, Internet resource, business process, content management system, content life cycle, Internet newspaper, Web site, visitors, Web page, number of visits, information search, percentage of visits, visiting conversion, conversion indicator, KPI.

REFERENCES

1. Clifton B. Advanced web metrics with Google Analytics. Indianapolis, John Wiley & Sons, 2012, 589 p.
2. Sulova S. A system for e-commerce website evaluation. Access mode: https://www.researchgate.net/profile/Snezhana-Sulova/publication/334734832_A_System_for_E-Commerce_Website_Evaluation/links/5d4549a0299bf1995b60d51f/A-System-for-E-Commerce-Website-Evaluation.pdf
3. Saura J. R., Palos-Sánchez P., Cerdá Suárez L. M. Understanding the digital marketing environment with KPIs and web analytics, *Future Internet*, 2019, 9(4), P. 76. DOI: <https://doi.org/10.3390/fi9040076>
4. García M. D. M. R., García-Nieto J., Aldana-Montes J. F. An ontology-based data integration approach for web analytics in e-commerce, *Expert Systems with Applications*, 2016, Vol. 63, pp. 20–34. DOI: <https://doi.org/10.1016/j.eswa.2016.06.034>
5. Heller D. Web analytics: functions, KPIs and reports in SMEs. Access mode: <https://kola.opus.hbz-nrw.de/opus45-kola/frontdoor/deliver/index/docId/1295/file/BachelorThesisDominiKeller.pdf>
6. Golyash I., Panasjuk V., Sachenko S. The performance audit of a corporate website as a tool for its internet marketing strategy, *EUREKA: Social and Humanities*, 2017, Vol. 5, pp. 57–66.
7. Rodello I. A., Dândolo V., Grande M. M. Evaluation of the impact of promotional campaign through a social networks on the key performance indicators of website for online of group-buying in Brazil, *European Journal of Management Issues*, 2016, Vol. 7, pp. 244–249.
8. Shaytura S. V., Kozhaye Y. P., Ordov K. V. Antonenkova A. V., Zhenova N. A. Performance evaluation of the electronic commerce systems, *Performance evaluation*, 2017, Vol. 38, pp. 1–11.
9. Lytvyn V., Vysotska V., Demchuk A., Demkiv I., Ukhanska, Hladun V., Kovalchuk R., Petruchenko O., Dzybyk L., Sokulska N. O. Design of the architecture of an intelligent system for distributing commercial content in the internet space based on SEO-technologies, neural networks, and Machine Learning, *Eastern-European Journal of Enterprise Technologies*, 2019, Vol. 2(2–98), pp. 15–34. DOI: <https://doi.org/10.15587/1729-4061.2019.164441>
10. Batiuk T., Vysotska V., Lytvyn V. Intelligent system for socialization by personal interests on the basis of SEO technologies and methods of machine learning, *Computational Linguistics and Intelligent Systems (COLINS 2020), 4th International Conference, Lviv, 23–24 April 2020 : CEUR workshop proceedings*. Aachen, CEUR-WS.org, 2020, Vol. 2604, pp. 1237–1250.
11. Antonyuk N., Medykovskyy M., Chyrun L., Dverii M., Oborska O., Krylyshyn M., Vysotsky A., Tsiura N., Naum O. Online Tourism System Development for Searching and Planning Trips with User's Requirements, *Advances in Intelligent Systems and Computing*, 2020, Vol. 1080, pp. 831–863. DOI: https://doi.org/10.1007/978-3-030-33695-0_55
12. Kuchkovskiy V., Andrunyk V., Krylyshyn M., Chyrun L., Vysotskyi A., Chyrun S., Sokulska N., Brodovska I. Application of Online Marketing Methods and SEO Technologies for Web Resources Analysis within the Region, *Computational Linguistics and Intelligent Systems (COLINS 2021), 5th International Conference, Lviv, 22–23 April 2021, CEUR workshop proceedings*. Aachen, CEUR-WS.org, 2021, Vol. 2870, pp. 1652–1693.
13. Rusyn B., Pohreliuk L., Kapshii O., Varetskyy J., Demchuk A., Karpov I., Gozhyj A., Gozhyj V., Kalinina I. An Intelligent System for Commercial of Information Products Distribution Based SEO and Sitecore CMS, *Computational Linguistics and Intelligent Systems (COLINS 2020), 4th International Conference, Lviv, 23–24 April 2020, CEUR workshop proceedings*. Aachen, CEUR-WS.org, 2020, Vol. 2604, pp. 760–777.
14. Orekhov S., Malyhon H., Liutenko I., Goncharenko T. Using Internet News Flows as Marketing Data Component, *Computational Linguistics and Intelligent Systems (COLINS 2020), 4th International Conference, Lviv, 23–24 April 2020, CEUR workshop proceedings*. Aachen, CEUR-WS.org, 2020, Vol. 2604, pp. 358–373.
15. Pavlenko O., Tymofeieva I. Search Query Data Analysis: Challenges and Opportunities, *Computational Linguistics and Intelligent Systems (COLINS 2020) : 4th International Conference, Lviv, 23–24 April 2020 : CEUR workshop proceedings*. Aachen, CEUR-WS.org, 2020, Vol. 2604, pp. 452–461.
16. Kliuiev O., Vnukova N., Hlibko S., Brynza N., Davydenko D. Estimation of the Level of Interest and Modeling of the Topic of Innovation Through Search in Google, *Computational Linguistics and Intelligent Systems (COLINS 2020) : 4th International Conference, Lviv, 23–24 April 2020 : CEUR workshop proceedings*. Aachen, CEUR-WS.org, 2020, Vol. 2604, pp. 523–535.
17. Radiuk P., Hrypynska N. A Framework for Exploring and Modelling Neural Architecture Search Methods, *Computational Linguistics and Intelligent Systems (COLINS 2020) : 4th International Conference, Lviv, 23–24 April 2020 : CEUR workshop proceedings*. Aachen, CEUR-WS.org, 2020, Vol. 2604, pp. 1060–1074.
18. Veres O., Rusyn B., Sachenko A., Rishnyak I. Choosing the Method of Finding Similar Images in the Reverse Search System, *Computational Linguistics and Intelligent Systems (COLINS 2018), 2nd International Conference, Lviv, 25–27 June 2018, CEUR workshop proceedings*. Aachen, CEUR-WS.org, 2018, Vol. 2136, pp. 99–107.
19. Basyuk T., Vasyliuk A., Lytvyn V. Mathematical Model of Semantic Search and Search Optimization, *Computational Linguistics and Intelligent Systems (COLINS 2019), 3rd International Conference, Kharkiv, 18–19 April 2019, CEUR workshop proceedings*. Aachen, CEUR-WS.org, 2019, Vol. 2362, pp. 96–105.
20. Adamuthe A., Nitave T. Adaptive harmony search for optimizing constrained resource allocation problem, *International Journal of Computing*, 2018, Vol. 17(4), pp. 260–269.
21. Aswani R., Kar A. K., Ilavarasan P. V., Dwivedi Y. K. Search engine marketing is not all gold: Insights from Twitter and SEOClerks, *International Journal of Information Management*, 2018, Vol. 38(1), pp. 107–116.

AUTOMATIC DETERMINATION OF THE NAVIGATORS MOTIVATION MODEL WHEN OPERATING WATER TRANSPORT

Nosov P. S. – PhD, Associate Professor of Navigation Department, Kherson State Maritime Academy, Ukraine.

Popovych I. S. – Dr. Sc., Professor of the Department of Psychology, Kherson State University, Ukraine.

Zinchenko S. M. – PhD, Associate Professor of Ship Handling Department, Head of the laboratory of electronic simulators, Kherson State Maritime Academy, Ukraine.

Kobets V. M. – Dr. Sc., Full Professor of the Department of Informatics, Software Engineering and Economic Cybernetics, Kherson State University, Ukraine.

Safonova A. F. – PhD, Associate Professor of the Department Fundamental disciplines, Kherson Polytechnic Special College of Odessa Polytechnic State University, Ukraine.

Appazov E. S. – PhD, Associate Professor of Innovative Technologies and Technical Devices of Navigation Department, Kherson State Maritime Academy, Ukraine.

ABSTRACT

Context. The article proposes an approach for automated identification of the navigators motivational model in the control of water transport. Algorithms for data extraction as a result of the man-machine interaction of navigator with the electronic control systems of the vessel during performing navigation operations of increased complexity are proposed.

Objective. The purpose of research is to apply formal and algorithmic approaches to extracting data on the motivational model of navigator to prevent accidents in water transport.

Method. The identification of manifestation determination of navigators' mental activity by means of the visual concept of the geometric group theory is proposed. This approach delivered the visual systematic-logical combining of diagnostic methods aimed at determining navigators motivational centers and the processes of professional activity like maneuver performing. The key indicator of identification is said to be the parameter of the navigator's activity as "rpm_port" having an impact on the vessel speed being a marker of intensification of the navigator's physiological activity. Such an approach is beneficial in time phase identification while maneuvering indicating explicitly at the stepping up of the navigator's physiological motivational state. It was proven to be correct based on the results due to Ward's dendrogram, several statistical methods and applied software. The obtained research results encourage the prediction of the navigator' motivational states in critical situations.

Results. In order to confirm the proposed formal-algorithmic approach, an experiment was carried out using the navigation simulator Navi Trainer 5000. Automated analysis of experimental ones made it possible to form a motivational map of the navigator and determine the decision-making model affecting in the processes of control vessel in difficult situations.

Conclusions. The proposed research approaches made it possible to automate the processes of extracting data indicating the principles of decision-making by navigator. The effectiveness of proposed approach was substantiated by the results of experimental data automated processing and the constructed tree-like decision-making spaces.

KEYWORDS: motivation identification systems, automated data processing systems, modeling of decision making models, computer simulators, analysis of the human factor, automated control systems.

ABBREVIATIONS

ECDIS is an Electronic Chart Display and Information System;
ARPA is an automatic radar plotting aid;
AIS is an Automatic Identification System;
GPS is a Global Positioning System;
NTPRO 5000 is a navigation simulator "Navi Trainer 5000";
ANOVA is an Analysis of Variance;
MMS is mental motivational states;
LNE is a level of navigator expectations;
LPC is a level of personality claims;
LNEc is a level of the navigator's expectations;
LAEc is a level of awareness of expected events;
LEAc is a level of expected attitude towards the participants of interpersonal inter-action;
LERc is a level of expected results of activity;
SPSS is a Statistical Package for the Social Sciences;
rpm_port is a vessel speed.

NOMENCLATURE

s is a motivational influencing factor;
 S is a set of factors;
 y is a control actions;
 Y is a set of control actions;
 x is a navigational situations;
 X is a set of navigation situations;
 w_{nav}^a is a navigator action utility function;
 R is a navigation results space;
 y'_{nav} is a navigator behavior alternatives;
 $w_{nav\max}^a$ is a maximum expected usefulness for the navigator;
 x_t is a moment of making decisions;
 $Gv \uparrow$ is an expansion of feature space;
 $Gv \downarrow$ is a narrowing of feature space;
 Re is a peer review function;
 K is a context of the navigation situation;
 δ is a structural diagnostics;

δ is a functional diagnostics;
 e_r is a stock of resources for decision making;
 Fr is a fragment to define MMS;
 I is an information structure of the MMS identification task;
 $\beta_0(S)$ is a prior distribution on the set of states;
 $q^{u,\gamma}$ is a structure transition function;
 φ_n is a safe vessel handling strategy;
 π_1^n is a decision making strategy;
 τ_1^n is a strategy for choosing alternatives;
 ζ_i is an alternatives of states;
 x' is a comparison x with MMS;
 R^S is a state space;
 θ_n is a navigation risk criterion;
 R_b is a binary relation;
 $\Gamma(x_p)$ is a preference function;
 Ω is a set matrix;
 s_p is a complexity of the project;
 u_q is an utility function;
 af is an expectation level;
 f is a subjective probability;
 μ is a selective probability;
 x_{ex} is an expenses;
 v is a scaling factor;
 γ is a reference level;
 $\Delta(f)$ is a decision checkpoint function;
 $P^*(s_p)$ is a probability measure;
 X_H is a better alternative;
 $\mathfrak{S}(x_{ex})$ are maximum values of the quality criterion relative to the strategy x_{ex} ;
 \mathfrak{S}_{opt} is an optimal solution;
 Ω_X is an area of agreement and compromise;
 Λ is an importance of criteria;
 $P_{U_{x_i}}$ is a probability of the expected utility of a solution;
 $P_{(C_{x_i} \geq U_{x_i})}$ is a probability of the decision being useful is greater than initially expected;
 ψ_t is an element of random behavior;
 a^3, b^2, c^6 are words of the mathematical system;
 G_1, G_2 are represent a group, for MMS;
 G' is a producing a new group;
 e_i is a starting point of the motivation identification system report;
 g_1, g_2 are points of the Cayley's graph;
 df are degrees of freedom;
 $\Pr(>F)$ is a probability distribution;
 Z is a decision-making task by the navigator;
 $Z^n(t)$ is an inner emotional experience;
 $Z^{ex}(t)$ is an external reactions, behavior model;
 $E(t) < \infty$ is a shortage of time resources;

t_z is a discrete variable step;
 Z_{Plan} is a how many tasks to be solved;
 \bigcirc is a future tense symbol;
 D_Z is a how many targets to be deleted;
 A_Z is a lots of added goals;
 K_Z is a bounded rationality ratio;
 Δ_Z is a bifurcation period of the motivation model;
 Fun is an input processing method;
 $M_{F_1 \dots F_7}$ is a navigator motivation factor.

INTRODUCTION

Contemporary experimental data analysis systems has the tendency of welcoming the structural submission of human factor essence being predominantly characterized as a multifactorial phenomenon [1–3], which is empirically studied in various ergatic systems [4–6]. Being aware of having a diverse range of publications aimed at building and shaping the structure of this phenomenon we are to take into consideration possible uncertainties in the system of concepts as well as in their conceptual connections. This negative experience seemed to allow the implementation of the effective application of the spoken above techniques to be undertaken in practice of vessel control. However, the increasing number of navigation devices of ECDIS, ARPA, AIS, GPS and information brings new requirements to this problem [7–9]. It goes without saying that this very issue would definitely be considered advantageous enough in identifying the sea-going navigator's motivational mental state [10–12].

For example, when performing typical maneuvers of entering the harbor of Istanbul, it was noted that navigators are guided by different motives in the same navigation situations (Fig. 1).

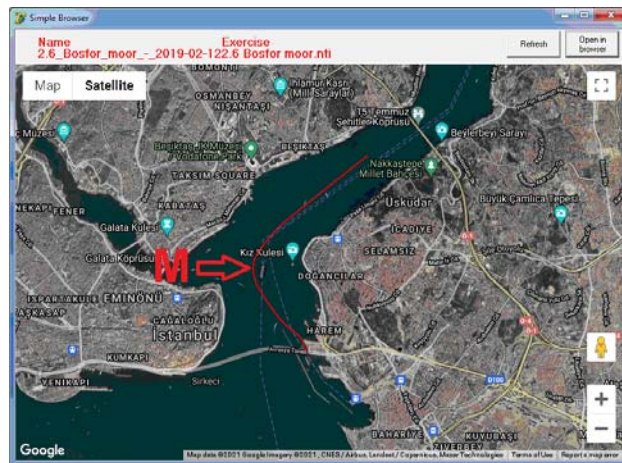


Figure 1 – Control of electronic navigation when performing navigation maneuvers

All this indicates the need to build mathematical and automated systems for identifying the element of human factor – the motivation of navigators in order to improve the safety of navigation.

Thus, the **object of research** is the human-machine interaction of the ergatic system subject for water transport.

The **subject of research** is the models and algorithms for automated identification of navigator's motivation factor in the vessel electronic control systems.

The **purpose of research** is to apply formal and algorithmic approaches to extracting data on the motivational model of navigator to prevent accidents in water transport.

The purpose of the article is chosen to be applying special purpose solution method embracing of following problems:

1. Analyze the existing formal approaches in order to describe the principles of decision-making by navigators. And also to determine the elements of motivational models of decision-making by navigators in human-machine systems.

2. To propose the principle of extracting data on motivation of navigator in process of vessel control based on the geometric theory of groups in the form of $G' = G_1 \circ G_2$.

3. To form a 3D information model of the navigators structural motivation of based on Cayley's graph elements: $Cay\langle b | b^2 = e \rangle$, $Cay\langle a | a^3 = e \rangle$, $Cay\langle c | c^6 = e \rangle$.

4. Based on the data electronic simulator, correlate the characteristic spaces of MMS navigator motivation: LNE, LPC, LNEc, LAEc, LEAc, LERc and the vessel control parameters.

5. Conduct an experiment using the navigation simulator Transas navigation simulator NTPRO 5000. Based on the data obtained and Ward's dendrogram and MMS motivation points, using a number of statistical ANOVA methods in RStudio, build a motivation map of navigators.

6. On the basis of experimental data, determine the correspondence of motivational decision-making models F1–F7 regarding the navigators teams.

The obtained motivational models will make it possible to identify the class of decision-making by navigators in difficult situations to ensure the safety of water transport control.

1 PROBLEM STATEMENT

The formal statement of the problem is to identify the motivation of the navigator by analyzing the tuple of input variables, $s \in S$, $y \in Y$, $x \in X$ to $|X| \leq |S|$ such that, $w_{nav}^s : (Y \times S \times X) \rightarrow R$ in terms of situational control [13, 14].

An additional condition for identification is maximization $w_{navmax}^s : (x^*, y^*)_{s_t} \in X \times Y$ within the framework such that $y' \succ y \Leftrightarrow w_{navmax}^s(y', s, x) > w_{nav}^s(y, s, x)$.

Identification is limited by terminal factors $(x_t | w_{navmax}^s)$:

$$\forall x_t \text{ Re}(x_t | K) = (x_t \rightarrow \delta, \underline{\delta} | K) \rightarrow \text{Re}(\underline{\delta}) = e(Gv \uparrow (\underline{\delta})) \otimes_e e(Fr(Gv \downarrow (\underline{\delta})))$$

in conditions:

$$I = \{S, \beta_0(S), X, [X_s \subset X, s \in S], Y, [Y_x \subseteq Y, x \in X], A, q^s(S | S \times Y), w_{nav}^s(Y \times (S \times X)), \zeta \in A\}.$$

The output variable is a $q^{s, \underline{\delta}}(S | S \times Y)_M$ and depends on M and $Con(x, x' | K) = 1 \Leftrightarrow x = x' : X_s \subset X, Y_x \subseteq Y$.

The main criterion of the task is a $\theta_n(\underline{\delta}_1^n | \pi_1^n, \tau_1^n)$ in the space R^S , which will determine the dependence w_{nav}^s on θ_n within the MMS scale.

The resulting factor is the navigation safety strategy $\Phi_n(\pi_1^n | \tau_1^n)_{\underline{\delta}} \rightarrow \max$, in the form of a scale:

$$\underline{\delta}^{-1}(x)_M = \cup \{s \in S : \underline{\delta}(s) = x\} \subset S, x \in X.$$

2 LITERATURE REVIEW

In a study aimed at modeling decision-making processes under the constraints of experiments [20], an approach based on use of R_b is proposed. Moreover, $\Gamma(x_{ex})$, determines the most preferable solutions from Ω . This study is of interest due to fact that, according the analogy with vessel control processes and the experimenter operates with a set of observed factors and criteria, such as in Fig. 2.

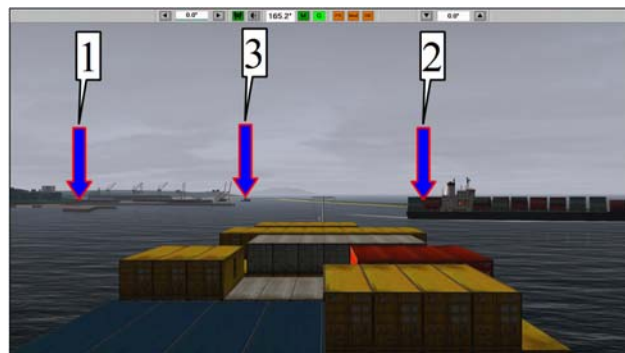


Figure 2 – Observed situation to vessel control

In the study [21], the processes of creating software based on recursion of data in the form of decisions made results are well considered. An important element is to take into account the risks associated with both stochastic uncertainty and s_t . In this case, it is important to accurately determine the factors determining u_q .

Also in the study [22], the solution to problem of decision-making synthesis models under conditions of fractal dimension is deeply and in detail considered. This study directs the scientific search for this article in field of logical-metric spaces, which allows us to consider the problem from a new angle. The paper considers methods for constructing feature spaces, as well as the principles of decision trees application in a practical aspect.

In turn, work [23] considers the aspect of taking into account subjective expectations when making decisions,

which is inapplicable in water transport management due to the high risk of disasters.

$$a_f = v^{-1}(Ev(f)),$$

$$U(f, \mu) = v(a_f) + \sum_{x \in \text{sup}(\mu)} u(x_{ex} - \gamma a_f) \mu(x).$$

However, it should be borne in mind that captains are subject to decision-making based on the expected utility factor, which also depends on the individual motivational model. This is how the work introduces $\Delta(f)$ and $P^*(s_p)$, which in our case predetermines the dependence on motivational factor:

$$U(f, \mu) = \sum_{s \in \text{sup}(P^*)} v(f(s_p)) P^*(s_p) + \sum_{x \in \text{sup}(\mu)} u(x_{ex} - \Delta(f)) \mu(x_{ex}).$$

It follows from this that an automated analysis of the captain's motivational factor can have a decisive impact on vessel control.

In the study [24], for decision-making, alternatives are considered that are decisive according to the introduced criteria:

$$X_H = \arg \max \mathfrak{I}(x_{ex}).$$

In this case, the optimality of decision taken is reduced to solving the chain of local goals, which forms the current strategy:

$$\mathfrak{I}_{opt} = \mathfrak{I}(X_{opt}) = \underset{X \in \Omega_X}{opt} [\mathfrak{I}(X), \Lambda].$$

However, in the case of performing navigation tasks and vessel control tasks, this approach may not be effective enough. This is due to the fact that, guided by motivational model, the captain can no longer change of strategy and follows predetermined principles until the last stage of the task.

In this work, the human factor is determined randomly [25]. In this case, a distinction is indicated according to the degree of random factors influence:

$$\text{if } P_{U_{x_i}} \geq P_{(C_{x_i} \geq U_{x_i})}$$

$$\text{then } 0 \leq \psi_t \leq 1 - \left(P_{U_{x_i}} - P_{(C_{x_i} \geq U_{x_i})} \right)$$

$$\text{or if } P_{U_{x_i}} < P_{(C_{x_i} \geq U_{x_i})}$$

$$\text{then } \left| P_{U_{x_i}} - P_{(C_{x_i} \geq U_{x_i})} \right| \leq \psi_t \leq 1 + \left| P_{U_{x_i}} - P_{(C_{x_i} \geq U_{x_i})} \right|.$$

However, in transport systems, random processes are unlikely. Marine companies try to take into account random factors in order to reduce the likelihood of unforeseen accidents and disasters.

Thus, the studies considered indicate the need for a scientific search for approaches to formalize and algorithmize the factors of motivating navigators when managing water transport.

3 MATERIALS AND METHODS

During the work with navigational devices, the records of the navigators' activity are saved on the server. The information received indicates the nature and settings of the vessel control model in the ergatic system.

Devices such as information navigation systems, radar and warning systems are synchronized. Their use cannot be simultaneous; therefore, it becomes possible to identify the motivation of the navigators.

To meet the above mentioned targets, an unambiguous approach is proposed to be implemented consisting of two essential stages: the first one is defining the navigator's motivational centers and their significance within the framework of the visual geometric concept and the second one is an analysis of decision making by navigators to determine their MMS.

Consider the stage of determining the navigator's motivational centers. We will establish the order of identifying extrema of the navigators' motivation. We will assume that these extrema were revealed in the course of observation and a series of specialized tests [15–17]. In order to measure their influence, it is necessary to analyze a number of indexes of MMS selectively using the methodology of social expectations of personality.

Application of psycho-diagnostic instruments for examining an individual's motivation. We will consider two methods as instruments for determining the influence of extrema: "LNE", modified and "LPC" that have correlations investigated in the study, and therefore they must be presented within the same metric space [18,19].

Having determined the extrema of MMS, we will choose the outlined methods relative to them. Each of them will – G_1 and – G_2 . Simultaneous participation of the methods will be presented as a product of these groups, producing a G' [26].

Since there are: a^3 , b^2 and c^6 in the group $G' = G_1 \circ G_2$ and each one is a geometric structure, we have Cayley's graphs for all of them (Fig. 3):

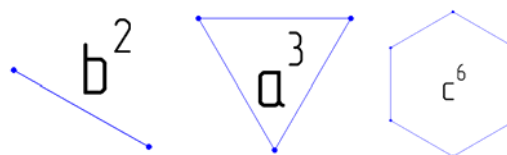


Figure 3 – The words: a^3 , b^2 and c^6 in the group

Then group $G' = G_1 \circ G_2 = \langle a, b, c \mid a^3 = b^2 = c^6 = e \rangle$, meaning that Cayley's graph of a free product implies "agglutination" of two graphs G_1 and G_2 .

In order to create Cayley's graph of the group G_1 we proceed from the following logic: G_1 has two generatrices, i. e. The number of edges equals to a doubled product of the generatrices – 4, but as far as b^2 is a biangle and "agglutinates" in one section, we have three edges – sections. Therefore, we put one edge b^2 (in sum – three) from each point of the apex (of the triangle) a^3 and have Cayley's graph for G_1 (see Fig. 4 a, b).

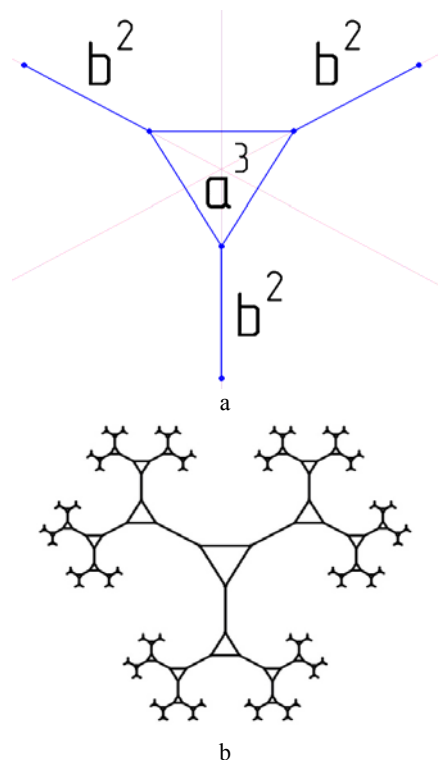


Figure 4 – The formation of Cayley's graph for G_1 :
 a – composition of generators for G_1 ; b – fractal of Cayley's graph representation

The move to the stage of the formation of the group G' . In order to obtain a product, we multiply G_1 by G_2 , represented by a hexahedron and we will determine the number of edges of the graph. Proceeding from the fact that there are three products, we will double and get 6, but there will be 5 because of "agglutination" b^2 . Consequently, the products $G' = G_1 \circ G_2$ will have Cayley's graph depicted in Fig. 5.

Each point e_i , synchronizes with the point – extrema of MMS and allows identifying their influence. The graph G' allows examining each extremum of MMS in a combinative way. We can see that e_1 originates from the application of the "motivational core" c^6 , then there is a move to the method "LNE", and the first step is to perform analysis of the index LNEc. Further, if the order of the methods proves to be not efficient enough, then the next round of the research starts from the point e^2 , begin-

ning with "LNE", without LNEc and further moves to the "motivational core" c^6 , and then to LNEc. This group can be represented in a 3D projection for visualization (see Fig. 6).

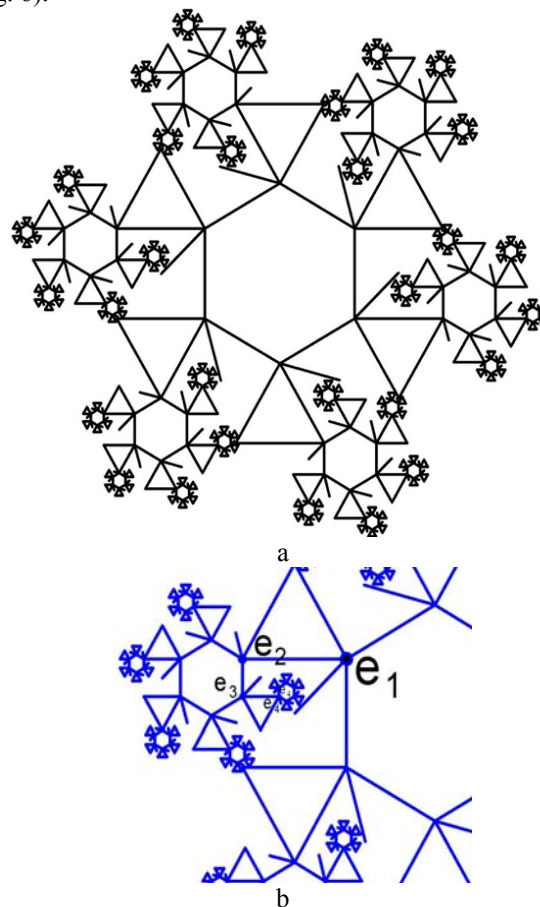


Figure 5 – Cayley's graph of the desired group G' :
 a – fractal space of $G' = G_1 \circ G_2$; b – representation of extremum points on a Cayley's graph space

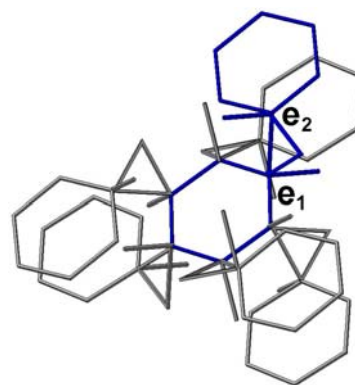


Figure 6 – 3D view of the group G'

The next task is to identify the influence of the extrema of motivation, determining the importance of the event projected onto a newly created group of the manifestation of MMS. The determination of the indexes relative to the extrema of motivation by the chosen methods implies the

use of a comprehensive approach [27–29]. Taking into account that the methods number 10 indexes and they correlate with each other to this or that extent that was proved by the research, we can maintain that they are indissoluble. Consequently, it is necessary to find the generalized value of the extremum of motivation summarizing the data of all the indexes. Therefore, we developed a program module that allows synchronizing the data in one Cartesian coordinate system C_i (see Fig. 7). For instance, there is the above-mentioned trajectory b^2ab , followed by the strategy b^{-1} “planning to move a rudder blade”, that is evident in the pause in the vessel route and a radical change in the course.

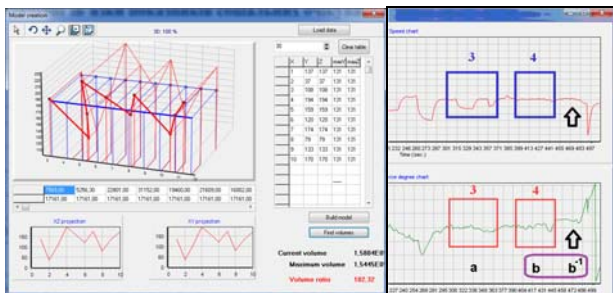


Figure 7 – Determining the influence of the points C_i of MMS

Having obtained the data of all the indexes we can use this module to find the generalized index of the navigator’s motivation. Following the trajectory of examining the points of MMS and determining the most effective approach we select an individual algorithm of interaction with each navigator within the framework of the research of his motivation. The accurate values of the points of MMS enable us to understand that the surface circumflexing them cannot be flat and it results in the curvature of the metric space of motivation. This curvature of the surface is nothing but an individual and unique imprint of the navigator’s perception of the situation in the light of his motivation [30–32].

In this way a quasi-isometric formal surface with the points of extremum is formed [33, 34]. The question arises: what correlation is between them and can they play the role of a connected graph? It is known that these lines are geodesic, i. e. the surface itself is not curvilinear. This research stage requires the application of the method making it possible to find these connections between the points-extrema. The connection between the points by the edge is only possible when the subsequent point emerges from the preceding one when multiplied by one letter of the set of the generatrixes. For instance, g_1 and g_2 are connected by the edge if $\exists a \in A; g_2 = g_1 \cdot a$ or $g_1 = g_2 \cdot a$ in the set of generatrixes.

It is important to find the principle of transfer, connecting these points by the edges, something common and essentially important within the framework of the methodology of social expectations of an individual. Grouping points-extrema of motivation is necessary to determine MMS, and, consequently, to achieve the main purpose of the research.

The analysis of the points-extrema gives us a reason to think that these points are heterogeneous, each of them having a number of parameters consisting of three categories: identifiers relative to Cayley’s graph; psychological methods “LNE” and “LPC”; factors relative to sailing through the Bosphorus strait.

4 EXPERIMENTS

Taking into consideration the specificity of the subject area and also the geometric orientation of the scientific research, we suggest clustering the data. Cluster analysis will allow dividing the points of motivation MMS into groups and determining their relationships geometrically.

In order to do this, we used clustering in the form of a dendrogram by Ward’s method for finding single connections of the points of MMS, by the principle of Euclidian distance (see Fig. 8). Having analyzed these geometric connections between the points of motivation of MMS, we are enabled to construct the desired map of the navigator’s motivation (see Fig. 9).

The represented geometric relationships of MMS allow visualizing how the manifestations of the navigators’ motivation are ordered in the desired conditions. It is necessary to mention that this order is characteristic only of a certain individual and determined by a variety of factors, affecting their motivation. It allows predicting navigators’ behavior in critical situations similar by the set of indexes that will make it possible to prevent disastrous situations on maritime transport. Hence, according to the involved approach, the relationship, the personality structure of some navigators’ motivation being actively engaged in complex maneuver operations is highly likely to be ascertained [35–37].

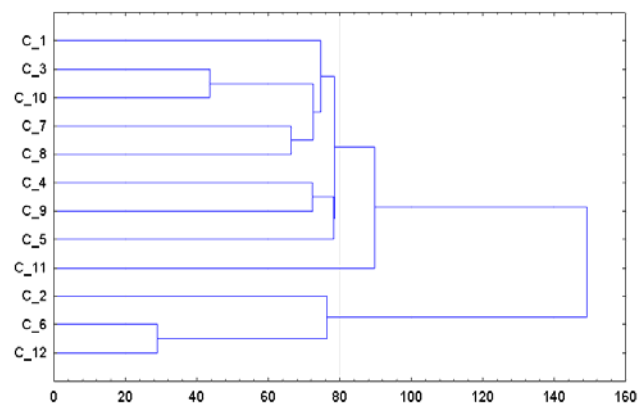


Figure 8 – The dendrogram by Ward’s method for the points of motivation of MMS

It must be emphasised that the foremost phase of this research is being able to figure out the method facing the challenge of having clusters of navigation situations according to the indications of navigators’ actions for correlating them with their motivational mental states.

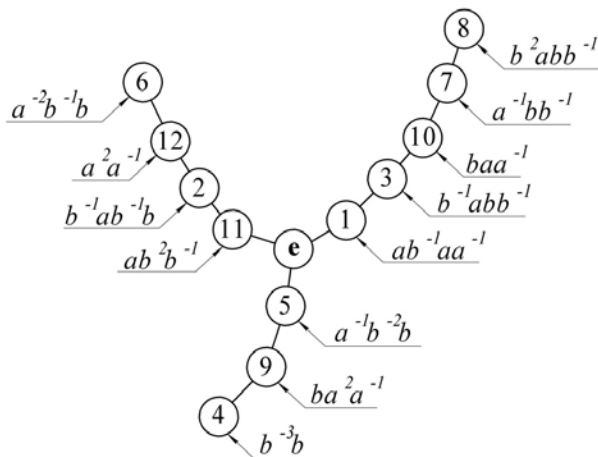


Figure 9 – The map of the navigator’s motivation

It should be underlined separately that we propose the concept; “the navigator’s mental motivational state” to be applied as “the general functional level of mental activity having direct impact on the conditions of his professional activity and individual psychological characteristics”. In the course of the research situational mental states connected with the definite location in the water area and navigation tasks at the corresponding time interval are put forward to be taken into consideration.

Being on the point of getting a passing, especially within the high intensity maritime transport areas, the navigator of the vessel is striving to endeavor to accomplish the navigation task in the best possible light. Therewith, the significant element of such kind of work is noticed to be the skill of the navigator’s motivational sphere management, knowledge of the structure and system of meaningful characteristics of motivational mental states. In addition, such properties as internality, externality, activity, passivity, openness, closeness, etc. are being able to facilitate this or that MMS discovery with sufficient accuracy. Thus, having been spoken about approaches would definitely further encourage us to come closer to the definition of characteristics welcoming the direct effect on the result of professional activity as well as to the relevant to the having been formed individual behavior patterns to be taken on.

The sequential search performed among test methods at the preliminary stage is noticed to have triggered off the opportunities to determine the individual prerequisites for the manifestation of motivational characteristics. Accordingly, these very issues indicate at MMS and an established behavior model. Thus, the following possibilities to be determined have turned out : an internal motive, a cognitive motive, a avoidance motive, a competition motive, a motive for a change in activity, a self-esteem motive, significance of results, complexity of the task, volitional effort, assessment of the level of achieved results, assessment of one’s potential, target level of mobilization of efforts, expected level of results, regularity of results and initiative.

In response to the mentioned above, the methods based on coping tests [38, 39] let eight ways of difficulties overcoming be lighted upon. These items probably

include the following to be worth mentioning: confrontation, distancing, self-control, seeking social support, taking responsibility, avoidance, planning a solution to the problem, positive reevaluation. The fractal organization of the above methods at key stages of the route trajectory conduce to delineate the correlation of the dominant motives to the situations.

It must be emphasised that the foregoing happened to have provided the possibility to significantly reduce the time for MMS analysis being predominantly concentrated on the likelihood of “bursts” of navigators’ motivational activity. So, this approach is said to have hastened the selection of the required methodology basing on three or four stages of navigators’ actions in typical situations.

Taking into account the evidence of vivid manifestation of navigators’ mental activity boost getting used to being accompanied by an increase in their motivation the conclusion is that the nature of the actions precisely indicates the navigators’ mental state.

Well, for instance, such a parameter as rpm_port at moderate rates (3–5 knots) is unable to be affecting the mental activity boost but, nevertheless, at high rates (8–12 knots) makes the navigators severely enhance their entire capacity-building. It goes without saying that this item is impossible to be looked at without motivation [40].

With an aim to analyze the influence of this parameter on the MMS shaping we want a range diagram (Boxplot) rpm_port (main rotor speed) to be introduced. For getting analysis done an experiment was carried out with 11 different navigation teams being involved in facing the same terms and conditions of the Bosphorus location. So, from the analysis of the diagram the sample differences are sure to be noticed. The medians are not noticed to be significantly different only in the first, second and third samples (rpm_port1, rpm_port2, rpm_port3). Sample data rpm_port1, rpm_port6, rpm_port7, rpm_port8, rpm_port9, rpm_port10 skewed to the left; rpm_port10, rpm_port11 – skewed to the right; rpm_port5 – symmetric; rpm_port2, rpm_port3 – very tight placement (Fig. 10).

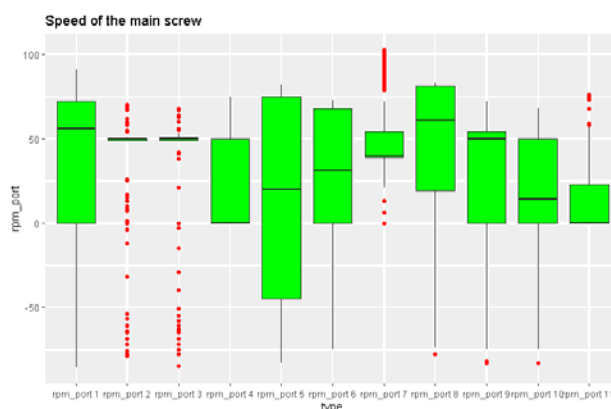


Figure 10 – Boxplot plot of the rpm_port parameter

The interquartile range does not visually differ too much in the samples rpm_port4, rpm_port6, rpm_port8, rpm_port9, rpm_port10. The samples rpm_port2, rpm_port3, rpm_port7 have a lot of outliers.

So, having been delivered results do indicate a clear difference in the navigators' behavior models. For example, the rpm_port speed parameter and its change along the trajectory of the routes are presenting significant differences. In order to achieve the target of finding the groups having alike patterns of behavior a confidence interval diagram was decided to be navigator.

It would be highly appreciated due to reflecting the presence of common points of intervals for the following groups of samples: rpm_port1, rpm_port2, rpm_port7, rpm_port8; rpm_port1, rpm_port3, rpm_port9; rpm_port3, rpm_port6, rpm_port9; rpm_port4, rpm_port5, rpm_port6, rpm_port10, rpm_port11; rpm_port5, rpm_port6, rpm_port9, rpm_port10. Average items closely spaced for the following groups of samples: rpm_port1, rpm_port2; rpm_port1, rpm_port3; rpm_port2, rpm_port7, rpm_port8; rpm_port4, rpm_port5, rpm_port6, rpm_port10, rpm_port11; rpm_port6, rpm_port9, rpm_port10 (Fig. 11).

Subsequently, to determine the homogeneity of variances, the Bartlett Test of Homogeneity of Variances and Computes Levene's test for homogeneity of variance across groups centered on the median are considered to be of great assistance.

Bartlett test of homogeneity of variances
 data: rpm_port by type
 Bartlett's K-squared = 428.15, df = 10, p-value < 2.2e-16
 Levene's Test for Homogeneity of Variance (center = median)
 Df F value Pr(>F), group 10 39.493 < 2.2e-16 ***
 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

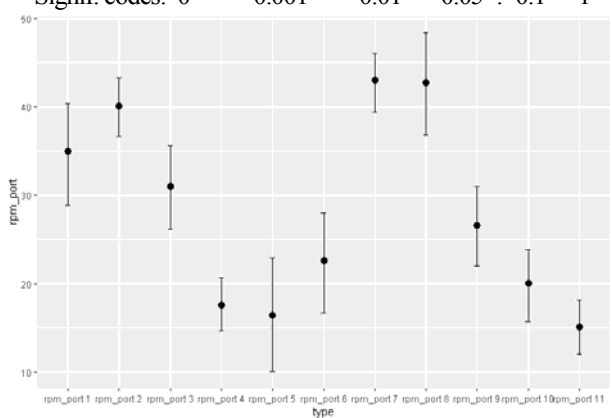


Figure 11 – Diagram of confidence intervals for the rpm_port

Accordingly, based on the results of both completed tests, we are definitely to discard the null hypothesis of equality of variances (since with $F_{10,2992} = 39.493$ p-value < .05). This issue indicates the evidence of variances to be different being admitted as solid proof that the populations from which the data is selected are not to be regarded as identical. Further, we are on the point of performing ANOVA aiming to compare the average samples. As a result, the number $F_{10,2992} = 20.23$, $p < 2e-16$ was obtained. The conclusion can be made, this issue means that there are statistically significant differences between the average samples of the rpm_port parameter. This item is believed to be the ground to have the hypothesis of the study of shaping navigator's behavior models with individual characteristics generally confirmed. Further, let's do posteriori comparisons with an aim to identify differences between the samples. As the test for equality of variances is noticed to have uncovered statistically substantial differences, the Games-Howell one being used for pairwise comparison of normally distributed samples with unequal variances is proposed to be introduced. The obtained results are presented in Table 1.

It is vividly seen that after having carried out analysis of the table five types of navigators' behavior models are highly likely to be identified.

Therefore, each of these issues represents the maximum number of samples between having no statistically significant differences [41–43]. To be precisely, no differences were able to have been observed in the middle of each behavior model.

Moreover, at least one statistically significant difference between the behavior models do exist. Group1 includes: rpm_port1–8; group3: rpm_port1, rpm_port3–9; group4: rpm_port4–6, rpm_port10–11; group5: rpm_port5, rpm_port6, rpm_port9, rpm_port10 (Fig. 12 a, b).

Hence, the analysis of the diagrams as well as the application of subsequent techniques delivered a beneficial opportunity for us to succeed in having the navigational watches grouped. No. 1–11 (Fig. 12 b).

Table 1 – The difference is a statistically significant difference between the sample

	rpm_port1	rpm_port2	rpm_port3	rpm_port4	rpm_port5	rpm_port6	rpm_port7	rpm_port8	rpm_port9	rpm_port10
rpm_port2	–									
rpm_port3	–	–								
rpm_port4	differ	differ	differ							
rpm_port5	differ	differ	differ	–						
rpm_port6	–	differ	–	–	–					
rpm_port7	–	–	differ	differ	differ	differ				
rpm_port8	–	–	–	differ	differ	differ	–			
rpm_port9	–	differ	–	differ	–	–	differ	differ		
rpm_port10	differ	differ	differ	–	–	–	differ	differ	–	
rpm_port11	differ	differ	differ	–	–	–	differ	differ	differ	–

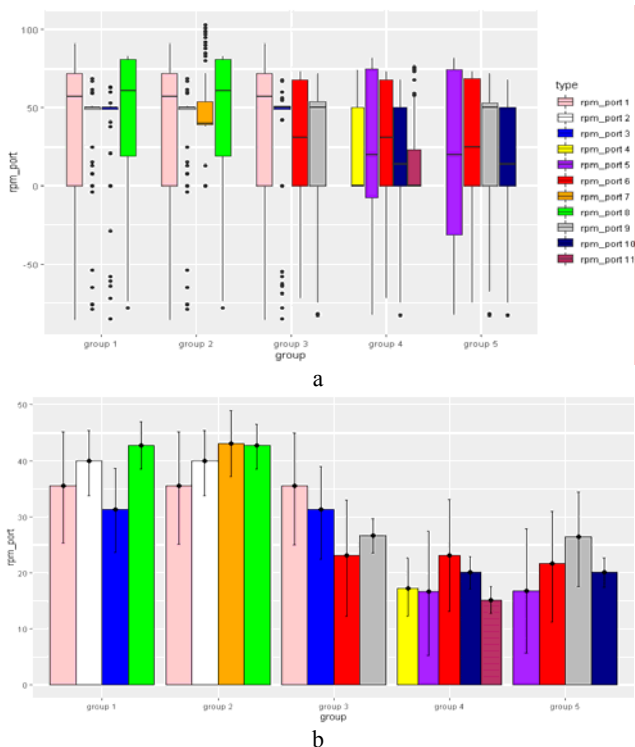


Figure 12 – Diagram of the groups of statistically significant difference between the behavior models: a – boxplot diagram of the groups of samples of the rpm_port parameter between which statistically significant differences were revealed; b – diagram of confidence intervals for the groups of samples of the rpm_port parameter between which statistically significant differences were revealed

Besides, the having been developed questionnaires contributed into process of getting empirical data obtained for each formed group.

Data processing with RStudio and “SPSS” v. 23.0 turned out to be favorable enough to apply to Spearman’s statistical correlation methods, the Promax component rotation one and etc. So, to calculate the minimum, maximum, arithmetic mean of the parameters and standard deviations became highly likely to be done.

5 RESULTS

The calculated results were a huge assistance of allowing us to correlate the survey data and the navigators’ mental motivational states (Table 2).

Thus, the conclusion can be made that the having been proposed approach assumes indirect navigators’ MMS determination by means of the parameters of the vessel’s movement. Then, with respect to the level of goal-setting, we define the model of the motivational factor influence on the water transport operation:

$$\begin{aligned}
 Z(t_z + 1) &= Z(t_z) \setminus D_Z(t_z + 1) \cup A_Z(t_z + 1) \text{ if } 0 < E(t) < \infty, \\
 \forall t \ Z(t) &= Z^{in}(t) \cup Z^{ex}(t), \\
 D_Z(t_z + 1) \ \& \ A_Z(t_z + 1) &= \arg \text{opt}_{Z(t)} K_Z(t_z + 1), \\
 Z_{Plan}(t_z + 0) &= Z_{Plan}(t_z + 0) \cup \{Z - \text{task}\}_t, \\
 \uparrow Z(t_z) &= [Z(t_z) \rightarrow \text{Chaos}(t_z + 1 - \Delta_z, \Delta_z)], \Delta_z \rightarrow \min, \\
 E(t + 1) &= \text{Fun}(E(t), Z(t), \{F(t)\}, M_{F_1 \dots F_7}(t)).
 \end{aligned}$$

Moreover, despite being a success, this approach is to be taken only at the initial stages of psychological diagnostics [44]. Much further analysis of likewise conclusions is sure to presume the involvement of a highly qualified specialist. This will allow the development of an intelligent assessment of diagnostics of experimental data in real time [45].

Table 2 – Mental motivational states of navigators

Commands	Mental motivational states of cadets
2, 8, 7	F3 “Motivation for obtaining” is considered to be a reflection of significant motivation levels to compete and motivation to succeed without having positive reassessment of the real situation. The motivational <i>mental state</i> of this group of cadets is having a lot to do with craving to achieve the desired result at any cost. F6 “Affiliation motivation” is said to be characterized as the desire of the individual to be better than others. The subject is being implicated into constant retrieval to perform complex tasks of professional activity seeking social support and encouragement in the environment sometimes in the eyes of beholder close friends.
1, 3	F2 “Confrontational motivation” is the individual’s confrontational regulatory ability to do something by his own initiative, to confront avoiding the search of other alternatives to cope with the problem as well as shirking the responsibility and positive reassessment of the situation. F4 “Pragmatic motivation” is said to be planning for solving a problem with having it not accompanied by a positive reassessment of the real situation. This motivational <i>mental state</i> needs to be defined mostly as the search for approval, a desire to hold on to one’s point of view and beliefs.
9, 6	F1 “Ego-motivation” is said to reflect the dependence of the individual components of the personality’s motivational structure, to be directly, internal and cognitive motive, volitional effort, assessment of one’s potential, the intended level of efforts mobilization. The significant negative correlation with confrontation and self-control in the context of professional activity performing is definitely to be taken into consideration. The effect of this factor seems to be characterized by internal and cognitive activity in professional activity. F7 “Motivation for responsibility” depicts the idea of getting all motivational mental states accompanied by the desire to perform all assigned tasks from the simplest to the most difficult ones appropriately. Such a motivational <i>mental state</i> is being thought about as inherited peculiarity for being prone towards order, and discipline in everything.
4, 5, 10, 11	F5 “Motivation for avoidance” includes variables, the psychological content of which is aimed at mirroring the desire to escape from responsibility, initiative and whatever. It seems to have a negatively evaluated significant relationship with the assessment of the level of achieved results. Such individuals have a tendency to totally circumvent no way having to deal with anything. They get used to being apt to “keep your head down”.

Hence, the proposed approaches are believed to have been implemented and tested in the course of an experiment using the Navi Trainer 5000 navigation simulator. They managed to have demonstrated the potential and the effectiveness of the proposed above research hypotheses. The experimental data, as well as the developed psychological methods for analyzing sea-going navigators' mental motivational states while performing complex maneuvers and navigation tasks, facilitate to find out important dependencies and come up as a result of scientific research.

As a result of the study, the expediency of the psychological and informational approaches used, as well as the principles of the formation of models of behavior of navigators in difficult conditions was confirmed.

Thereby, the psychological and formal-analytical approaches presented in the study, as well as the developed software tools, are named to be worth being delivered to apply the classification of the navigators' mental motivational states in the form of a geometrically defined space of states. Accordingly, the results of the study might be of great interest and assistance for the sake of predicting and preventing catastrophic consequences in maritime transport due to the human factor.

CONCLUSIONS

We analyzed formal approaches to the practical aspects of the decision theory application in situations of stochastic uncertainty and resource constraints. As a result, an approach was proposed that is closest to the problem of research based on the models formation of motivation when making decisions by navigators in human-machine systems.

On the basis of the indicated approach, the principle of vessel control converting information parameters from the ECDIS server was determined. This made it possible to structure the data on the motivation the navigator on the basis of geometric product of groups $G' = G_1 \circ G_2$.

A 3D information model the structural motivation of navigators was formed based on the Cayley's graph elements: a^3, b^2, c^6 . This made it possible to identify the extremum points of motivation, to determine the degree of importance in relation to MMS.

On the basis of electronic simulator data the periods of extrema were synchronized with respect to the speed-power parameters of the vessel's control. This made it possible to form the characteristic spaces of motivation the MMS navigator: LNE, LPC, LNEc, LAEc, LEAc, LERc.

To achieve the main goal of the study, an experiment was carried out, as a result of which data were collected on the motivational models of navigators when performing a maneuver of entering the harbor of Istanbul. The data were obtained in the course of automated analysis of NTPRO 5000 navigation simulator log files with subsequent processing by statistical ANOVA methods in RStudio. As a result of calculations, a motivation map for

navigators was built based on Ward's dendrogram and MMS points.

On the basis of experimental data, motivational models of decision-making F1–F7 of navigators teams were determined.

The obtained motivational models will make it possible to identify the class of decision-making by navigators in difficult situations to ensure the safety of water transport control.

Thus, this study made it possible to significantly approach the solution of the problems to identifying motivational models of water transport navigators to ensure safety.

ACKNOWLEDGEMENTS

The team of authors is grateful to the management of the Kherson State Maritime Academy (Ukraine) for the opportunity to conduct research on the basis of the research laboratory "Development of decision support systems, ergatic and automated vessel control systems" using the certified Navi Trainer 5000 navigation simulator. The study was carried out within the framework of the research topic No. 0121U109680 "Development of latest models and software for automated control systems for civil and special-purpose marine vessels".

REFERENCES

1. Yundong H. Work motivation and operational risk assessment. A new direction for organizational behavior studies, SSRN, 2016, *Electronic Journal*. DOI: 10.2139/ssrn.2758698.
2. Edwards T., Poling, A. (2020). Motivating operations and negative reinforcement. *Perspectives on behavior science*, 2020, Vol. 4, No. 43, pp. 761–778. DOI:10.1007/s40614-020-00266-8.
3. Ahmad M. A. A., Bakry H. El., Mohamed A. E. D. et al. The impact of using virtual reality on student's motivation for operating system course learning, *Journal of E-Learning and Knowledge Society*, 2020, Vol. 2, No. 16, pp. 25–33. DOI:10.20368/1971-8829/1135076.
4. Nosov P. S., Ben A. P., Mateichuk V. N. et al. Identification of "Human error" negative manifestation in maritime transport, *Radio Electronics, Computer Science, Control*, 2018, Vol. 4, No. 47, pp. 204–213. DOI:10.15588/1607-3274-2018-4-20.
5. Nosov P., Ben A., Safonova A. et al. Approaches going to determination periods of the human factor of navigators during supernumerary situations, *Radio Electronics, Computer Science, Control*, 2019, Vol. 2, No. 49, pp. 140–150. DOI:10.15588/1607-3274-2019-2-15.
6. Nosov P. S., Zinchenko S. M., Popovych I. S. et al. Diagnostic system of perception of navigation danger when implementation complicated maneuvers, *Radio Electronics, Computer Science, Control*, 2020, Vol. 1, pp. 146–161. DOI:10.15588/1607-3274-2020-1-15.
7. Zinchenko S. M., Ben A. P., Nosov P. S. et al. Improving the accuracy and reliability of automatic vessel motion control systems, *Radio Electronics, Computer Science, Control*, 2020, Vol. 2, pp. 183–195. DOI:10.15588/1607-3274-2020-2-19.

8. Frisson C., Malacria S., Bailly G. et al. A system to analyze users behaviors in their applications, *InspectorWidget*, 2016, pp. 1548–1554. DOI:10.1145/2851581.2892388.
9. Zinchenko S.M., Nosov P.S., Mateychuk V.M. et al. Automatic collision avoidance with multiple targets, including maneuvering ones, *Radio Electronics, Computer Science, Control*, 2019, Vol. 4, pp. 211–221. DOI: 10.15588/1607-3274-2019-4-20.
10. Shevchenko R., Popovych I., Spytka L. et al. Comparative analysis of emotional personality traits of the students of maritime science majors caused by long-time staying at sea, *Revista Inclusiones*, 2020, Vol. 7, num Especial, 538–554.
11. Shevchenko R., Cherniavskiy V., Zinchenko S. et al. Research of psychophysiological characteristics of response to stress situations by future sailors, *Revista Inclusiones*, 2020, Vol. 7, num Especial, 566–579.
12. Solmaz M., Özsever B., Güllü A. et al. Development of evaluation procedures for watchkeeping officers using bridge simulator, *TransNav, the International journal on marine navigation and safety of sea transportation*, 2020, Vol. 14, pp. 565–571. DOI:10.12716/1001.14.03.07.
13. Prokopchuk Y. A. Sketch of the Formal Theory of Creativity. Dnepr, PSACEA Press, 2017, 452 p.
14. Baranov V. V. Processes for making management decisions motivated by interests, *Fizmatlit*, 2005, 296.
15. Irodov V. F., Barsuk R. V. et al. Decision-making during limited of experiments with multiple criteria, *Radio Electronics, Computer Science, Control*, 2020, Vol. 1, pp. 200–208. <https://doi.org/10.15588/1607-3274-2020-1-20>.
16. Melnyk K. V., Hlushko V. N., Borysova, N. V. et al. Decision support technology for sprint planning, *Radio Electronics, Computer Science, Control*, 2020, Vol. 1, pp. 135–145. <https://doi.org/10.15588/1607-3274-2020-1-14>.
17. Subbotin S. A., Gofman Y. A. The fractal analysis of sample and decision tree model, *Radio Electronics, Computer Science, Control*, 2020, Vol. 1, pp. 98–107. <https://doi.org/10.15588/1607-3274-2020-1-11>.
18. He Ying, Dyer James, Butler John. A decision-making model with utility from anticipation and disappointment, *Journal of Multi-Criteria Decision Analysis*, 2019, Vol. 26. 10.1002/mcda.1657.
19. Lu, Jin-Biao, Liu, Zhi-Jiang, Tulenty, Dmitry et al. Implementation of Stochastic Analysis in Corporate Decision-Making Models, *Mathematics*, 2021, Vol. 9. DOI:10.3390/math9091041.
20. Acevedo Rafael, Fernández Pedro, Mora Jose et al. Rational Irrationality: A Two-Stage Decision-Making Model. *Advances in Decision Sciences*, 2021, Vol. 25. DOI:10.47654/v25y2021i1p1-39.
21. Pleskacz K., Uriasz J. Understanding of navigational information systems, *Annual of navigation*, 2012, 19, pp. 121–132. DOI:10.2478/v10367-012-0010-z.
22. Wenting L. Tian S., Hong Z. et al. Interface design in the ship navigation information system, *IEEE 11th International Conference on Computer-Aided Industrial Design & Conceptual Design*, 2010, Vol. 1, pp. 395–400. DOI:10.1109/CAIDCD.2010.5681326.
23. Wang C., Zhang J., Yu H., Wang D. et al. Experimental study of optimized interface displays of navigation information system, *Lecture Notes in Electrical Engineering*, 2015, DOI:10.1007/978-3-662-48224-7_45.
24. Nosov P.S., Popovych I.S., Cherniavskiy V.V. et al. Automated identification of an operator anticipation on marine transport, *Radio Electronics, Computer Science, Control*, 2020, Vol. 3, pp. 158–172. DOI:10.15588/1607-3274-2020-3-15.
25. Stillwell J. Group Theory. *In book: Mathematics and Its History*, 2020, pp. 257–282. DOI:10.1007/978-3-030-55193-3_14.
26. Malyutin A., Netsvetaev N. Geometry and topology. Part 10, *Journal of Mathematical Sciences*, 2009, Vol. 3.
27. Huebner M., Vach W. Cessie le S. A systematic approach to initial data analysis is good research practice, *The journal of thoracic and cardiovascular surgery*, 2015, P. 151. DOI:10.1016/j.jtcvs.2015.09.085.
28. Clarke W. Approaches to data analysis, *Clinical trials in neurology. Design, Conduct, Analysis*, 2012, pp. 52–68. DOI:10.1017/CBO9781139032445.007.
29. Weir J.P. Statistical approaches to data analysis, *ACSM's Research Methods*, 2015, pp. 325–342.
30. Lehtola V., Montewka J., Salokannel J. Sea captains' views on automated ship route optimization in ice-covered waters, *Journal of Navigation*, 2019, Vol. 73, pp. 1–20. DOI: 10.1017/S0373463319000651.
31. Picu C., Rusu E. Multiple physical stress exposures of sailors on several ships – a longitudinal study, *Physics, Theoretical Mechanics*, 2018, Vol. 41. pp. 84–93. DOI:10.35219/ann-ugal-math-phys-mec.2018.1.12.
32. Šlišković A., Penezic Z. Lifestyle factors in Croatian seafarers as relating to health and stress on board, *Work*, 2017, Vol. 56. pp. 1–10. DOI:10.3233/WOR-172501.
33. Degtyarev A., Itenberg I., Kharlamov V. Real surfaces: the topological aspects, *Lecture Notes in Mathematics book series*, 2000, Vol. 1746. DOI:10.1007/BFb0103965.
34. Gregori V., Miñana J., Miravet D. A duality relationship between fuzzy partial metrics and fuzzy quasi-metrics. *Mathematics*, 2020, Vol. 8, No 9, P. 1575. DOI:10.3390/math8091575.
35. Chaudhari M., Sarkar S., Sharma D. Analyzing risky behavior in traffic accidents, *Navigation and transport*, 2020, pp. 464–471. DOI:10.1109/SMC42975.2020.9283330.
36. Smolarek L. Dimensioning the Navigational Safety in Maritime Transport, *Journal of Konbin*, 2010, Vol. 1415. pp. 271–280. DOI:10.2478/v10040-008-0184-6.
37. Kriukova T.L., Kuftiak Ye.V. The questionnaire of controlling (The Adaptation of the Methods WCQ), *Journal of an applied psychology specialist*, 2007, Vol. 3, No. 93. pp. 102–112.
38. Lazarus R. S., Folkman S. Stress, Appraisal and coping. New York, Springer Publishing Company, 1984, URL:<https://www.worldcat.org/title/stress-appraisal-and-coping/oclc/609341596>.
39. Quadri C., Zignani M., Gaito, S. et al. Gathering behavior of groups of people in a city. *IEEE International Conference on Smart Computing, SMARTCOM*, 2018, DOI:10.1109/SMARTCOMP.2018.00025.
40. Pazouki K., Zaman I., Norman R., et al. Development of automatic mode detection system by implementing the statistical analysis of ship data to monitor the performance, *The International Journal of Maritime Engineering*, 2017, Vol. 159, pp. 225. DOI:10.3940/ rina.ijme.2017.a3.411.
41. Guo B., Bitner-Gregersen E. M., Sun H. et al. Statistics analysis of ship response in extreme seas, *Ocean Engineering*, 2016, Vol. 119. DOI:10.1016/ j.oceaneng.2016.03.060.
42. Eliopoulou E., Papanikolaou A., Voulgarellis M. Statistical analysis of ship accidents and review of safety level, *Safety Science*, 2016, Vol. 85, pp. 282–292. DOI:10.1016/j.ssci.2016.02.001.

43. Nosov P. S., Cherniavskiy V. V., Zinchenko S. M. et al. Identification of marine emergency response of electronic navigation operator, *Radio Electronics, Computer Science, Control*, 2021, Vol. 1. pp. 208–223. DOI:10.15588/1607-3274-2021-1-20.
44. Pinakpani P., Polisetty A., Bhaskar G. et al. An Algorithmic Approach for Maritime Transportation, *International Journal of Advanced Computer Science and Applications*, 2020, P. 11. DOI:10.14569/IJACSA.2020.0110296.
45. Bazaras D., Palšaitis R., Petraška A. et al. Criteria System of Emergency Situations Risks Assessment in the Baltic Sea Ports. *Transport and Telecommunication Journal*, 2017, Vol. 18. DOI:10.1515/tjt-2017-0024.

Received 14.06.2021.
Accepted 03.08. 2021.

УДК 004.942:656.61.052

АВТОМАТИЧНЕ ВИЗНАЧЕННЯ МОДЕЛІ МОТИВАЦІЇ НАВІГАТОРА ПРИ ЕКСПЛУАТАЦІЇ ВОДНОГО ТРАНСПОРТУ

Носов П. С. – канд. техн. наук, доцент кафедри судноводіння Херсонської державної морської академії, Україна.

Попович І. С. – д-р псих. наук, професор кафедри психології Херсонського державного університету, Україна.

Зінченко С. М. – канд. техн. наук, доцент кафедри управління судном, завідувач лабораторією електронних симуляторів Херсонської державної морської академії, Україна.

Кобець В. М. – д-р економ. наук, професор кафедри інформатики, інженерії програмного забезпечення та економічної кібернетики Херсонського державного університету, Україна.

Сафонова Г. Ф. – канд. техн. наук, доцент кафедри фундаментальних дисциплін Херсонського політехнічного фахового коледжу державного університету «Одеська політехніка», Україна.

Аппазов Е. С. – канд. техн. наук, доцент кафедри інноваційних технологій та технічних засобів судноводіння, Херсонська державна морська академія, Україна.

АНОТАЦІЯ

Актуальність. У статті запропоновано підхід автоматизованої ідентифікації мотиваційної моделі навігаторів при управлінні морським транспортом. Запропоновано алгоритми вилучення даних в результаті людино-машинної взаємодії навігатора з електронними системами управління судном при виконанні навігаційних операцій підвищеної складності.

Мета. Метою дослідження є застосування формального і алгоритмічного підходів до вилучення даних мотиваційної моделі навігатора для запобігання аварій на водному транспорті.

Метод. Пропонується ідентифікація детермінованих проявів розумової діяльності навігаторів за допомогою візуальної концепції геометричної теорії груп. Такий підхід забезпечив наочне системо-логічне поєднання діагностичних методів що спрямовані на визначення мотиваційних центрів штурмана і процесів професійної діяльності, наприклад при виконанні маневрів. Ключовим показником ідентифікації вважається параметр активності штурмана «*grm_port*», що впливає на швидкість судна і є маркером посилення його фізіологічної активності. Такий підхід корисний для ідентифікації тимчасових фаз при маневруванні, що явно вказують на зміну мотиваційного стану навігатора. Даний аспект був доведений на підставі результатів дендрограми Уорда, кількох статистичних методів і прикладного програмного забезпечення. Отримані результати досліджень дозволяють прогнозувати мотиваційні стани навігатора у критичних ситуаціях.

Результат. З метою підтвердження запропонованого формально-алгоритмічного підходу був проведений експеримент з використанням навігаційного симулятора Navi Trainer 5000. Автоматизований аналіз експериментальних даних дозволив сформувати мотиваційну карту навігатора і визначити модель прийняття рішень що впливають на процеси управління судном у складних ситуаціях.

Висновок. Запропоновані підходи дослідження дозволили автоматизувати процеси вилучення даних що вказують на принципи прийняття рішень навігатором. Результативність запропонованого підходу була обґрунтована за результатами автоматизованої обробки експериментальних даних і побудованих ознакових деревоподібних просторів прийняття рішень.

КЛЮЧОВІ СЛОВА: системи ідентифікації мотивації, системи автоматизованої обробки даних, моделювання моделей прийняття рішень, комп'ютерні стимулятори, аналіз людського фактора, автоматизовані системи управління.

УДК 004.942:656.61.052

АВТОМАТИЧЕСКОЕ ОПРЕДЕЛЕНИЕ МОДЕЛИ МОТИВАЦИИ НАВИГАТОРА ПРИ ЭКСПЛУАТАЦИИ ВОДНОГО ТРАНСПОРТА

Носов П. С. – канд. техн. наук, доцент кафедры судовождения Херсонской государственной морской академии, Украина.

Попович И. С. – д-р псих. наук, профессор кафедры психологии Херсонского государственного университета, Украина.

Зинченко С. М. – канд. техн. наук, доцент кафедры управления судном, заведующий лабораторией электронных симуляторов Херсонской государственной морской академии, Украина.

Кобец В. М. – д-р эконом. наук, профессор кафедры информатики, инженерии программного обеспечения экономической кибернетики Херсонского государственного университета, Украина.

Сафонова А. Ф. – канд. техн. наук, доцент кафедры фундаментальных дисциплин Херсонского политехнического профессионального колледжа государственного университета «Одесская политехника», Украина.

Аппазов Э. С. – канд. техн. наук, доцент кафедры инновационных технологий и технических средств судовождения, Херсонская государственная морская академия, Украина.

АННОТАЦІЯ

Актуальність. В статті пропонується підхід автоматизованої ідентифікації мотиваційної моделі навігаторів при управлінні морським транспортом. Предполагается алгоритмы извлечения данных в результате человеко-машинного взаимодействия навігатора с електронними системами управління судном при виконанні навігаційних операцій підвищеної складності.

Цель. Целью исследования является применение формального и алгоритмического подходов к извлечению данных мотивационной модели навігатора для предотвращения аварий на водном транспорте.

Метод. Предлагается идентификация детерминированных проявлений мыслительной деятельности навігаторов с помощью наглядной концепции геометрической теории групп. Такой подход обеспечил наглядное систематико-логическое сочетание диагностических методов, направленных на определение мотивационных центров штурмана и процессов профессиональной деятельности, например при выполнении маневров. Ключевым показателем идентификации считается параметр активности штурмана («rpm_port»), влияющий на скорость судна и являющийся маркером усиления его физиологической активности. Такой подход полезен для идентификации временной фазы при маневрировании, что явно указывает на изменение мотивационного состояния навігатора. Данный аспект был доказан на основании результатов дендрограммы Уорда, нескольких статистических методов и прикладного программного обеспечения. Полученные результаты исследований позволяют прогнозировать мотивационные состояния навігатора в критических ситуациях.

Результат. С целью подтверждения предложенного формально-алгоритмического подхода был проведен эксперимент с использованием навігационного симулятора Navi Trainer 5000. Автоматизированный анализ экспериментальных данных позволил сформировать мотивационную карту навігатора и определить модель принятия решений влияющих на процессы управления судном в сложных ситуациях.

Вывод. Предложенные подходы исследования позволили автоматизировать процессы извлечения данных указывающих на принципы принятия решений навігатором. Результативность предложенного подхода была обоснована по результатам автоматизированной обработки экспериментальных данных и построенных признаковых древовидных пространств принятия решений.

КЛЮЧЕВЫЕ СЛОВА: системы идентификации мотивации, системы автоматизированной обработки данных, моделирование моделей принятия решений, компьютерные симуляторы, анализ человеческого фактора, автоматизированные системы управления.

ЛІТЕРАТУРА / LITERATURA

1. Yundong H. Work motivation and operational risk assessment. A new direction for organizational behavior studies / H. Yundong // SSRN Electronic Journal. – 2016. DOI:10.2139/ssrn.2758698.
2. Edwards T. Motivating operations and negative reinforcement / T. Edwards, A. Poling // Perspectives on behavior science. – 2020. – Vol. 4, No. 43. – P. 761–778. DOI: 10.1007/s40614-020-00266-8.
3. The impact of using virtual reality on student’s motivation for operating system course learning / [M. A. A. Ahmad, H. El. Bakry, A. E. D. Mohamed, M. B. Senouy] // Journal of E-Learning and Knowledge Society. – 2020. – Vol. 2, No. 16. – P. 25–33. DOI:10.20368/1971-8829/1135076.
4. Identification of “Human error” negative manifestation in maritime transport / [P. S. Nosov, A. P. Ben, V. N. Mateichuk, M. S. Safonov] // Radio Electronics, Computer Science, Control. – 2018. – Vol. 4, No. 47. – P. 204–213. DOI:10.15588/1607-3274-2018-4-20.
5. Approaches going to determination periods of the human factor of navigators during supernumerary situations / [P. Nosov, A. Ben, A. Safonova, I. Palamarchuk] // Radio Electronics, Computer Science, Control. – 2019. – Vol. 2, No. 49. – P. 140–150. DOI:10.15588/1607-3274-2019-2-15.
6. Nosov P. S. Diagnostic system of perception of navigation danger when implementation complicated maneuvers / [P. S. Nosov, S. M. Zinchenko, I. S. Popovych et al.] // Electronics, Computer Science, Control. – 2020. – Vol. 1. – P. 146–161. DOI:10.15588/1607-3274-2020-1-15.
7. Zinchenko S. M. Improving the accuracy and reliability of automatic vessel motion control systems / [S. M. Zinchenko, A. P. Ben, P. S. Nosov et al.] // Radio Electronics, Computer Science, Control. – 2020. – Vol. 2. – P. 183–195. DOI:10.15588/1607-3274-2020-2-19
8. A system to analyze users behaviors in their applications / [C. Frisson, S. Malacria, G. Bailly, T. Dutoit] // Inspector-Widget. – 2016. – P. 1548–1554. DOI:10.1145/2851581.2892388.
9. Automatic collision avoidance with multiple targets, including maneuvering ones / [S. M. Zinchenko, P. S. Nosov, V. M. Mateychuk et al.] // Radio Electronics, Computer Science, Control. – 2019. – Vol. 4. – P. 211–221. DOI:10.15588/1607-3274-2019-4-20.
10. Comparative analysis of emotional personality traits of the students of maritime science majors caused by long-time staying at sea / [R. Shevchenko, I. Popovych, L. Spytyska et al.] // Revista Inclusiones. – 2020. – Vol. 7. – P. 538–554.
11. Research of psychophysiological characteristics of response to stress situations by future sailors / [R. Shevchenko, V. Cherniavskiy, S. Zinchenko, M. Palchynska et al.] // Revista Inclusiones. – 2020. – Vol. 7. – P. 566–579.
12. Development of evaluation procedures for watchkeeping officers using bridge simulator / [M. Solmaz, B. Özsever, A. Güllü, C. Meşe] // TransNav, the International journal on marine navigation and safety of sea transportation. – 2020. – Vol. 14. – P. 565–571. DOI:10.12716/1001.14.03.07.
13. Prokopchuk Y. A. Sketch of the formal theory of creativity / Y. A. Prokopchuk. – Dnepr : PSACEA, 2017. – P. 452.
14. Баранов В. В. Процессы принятия управляющих решений, мотивированных интересами / В. В. Баранов. – Физматлит, 2005. – 296 с.
15. Irodov V. F. Decision-making during limited of experiments with multiple criteria / V. F. Irodov, R. V. Barsuk // Radio Electronics, Computer Science, Control. – 2020. – Vol. 1. – P. 200–208. DOI:10.15588/1607-3274-2020-1-20.
16. Melnyk K. V. Decision support technology for sprint planning / K. V. Melnyk, V. N. Hlushko, N. V. Borysova // Radio Electronics, Computer Science, Control. – 2020. – Vol. 1. – P. 135–145. DOI:10.15588/1607-3274-2020-1-14.
17. Subbotin S. A. The fractal analysis of sample and decision tree model / S. A. Subbotin, Y. A. Gofman // Radio Electronics, Computer Science, Control. – 2020. – Vol. 1. – P. 98–107. DOI: 10.15588/1607-3274-2020-1-11.

18. He Y. A decision-making model with utility from anticipation and disappointment / Y. He, J. Dyer, J. Butler // *Journal of Multi-Criteria Decision Analysis*. – 2019. – P. 26. DOI: 10.1002/mcda.1657.
19. Lu J-B. Implementation of Stochastic Analysis in Corporate Decision-Making Models / [J-B Lu, Z-J Liu, D. Tulenty et al.] // *Mathematics*. – 2021. – P. 9. DOI: 10.3390/math9091041.
20. Rational irrationality a two-stage decision-making model / [R. Acevedo, P. Fernández, J. Mora, E. Aponte] // *Advances in Decision Sciences*. – 2021. – P. 25. DOI: 10.47654/v25y2021i1p1-39.
21. Pleskacz K. Understanding of navigational information systems / K. Pleskacz, J. Uriasz // *Annual of navigation*. – 2012. – Vol. 19. – P. 121–132. DOI:10.2478/v10367-012-0010-z.
22. Interface design in the ship navigation information system / [L. Wenting, S. Tian, Z. Hong, Y. Ying] // *IEEE 11th International Conference on Computer-Aided Industrial Design & Conceptual Design*. – 2010. – Vol. 1. – P. 395–400. DOI: 10.1109/CAIDCD.2010.5681326.
23. Experimental study of optimized interface displays of navigation information system / [C. Wang, J. Zhang, H. Yu et al.] // *Lecture Notes in Electrical Engineering*. – 2015. DOI:10.1007/978-3-662-48224-7_45.
24. Automated identification of an operator anticipation on marine transport / [P. S. Nosov, I. S. Popovych, V. V. Cherniavskiy et al.] // *Radio Electronics, Computer Science, Control*. – 2020. – Vol. 3. – P. 158–172. DOI:10.15588/1607-3274-2020-3-15.
25. Stillwell J. Group Theory / J. Stillwell // *In book: Mathematics and Its History*. – 2020. – P. 257–282. DOI:10.1007/978-3-030-55193-3_14.
26. Malyutin A. Geometry and topology / Malyutin A., Netsvetayev N. // *Journal of Mathematical Sciences*. – 2009. – Vol. 3. – P. 10.
27. Huebner M. A systematic approach to initial data analysis is good research practice / M. Huebner, W. Vach, le S. Cessie // *The journal of thoracic and cardiovascular surgery*. – 2015. – P. 151. DOI:10.1016/j.jtcvs.2015.09.085.
28. Clarke W. Approaches to data analysis / W. Clarke // *Clinical trials in neurology, Design, Conduct, Analysis*. – 2012. – P. 52–68. DOI:10.1017/CBO9781139032445.007.
29. Weir J.P. Statistical approaches to data analysis / In L. E. Armstrong & W. J. Kraemer (Eds.) // *ACSM's Research Methods*. – 2015. – P. 325–342.
30. Lehtola V. Sea captains' views on automated ship route optimization in ice-covered waters / V. Lehtola, J. Montewka, J. Salokannel // *Journal of Navigation*. – 2019. – Vol. 73. – P. 1–20. DOI:10.1017/S0373463319000651.
31. Picu C. Multiple physical stress exposures of sailors on several ships – a longitudinal study / C. Picu, E. Rusu // *Physics, Theoretical Mechanics*. – 2018. – Vol. 41. – P. 84–93. DOI:10.35219/ann-ugal-math-phys-mec.2018.1.12.
32. Slišković A. Lifestyle factors in Croatian seafarers as relating to health and stress on board / A. Slišković, Z. Penezic // *Work*. – 2017. – Vol. 56. – P. 1–10. DOI:10.3233/WOR-172501.
33. Degtyarev A. Real surfaces: the topological aspects / A. Degtyarev, I. Itenberg, V. Kharlamov // *Lecture Notes in Mathematics book series*. – 2000. – Vol. 1746. DOI: 10.1007/BFb0103965.
34. Gregori V. A duality relationship between fuzzy partial metrics and fuzzy quasi-metrics / V. Gregori, J. Miñana, D. Miravet // *Mathematics*. – 2020. – Vol. 8(9). – P. 1575. DOI:10.3390/math8091575.
35. Chaudhari M. Analyzing risky behavior in traffic accidents / M. Chaudhari, S. Sarkar, D. Sharma // *Navigation and transport*. – 2020. – P. 464–471. DOI:10.1109/SMC42975.2020.9283330.
36. Smolarek, L. Dimensioning the Navigational Safety in Maritime Transport / L. Smolarek // *Journal of Konbin*. – 2010. – Vol. 1415. – P. 271–280. DOI:10.2478/v10040-008-0184-6.
37. Kriukova T. L. The questionnaire of controlling (The Adaptation of the Methods WCQ) / T. L. Kriukova, Ye. V. Kuftiak // *Journal of an applied psychology specialist*. – 2007. – Vol. 3, No. 93. – P. 102–112.
38. Lazarus R.S. Stress, Appraisal and coping / R. S. Lazarus, S. Folkman // *New York, Springer Publishing Company*. – 1984. URL:https://www.worldcat.org/title/stress-appraisal-and-coping/oclc/609341596.
39. Gathering behavior of groups of people in a city / [C. Quadri, M. Zignani, S. Gaito, G. Rossi] // *IEEE International Conference on Smart Computing, SMARTCOMP*. – 2018. DOI:10.1109/SMARTCOMP.2018.00025.
40. Pazouki K. Development of automatic mode detection system by implementing the statistical analysis of ship data to monitor the performance / [K. Pazouki, I. Zaman, R. Norman et al.] // *The International Journal of Maritime Engineering*. – 2020. – Vol. 159. – P. 225. DOI:10.3940/rina.ijme.2017.a3.411.
41. Guo B. Statistics analysis of ship response in extreme seas / B. Guo, E. M. Bitner-Gregersen, H. Sun, J. Helmers // *Ocean Engineering*. – 2020. – Vol. 119. DOI:10.1016/j.oceaneng.2016.03.060.
42. Eliopoulou E. Statistical analysis of ship accidents and review of safety level / E. Eliopoulou, A. Papanikolaou, M. Voulgarellis // *Safety Science*. – 2016. – Vol. 85. – P. 282–292. DOI:10.1016/j.ssci.2016.02.001.
43. Identification of marine emergency response of electronic navigation operator / [P. S. Nosov, V. V. Cherniavskiy, S. M. Zinchenko et al.] // *Radio Electronics, Computer Science, Control*. – 2020. – Vol. 1. – P. 208–223. DOI: 10.15588/1607-3274-2021-1-20.
44. Pinakpani, P. An Algorithmic Approach for Maritime Transportation / [P. Pinakpani, A. Polisetty, G. Bhaskar et al.] // *International Journal of Advanced Computer Science and Applications*. – 2020. – P. 11. DOI:10.14569/IJACSA.2020.0110296.
45. Criteria System of Emergency Situations Risks Assessment in the Baltic Sea Ports / [D. Bazaras, R. Palšaitis, A. Petraška, A. Zvaigzne] // *Transport and Telecommunication Journal*. – 2017. – P. 18. DOI: 10.1515/tjt-2017-0024.

MULTI-AGENT LATENT SEMANTIC INTERNET TECHNOLOGY FOR THE FORMATION OF A SUBJECT-ORIENTED KNOWLEDGE MODEL

Stenin A. A. – Dr. Sc., Professor of the Department of technical Cybernetics, Kiev Polytechnic Institute. Igor Sikorsky, Kiev, Ukraine.

Pasko V. P. – PhD, Associate Professor, Department of technical Cybernetics, Kyiv Polytechnic Institute. Igor Sikorsky, Kiev, Ukraine.

Soldatova M. A. – PhD, Senior lecturer of the Department of technical Cybernetics, Kiev Polytechnic Institute. Igor Sikorsky, Kiev, Ukraine.

Drozdovich I. G. – PhD, Senior Scientific Associate at the Institute of telecommunications and global information space of the National Academy of Sciences of Ukraine.

ABSTRACT

Context. The article proposes a latent-semantic technology for extracting information from Internet resources, which allows processing information in natural language, as well as a multi-agent search algorithm based on it. The relevance of this approach to the search for subject-oriented information determined by the fact that currently a direct lexical comparison of queries with document indexes does not fully satisfy the developer. The object of the study is a multi-agent latent-semantic algorithm for searching for subject-oriented information.

Objective. The work is to increase the efficiency of forming a knowledge model that is adequate for this subject area.

Method. A latent semantic technology based on the weighted descriptor method developed by the authors is proposed. The main difference from the existing methods is that the analysis of words occurring in the text both in frequency and taking into account semantics carried out by selecting the appropriate descriptors, which improves the quality of the information found.

Results. The developed latent-semantic technology of information search tested in the task of constructing a knowledge model of automated decision support systems for operational and dispatching control of urban engineering networks. The conducted modeling of the search for subject-oriented information in this subject area showed the effectiveness of the developed approach.

Conclusions. Improving the efficiency of search and semantic content of subject-oriented information of the knowledge model of this subject area achieved by using the weighted descriptor method based on Zipf's laws in this technology. The prospects for further research are to build evolutionary models of knowledge and improve the quality of updated information.

KEYWORDS: Internet resources, information search, Zipf's laws, Grebner bases, intelligent agents, weighted descriptors, latent semantic analysis, multi-agent automatic search procedure.

ABBREVIATIONS

DSS – decision support systems;
IRS – information retrieval systems;
LSA – latent semantic analysis;
UEN – urban engineering networks.

NOMENCLATURE

f_i is a frequency of occurrence of the i -th linguistic variable;

k_i is a number of documents with the i -th linguistic variable;

N is a significant set of documents;

N_0 is a total number of documents under consideration;

f is a frequency of occurrences of words in texts;

k is a rank of a frequency;

w_{ij} is a frequency significance coefficient;

α_i is a semantic significance coefficient;

a_{ij} is a frequency of appearance of the i -th descriptor in the j -th document $i=1, \dots, m, j=1, \dots, N_0$;

a_{ik} is a partial private evaluation of the i -th descriptor significance;

k is a number of expert;

S_q is q -th current situation;

x_i is i -th linguistic variable (descriptor);

$M(x_i)$ is a function of belonging of i -th descriptor;

T_q^i is a basis, which containing i -th descriptor of q -th situation.

INTRODUCTION

With the development of Internet technologies, a new giant source of information resources has appeared. Thanks to the widespread development and application of computer technology, we can get information in electronic form in all areas of human activity, such as science, production, commerce, literature, entertainment, etc. [1]. The Internet is compatible with various electronic networks and databases and allows easy access to almost any kind of information. However, the development of the Internet as an information repository took place without taking into account the need to search for the documents. As a result, on the Internet, in contrast to traditional IRS, where the document storage system is focused on active search [2–4], the Internet document storage system is not a given a priori with respect to the task of information retrieval, i.e. poorly structured. The Internet is a decentralized document repository with no of the single management and organization and of development. The Internet is heterogeneous, as not only different platforms uses, but also different standards of information presentation. The Internet brings together both modern

and legacy systems. Piece of information is stored in a form other than text (multimedia)

It follows that the task of extracting information from the Internet is complex since it is necessary to extract not only the type of data scheme, but also the semantic information associated with it. In addition, given the need to search for the specific information of the selected subject area, there is a need to process a huge quantity of documents. Thus, it is also important to automate the search process for the effective selection of the most informative content. The process of search in an IRS can be represented in the form of the scheme shown in Fig. 1 [4].

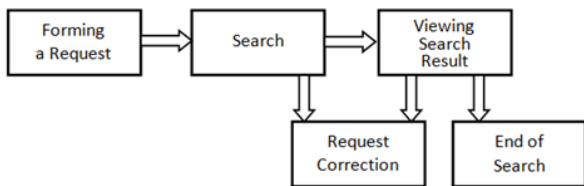


Figure 1 – Diagram of the search process in Information Retrieval Systems

The object of research is a multi-agent procedure for searching the Internet for the necessary information, taking into account its semantic value for a given subject area.

The subject of research is the latent-semantic method of weighted descriptors, which allows extracting the most significant documents in terms of meaning and meaning that are very close to this subject area.

The aim of the research is to form an effective query and automated procedure for analyzing information extracted from the Internet for a given subject area.

1 PROBLEM STATEMENT

Taking into account the above, we can say that the main task of the development of IRS on the Internet is the development of methods and tools for semantic analysis of the text in a natural language using a multi-agent approach to reduce the time of searching for the necessary information and increase its semantic value [5].

The formalized task of searching for innovative subject-oriented document can be presented in the form of matrix “descriptors-documents”

$$A = \{T, W, R, \Lambda\} \quad (1)$$

Necessary:

1. Define a set of descriptors $W = \{w_i\}$, that reflect the semantic content of the terms of reference and the current situation S_i at the time of innovative design of the technical object

$$S_q = \left\{ M_{M_{S_q}(x_i)}(T_q^i) / T_q^i \right\} q = 1, \dots, Q; i = 1, \dots, N_q; x_i \in X.$$

2. Based on (1) and (2), determine the initial matrix “descriptors -documents” A using the relevance criteria

and expert assessments, where m is the number of descriptors, N_0 is the number of documents found.

3. Perform LSA k -approximation of matrix A in order to determine the most informative documents, i.e. find the matrix

$$\tilde{A} \approx USV^T \quad (2)$$

2 REVIEW OF THE LITERATURE

Currently, in the sense of automation of the IRS, actively working on the development of algorithms, which automatically generate intermediary programs (intelligent agents) in the search for information from the Internet.

Achieving full automation in this matter is unlikely, and we can only talk about the creation of automated methods and systems for extracting information from the Internet. Actual research in the field of work with poorly structured information on the basis of “intelligent agents” led to the emergence of a large quantity of alternative tools for their creation [4, 8].

The main approaches to solving the problem of data extraction from the Internet are borrowed from such areas as data processing in a natural language, machine learning, ontology, etc. In particular, the main task of extracting data from the Web is to obtain certain pieces of information (fields) from HTML documents [6, 7].

This task is close to the task of automatic clustering and is to find the decomposition of HTML documents into classes that contain documents with a similar structure. The task of displaying applied objects in points of multidimensional space is to determine the basis of features that form a multidimensional space, and the method of decomposition of the document on this basis, that is, the calculation of the coordinates.

Different approaches were used to determine the coordinates of the document in the space of basic features. In particular, the authors of [7] propose to use the approach popular in calculating the weights of terms in the IRS, using a vector model of presentation of documents. The coordinates of the document are determined by the formula:

$$w_i = f_i / \log(N / k_i) \quad (3)$$

Often, an entropic measure is used to assess the quality of clustering. However, this approach, which determines the significance of the term only in frequency does not guarantee the significance of the document in meaning.

To overcome this contradiction, the W3C Interest Group Note consortium develops Web semantic [8, 9]. According to the idea of its creators, the implementation of this paradigm on the Internet will allow information systems to understand the content of information to some extent and act as intellectual intermediaries capable of manipulating it on the instructions of a person [10, 11].

In this sense, the use of George Zipf’s laws is relevant [12]. Zipf found that if you multiply the probability of

finding a word in the text by the rank of frequency, the resulting value is approximately constant for all texts in one language (for English texts $C \approx 1$, for Russian texts $C \approx 0.06 - 0.07$):

$$C = (f \cdot x) / N \approx \text{const}, \quad (4)$$

frequency of occurrence of the word in the text has the form of a hyperbola (Fig. 2). The rank is permanent for texts of the same language.

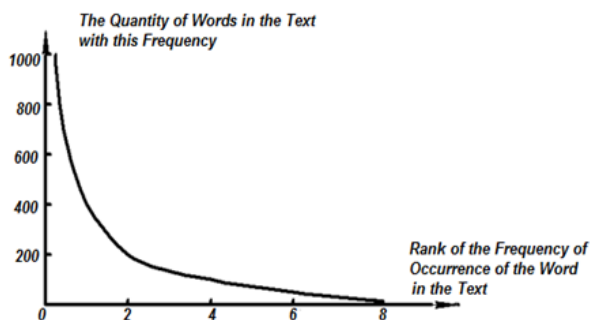


Figure 2 – Zipf curve

The discovery of Zipf's laws was the beginning of infometrii, the science of measuring the quantitative characteristics of information. Explanation of Zipf's laws based on correlation properties of additive Markov chains with step memory function [13]. Zipf's laws are universal. They, in particular, meet the characteristics of the popularity of Internet sites.

3 MATERIAL AND METHODS

As mentioned above, the semantic approach is currently one of the main ways to improve the IRS, as direct lexical comparison of requests with document indexes does not fully satisfy the developer. This is due to found documents have either polysemy (i.e, lot of extra words) or synonymy (i.e, not all meaningful words extracted). Therefore, within the framework of the semantic approach, a latent-semantic method of weighted descriptors, which allows extracting the most significant in meaning and significance documents that are very close to the subject area, is proposed.

The method based on the idea of Groebner basis [14], which statistically constructed conceptual descriptors. This method assumes that the conceptual descriptors in the sentences have an underlying, "latent" meaning, which obscured by the use of different words. "Ideal" in determining the basis of Groebner consider the terms of reference for the implementation of some innovative project. To obtain meaningful conceptual descriptors, we will use Zipf's laws described above.

Zipf curve analysis (Fig. 2) shows that the most significant words, and, consequently, the meaningful conceptual descriptors constructed on them, lie in the middle part of the Zipf curve. In the same time, in the Russian language the most common prepositions, pronouns, etc., and in English – service words and others. They answer by the left part of the diagram. The right part

of the Zipf curve corresponds to words that have no decisive semantic meaning and are not of interest in the formation of conceptual descriptors.

Therefore, the success of the IRS depends on how the range of the most significant words and conceptual descriptors constructed on their basis will be determined. In many ways, the definition of the range large ($N_0 \gg 0$) depends on the correct compilation of the special dictionaries – the thesaurus of a subject area and the "stop-dictionary". The thesaurus of a subject area gives you the opportunity to correctly determine the set of concept descriptors from the technical specifications and the most important concepts in this subject area. Stop dictionary cuts off "interference" in the form of "extra" words, i.e. for the Russian language – it particles, prepositions, pronouns, etc.

The choice of the quantity of descriptors determined by the quantity of constructed Groebner basis, the totality of which fully reflects the "ideal", i.e. an innovative project reflected in the terms of reference.

Let us initially construct and select n descriptors. Then on their request to the Internet, we will get a rectangular matrix "descriptors-documents" $A_1 = \{a_{ij}\}$ with dimension $m \times N_0$.

Usually, the Internet search for the selected descriptors the initial quantity of documents turns out to be quite large. In this case, the work with large-dimensional matrices within the LSA is significantly complicated. Therefore, in this paper proposed to determine the most significant set of documents from the set in two stages.

At the first stage, we rank the set of N_0 found documents by the frequency of using basic descriptors in them and by their semantic significance. To do this, we construct a criterion of the relevance of the following form:

$$R_j = \sum_{i=1}^n \alpha_i w_{ij}, \quad j = \overline{1, N}. \quad (5)$$

As a frequency significance coefficient can use the generally accepted formula (3), in which we understand the frequency of appearance of the descriptor in the document.

The semantic coefficient of significance determined by the method of expert evaluations according to the following formula:

$$\alpha_i = \sum_{k=1}^K \alpha_{ik} / \sum_{i=1}^m \sum_{k=1}^K \alpha_{ik}. \quad (6)$$

Private estimates are determined by their ranking by a natural quantity series, in this case, from 1 to m . A larger quantity corresponds to a large assessment of the significance of the descriptor.

As a result, we form the set of N significant documents and the initial matrix $A(m \times n)$ for LSA, where $n=N \ll N_0$. As a rule, the sample set N depends on the concrete problem and chosen empirically. In this case, we accept $a_{ij} = 1$, if the i -th descriptor is in the j -th document, and $a_{ij} = 0$, otherwise.

At the second stage, we use singular value decomposition of matrix A , which is described below.

The LSA method based on the principles of factor analysis, in particular, the identification of latent relationships of the studied phenomena or objects [15]. LSA can be compared with a simple view of a neural network consisting of three layers: the first layer contains a set of words, the second – a set of documents, and the third, the middle (hidden layer), is a set of nodes with different weight coefficients connecting the first and second layers.

The basic idea of k -approximation of the latent-semantic approach to the matrix A is to replace a matrix \tilde{A} containing only k – the first linearly-independent components of the matrix, and reflects the basic structure of the various dependencies presented in the A .

More formally, according to the singular decomposition theorem [16], a rectangular real matrix can be decomposed into a product of three matrices:

$$A=USV^T. \quad (7)$$

In this transformation U and V specially constructed orthogonal matrices U and V and S is the diagonal matrix whose diagonal values represent the singular values of the A matrix. This decomposition has a remarkable feature. If matrix S leave only k the largest singular values, in the matrices U and V leave only the columns corresponding to these values, then the product of the resulting matrices S , U and V will be the best approximation of the original matrix A to the matrix (2) with a k rank [17].

The well-known MATCAD software package [18,19] provides an **svd** (A) function that implements the singular value decomposition of matrix A .

4 EXPERIMENTS

As an example of the use of LSA, the terms of reference for the development of intelligent DSS software for operational dispatching control of urban engineering networks (UEN) are taken. The analysis of the semantic content of the text of the technical task made 13

$$V:=M_3 = \begin{pmatrix} -0426 & -0046 & -6326 \times 10^3 & -0535 & -4818 \times 10^3 & -0372 & -0011 & -0625 & -3862 \times 10^3 \\ 4892 \times 10^3 & -0022 & -0653 & 6277 \times 10^3 & -0585 & 1894 \times 10^3 & -0089 & 7407 \times 10^3 & -0471 \\ -0028 & 0701 & 0037 & -0064 & -01 & 0162 & 0671 & -0086 & -0088 \\ 0217 & 0244 & -0148 & -0277 & 0115 & 0684 & -0436 & -0327 & 013 \\ 0692 & -0325 & 0224 & -0489 & -0123 & -0031 & 0271 & -0016 & -0194 \\ -0273 & -0336 & 0438 & 0293 & -0134 & 0453 & 0115 & -0312 & -0448 \\ 0031 & -0298 & -0141 & 017 & -0444 & 015 & 0296 & -0239 & 0705 \\ 0303 & -0119 & -0404 & 0391 & 0528 & -0122 & 0228 & -0464 & -0133 \\ 0354 & 0354 & 0354 & 0354 & -0354 & -0354 & -0354 & -0354 & 0 \end{pmatrix}$$

descriptors ($m = 13$) as Groebner bases. They are the following words and phrases which are most often found, bearing the main semantic significance of the essence of the technical task: intelligent DSS, UEN, dispatcher, operational management, emergencies, fuzzy logic, situational uncertainty, fuzzy situational network, probabilistic transitions, multi-agent technology, decision-making, etc. [20].

In the formation of descriptors, the operation stemming is used [21]. In addition, of the descriptors were deleted stop characters, i.e. all conjunctions, particles, prepositions, etc.

5 RESULTS

On the descriptors, a search of documents was carried out on the Internet, the quantity of which after ranking and truncation at the first stage was $m=13$. As a result, we got matrix ($m \times n$):

$$A := \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Next, we perform a singular value decomposition of the resulting matrix in MATHCAD, i.e. we present A as matrix “descriptors-documents” with the diagonal elements of the S matrix (singular values) ordered in descending order:

$$M := svd2(A)$$

$$M = \begin{pmatrix} \{9,1\} \\ \{13,9\} \\ \{9,9\} \end{pmatrix}$$

	1	2	3	4	5	6	7	8	9
1	-0.574	6.205 10 ⁻³	-6.205 10 ⁻³	0.199	0.13	0.163	0.157	0.251	-0.539
2	-0.34	4.148 10 ⁻³	-0.066	-0.405	-0.423	-0.019	-0.096	-0.166	-0.402
3	-0.34	4.148 10 ⁻³	-0.066	-0.405	-0.423	-0.096	-0.096	-0.166	0.402
4	-4.395 10 ⁻³	-0.518	-0.067	0.065	-0.077	-0.147	0.168	-0.019	0.193
5	-4.395 10 ⁻³	-0.518	-0.067	0.065	-0.077	-0.147	0.168	-0.019	0.039
6	-0.574	-6.205 10 ⁻³	-6.352 10 ⁻³	0.199	0.13	0.163	0.157	0.251	0.539
7	-0.308	3.728 10 ⁻³	-0.05	-0.074	0.565	-0.595	-0.292	-0.371	0
8	-4.395 10 ⁻³	-0.518	-0.067	0.065	-0.077	-0.147	0.167	-0.019	-0.232
9	-0.017	-0.034	0.605	-0.129	-0.045	-0.225	-2.822 10 ⁻³	0.252	-0.048
10	-5.15 10 ⁻³	-0.225	0.312	-0.392	0.414	0.562	0.217	-0.404	0
11	-0.122	-6.162 10 ⁻³	0.381	0.622	-0.298	0.119	-0.207	-0.55	0
12	-0.017	-0.034	0.605	-0.129	-0.045	-0.225	-2.822 10 ⁻³	0.252	0.048
13	-3.264 10 ⁻³	-0.375	-0.028	-0.023	0.085	0.309	-0.82	0.287	0

$$S := \text{diag}(M_1) = \begin{pmatrix} 3.414 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3.299 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2.268 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.491 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.195 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.983 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.714 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.434 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$S2_sing := \text{submatrix}(S, 1, 2, 1, 2) = \begin{pmatrix} 3.414 & 0 \\ 0 & 3.299 \end{pmatrix}.$$

$$S2_sing := \text{submatrix}(M_3, 1, 2, 1, 9) = \begin{pmatrix} -0.426 & -0.046 & -6.326 \times 10^{-3} & -0.535 & -4.818 \times 10^{-3} & -0.372 & -0.011 & -0.625 & -3.862 \times 10^{-3} \\ 4.892 \times 10^{-3} & -0.022 & -0.653 & 6.277 \times 10^{-3} & -0.585 & 1.894 \times 10^{-3} & -0.089 & 7.407 \times 10^{-3} & -0.471 \end{pmatrix}$$

	1	2
1	-0.574	6.205 · 10 ⁻³
2	-0.34	4.148 · 10 ⁻³
3	-0.34	4.148 · 10 ⁻³
4	-4.395 · 10 ⁻³	-0.518
5	-4.395 · 10 ⁻³	-0.518
6	-0.574	6.205 · 10 ⁻³
7	-0.308	3.728 · 10 ⁻³
8	-4.395 · 10 ⁻³	-0.574
9	-0.017	-0.034
10	-5.15 · 10 ⁻³	-0.225
11	-0.122	-6.162 · 10 ⁻³
12	-0.017	-0.034
13	-3.264 · 10 ⁻³	-0.375

By according rules matrix multiplication, columns and rows corresponding to smaller singular values give the least contribution to the final result. In our case, according to (6) a two-dimensional singular value decomposition ($k = 2$) is conducted.

$$U2_sing := \text{submatrix}(M_2, 1, 13, 1, 2) =$$

$$X =$$

	1	2	3	4	5	6	7	8	9
1	0.835	0.089	-9.818 · 10 ⁻³	1.049	-2.541 · 10 ⁻³	0.729	0.02	1.225	-2.084 · 10 ⁻³
2	0.495	0.053	-1.598 · 10 ⁻³	0.621	-2.541 · 10 ⁻³	0.432	0.012	0.726	-1.968 · 10 ⁻³
3	0.495	0.053	-1.598 · 10 ⁻³	0.621	-2.541 · 10 ⁻³	0.432	0.012	0.726	-1.968 · 10 ⁻³
4	-1.968 · 10 ⁻³	0.039	1.117	-2.701 · 10 ⁻³	1.001	2.343 · 10 ⁻³	0.152	-3.28 · 10 ⁻³	0.806
5	-1.968 · 10 ⁻³	0.039	1.117	-2.701 · 10 ⁻³	1.001	2.343 · 10 ⁻³	0.152	-3.28 · 10 ⁻³	0.806
6	0.835	0.089	-9.818 · 10 ⁻⁴	1.049	-2.541 · 10 ⁻³	0.729	0.02	1.225	-2.084 · 10 ⁻³
7	0.448	0.048	-1.382 · 10 ⁻³	0.563	-2.13 · 10 ⁻³	0.391	0.011	0.658	-1.736 · 10 ⁻³
8	-1.968 · 10 ⁻³	0.039	1.117	-2.701 · 10 ⁻³	1.001	2.343 · 10 ⁻³	0.152	-3.28 · 10 ⁻³	0.806
9	0.024	5.049 · 10 ⁻³	0.073	0.03	0.065	0.021	0.01	0.035	0.052
10	3.866 · 10 ⁻³	0.017	0.485	4.754 · 10 ⁻³	0.434	5.134 · 10 ⁻³	0.066	5.502 · 10 ⁻³	0.35
11	0.178	0.019	0.016	0.223	0.014	0.155	6.499 · 10 ⁻³	0.261	0.011
12	0.024	5.049 · 10 ⁻³	0.073	0.03	0.065	0.021	0.01	0.035	0.052
13	-1.308 · 10 ⁻³	0.028	0.809	-1.809 · 10 ⁻³	0.725	1.799 · 10 ⁻³	0.11	-2.204 · 10 ⁻³	0.584

Matrix $X = \tilde{A}$ (formula (6)). It is seen that it is a good approximation of the original A matrix.

Now, on the descriptors plane (X, Y) , we get the points corresponding to the individual documents $d_i(x_i, y_i)$, where $i=1, 2, \dots, 9$ (Fig. 3). The coordinates of the documents taken from the V_2 sign matrix sign for values $\{x_i, y_i\}$ ($i=1, 2, \dots, 9$).

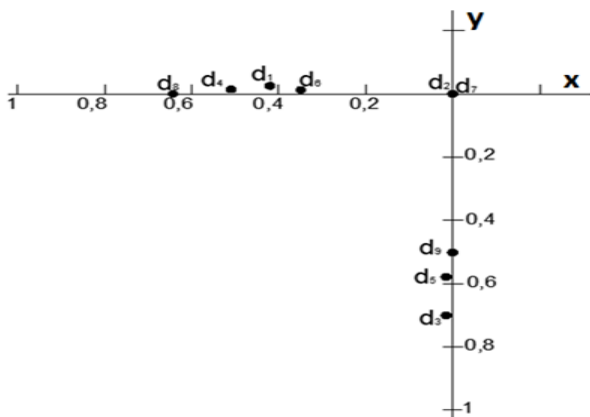


Figure 3 – Placing documents on the descriptors plane (X, Y)

From Fig. 3 it can be seen that the found documents are several groups (in our case 3), which are united by a group of similar keywords from separate descriptors. This made it possible to narrow down the quantity of documents to find specific information on the introduction of innovative elements in the developed software of intelligent DSS for operational dispatching control of urban engineering networks.

6 DISCUSSION

In practice, obviously, the quantity of groups will be much larger (even during the first stage), the phase space will be multidimensional, but the idea of LSA remains the same. In fact, this is a factor analysis that allows to determine the degree of correlation of the found documents with the constructed descriptors, and, consequently, with the technical task.

Given the possibility of using a multi-agent approach to reduce the time to search for the necessary information and increase its semantic value, the process of searching for Internet information in the IRS can be represented by the following scheme, shown in Fig. 4 [22].

As a result, the search strategy for the proposed latent-semantic method of weighted descriptors can be formulated in the form of an algorithm as follows:

Step 1. To take as an “ideal” the text of the terms of reference for the implementation of some innovative project in a specific subject area.

Step 2. To remove superfluous words, which described above, by using “stop-dictionary”.

Step 3. To construct conceptual descriptors with the help of the thesaurus of this subject area and the technical task.

Step 4. To arrange the conceptual descriptors in descending order of their frequency.

Step 5. To determine the frequency, range of the most significant descriptors (usually 10–20 descriptors).

Step 6. To make a request and get a rectangular matrix “descriptors-documents”.

Step 7. According to the formula (5), to arrange the documents in descending order of relevance.

Step 8. To spend on the LSA k -approximation.

Step 9. To put the selected documents in the knowledge model of the subject area.

Steps 6–9 are implemented by intelligent agents automatically and periodically in the search engines allocated to them.

Developed a method of extracting object-oriented information from Internet on based weighted descriptors with the use of Groebner basis. An automated multi-agent procedure for extracting information from the Internet with a semantic analysis of its semantic content developed.

CONCLUSION

One of the key points of the search process is the formation of an effective query and an automated procedure for the analysis of information extracted from the Internet for this subject area.

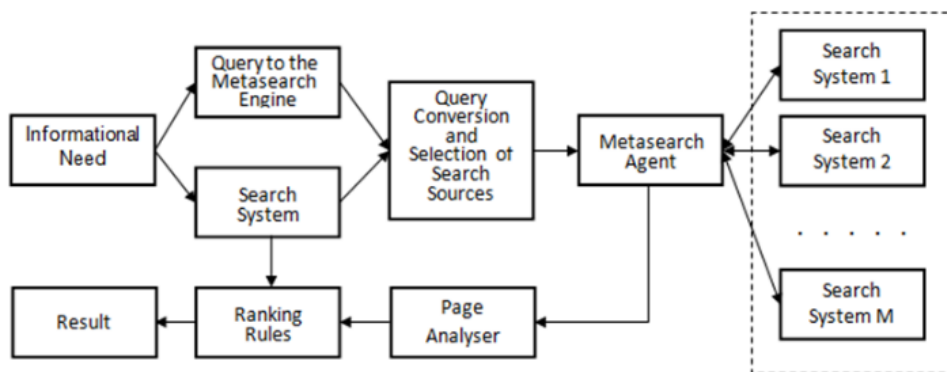


Figure 4 – Multi-agent Search Scheme in Internet Retrieval Systems

The scientific novelty of the results is that for the first time a latent-semantic method of weighted descriptors is proposed, which allows extracting the most significant documents in terms of meaning and meaning, very close to the subject area. The latent-semantic method of weighted descriptors proposed, which allows to extract the most significant in meaning and significance documents that are very close to the subject area. The method based on the idea of Groebner basis, which statistically constructed conceptual descriptors. The choice of the number of descriptors determined by the number of constructed Groebner basis. As the Internet search by the selected of descriptors the initial number of documents turns out to be quite large, used singular value decomposition by k -approximation of latent-semantic analyses.

The practical significance of the obtained results lies in the fact that the use of the latent-semantic method of weighted descriptors proposed in the work can significantly reduce the search time for the necessary information, taking into account its semantic value for this subject area. To select the documents closest to this subject area carried out on the k -approximation of latent semantic analysis.

Prospects for further research proposed by the authors in [23]. They consist in the construction of evolutionary models of knowledge based on the logical evaluation of data obtained from a neural network with neurons having memory and integrated logic. At the same time, the integrated logic is implemented on the basis of a genetic algorithm that processes the updated knowledge model and improves each next generation of “genes” by weighing semantic data based on the superposition of the reference reaction to the situation and the assessment of the situation by the current generation of “genes”. It should note that the information obtained in this way serves as the basis for the correct choice of research methods for the innovative development of organizational and technical systems in this subject area [24].

ACKNOWLEDGEMENT

The work supported by the State Budget Research Project of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute” “Developing specialized knowledge bases for recursive parser of weakly connected originated natural language text information and Web-applications” (a state registration number 0110U002409).

REFERENCES

1. Chu H., Rosenthal M. Search engines for the World Wide Web: A comparative study and evaluation methodology, *Proceedings of the annual meeting-american society for information science: journal*, 2009, Vol. 33, pp. 127–135.
2. Singhal A. Modern Information Retrieval: “A brief Overview”, *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, 2001, Vol. 24, No. 4, pp. 35–43.
3. Gandal N. The dynamics of competition in the internet search engine market, *International journal of industrial organization*, 2001, Vol. 19, pp. 1103–1117. doi:10.1016/S0167-7187(01)00065-0
4. Tarakeswar M. K., Kavitha M. D. Search Engines: A Study // *Journal of Computer Applications (JCA) : journal*, 2011, Vol. 4, No. 1, pp. 29–33.
5. Mikhalev A. I., Stenin A. A., Shitikova I. G., Lemeshko V. A. Intellectual multi-agent system of formation of the subject-oriented evolutionary model of knowledge, *System technologies*, 2018, No. 3 (116), pp. 57–63.
6. Agirre E., Cer D., Diab M. et al. A pilot on semantic textual similarity, *The 6-th International Workshop on Semantic Evaluation, Atlanta, USA*, 2012, pp. 385–393.
7. Bao J., Shen J., Liu X., et al. Semantic Sequence Kin: A Method of Document Copy Detection, *Advances In Knowledge Discovery and Data Mining. Lecture Notes in Artificial Intelligence (LNAI)*. Sydney, Australia, 2004, Vol. 3056, pp. 529–538.
8. Floridi L. Semantic Web, *A Philosophical Assessment, Episteme*, 2009, Vol. 6, No. 1, pp. 25–37.
9. Berners-Lee T., Hendler J., Lassila O. The semantic web, *Scientific American*, 2001, pp. 29–37.
10. Kalchenko D. Intelligent agents of semantic Web, *Computer press-confer*, 2004, No. 10, pp. 26–32.
11. Etzioni O., Weld D., “Intelligent agents on the internet/ O.Etzioni Weld, *Fact, Fiction, and Forecast*”, *IEEE Expert*, No. 4, 1995, pp. 44–49.
12. Wentia Li. Random Texts Exhibit Zipf’s Law, *Like Word Frequency Distribution Santa Fe institute. NM 87501*, 1992, Vol. 38, No. 6, pp. 1842–1845.
13. Kechedzhy K. E., Ustenko O. V., Yampol’ski V. A. Rank distributions of words in additive many-step Markov chains and the Zipf, *Physical review*, 2005, Vol. 72, pp. 1–6.
14. Gerdt V. P. Groebner bases and innovative methods for algebraic and differential equations, *Mathematics and computers in modelling*, 1997, Vol. 25, No. 8/9, pp. 75–90
15. Orlov A.I. Organizational and economic modeling. P.2: Expert estimations. Moscow, Bauman Moscow State Technical University 2011, 486 p.
16. Golub, J. Matrix calculus. Moscow, Mir, 1999, 548 p.
17. Alston S. Hausholder Unitary triangularization of an asymmetric matrix, *Journal of New Technologies in Computational Systems*, 1958, ACM, 5 (4), pp. 339–342. DOI:10.1145/320941.320947
18. Jones K. S. Statistical interpretation of term specificity and its application to search, *Journal. MCB University Documentation*, 2004. Vyp. 60, № 5, pp. 493–502.
19. Matthews D., Curtis D., Fink K. Numerical Methods. Using MATLAB. Numerical Methods: Using MATLAB. 3rd ed. Moscow, Williams Publisher, 2001, 720 p.
20. Charles Henry Edwards Penney, David E. Differential Equations and the Eigenvalue Problem: Modeling and Computation with Mathematica, Maple and MATLAB. 3rd edition. Moscow, Williams Publishing House, 2007, 1104 p.
21. Alexa M., Zuell C. Text Analysis Software: Commonalities, Differences and Limitations, The Results of a Review, *Springer Netherlands*, 2000, Vol. 34 (3), pp. 299–321.
22. Dubinsky A. G. Model of multi-agent information retrieval system in the global network, *Artificial intelligence*, 1999, No. 3, pp. 271–279.
23. Stenin A. A., Pasko V. P., Lemeshko V. A. Neurosemantic approach to building automated information retrieval systems, *Adaptive automatic control systems*, 2019, No. 1(34), pp. 125–130.

УДК 004.91

МУЛЬТИАГЕНТНА ЛАТЕНТНО-СЕМАНТИЧНА ІНТЕРНЕТ-ТЕХНОЛОГІЯ ФОРМУВАННЯ ПРЕДМЕТНО-ОРІЄНТОВАНОЇ МОДЕЛІ ЗНАНЬ

Стенін О. А. – д-р техн. наук, професор кафедри технічної кібернетики Київського політехнічного інституту. Ігор Сікорський, Київ, Україна.

Пасько В. П. – канд. техн. наук, доцент кафедри технічної кібернетики Київського політехнічного інституту. Ігор Сікорський, Київ, Україна.

Солдатова М. А. – канд. техн. наук, старша викладачка кафедри технічної кібернетики Київського політехнічного інституту. Ігор Сікорський, Київ, Україна.

Дроздович І. Г. – канд. техн. наук, старший науковий співробітник Інституту телекомунікацій і глобального інформаційного простору НАН України.

АНОТАЦІЯ

Актуальність. У статті пропонується латентно-семантична технологія вилучення інформації з інтернет-ресурсів, що дозволяє обробляти інформацію природною мовою, а також заснований на ній мультиагентний алгоритм пошуку. Актуальність даного підходу до пошуку предметно-орієнтованої інформації визначається тим, що в даний час пряме лексичне порівняння запитів з індексами документів не повною мірою задовольняє розробника. Об'єкт дослідження – мультиагентний латентно-семантичний алгоритм пошуку предметно-орієнтованої інформації. Мета роботи – підвищення ефективності формування адекватної даної предметної області моделі знань.

Метод. Запропонована латентно-семантична технологія, заснована на розробленому авторами методі зважених дескрипторів. Основна відмінність від існуючих методів полягає в тому, що аналіз слів, що зустрічаються в тексті як по частотності, так і з урахуванням семантики, здійснюється шляхом підбору відповідних дескрипторів, що підвищує якість знайденої інформації.

Результати. Розроблена латентно-семантична технологія пошуку інформації апробована в задачі побудови моделі знань автоматизованої СППР для оперативно-диспетчерського управління міськими інженерними мережами (ГІС). Проведене моделювання пошуку предметно-орієнтованої інформації даної предметної області показало ефективність розробленого підходу.

Висновки. Підвищення ефективності пошуку і семантичного наповнення предметно-орієнтованої інформації моделі знань даної предметної області досягається за рахунок використання в даній технології методу зважених дескрипторів, заснованого на законах Зіпфа. Перспективи подальших досліджень полягають у побудові еволюційних моделей знань і підвищення якості оновлюваної інформації.

КЛЮЧОВІ СЛОВА: інтернет-ресурси, інформаційний пошук, закони Зіпфа, базиси Гребнера, інтелектуальні агенти, зважені дескриптори, латентно-семантичний аналіз, мультиагентна процедура автоматичного пошуку.

УДК 004.91

МУЛЬТИАГЕНТНЫЕ ЛАТЕНТНОГО СЕМАНТИЧЕСКОГО ИНТЕРНЕТ-ТЕХНОЛОГИЯ ФОРМИРОВАНИЯ ПРЕДМЕТНО-ОРИЕНТИРОВАННАЯ МОДЕЛЬ ЗНАНИЙ

Стенін А. А. – д-р техн. наук, профессор кафедры технической кибернетики Киевского политехнического института. Игорь Сикорский, Киев, Украина.

Пасько В. П. – канд. техн. наук, доцент кафедры технической кибернетики Киевского политехнического института. Игорь Сикорский, Киев, Украина.

Солдатова М. А. – канд. техн. наук, старший преподаватель кафедры технической кибернетики Киевского политехнического института. Игорь Сикорский, Киев, Украина.

Дроздович И. Г. – канд. техн. наук, старший научный сотрудник Института телекоммуникаций и глобального информационного пространства НАН Украины.

АННОТАЦИЯ

Актуальность. В статье предлагается латентно-семантическая технология извлечения информации из интернет-ресурсов, позволяет обрабатывать информацию на естественном языке, а также основанный на ней Мультиагентный алгоритм поиска. Актуальность данного подхода к поиску предметно-ориентированной информации определяется тем, что в настоящее время прямое лексическое сравнение запросов с индексами документов не в полной мере удовлетворяет разработчика. Объект исследования – мультиагентный латентно-семантический алгоритм поиска предметно-ориентированной информации. Цель работы – повышение эффективности формирования адекватной данной предметной области модели знаний.

Метод. Предложенная латентно-семантическая технология, основанная на разработанном авторами методе взвешенных дескрипторов. Основное отличие от существующих методов состоит в том, что анализ слов, встречающихся в тексте как по

частотности, так и с учетом семантики, осуществляется путем подбора соответствующих дескрипторов, повышает качество найденной информации.

Результаты. Разработанная латентно-семантическая технология поиска информации апробирована в задаче построения модели знаний автоматизированной СППР для оперативно-диспетчерского управления городскими инженерными сетями (ГИС). Проведенное моделирование поиска предметно-ориентированной информации данной предметной области показало эффективность разработанного подхода.

Выводы. Повышение эффективности поиска и семантического наполнения предметно-ориентированной информации модели знаний данной предметной области достигается за счет использования в данной технологии метода взвешенных дескрипторов, основанного на законах Zipf'a. Перспективы дальнейших исследований заключаются в построении эволюционных моделей знаний и повышения качества обновляемой информации.

КЛЮЧЕВЫЕ СЛОВА: интернет-ресурсы, информационный поиск, законы Zipf'a, базы Гребнера, интеллектуальные агенты, взвешенные дескрипторы, латентно-семантический анализ, мультиагентная процедура автоматического поиска.

ЛІТЕРАТУРА / LITERATURE

1. Чу Х. Поисквые системы для Всемирной паутины: сравнительное исследование и методология оценки / Х. Чу, М. Розенталь // Труды ежегодного совещания. Американское общество информационных наук: науч.-техн. сб. – 2009. – Том 33. – С. 127–135
2. Сингхал Амит Современный информационный поиск: «Краткий обзор» / Сингхал Амит // Бюллетень Технического комитета IEEE. Компьютерное общество по инженерии данных. – 2001. – Том 24, № 4. – С. 35–43.
3. Гэндал Н. Динамика конкуренции на рынке поисковых систем Интернета / Н. Гэндал // Международный журнал промышленной организации. – 2001. – Том 19. – С. 1103–1117. DOI: 10.1016/S0167-7187(01)00065-0
4. Tarakeswar M. K. Поисквые системы: Исследование / М. К. Tarakeswar, М. D. Kavitha // Журнал компьютерных приложений. – 2011. – Том 4, № 1. – С. 29–33.
5. Интеллектуальная мультиагентная система формирования предметно-ориентированной эволюционной модели знаний / [А. И. Михалев, А. А. Стенин, И. Г. Шитикова, В. А. Лемешко] // Системные технологии: науч.-техн. сб. – 2018. – Вып. 3(116). – С. 57–63.
6. Пилотный проект по семантическому текстовому сходству / [Э. Агирре, Д. Сер, М. Диаб и др.] // Семантические оценки: 6-й Международный семинар, Атланта, США, 2012. – С. 385–393
7. Семантическая последовательность Kin: Метод обнаружения копий документов / [Дж. Бао, Д. Шен, Х. Лью, Х. Ли, и др.] // Достижения в области обнаружения знаний и интеллектуального анализа данных: Лекционные заметки по искусственному интеллекту. – Сидней, Австралия, 2004. – Том 3056. – С. 529–538.
8. Флориди Л. Веб 2.0 против Семантической сети / Л. Флориди // Философская оценка. – 2009. – Вып. 6, № 1. – С. 25–37.
9. Бернерс-Ли Т. Семантическая паутина / Т. Бернерс-Ли, Д. Хендлер, О. Лассила // Американская Наука. – 2001. – С. 29–37.
10. Кальченко Д. Интеллектуальные агенты семантического Web'a / Д. Кальченко // Компьютер Пресс. – 2004. – Вып. 10. – С. 26–32.
11. Эциони О. Интеллектуальные агенты в Интернете / О. Эциони, Д. Велд // Факты, вымысел и прогноз, IEEE Эксперт. – 1995. – № 4. – С. 44–49.
12. Wentain Li. Random Texts Exhibition Zipf's Law / Li. Wentain // Like Word Frequency Distribution. Santa Fe institute. NM 87501. – 1992. – Vol. 38, №6. – P. 1842–1845.
13. Кечеджи К. Е. Ранговые распределения слов в аддитивной форме. Многошаговые цепи Маркова и закон Ципфа / К. Е. Кечеджи, О. В. Устенко, В. А. Ямпольский // Физическое обозрение. – 2005. – Вып. 72, – С. 1–6.
14. Гердт В. П. Основы Гребнера и инвойтивные методы для алгебраических и дифференциальных уравнений / В. П. Гердт // Математика и компьютеры в моделировании – 1997. – Вып 25, № 8/9. – С 75–90.
15. Орлов А. И. Организационно-экономическое моделирование / А. И. Орлов. – Ч. 2: Экспертные оценки. – М. : МГТУ им. Н. Э. Баумана, 2011. – 486 с.
16. Голуб Дж. Матричные исчисления / Дж. Голуб. – М. : Мир, 1999. – 548 с.
17. Олстон С. Унитарная триангуляризация несимметричной матрицы / Олстон С. Хаусхолдер // Журнал о новых технологиях в вычислительных системах – 1958. – АСМ, 5 (4). – С. 339–342. DOI: 10.1145/320941.320947
18. Джонс К. С. Статистическая интерпретация специфичности термина и ее применение в поиске / К. С. Джонс // Журнал. Документация «МСВ University. 2004. – Вып. 60, № 5. – С. 493–502.
19. Мэтьюз Д. Численные методы. Использование MATLAB / Д. Мэтьюз, Д. Куртис, К. Финк. – Numerical Methods: Using MATLAB. – 3-е изд. – М. : Изд-во «Вильямс», 2001. – 720 с.
20. Чарльз Генри Эдвардс Пенни. Дифференциальные уравнения и проблема собственных значений: моделирование и вычисление с помощью Mathematica / Чарльз Генри Эдвардс, Дэвид Э. – Maple и MATLAB. 3-е издание. – М. : Изд-во «Вильямс», 2007. – 1104 с.
21. Алекса М. Программное обеспечение для анализа текста: общие черты, различия и ограничения: Результаты обзора / М. Алекса, С. Zuell // Springer Netherlands. – 2000. – Вып. 34 (3). – С. 299–321.
22. Дубинский А. Г. Модель многоагентной информационно-поисковой системы в глобальной сети / А. Г. Дубинский // Искусственный интеллект. – 1999. – № 3. – С. 271–279.
23. Стенин А. А. Нейросемантический подход к построению автоматизированных информационно-поисковых систем / А. А. Стенин, В. П. Пасько, В. А. Лемешко // Адаптивные системы автоматического управления. – 2019. – №1 (34). – С. 125–130.
24. Спеціальні методи наукових досліджень / П. О. Киричок, С. В. Струтинський, В. Г. Олійник ; Нац. техн. ун-т України «Київ. політехн. ін-т». – Київ : АртЕк, 2016. – 592 с. ISBN 978-617-7264-28-5

УПРАВЛІННЯ У ТЕХНІЧНИХ СИСТЕМАХ

CONTROL IN TECHNICAL SYSTEMS

УПРАВЛЕНИЕ В ТЕХНИЧЕСКИХ СИСТЕМАХ

УДК 519.85

ОПТИМІЗАЦІЯ РОЗКЛАДУ НАВЧАННЯ В УНІВЕРСИТЕТІ

Косолап А. І. – д-р фіз.-мат. наук, професор, завідувач кафедри спеціалізованих комп'ютерних систем Українського державного хіміко-технологічного університету, Дніпро, Україна.

Дубовик Т. М. – старший викладач кафедри спеціалізованих комп'ютерних систем Українського державного хіміко-технологічного університету, Дніпро, Україна.

АНОТАЦІЯ

Актуальність. В роботі розглядається відома задача складання розкладу навчання в університеті. Такі задачі розв'язуються декілька разів на рік в кожному навчальному закладі. Не дивлячись на багаточисленні дослідження в даній галузі, проблема побудови оптимального розкладу залишається відкритою. Це пов'язано зі складністю відповідної оптимізаційної задачі, зокрема її значною розмірністю, що затрудняє чисельне розв'язування такої задачі існуючими методами оптимізації. Вдосконалення потребують також оптимізаційні моделі складання розкладів. Таким чином, оптимізація розкладу є складною обчислювальною проблемою і потребує розробки нових методів її розв'язування.

Мета роботи. Вдосконалення оптимізаційних моделей складання розкладів навчання в університеті та використання нових ефективних методів для їх розв'язування.

Метод. Ми використовуємо метод точної квадратичної регуляризації для розв'язування оптимізаційних задач складання університетського розкладу навчання. Точна квадратична регуляризація дозволяє перетворити складні оптимізаційні моделі з булевими змінними до задачі максимуму норми вектору на опуклій множині. Для розв'язування цієї задачі ми використовуємо ефективний прямо-двоїтий метод внутрішньої точки та метод дихотомії. Цей метод показав значно кращі результати при розв'язуванні багатьох складних мультимодальних задач. Це підтверджується багатьма порівняльними обчислювальними експериментами. Ще більшу ефективність метод точної квадратичної регуляризації демонструє при розв'язуванні задач складання розкладу навчання. Цей метод оптимізації використовується вперше для даного класу задач, тому він потребував розробки відповідного алгоритмічного забезпечення.

Результати. Побудована нова більш проста оптимізаційна модель складання розкладу, яка легко реалізується програмно в пакеті Excel при наявності надбудов OpenSolver, RiskSolver та інших. Приведений невеликий приклад побудови розкладу та описана пошарова інструкція отримання оптимального розв'язку.

Висновки. Розроблена нова ефективна технологія складання розкладу навчання в університеті, яка виділяється простою реалізацією і не потребує розробки спеціального програмного забезпечення. Ефективність забезпечується використанням нового методу точної квадратичної регуляризації.

КЛЮЧОВІ СЛОВА: розклад навчання, оптимізація, задачі з булевими змінними, метод точної квадратичної регуляризації.

АБРЕВІАТУРИ

УДХТУ – Український державний хіміко-технологічний університет;

Excel – офісний пакет Microsoft;

OpenSolver – надбудова Excel;

RiskSolver – ще одна надбудова Excel;

EQR – точна квадратична регуляризація;

V_i – i -й викладач;

P_i – i -й предмет.

НОМЕНКЛАТУРА

a_{kj} – лекційне навантаження в годинах по k -му курсу та j -му предмету;

b_{kj} – лабораторне навантаження в годинах по k -му курсу та j -му предмету;

S_k – множина предметів k -го курсу;

V_k – множини індексів занять k -го викладача;

R_k – тижнева кількість годин на k -му курсі;

A_k – множина лекцій k -го курсу;

B_k – множина лабораторних занять k -го курсу;

r – кількість комп'ютерних класів;

K – кількість викладачів;

x_{ij} – булева змінна, яка дорівнює одиниці, якщо j -те заняття проводиться на i -й парі;

z – допоміжна змінна перетворення EQR;

s – параметр точної квадратичної регуляризації;

d – змінна методу EQR;
 $\|x\|^2$ – квадрат Евклідової норма вектору x ;
 I_1 – множина навчальних пар по чисельнику (початок інтервалу сумування);
 J_1 – множина навчальних пар по чисельнику (кінець інтервалу сумування);
 I_2 – множина навчальних пар по знаменнику (початок інтервалу сумування);
 J_2 – множина навчальних пар по знаменнику (кінець інтервалу сумування);
 i – змінна індексів сум, параметрів та змінних;
 j – змінна індексів сум, параметрів та змінних;
 k – змінна індексів сум, параметрів та змінних;
 l – змінна індексів сум, параметрів та змінних.

ВСТУП

Проблема складання розкладу занять виникає 4–8 разів на рік в кожному університеті. Це досить складна комбінаторна проблема, розв’язок якої потребує досить багато часу. Тому цій проблемі присвячено багато публікацій. Немає єдиного розуміння який розклад вважати кращим. Як правило, задача складання розкладу зводиться до оптимізаційної комбінаторної задачі великої розмірності з булевими змінними та з безліччю обмежень. Така задача є складною обчислювальною проблемою. Тому існуючі програмні пакети складання розкладу не знаходять оптимальні розв’язки, а тільки допомагають диспетчеру знайти допустимий розв’язок. Деякі програмні пакети містять розв’язування оптимізаційних задач. Для розв’язування таких задач частіше використовуються методи розгалужень та границь, які потребують досить багато часу, або генетичні та еволюційні алгоритми, які теж потребують багато часу і не гарантують отримання оптимальних розв’язків, а іноді і допустимих розв’язків. Використовують також евристичні алгоритми, які часто знаходять розв’язки далекі від оптимальних. Тому проблема побудови оптимальних розкладів навчання залишається відкритою [1]. В даній роботі пропонується нова проста оптимізаційна модель складання розкладу навчання та вперше для даного класу задач використовується ефективний метод точної квадратичної регуляризації для її розв’язування [2], який потребував розробки нового алгоритмічного забезпечення, враховуючи специфіку даної задачі.

Об’єктом дослідження даної роботи є процеси побудови оптимальних розкладів занять та їх математичне моделювання.

Предметом дослідження є оптимізаційні комбінаторні задачі великої розмірності з булевими змінними, до яких належать задачі складання розкладу навчання.

Метою даного дослідження є вдосконалення оптимізаційної моделі для складання розкладу навчання та розробка простого та ефективного алгоритму знаходження оптимального розкладу.

1 ПОСТАНОВКА ЗАДАЧІ

Відомими даними для задачі складання розкладу є S_k – множина предметів k -го курсу; K – кількість викладачів, що забезпечують навчальний процес, їх навчальне навантаження a_{kj} в годинах на тиждень по лекціям, лабораторним та практичним заняттям b_{ij} ; кількість можливих пар занять кожного дня тижня. Задача складання розкладу полягає в визначенні для кожного викладача змінної x_{ij} , яка дорівнює одиниці якщо i -те заняття проводиться на j -й парі тижня інакше ця змінна дорівнює нулю. Змінні x_{ij} утворюють матрицю, строчкам якої відповідають пари занять, а стовпцям – предмети. Оптимальним розкладом вважаємо такий, для якого дисперсія кількості занять студентів кожного дня тижня є мінімальною. Це означає, що сума квадратів сум x_{ij} для кожного дня тижня кожної групи буде мінімальною. Мінімум потрібно знайти при виконанні наступних обмежень. Студенти однієї групи можуть бути одночасно присутні тільки на одному занятті (суми x_{ij} по строках не більше одиниці), кожен викладач одночасно проводить тільки одне заняття (суми x_{ij} по строках по кожному викладачеві не більше одиниці), існує обмеження по кількості комп’ютерних класів для проведення лабораторних робіт, по кількості пар занять кожного дня, у студентських заняттях не повинно бути вікон, заняття проводяться згідно з навчальному плану. Дана постановка задачі складання розкладу відрізняється від існуючих вибором критерія оптимальності та обмеженнями на відсутність вікон в заняттях студентів.

Для того, щоб скласти розклад занять на тиждень тільки по одній спеціальності необхідно заповнити таблицю нулями та одиницями, де елементи таблиці відповідають парам занять. В цій таблиці буде 50 рядків (кількість можливих пар занять протягом тижня), а по стовпцям будуть предмети для кожного курсу (лекції, практичні та лабораторні роботи). В реальному розкладі така таблиця для кожної групи об’єднується в один стовпчик і тоді на місце одиниці записують назву предмету та викладача. Поряд з назвами предметів в шапці таблиці необхідно вказати прізвище викладача. Число рядків таблиці дорівнює числу можливих занять протягом тижня (кожен день можна проводити заняття з першої по п’яту пару) з урахуванням чисельника та знаменника. Таким чином, якщо на кожному курсі (всього п’ять курсів) навчається тільки одна група і кількість предметів в семестрі, наприклад, дорівнює 6, то отримуємо таблицю розміром 50×60 (по кожному предмету маємо лекцію та практичне або лабораторне заняття). Тоді в розглянутій задачі кількість невідомих булевих змінних буде дорівнювати 3000. Це досить складна задача для сучасних методів оптимізації.

Для побудови оптимізаційної моделі даної задачі будемо враховувати наступне. Кожний день у групи студентів повинно бути не менше 2-х та не більше 4-х занять (пар). Як правило, заняття студентів можуть проводитись як по чисельнику так і по знаменнику.

Будемо допускати, що кожен день заняття можуть проводитися з першої по п'яту пару. Якщо заняття проводяться по чисельнику та знаменнику, то можлива кількість пар кожен день дорівнює 10. При дефіциті аудиторій кількість пар кожного дня може бути більшою. Крім того, кількість занять на тиждень, як правило, відоме та для кожного курсу визначається навчальним планом. В даний момент часу студентська група може бути присутня тільки на одному занятті, а викладач теж може проводити заняття тільки по одному предмету. Лабораторні роботи, як правило, проводяться в комп'ютерних класах кількість яких обмежена, тому одночасно таких занять може бути не більше ніж комп'ютерних класів. Крім того, в розкладі не повинно бути вікон для студентів. З двох розкладів кращим буде такий в якому дисперсія кількості занять кожного дня протягом тижня буде мінімальною. Такий розклад створює найкращі умови для якісного навчання студентів. Можна допустити обмеження на відсутність вікон і для викладачів, але значне збільшення обмежень задачі може привести до того, що допустима множина задачі стане пустою.

2 ОГЛЯД ЛІТЕРАТУРИ

Перші публікації присвячені складанню розкладів занять з'явилися ще в 60-ті роки минулого століття. Але в ті роки значних успіхів в розв'язанні цієї проблеми досягнуто не було. Незважаючи на відсутність успіхів та враховуючи актуальність проблеми, дослідження пошуку ефективних методів її розв'язування інтенсивно продовжуються. З'являються перші оглядові роботи присвячені розкладу навчання в університетах [1, 3, 4]. В цих роботах приводиться багато оптимізаційних моделей та методів і робиться висновок, що потрібно провести ще багато досліджень щоб розв'язати дану проблему. В наш час також з'являються оглядові публікації присвячені даній проблемі [5]. Зокрема огляд моделей складання розкладу розглянутих в роботі [6], а в роботах [7–9] приведений огляд сучасних методів розв'язування даної проблеми. В цих публікаціях стверджується, що на сьогодні проблема складання розкладу навчання для університету як і раніше залишається відкритою. Розробляється також програмне забезпечення, яке допомагає диспетчеру складати розклад. Але оптимізація такого розкладу потребує розробки нових методів. Як стверджується в роботі [5] сучасне програмне забезпечення орієнтовано в першу чергу на графічний інтерфейс, а не на ефективні алгоритми оптимізації.

Детальний аналіз існуючих технологій складання розкладів розглянуто в роботах [8, 9]. Всі ці дослідження свідчать про те, що в даний час не існує ефективних методів оптимізації складання розкладу в університетах.

3 МАТЕРІАЛИ І МЕТОДИ

Побудуємо оптимізаційну модель задачі складання розкладу навчання в університеті. З навчального плану відомо a_{kj} – кількість годин лекцій j -го предмету

кожного тижня, b_{kj} – практичних чи лабораторних робіт на k -му курсі протягом тижня. Відома також кількість викладачів, що забезпечують навчальний процес, а також V_k – множини індексів занять k -го викладача, S_k – множина занять k -го курсу (лекції та лабораторні чи практичні заняття). Введемо змінні моделі x_{ij} , які дорівнюють одиниці якщо i -те заняття проводиться на j -й парі тижня, інакше ця змінна дорівнює нулю.

В якості цільової функції візьмемо наступну

$$\min \left\{ \sum_{k=1}^5 \sum_{l=1}^5 \left(\sum_{i=10(l-1)+1}^{10l} \sum_{j \in S_k} x_{ij} \right)^2 \right\}. \quad (1)$$

Це сума часів занять студентів кожного дня в квадраті для кожного курсу і вона повинна бути мінімальною. Якщо ця сума буде мінімальною, то дисперсія кількості пар занять кожного тижня теж буде мінімальною. Далі введемо наступні обмеження.

1. Тижнева кількість навчань студентів дорівнює

$$\sum_{i=1}^{50} \sum_{j \in S_k} x_{ij} = R_k, k = 1, \dots, 5. \quad (2)$$

2. Кількість занять на кожний день по кожному курсу обмежена

$$2 \leq \sum_{i=10(l-1)+1}^{10l} \sum_{j \in S_k} x_{ij} \leq 4, l = 1, \dots, 5, k = 1, \dots, 5, \quad (3)$$
$$2 \leq \sum_{i=10(l-1)+2}^{10l} \sum_{j \in S_k} x_{ij} \leq 4, l = 1, \dots, 5, k = 1, \dots, 5.$$

що означає мінімум дві пари та максимум чотири пари на день. Так як заняття проводяться по чисельникам та знаменникам, то в першому обмеженні індекс i приймає тільки непарні значення, а в другій нерівності – тільки парні значення. При збільшенні кількості обмежень задачі умови (3) стають суттєвими.

3. Кожне заняття проводиться

$$\sum_{i=1}^{50} x_{ij} = a_{kj}, j \in A_k, \sum_{i=1}^{50} x_{ij} = b_{kj}, j \in B_k, k = 1, \dots, 5. \quad (4)$$

4. Кількість одночасного проведення лабораторних робіт не перевищує кількості комп'ютерних класів

$$\sum_{j \in B} x_{ij} \leq r, i = 1, \dots, 50, B = \bigcup_{k=1}^5 B_k. \quad (5)$$

5. Одна група може бути присутня одночасно тільки на одному занятті

$$\sum_{j \in S_k} x_{ij} \leq 1, i = 1, \dots, 50, k = 1, \dots, 5. \quad (6)$$

6. Один викладач теж може проводити одночасно не більше одного заняття

$$\sum_{j \in V_k} x_{ij} \leq 1, i = 1, \dots, 50, k = 1, \dots, K. \quad (7)$$

7. В розкладі не повинно бути студентських вікон. Це рівносильно наступним обмеженням

$$\begin{aligned} & x_{10(l-1)+i} + x_{10(l-1)+j} - 2 \sum_{k=10(l-1)+i+1}^{10(l-1)+j-1} x_{10(l-1)+k} \leq 1, \\ & i \in I_1 = \{1, 3, 5, 1, 3, 1\}, j \in J_1 = \{5, 7, 9, 7, 9\}, l = 1, 2, 3, 4, 5, \\ & x_{10(l-1)+i} + x_{10(l-1)+j} - 2 \sum_{k=10(l-1)+i+1}^{10(l-1)+j-1} x_{10(l-1)+k} \leq 1, \\ & i \in I_2 = \{2, 4, 6, 2, 4, 2\}, j \in J_2 = \{6, 8, 10, 8, 10, 10\}, l = 1, 2, 3, 4, 5, \end{aligned} \quad (8)$$

де індекси i, j приймають тільки однакові місця в заданих множинах I_1, J_1 та I_2, J_2 .

Перше обмеження (8) виключає вікна по чисельнику, а друге – по знаменнику. Як правило, обмеження (8) в існуючих моделях розкладу не враховується. Розглянемо обмеження (8) більш детально. Якщо в розкладі є вікна, то в ньому будуть наступні послідовності нулів та одиниць 101, 1001, 10001 одного дня по чисельнику або знаменнику. Для цих послідовностей сума крайніх одиниць мінус подвоєна сума не крайніх нулів буде більше одиниці, тобто обмеження (8) буде порушуватись. Для всіх інших послідовностей відповідна сума буде не більше одиниці. Наприклад, для послідовностей 100, 110, 111, 1101, 11011 відповідно отримуємо $1 \leq 1, -1 \leq 1, 0 \leq 1, 0 \leq 1, 0 \leq 1$. В обмеженнях (8) враховуються послідовності з трьох, чотирьох та п'яти чисел підряд для кожного дня по чисельнику та знаменнику. В попередньому прикладі ми спостерігаємо вікна для двох останніх послідовностей, які містять послідовності 101, тому для них обмеження (8) будуть порушені і послідовності 1101, 11011 будуть недопустимими.

Якщо допустити, що на кожному курсі навчається декілька груп, тоді лекції часто проводяться одночасно для всіх груп. Ця умова задається простими обмеженнями $x_{ij} = x_{ik}, i = 1, \dots, 50$ (j -та та k -та лекція проводиться одночасно). Можна додати також обмеження на аудиторний фонд, але для багатьох університетів дефіцит аудиторного фонду не є критичним. Інколи j -й викладач на i -й парі не може бути присутнім (наприклад, працює в цей час в другому місці), тоді відповідна змінна $x_{ij} = 0$.

Запропонована модель відноситься до класу квадратичних з лінійними обмеженнями та булевими змінними. Всі обмеження є лінійними, а цільова функція квадратична. Розглянуту модель ми можемо спростити. Так, обмеження (2) визначається навчальним планом, тож його можна опустити. Всі інші обмеження є суттєві.

Отримана модель має велику розмірність не тільки по змінним але й по обмеженням. Якщо припустити, що кількість викладачів дорівнює 15, на кожному курсі навчається тільки одна група і кількість занять на кожному курсі (в семестрі) дорівнює 10, то отримуємо задачу з 2500 булевими змінними та 1400 обмеженнями. Розв'язати таку задачу методом розгалужень та границь неможливо. Якщо булеві змінні замінити неперервними на відрізьку $[0, 1]$, то задача (1)–(8) буде розв'язана прямо-двоїтим методом внутрішньої точки за 1–2 хвилини. Але отриманий розв'язок буде недопустимий.

Умову булевості змінних в задачі (1)–(8) можна замінити обмеженнями

$$\sum_{i=1}^{50} \sum_{k=1}^5 \sum_{j \in S_k} x_{ij} (1 - x_{ij}) \leq 0, 0 \leq x \leq 1, \quad (9)$$

але ці обмеження є неопуклими, а задача (1)–(9) стає мультимодальною. Для її розв'язування можна використати генетичні чи еволюційні алгоритми. Але ці алгоритми не гарантують отримання навіть допустимого розв'язку задачі (1)–(9). Єдиною альтернативою існуючим алгоритмам є метод точної квадратичної регуляризації (EQR) [2]. В цьому методі квадратична регуляризація використовується для перетворення задачі (1)–(9) до наступної

$$\max \{ \|x\|^2 \mid \sum_{k=1}^5 \sum_{l=1}^5 \left(\sum_{i=10(l-1)+1}^{10l} \sum_{j \in S_k} x_{ij} \right)^2 + s + \|x\|^2 \leq d \}, \quad (10)$$

(формула (10) містить додаткове обмеження) при обмеженнях (3)–(8), а обмеження (9) замінюємо наступним

$$2 \|x\|^2 - \sum_{i=1}^{50} \sum_{k=1}^5 \sum_{j \in S_k} x_{ij}^2 + \sum_{k=1}^5 \sum_{j \in A_k} a_{kj} + \sum_{k=1}^5 \sum_{j \in B_k} b_{kj} \leq d. \quad (11)$$

Таким чином, мультимодальна задача (1)–(9) звелась до максимізації норми вектору на опуклій множині. Обмеження (10) містить скалярний параметр s , який повинен задовольняти умові

$$s \geq \sum_{k=1}^5 \sum_{j \in A_k} a_{kj} + \sum_{k=1}^5 \sum_{j \in B_k} b_{kj} - \sum_{k=1}^5 \left(\frac{R_k}{5} \right), \quad (12)$$

при якій обмеження (10) буде активним, а квадрат норми вектору містить ще одну неперервну змінну z

$$\|x\|^2 = \sum_{i=1}^{50} \sum_{k=1}^5 \sum_{j \in S_k} x_{ij}^2 + z^2.$$

Таким чином, задача (1)–(9) точною квадратичною регуляризацією перетворена до еквівалентної задачі (3)–(8), (10)–(11). При такому перетворенні розмірність задачі зростає всього на одиницю. Далі будемо розв'язувати перетворену задачу розв'язком якої співпадає з розв'язком задачі (1)–(9).

В задачі (3)–(8), (10)–(11) необхідно знайти мінімальне значення скалярної величини d , для якої виконується умова $2\|x\|^2 - d = 0$. При фіксованих значеннях d задача (3)–(8), (10)–(11) розв'язується прямо-двоїстим методом внутрішньої точки [10]. Значення d будемо знаходити методом дихотомії. Це приблизно 20–30 розв'язувань даної задачі при фіксованих значеннях d . Таким чином, перетворена задача (3)–(8), (10)–(11) може бути розв'язана програмою OpenSolver за 30 хвилин для практичної задачі складання розкладу.

Ми пропонуємо наступний алгоритм для розв'язування задачі (3)–(8), (10)–(11).

Крок 1. Обираємо початкові дані для змінних $x_{ij} = 0$, $\forall i, j$, $z = 1$ та визначаємо параметр s за формулою (12).

Крок 2. Розв'язуємо задачу опуклої оптимізації

$$\min \{d \mid \sum_{k=1}^5 \sum_{l=1}^5 \left(\sum_{i=10(l-1)+1}^{10l} \sum_{j \in S_k} x_{ij} \right)^2 + s + \|x\|^2 \leq d\},$$

з обмеженнями (3)–(8), (11) програмою OpenSolver, що реалізує прямо-двоїстий алгоритм внутрішньої точки та знаходимо мінімально можливе значення змінної d .

Крок 3. Послідовно збільшуємо значення змінної d на величину, що пропорційна значенню $2\|x\|^2 - d$ по абсолютній величині. Для кожного такого фіксованого значення d розв'язуємо задачу (3–8), (10–11) програмою OpenSolver. При збільшенні d значення величини $2\|x\|^2 - d$ буде спадати до нуля по абсолютній величині.

Крок 4. Якщо $2\|x\|^2 - d = 0$ з заданою точністю, то задача складання розкладу розв'язана. Інакше переходимо до кроку 3.

4 ЕКСПЕРИМЕНТИ

Побудувати розклад занять для однієї спеціальності по запропонованій моделі досить легко, використовуючи тільки пакет Excel та надбудову OpenSolver для цього пакету. На лист Excel в один рядок вводимо назви предметів та прізвища їх викладачів з пропуском двох стовпців між групами (курсами). По кожній групі спочатку вводяться лекції, потім лабораторні чи практичні заняття. В рядок нижче вводиться годинне навантаження на тиждень по кожному предмету. Нижче для кожної групи виділяємо 50 рядків, частина з яких буде заповнена одиницями при оптимізації даної моделі. Тож спочатку ці рядки заповнимо нулями.

Це вся інформація для розрахунку розкладу. Далі необхідно ввести формули для цільової функції та обмежень задачі. В першій стовпець між групами вводимо суму по рядкам кожної групи. Очевидно, що ця сума не повинна бути більше одиниці (група може бути присутня одночасно тільки на одному занятті). Нижче в цьому стовпці вводимо формули сум цього стовпця по кожному дню тижня. Ці значення не повинні бути менше 2-х і більше 4-х одиниць для чисельника та знаменника. Ці формули використовуємо та-

кож для введення формули цільової функції. Вона буде дорівнювати сумі квадратів значень, обчислених цими формулами. В другий стовпець між групами вводимо формули, що виключають вікна між заняттями в групі. В рядок нижче вхідних даних вносимо формули сум по стовпцям частини таблиці, що заповнена нулями (з понеділка по п'ятницю), вони повинні дорівнювати навантаженню в годинах по кожному предмету. Справа даної таблиці вносимо формули сум рядків по кожному викладачеві по всім групам. Кожна така сума не повинна бути більше одиниці (викладач одночасно може бути присутнім не більше ніж на одній парі). На кінець, в останній стовпець вводимо суми стовпців лабораторних робіт по всім групам. Сума чисел кожного рядка не повинна бути більше кількості комп'ютерних класів. Далі викликаємо діалогове вікно надбудови Solver куди вводимо адресу комірок цільової функції, комірок, що змінюються та адреси комірок з формулами для обмежень задачі. Можна також вказати, що змінні булеві. Але розв'язати цю задачу програмою Solver ми не зможемо, так як вона дозволяє розв'язувати задачі з кількістю змінних не більше 200. Тому викликаємо надбудову OpenSolver, яка може розв'язати цю задачу, але для цього може знадобиться більше однієї доби. Ми для розв'язування задачі складання розкладу будемо використовувати метод точної квадратичної регуляризації, для якого був розроблений приведенний вище алгоритм з урахуванням специфіки даної задачі. Для цього необхідно додати декілька формул, зокрема додати комірку для неперервної змінної z , змінної d , ввести формули (10)–(11). Адреси комірок цих змінних та формул обмежень додаємо в діалогове вікно програми Solver та видаляємо обмеження булевих змінних. Спочатку будемо розв'язувати задачу мінімізації змінної d . Це досить проста опукла задача, яку програма OpenSolver розв'язує за 1 хвилину. Після отримання мінімального значення d перевіряємо умову $2\|x\|^2 - d = 0$ (формулу для цього виразу теж необхідно занести в комірку). Звичайно буде виконуватися нерівність $2\|x\|^2 - d < 0$, тому в діалогове вікно програми Solver занесемо наступні зміни. Цільову комірку замінимо на комірку з формулою $\|x\|^2$, яку будемо максимізувати, а з адрес комірок, що змінюються, видалимо адресу комірки змінної d . Значення цієї комірки збільшимо та розв'яжемо задачу програмою OpenSolver.

Після розв'язування задачі побачимо, що значення $2\|x\|^2 - d$ зменшилось по абсолютній величині. Далі знову збільшимо значення d та розв'яжемо задачу. Після кожного розв'язування задачі значення $2\|x\|^2 - d$ буде наближатись до нуля по абсолютній величині. В залежності від кроку зміни величини d , потрібно буде 20–30 розв'язувань задачі до виконання умови $2\|x\|^2 - d = 0$ з заданою точністю. Виконання цієї умови означатиме, що знайдено оптимальний розклад занять і всі змінні будуть булеві та всі обмеження задачі будуть виконуватися.

Для демонстрації ефективності методу точної квадратичної регуляризації, для знаходження оптимального розкладу занять, ми виберемо дещо спрощений варіант розглянутої задачі. Складемо розклад тільки для двох груп, в кожній групі викладається шість різних предметів (шість лекцій та шість лабораторних робіт). Кількість викладачів дорівнює шести. Заняття проводяться без чисельника та знаменника (це дозволяє скоротити кількість рядків таблиці з 50 до 25). Така задача буде мати 600 булевих змінних, що також є складним для існуючих методів оптимізації. Вхідні дані задачі приведені в табл. 1. для першої групи, вони співпадають з даними для другої групи.

Спочатку задача розв'язувалась без обмежень на відсутність вікон і було отримано розклад, де майже кожен день були вікна. Далі були задані обмеження на відсутність вікон. При відсутності обмеження на кількість комп'ютерних класів було отримано наступний розв'язок (табл. 2).

Якщо викладач в деякі дні або часи зайнятий на другій роботі, то достатньо присвоїти відповідним змінним даного викладача на даній парі значення нуль. Як показали експерименти при такій умові графік роботи всіх викладачів може суттєво змінюватись.

Якщо припустити, що ці дві групи мають тільки один комп'ютерний клас для проведення лабораторних робіт то розподіл занять по дням тижня де що змінюється (табл. 3). Значення цільової функції при цьому збільшилось на дві одиниці.

Час розв'язування даної задачі на комп'ютері Intel(R) Core(TM) i5-8300H CPU 2.30 GHz склав до 5 хвилин часу. Метод EQR не сильно чутливий до зростання розмірності задачі. Ступінь зростання часу розв'язування задачі буде визначатись прямо-двоїстим методом внутрішньої точки. Як відомо, останній метод є поліноміальним. Тому даний алгоритм можна використовувати для розв'язування задач складання розкладу великої розмірності.

Наведемо отриманий запропонованим алгоритмом оптимальний розклад для даної задачі (табл. 4). В цій таблиці одиницями позначені заняття, що проводяться. З цієї таблиці бачимо, що вікна в розкладі занять для студентів кожної групи відсутні. Кількість пар занять для першої групи по дням тижня відповідно дорівнює 2, 3, 4, 4, 3, а для другої групи маємо 4, 3, 3, 3, 3. Коли розв'язувалась задача без обмеження на кількість комп'ютерних класів, то відповідні дані дорівнювали теоретично мінімальній дисперсії (3, 3, 4, 3, 3) та (3, 3, 3, 3, 4). Додаткові обмеження незначно зменшили значення дисперсії. З табл. 4 видно, що викладачі мають вікна в своїх заняттях. Це тому, що відповідні обмеження в моделі не враховувалися. Внесення таких обмежень призведе до зменшення цільової функції задачі і відповідно зменшить якість навчання. Автори спочатку розглядали в якості критерію задачі максимум дисперсії занять викладачів, але такий критерій не відповідає високій якості навчання студентів.

5 РЕЗУЛЬТАТИ

Результати розв'язування задачі складання розкладу запропонованим алгоритмом наведені в табл. 4. Отриманий розв'язок задовольняє всім обмеженням задачі та містить рівномірне навантаження занять на кожний день тижня. Задача розв'язувалась в добре відомому офісному пакеті Excel 10 з використанням її надбудови OpenSolver. Процедура розв'язування елементарна та не потребує спеціальних знань. В той же час сучасні методи оптимізації стикаються зі значними обчислювальними проблемами при розв'язуванні задач оптимального складання розкладу навчання.

6 ОБГОВОРЕННЯ

Задачі складання розкладу навчання виникають по декілька разів на рік в кожному університеті. Кожен університет має свою специфіку, яка закладається у відповідну модель. На жаль не існує стандартних моделей на яких можна було б порівнювати різні алгоритми розв'язку. В деякій мірі стандартною можна вважати запропоновану модель, так як вона містить тільки природні обмеження та цільову функцію, яка забезпечує якість навчання. Щодо існуючих методів розв'язування задач оптимізації з булевими змінними, то пропонуються переважно тільки методи розгалужень та границь і еволюційні алгоритми, що використовують випадковий пошук. Ці методи потребують занадто багато часу навіть при розв'язуванні задач невеликої розмірності. Зі зростанням розмірності задачі час розв'язування експоненційно зростає. Тому при розв'язуванні практичних задач складання розкладу ці методи будуть вимагати декілька днів чи тижнів неперервної роботи комп'ютера, або завершувати роботу раніше з отриманням розв'язку далекого від оптимального та часто недопустимого. На даний час використання методу EQR та запропонованого алгоритму для складання розкладу не має альтернатив. Це підтверджується також значним обсягом порівняльних обчислювальних експериментів для тестових мультимодальних задач [11]. Крім того, даний метод легко реалізується в пакеті Excel за допомогою надбудови OpenSolver практично для довільної розмірності.

Для підтвердження ефективності запропонованого алгоритму були проведені обчислювальні експерименти. Враховуючи те, що запропонована модель складання розкладу є новою, порівнювались результати розв'язування даної моделі розробленим алгоритмом та існуючими програмами. Для обчислювального експерименту була скорочена матриця невідомих булевих змінних розглянутого прикладу до розміру 15×8 . В якості альтернативи розробленому алгоритму використовувалось існуюче програмне забезпечення, це надбудови Solver та OpenSolver пакету Excel 10 в яких реалізовані алгоритми еволюційні та з булевими змінними. Спочатку розв'язувалась задача (1)–(8) з булевими змінними програмами Solver та OpenSolver в яких реалізовані алгоритми методу розгалужень та границь. Кожна з цих програм протягом 5 годин не змогла завершити роботу. Зауважимо, що

якщо відкинути умову булевості змінних, то дана задача буде розв'язана за долі секунди.

Далі для даного прикладу розв'язувалась задача (1)–(9) еволюційним пошуком. Але для цього прикладу даний пошук не зміг знайти допустимий розв'язок задачі. В той же час, для розв'язування задачі (3)–(8),

(10)–(11) запропонованим алгоритмом потребувалось менше однієї хвилини машинного часу. Таким чином, запропонований алгоритм для даного класу задач, як мінімум в 300 разів ефективніше існуючих алгоритмів, причому зі зростанням розмірності задачі ця ефективність також буде зростати.

Таблиця 1 – Дані для складання розкладу

Заняття	Лекції						Лабораторні роботи					
	В1	В2	В3	В4	В5	В6	В1	В2	В3	В4	В5	В6
Викладачі	П1	П2	П3	П4	П5	П6	П1	П2	П3	П4	П5	П6
Предмети	1	1	1	1	1	1	2	2	1	2	2	1
Години	1	1	1	1	1	1	2	2	1	2	2	1

Таблиця 2 – Розподіл занять по дням тижня

	Понеділок					Вівторок					Середа					Четвер					П'ятниця					
Гр.1	0	1	1	1	0	0	1	1	1	0	0	1	1	1	1	0	1	1	1	0	0	1	1	1	1	0
Гр.2	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	1	1	1	1	1	0

Таблиця 3 – Розподіл занять по дням тижня (з урахування комп'ютерних класів)

	Понеділок					Вівторок					Середа					Четвер					П'ятниця					
Гр.1	0	0	1	1	0	1	1	1	0	0	0	1	1	1	1	0	1	1	1	1	0	0	1	1	1	1
Гр.2	1	1	1	1	0	0	0	1	1	1	0	1	1	1	0	0	0	1	1	1	1	1	1	0	0	0

Таблиця 4 – Оптимальний розклад

	Група 1												Група 2											
	В	1	2	3	4	5	1	6	2	3	4	6	1	2	3	4	5	1	6	2	3	4	6	6
Г	1	1	1	1	1	1	2	2	1	2	2	1	1	1	1	1	1	1	2	2	1	2	2	1
	Лекції						Лабораторні						Лекції						Лабораторні					
П																								1
																								1
											1							1						
								1										1						
В																								
			1																					
																								1
С																								
						1													1					
							1															1		
	1																				1			
Ч																								
П																								
																								1

ВИСНОВКИ

В роботі вирішено актуальну науково-технічну проблему знаходження оптимального розкладу навчання в університетах.

Розглянута задача складання розкладу для студентів однієї спеціальності. Побудована нова оптимізаційна модель цієї задачі, що містить нову цільову функцію та обмеження на відсутність вікон в заняттях студентів. Цю модель легко узагальнити для факультету чи всього університету. При необхідності обмеження задачі можна доповнити.

Наукова новизна роботи полягає в тому, що вперше за 70 років досліджень складання розкладів пропонується простий та ефективний алгоритм знаходження оптимального розкладу. Існуючі моделі та методи або досить складні і не дозволяють отримати оптимальні розв'язки існуючими методами, або використовують евристичні алгоритми, що знаходять розв'язки часто далекі від оптимальних.

Практична цінність роботи полягає в тому, що ми вже сьогодні можемо впроваджувати дану модель та алгоритм в диспетчерські відділи кожного університету.

Майбутні напрями дослідження та розробки пов'язані з дачею практичних рекомендацій диспетчерським відділам по використанню запропонованої моделі та алгоритму. Ми також продовжуємо обчислювальні експерименти для нової моделі складання розкладів, а також для її узагальнень. Актуальною залишається проблема обчислювальних експериментів для складання розкладу факультету чи всього університету.

ПОДЯКА

Автори даної статті щиро вдячні співробітникам відділу диспетчеризації УДХТУ за змістовну постановку задачі складання розкладу навчання в університеті.

УДК 519.85

ОПТИМИЗАЦИЯ РАСПИСАНИЯ ОБУЧЕНИЯ В УНИВЕРСИТЕТЕ

Косолап А. І. – д-р физ.-мат. наук, профессор, заведующий кафедрой специализированных компьютерных систем Украинского государственного химико-технологического университета, Днепр, Украина.

Дубовик Т.М. – старший преподаватель кафедры специализированных компьютерных систем Украинского государственного химико-технологического университета, Днепр, Украина.

АННОТАЦИЯ

Актуальность. В этой работе рассматривается известная задача составления расписания обучения в университете. Такие задачи решаются несколько раз в год в каждом учебном заведении. Несмотря на многочисленные исследования в данной области, проблема построения оптимального расписания остается открытой. Это связано со сложностью соответствующей оптимизационной задачи, в частности ее значительной размерностью, что затрудняет численное решение такой задачи существующими методами оптимизации. В совершенствовании нуждаются также оптимизационные модели составления расписаний. Таким образом, оптимизация расписания является сложной вычислительной проблемой и требует разработки новых методов ее решения.

Цель работы. Совершенствование оптимизационных моделей составления расписаний обучения в университете и использования новых эффективных методов для их решения.

Метод. Мы используем метод точной квадратичной регуляризации для решения оптимизационных задач составления университетского расписания. Точная квадратичная регуляризация позволяет преобразовать сложные оптимизационные модели с булевыми переменными к задаче максимума нормы вектора на выпуклой множестве. Для решения этой задачи мы используем эффективный прямо-двойственный метод внутренней точки и метод дихотомии. Этот метод показал значительно лучшие результаты при решении многих сложных мультимодальных задач. Это подтверждается многими сравнительными

ЛІТЕРАТУРА / ЛИТЕРАТУРА

1. de Werra D. An introduction to timetabling / D. de Werra // European Journal of Operational Research. – 1985. – 19. – P. 151–162.
2. Kosolap A. A new method for global optimization / A. Kosolap // ESAIM: Proceedings and surveys. – 2021. – Vol. 71. – P. 121–130.
3. Junginger W., Timetabling in Germany – a survey [Электронный ресурс] // In Interfaces. – 1986. – Vol. 16, No. 4. – P. 66–74. <https://pubsonline.informs.org./doi/abs/10.1287/inte.16.4.66>.
4. Shraerf A. A survey of automated timetabling // Journal Artificial Intelligence Review. – 1999. – Volume 13, Issue 2. – P. 1–42. DOI:10.1023/A:1006576209967.
5. Practices in timetabling in higher education institutions: a systematic review / [R. A. Oude Vrielink, E. A. Jansen, E. W. Hans, J. van Hillegersberg] // Ann. Oper. Res. – 2019. – 275. – P. 145–160.
6. Aziz N.L.A. A brief review on the features of university course timetabling problem / N.L.A. Aziz, N.A.H. Aizam // AIP Conference Proceedings, Volume 2016, Issue 1. – P. 1–8. DOI.org/10.1063/1.5055403.
7. Qu R. A survey of search methodologies and automated system development for examination timetabling / R. Qu, E.K. Burke // Journal of scheduling. – 2009. – Volume 12, Number 1. – P. 55–89.
8. An overview of curriculum-based course timetabling / [A. Bettinelli, V. Cacchiani, R. Roberti, P. Toth] // TOP. – 2015. – No 23. – P. 313–349. DOI: 10.1007/s11750-015-0366-z.
9. Alghamdi H. A Review of Optimization Algorithms for University Timetable Scheduling Engineering / H. Alghamdi, T. Alsubait, H. Alhakami, A. Baz // Technology & Applied Science Research. – 2020. – Vol. 10, No. 6. – P. 6410–6417.
10. Nocedal J. Numerical optimization / J. Nocedal, S. J. Wright. – Springer, 2006. – 685 p.
11. Kosolap A. Practical Global optimization / A. Kosolap. – Dnipro : Publisher Bila K.O., 2020. – 192 p.

Стаття надійшла до редакції 25.06.2021.
Після доробки 16.08.2021.

ми вычислительными экспериментами. Еще большую эффективность метод точной квадратичной регуляризации демонстрирует при решении задач составления расписания обучения. Этот метод оптимизации используется впервые, для данного класса задач, поэтому он потребовал разработки соответствующего алгоритмического обеспечения.

Результаты. Построена новая более простая оптимизационная модель составления расписания, которая легко реализуется программно в пакете Excel при наличии надстроек OpenSolver, RiskSolver и других. Приведен небольшой пример построения расписания и описана пошагово инструкция получения оптимального решения.

Выводы. Разработана новая эффективная технология составления расписания обучения в университете, которая отличается простотой реализации и не требует разработки специального программного обеспечения. Эффективность обеспечивается использованием нового метода точной квадратичной регуляризации.

КЛЮЧЕВЫЕ СЛОВА: университетское расписание, оптимизация, задачи по булевыми переменными, метод точной квадратичной регуляризации.

UDC 519.85

OPTIMIZATION OF TIMETABLE AT THE UNIVERSITY

Kosolap A. I. – Dr. Sc., Professor, head of the Department of Specialized Computer Systems of the Ukrainian State University of Chemical Technology, Dnipro, Ukraine.

Dubovik T. M. – Senior Lecturer of the Department of Specialized Computer Systems of the Ukrainian State Chemical-Technological University, Dnipro, Ukraine.

ABSTRACT

Context. In this paper, we consider a well-known university scheduling problem. Such tasks are solved several times a year in every educational institution. The problem of constructing an optimal schedule remains open despite numerous studies in this area. This is due to the complexity of the corresponding optimization problem, in particular, its significant dimension. This complicates its numerical solution with existing optimization methods. Scheduling models also need improvement. Thus, schedule optimization is a complex computational problem and requires the development of new methods for solving it.

Objective. Improvement of optimization models for timetabling at the university and the use of new effective methods to solve them.

Method. We use the exact quadratic regularization method to solve timetabling optimization problems. Exact quadratic regularization allows transforming complex optimization models with Boolean variables into the problem of maximizing the vector norm on a convex set. We use the efficient direct dual interior point method and dichotomy method to solve this problem. This method has shown significantly better results in solving many complex multimodal problems. This is confirmed by many comparative computational experiments. The exact quadratic regularization method is even more effective in solving timetabling problems. This optimization method is used for the first time for this class of problems, so it required the development of adequate algorithmic support.

Results. We propose a new, simpler timetabling optimization model that can be easily implemented software in Excel with the OpenSolver, RoskSolver, and others. We give a small example of building a schedule and describe step-by-step instructions for obtaining the optimal solution.

Conclusions. An efficient new technology developed for university timetable, which is simple to implement and does not require the development of special software. The efficiency of the technology is ensured by the use of a new method of exact quadratic regularization.

KEYWORDS: university timetabling, optimization, problems with boolean variables, exact quadratic regularization method.

REFERENCES

1. de Werra, D. An introduction to timetabling, *European Journal of Operational Research*, 1985, 19, pp. 151–162.
2. Kosolap A. A new method for global optimization, *ESAIM: Proceedings and surveys*, 2021, Vol. 71. pp. 121–130.
3. Junginger W. Timetabling in Germany – a survey [Электронный ресурс], *In Interfaces*, 1986, Vol. 16, No. 4, pp. 66–74. <https://pubsonline.informs.org>.
4. Shraerf A. A survey of automated timetabling, *Journal Artificial Intelligence Review*, 1999, Volume 13, Issue 2, pp. 1–8.
5. Oude Vrielink R. A., Jansen E. A., Hans E. W., van Hillegersberg J. Practices in timetabling in higher education institutions: a systematic review. *Ann. Oper. Res.*, 2019, 275, pp. 145–160.
6. Aziz N. L. A., Aizam N. A. H. A brief review on the features of university course timetabling problem. *AIP Conference Proceedings*, Volume 2016, Issue 1, pp. 1–8.
7. Qu R., Burke E.K. A survey of search methodologies and automated system development for examination timetabling. *Journal of scheduling*, 2009, Volume 12, Number 1, pp. 55–89.
8. Bettinelli A., Cacchiani V., Roberti R., Toth P. An overview of curriculum-based course timetabling, *TOP*, 2015, No 23, pp. 313–349. DOI: 10.1007/s11750-015-0366-z.
9. Hayat Alghamdi, Tahani Alsubait, Hosam Alhakami, Abdullah Baz A Review of Optimization Algorithms for University Timetable Scheduling Engineering, *Technology & Applied Science Research*, 2020, Vol. 10, No. 6, pp. 6410–6417.
10. Nocedal J., Wright S. J. Numerical optimization. Springer, 2006. 685 p.
11. Kosolap A. Practical Global optimization. Dnipro, Publisher Bila K. O., 2020, 192 p.

Наукове видання

**Радіоелектроніка,
інформатика,
управління**

№ 3/2021

Науковий журнал

Головний редактор – д-р техн. наук С. О. Субботін

Заст. головного редактора – д-р техн. наук Д. М. Піза

Комп'ютерне моделювання та верстання
Редактор англійських текстів

С. В. Зуб
С. О. Субботін

Оригінал-макет підготовлено у редакційно-видавничому відділі НУ «Запорізька політехніка»

Свідоцтво про державну реєстрацію
КВ № 24220-14060 ПР від 19.11.2019.

*Підписано до друку 26.08.2021. Формат 60×84/8.
Папір офс. Різогр. друк. Ум. друк. арк. 21,39.
Тираж 300 прим. Зам. № 679.*

69063, м. Запоріжжя, НУ «Запорізька політехніка», друкарня, вул. Жуковського, 64

Свідоцтво суб'єкта видавничої справи
ДК № 6952 від 22.10.2019.