9. Lisitskaya I. V., Oleshko O. I., Rudenko S. N., Drobat'ko E.V., Grigor'ev A.V. Kriptograficheskyie svoistva umenshenoi versii shifra «Myhomor», *Spetsialni telekomunicatsiini systemy'ta zahist informatsii. Zbirnuk naykovy'h prats'*, Kiiv, 2010, Issue 2(18), pp. 33–42.
10. Dolgov V. I., Oleinikov R. V., Bol'shakov A. Ю., Grigor'ev A.V., Drobat'ko E.V. Kriptograficheskie svoistva umenshenoi versii shifra «Kalina», *Prikladnaya radioelektronika*, 2010, vol. 9, No. 3, pp. 349-354.
11. Dolgov V. I., Makarchuk IA. A., Grigoriev A. V., Drobat'ko E. V. Issledovanie kriptograficheskih pokazateley umensheny'h modeley shifrov GOST i DES, *Prikladnaya radioelektronika*, 2011, Vol. 10, No. 2, pp. 127–134.
12. Dolgov V. I., Lisitskaya I. V., Grigoriev A. V., Shirokov A. V. Issledovanie tsiklicheskih i differentsialny'h svoistv umenshenoi modeli shifra Labirint, *Prikladnaya radioelektronika*, 2009, Vol. 8, No. 3, pp. 283–289.
13. Oleinikov R. V., Oleshko O. I., Lisitskay K. E., Tiviashev K. E. Differentsialny'e svoistva podstanovok, *Prikladnaya radioelektronika*. – 2010, Vol. 9, No. 3, pp. 326–333.
14 O'Connor L. J. On the Distribution of Characteristics in Bijective Mappings. Advances in Cryptology. EUROCRIPT 93, Lecture Notes in Computer Science, vol. 795, T. Hellesethed., Springer-Verlag, 1994, pp. 360–370.
15. *Lisitskaya* I.V. Svoistva zakonov raspredelenia XOR tablits i tablits lineiny'h approksimatsy sluchainy'h podstanovok. [Text], *Visnyk Charkivs'kogo natsionalnogo universitety imeni V. N. Karazina*, 2011, No. 960, Issue. 16, pp. 196–206.
16. *Shnaer* B. Prikladnaya kriptografiya. Protokoly', algoritmy', ishodny'e teksty' na iazuke Si. [Text]. Moscow, TRIUMF, 2002, 816 p.

УДК 004.9

Пшеничний О. Ю.

Аспірант Національного університету «Львівська політехніка»

ВЛАСТИВОСТІ АСОЦІАТИВНИХ ЗАЛЕЖНОСТЕЙ У АНАЛІЗІ ДАНИХ

У статті наведено результати дослідження властивостей асоціативних залежностей та можливостей їх ефективного агрегування. Розрозбено метод виявлення асоціативних залежностей широкого класу у великих наборах даних.

Ключові слова: асоціативна залежність, функціональна залежність, залежності даних, аналіз даних.

ВСТУП

Аналіз даних та отримання з них додаткової інформації про предметну галузь (Data Mining) є на сьогодні великою галуззю комп'ютерних наук, яка активно розвивається і збагачується новими методами, алгоритмами та програмними засобами, що їх реалізовують. Охопити всю структуру та різноманітність підходів даної галузі неможливо.

У даній роботі розглядається задача виявлення асоціативних залежностей у великих обсягах даних та її проблематика, вивчаються можливості оптимізації пошуку асоціативних залежностей та їх властивості.

Аналіз даних на предмет виявлення залежностей та кореляцій широко застосовується у соціології, психології, політології, фізиці, енергетиці, астрономії, комп'ютерних науках та безлічі інших прикладних дисциплін. Задача виявлення асоціативних залежностей в даних соціологічних опитувань розглядається в [1]. Даний напрям аналізу даних відносно не новий, проте в цій галузі до цих пір проводяться активні дослідження. Наприклад, у роботі [2] описується метод побудови агрегованих асоціативних правил на основі простіших залежностей. Пояснити такий інтерес до виявлення залежностей в даних можна стрімким злетом обчислювальної потужності комп'ютерної техніки, а також зростанням обсягів накопичених даних у багатьох галузях життя суспільства до таких обсягів, що аналіз їх експертним шляхом або неможливий, або неповний. Сучасні обчислювальні засоби дозволяють реалізовувати все складніші алгоритми та застосовувати їх до даних великих обсягів. Це стимулює науковців до розробки таких алгоритмів, а власників великих баз та сховищ даних – до розробки програмних засобів аналізу накопиченої інформації.

На даний час деякі науково-технічні галузі вже мають потужні методи аналізу даних, спеціалізовані до своїх потреб та структури даних. Серед них можна виділити програмні засоби CLASSIFI (Department of Pathology, UT Southwestern Medical Center) [3], BiNGO (Department of Plant Systems Biology, VIB/Ghent University) [4] та EASE (National Institute of Allergy and Infectious Diseases) [5]. Проте більшість науково-дослідних установ не можуть дозволити собі розробку подібних систем і потребують загальнодоступного методу широкого застосування.

Отже, ефективний пошук асоціативних залежностей в багатоатрибутних даних є актуальною задачею сучасного аналізу даних.

Варто зазначити, що data mining – дуже широка галузь аналізу даних і пошук асоціативних залежностей – лише її частина.

Метою даної роботи є вивчення властивостей асоціативних залежностей, що дозволять реалізувати ефективні алгоритми пошуку таких залежностей в реляційних базах даних.

ЗВ'ЯЗОК ТА МІСЦЕ РОБОТИ В ІСНУЮЧИХ МЕТОДАХ АНАЛІЗУ ДАНИХ

Технології data mining передбачають виявлення залежностей в даних виду «якщо … то …» або «для … справджується …». Такі залежності представляються імплікаціями, тобто продукційними правилами чи асоціативними правилами.

Data Mining включає широкий набір математичних та алгоритмічних засобів, що включають нейронні мережі, еволюційні алгоритми, дерева рішень та ін. Проте сучасні дослідження все більше роблять акцент на напрямку пошуку логічних залежностей в даних. За їх допомогою вирішуються задачі класифікації, прогнозування, формування образів на підставі формальних логік та ін. [2].

Очевидно, що в базах даних, що містять мільйони об'єктів, можна побудувати неймовірно велику кількість асоціативних залежностей і усі ці залежності не те що вивчати, а навіть зберігати неможливо. Але на щастя такого завдання ніхто не ставить. Натомість дійсно важливим є пошук таких асоціативних залежностей, що мають достатній рівень статистичної обґрунтованості.

Основні недоліки сучасних методів пошуку асоціацій в даних:

– працюють тільки з бінарними ознаками об'єктів;

– «не знаходять» асоціативних залежностей з малою підтримкою;

– не дозволяють ефективно реалізувати додавання нових записів у джерело даних та консолідувати дані, отримані з різних джерел;

– неефективно працюють з багатоатрибутними залежностями;

– недостатньо гнучкі в плані критеріїв відбору шуканих залежностей.

Одним з шляхів усунення цих недоліків є побудова агрегованих асоціативних правил. У роботі [2] пропонується система 4-х параметрів асоціативного правила, що описують його властивості. Безумовно така система характеристик є гнучкішою за єдиний параметр інтенсивності асоціації, що використовується в [1], проте факт достатності наведеної системи залишається під питанням.

З урахуванням вищезазначеного, у даній роботі розглядається актуальна науково-технічна задача вдосконалення методів побудови та оцінки агрегованих асоціативних правил в базах даних великого розміру.

СПЕЦИФІКА ДОСЛІДЖУВАНОЇ ЗАДАЧІ

Пошук довільних асоціативних правил $P(x) \to Q(x), x \in r(R)$ у відношенні $r(R)$ є дуже широкою задачею, вирішення якої поки в майбутньому і ця задача не є об'єктом даного дослідження. У даній роботі пропонується вивчення властивостей асоціативних залежностей, у яких умовний та результуючий предикат мають вигляд:

$$P = P_1^e \vee P_2^e \vee \ldots \vee P_h^e = \bigvee_{k=1}^{h} P_k^e,$$

$$P_k^e = A_{i_1} \in \left\{a_{(i_1)(j_1)}\right\} \wedge A_{i_2} \in \left\{a_{(i_2)(j_2)}\right\} \wedge A_{i_k} \in \left\{a_{(i_k)(j_k)}\right\},$$

$$\forall l = \overline{1..k} : A_{i_l} \in R, \forall m = \overline{1..\left|\left\{a_{(i_l)(j_m)}\right\}\right|} : a_{(i_l)(j_m)} \in dom\left(A_{i_l}\right),$$

$$\forall i, j \in \{1..h\} : \arg\left(P_i^e\right) = \arg\left(P_j^e\right). \quad (1)$$

$$F_I : \bigvee_{k=1}^{s} P_k^e \to \bigvee_{l=1}^{t} Q_l^e. \quad (2)$$

Позначення $\arg(P)$ використано, як оператор отримання множини атрибутів-аргументів предиката P.

Назвимо такі асоціативні залежності окремим терміном – імовірнісні продукційні залежності (ІПЗ). Введення окремого терміну необхідне для уникнення помилкового трактування викладень та розширення їх на довільні асоціативні залежності. Термін містить слово «ймовірнісна» тому, що основною характеристикою таких залежностей, як буде показано далі, є ймовірність її виконання для нового об'єкта заданої схеми, отриманого з випадкового процесу збору даних в предметній галузі (за умови відсутності довготермінових тенденцій зміни параметрів середовища). В літературі використовують різні позначення, інтерпретації та трактування даного поняття. Наприклад, у [2] використовується термін «інтенсивність асоціації», в [1] – «рівень довіри». Проте дані позначення стосуються більше факту, що має місце у наявних даних, а не в даних, з якими потрібно працювати в поточний момент часу. До того ж, якщо розглядати систему, як статичну, що не поповняється новими знаннями, втрачається зміст застосування таких технік, як наприклад, згладженя Лапласа [6] та інших методів захисту від шуму та невизначеності даних. Саме для наголошення на дослідженні випадкових процесів, їх динамічності та проблемах і введено термін «імовірність» у термін форми залежностей, що досліджуються. Друга частина терміну – «продукційні» особливого пояснення не потребує, оскільки в основі залежності лежить продукційне правило.

Отже, ймовірнісна продукційна залежність – це продукційне правило виду в селекції основного відношення, яке справджується для значущої кількості об'єктів цієї селекції. Поріг значущості повинен визначатись експертним шляхом, або виходячи з розрахунків імовірності помилкового виділення цієї залежності.

Запишемо позначення формул (1), (2) у термінах реляційної алгебри:

$$P^e(x) = \pi_{A_{i_1} A_{i_2} \ldots A_{i_l}}(x) \in \left\{a_{(i_1)(j_1)}\right\} \times \left\{a_{(i_2)(j_2)}\right\} \times \ldots \times \left\{a_{(i_k)(j_k)}\right\},$$

$$x \in r(R), \forall l = \overline{1..k} : A_{i_l} \in R, \forall m = \overline{1..\left|\left\{a_{(i_l)(j_m)}\right\}\right|} : a_{(i_l)(j_m)} \in dom\left(A_{i_l}\right). \quad (3)$$

$$P(x) = \pi_{A_{i_1}A_{i_2}\ldots A_{i_l}}(x) \in \left\{a_{(i_1)(j_{1,1})}\right\} \times \left\{a_{(i_2)(j_{1,2})}\right\} \times \ldots \times \left\{a_{(i_k)(j_{1,k})}\right\} \cup$$
$$\cup \left\{a_{(i_1)(j_{2,1})}\right\} \times \left\{a_{2,(i_2)(j_{2,2})}\right\} \times \ldots \times \left\{a_{(i_k)(j_{2,k})}\right\} \cup \ldots \cup$$
$$\cup \left\{a_{(i_1)(j_{h,1})}\right\} \times \left\{a_{2,(i_2)(j_{h,2})}\right\} \times \ldots \times \left\{a_{(i_k)(j_{h,k})}\right\}. \quad (4)$$

Тобто, предикати ІПЗ можна представляти як предикати, визначені на кортежах відношення $r(R)$, а не лише на множині атрибутів.

Поріг значущості ІПЗ може визначатись на основі довільної функції оцінки важливості знайденої залежності. Проте найчастіше використовуваними є показники рівня підтримки та рівня довіри. У роботі [2] показано, що цих параметрів недостатньо для адекватного опису залежностей предметної галузі і пропонується використовувати додаткові характеристики: рівень покращення та повну взаємну інформацію.

Розглянемо ці показники детальніше.

Рівень підтримки – характеристика предиката селекції на відношенні, що обчислюється як відношення кількості об'єктів, які задовольняють предикат P до загальної кількості об'єктів у відношенні:

$$Sup(P) = \frac{|\sigma_P(r)|}{|r|}. \quad (5)$$

У випадку обчислення рівня підтримки для ІПЗ умовний та результуючий предикат залежності об'єднуються знаком кон'юнкції:

$$Sup(S \to T) = Sup(S \wedge T) = \frac{|\sigma_{S \wedge T}(r)|}{|r|}. \quad (6)$$

Рівень довіри – відношення кількості об'єктів, для яких має місце така ІПЗ до кількості об'єктів в селекції:

$$Conf(S \to T) = P(S \to T) = \frac{|\sigma_{S \wedge T}(r)|}{|\sigma_S(r)|}. \quad (7)$$

З використанням поняття рівня підтримки, рівень довіри можна обчислити, як

$$Conf(S \to T) = \frac{Sup(S \to T)}{Sup(S)}. \quad (8)$$

Рівень покращення обчислюється, як відношення рівнів довіри та підтримки ІПЗ:

$$Imp(S \to T) = \frac{Conf(S \to T)}{Sup(T)} = \frac{Sup(S \wedge T)}{Sup(S) \cdot Sup(T)}. \quad (9)$$

Повна взаємна інформація в загальному випадку обчислюється, як

$$I_{X \leftrightarrow Y} = \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \log_2 \frac{p_{ij}}{p_i r_j}, \quad (10)$$

де $p_{ij} = P(X = x_i \wedge Y = y_j)$ – ймовірність того, що X знаходиться в стані x_i, а Y – в стані y_j; $p_i = P(X = x_i)$ – ймовірність того, що X знаходиться в стані x_i; $r_j = P(Y = y_j)$ – ймовірність того, що Y знаходиться в стані y_j.

Для асоціативних правил взаємна інформація буде визначатися як

$$I_{X \leftrightarrow Y} = \sum_{i=1}^{n} \sum_{j=1}^{m} Sup(x_i \to y_j) \log_2 Imp(x_i \to y_j). \quad (11)$$

АНАЛОГИ ПРАВИЛ ВИВЕДЕННЯ ФУНКЦІОНАЛЬНИХ ЗАЛЕЖНОСТЕЙ ДЛЯ ІПЗ

Як і у випадку з F-залежностями, множину ІПЗ, що мають місце в заданому відношенні, можна представити деякою їх підмножиною, з якої за допомогою правил виведення можна отримати усі ІПЗ даного відношення. Оскільки ІПЗ є розширенням F-залежностей, то варто розглянути трансформації аксіом виведення функціональних залежностей $F1$–$F6$ для ІПЗ.

ІПЗ характеризуються багатьма параметрами, найважливіші з яких наведені у формулах (5)–(11). Проте найчастіше використовуваним і найпростішим для розуміння є параметр рівня довіри. На ньому ґрунтуються більш складні характеристики залежностей. В ході досліджень було встановлено:

– фільтрація по рівню підтримки не дозволить здійснювати виведення ІПЗ з малих часткових залежностей – рівень підтримки неодмінно зростає при об'єднанні ІПЗ і тому неможливо здійснити відсікання деяких груп залежностей на основі даного параметра;

– рівень покращення є нелінійною характеристикою ні за потужністю множини кортежів, що задовольняють умовну частину, ні за потужністю множини кортежів, що задовольняють результуючу частину ІПЗ і теж не дає можливості здійснювати відсікання генерації ІПЗ на основі множини наявних залежностей;

– повна взаємна інформація аналогічно є нелінійною за потужністю селекції обома предикатами ІПЗ.

Отже, будувати правила виведення для параметрів рівня покращення та повної взаємної інформації немає змісту, оскільки вони обчислюються через рівень підтримки та рівень довіри, а також дані характеристики нелінійно залежать від кількості кортежів, які відповідають залежностям.

Строге доведення даного факту не наводиться, оскільки він інтуїтивно зрозумілий з вищенаведених міркувань, а формальне доведення дуже громіздке.

Вказані характеристики ІПЗ мають зміст лише в розгляді окремих залежностей, як додатковий параметр опису типу та сили залежності. Але при виведенні нових ІПЗ достатньо оперувати поняттями рівня підтримки та рівня довіри. Вони дозволяють достатньо обмежити набір знайдених залежностей, щоб потім можна було проводити другий етап відбору ІПЗ по довільних критеріях, які необхідні в тій чи іншій предметній галузі (в числі цих критеріїв можуть бути і параметри рівня покращення, повної взаємної інформації та ін.). Таким чином долається проблема гнучкості фільтрації шуканих асоціативних залежностей – використання двоетапної фільтрації дозволяє використовувати довільні критерії якості ІПЗ на другому етапі.

Отож, розглянемо трансформацію правил виводу функціональних залежностей для ІПЗ.

Рефлексивність рівня довіри.

$Conf\left(s \in S \to s \in S\right) = 1$ для будь-якого відношення $r(R)$.

Доведення:

$$Conf\left(s \in S \to s \in S\right) = \frac{\left|\sigma_{s\in S \wedge s\in S}\right|}{\left|\sigma_{s\in S}\right|} = \frac{\left|\sigma_{s\in S}\right|}{\left|\sigma_{s\in S}\right|} = 1 \ .$$

Поповнення рівня довіри.

Якщо $Conf\left(s \in S \to t \in T\right) = p$, то $Conf\left(s \in S \wedge w \in D(W) \to t \in T\right) = p$, де $D(W)$ – домен атрибута W відношення $r(R)$.

Доведення:

$$Conf\left(s \in S \wedge w \in D(W) \to t \in T\right) = \frac{\left|\sigma_{s\in S \wedge w\in D(W) \wedge t\in T}(R)\right|}{\left|\sigma_{s\in S \wedge w\in D(W)}(R)\right|} =$$

$$= \left|\forall x \in r : q = \pi_{W=w}(x) \in D(W) \Rightarrow w \in D(W)\right| =$$

$$= \frac{\left|\sigma_{s\in S \wedge t\in T}(R)\right|}{\left|\sigma_{s\in S}(R)\right|} = Conf\left(s \in S \to t \in T\right) = p.$$

Приклад 1.

A	B	C	D
a_1	b_1	c_1	d_1
a_1	b_1	c_2	d_1
a_1	b_2	c_2	d_1
a_2	b_2	c_2	d_1
a_2	b_2	c_2	d_2
a_2	b_2	c_2	d_2
a_2	b_3	c_3	d_2
a_2	b_2	c_2	d_2
a_3	b_3	c_3	d_2

$$Conf\left(a \in \{a_1, a_2\} \to d \in \{d_1\}\right) = \frac{4}{8} = 0,5$$

$$Conf\left(a \in \{a_1, a_2\} \wedge b \in \{b_1, b_2, b_3, b_4\} \to d \in \{d_1\}\right) =$$

$$= \frac{\left|\sigma_{a\in\{a_1,a_2\}\wedge b\in\{b_1,b_2,b_3,b_4\}\wedge d\in\{d_1\}}(R)\right|}{\left|\sigma_{a\in\{a_1,a_2\}\wedge b\in\{b_1,b_2,b_3,b_4\}}(R)\right|} = \frac{4}{8} = 0,5 \ .$$

Адитивність рівня довіри.

Якщо $Conf\left(s \in S \to t \in T\right) = p$ і $Conf\left(s \in S \to w \in W\right) = 1$, то $Conf\left(s \in S \to t \in T \wedge w \in W\right) = p$.

Доведення:

$$Conf\left(s \in S \to t \in T \wedge w \in W\right) = \frac{\left|\sigma_{s\in S \wedge t\in T \wedge w\in W}\right|}{\left|\sigma_{s\in S}\right|} =$$

$$= \left|s \in s \to w \in W\right| = \frac{\left|\sigma_{s\in S \wedge t\in T}\right|}{\left|\sigma_{s\in S}\right|} = Conf\left(s \in S \to t \in T\right) = p \ .$$

Приклад 2:

Розглянемо ІПЗ з прикладу 1:

$$Conf\left(A \in \{a_1\} \to B \in \{b_1\}\right) = \frac{2}{3} \cdot$$

$$Conf\left(A \in \{a_1\} \to D \in \{d_1\}\right) = 1$$

.

По них можна зробити висновок, що $Conf\left(A \in \{a_1\} \to B \in \{b_1\} \wedge D \in \{d_1\}\right) = \frac{2}{3}$. Це підтверджують обрахунки по формулі (7):

$$Conf\left(A \in \{a_1\} \to B \in \{b_1\} \wedge D \in \{d_1\}\right) =$$

$$= \frac{\left|\sigma_{A\in\{a_1\}\wedge B\in\{b_1\}\wedge D\in\{d_1\}}\right|}{\left|\sigma_{A\in\{a_1\}}\right|} = \frac{2}{3}.$$

Проективність рівня довіри.

Якщо $Conf\left(s \in S \to t \in T \wedge w \in W\right) = p$ і $Conf\left(s \in S \to w \in W\right) = 1$, то $Conf\left(s \in S \to t \in T\right) = p$.

Доведення:

$$Conf\left(s \in S \to t \in T\right) = \frac{\left|\sigma_{s\in S\wedge t\in T}\right|}{\left|\sigma_{s\in S}\right|} =$$

$$= \frac{\left|\sigma_{s\in S\wedge t\in T\wedge w\in W}\right|}{\left|\sigma_{s\in S}\right|}\Big| s \in S \to w \in W = Conf\left(s \in S \to t \in T \wedge w \in W\right) = p.$$

Приклад 3:

Розглянемо приклад, поданий в попередньому пункті, в зворотному варіанті.

З ІПЗ $Conf\left(A \in \{a_1\} \to B \in \{b_1\} \wedge D \in \{d_1\}\right) = \frac{2}{3}$ та $Conf\left(A \in \{a_1\} \to D \in \{d_1\}\right) = 1$ можна зробити висно-

вок, що $Conf\left(A\in\{a_1\}\to B\in\{b_1\}\right)=\frac{2}{3}$. Перевірка за формулою (7):

$$Conf\left(A\in\{a_1\}\to B\in\{b_1\}\right)=\frac{\left|\sigma_{A\in\{a_1\}\wedge B\in\{b_1\}}\right|}{\left|\sigma_{A\in\{a_1\}}\right|}=\frac{2}{3}.$$

Транзитивність рівня довіри.

Якщо $Conf\left(s\in S\to t\in T\right)=p$ і $Conf\left(t\in T\to w\in W\right)=1$, то $Conf\left(s\in S\to w\in W\right)\geq p$.

Доведення:

$$Conf\left(s\in S\to w\in W\right)=\frac{\left|\sigma_{s\in S\wedge w\in W}\right|}{\left|\sigma_{s\in S}\right|}.$$

$$Conf\left(s\in S\to t\in T\right)=\frac{\left|\sigma_{s\in S\wedge t\in T}\right|}{\left|\sigma_{s\in S}\right|}=p.$$

Таким чином, оскільки $\left|\sigma_{s\in S}\right|\geq 0$, $\left|\sigma_{s\in S\wedge w\in W}\right|\geq 0$ і $\left|\sigma_{s\in S\wedge t\in T}\right|\geq 0$ (випливає з означення реляційної операції селекції), то для доведення нерівності $Conf\left(s\in S\to w\in W\right)\geq Conf\left(s\in S\to t\in T\right)$ необхідно довести, що $\left|\sigma_{s\in S\wedge w\in W}\right|\geq\left|\sigma_{s\in S\wedge t\in T}\right|$.

Розглянемо змінну-кортеж x відношення $r(R)$, таку, що $\pi_s(x)\in S$ і $\pi_t(x)\in T$. Згідно умови $Conf\left(t\in T\to w\in W\right)=1$, якщо $\pi_T(x)\in T$, то $\pi_w(x)\in W$. Отже

$$Conf\left(s\in S\to t\in T\right)=p\wedge Conf\left(t\in T\to w\in W\right)=1:$$

$$:\forall x\in r(R):\pi_s(x)\in S\wedge\pi_t(x)\in T\Rightarrow\pi_w(x)\in W.$$

Звідси отримуємо ряд наслідків:

$$\sigma_{s\in S\wedge w\in W}\subseteq\sigma_{s\in S\wedge t\in T};$$

$$\left|\sigma_{s\in S\wedge w\in W}\right|\geq\left|\sigma_{s\in S\wedge t\in T}\right|;$$

$$Conf\left(s\in S\to w\in W\right)\geq Conf\left(s\in S\to t\in T\right)=p;$$

$$Conf\left(s\in S\to w\in W\right)\geq p.$$

Таким чином, транзитивність рівня довіри ІПЗ доведено.

Приклад 4:

З даних прикладу 2 можна побудувати такі ІПЗ:

$$Conf\left(C\in\{c_2\}\to B\in\{b_1\}\right)=\frac{1}{6};$$

$$Conf\left(B\in\{b_1\}\to D\in\{d_1\}\right)=1.$$

З правила транзитивності рівня довіри ІПЗ отримуємо, що $Conf\left(C\in\{c_2\}\to D\in\{d_1\}\right)\geq\frac{1}{6}$. Перевіримо це, обчисливши $Conf\left(C\in\{c_2\}\to D\in\{d_1\}\right)$ за формулою :

$$Conf\left(C\in\{c_2\}\to D\in\{d_1\}\right)=\frac{\left|\sigma_{C\in\{c_2\}\wedge D\in\{d_1\}}\right|}{\left|\sigma_{C\in\{c_2\}}\right|}=\frac{3}{6}=\frac{1}{2}.$$

$$Conf\left(C\in\{c_2\}\to D\in\{d_1\}\right)=\frac{1}{2}\geq\frac{1}{6}=$$

$$=Conf\left(C\in\{c_2\}\to B\in\{b_1\}\right).$$

Таким чином, підтверджується транзитивність рівня довіри ІПЗ.

Транзитивність рівня довіри ІПЗ є потужним правилом для висування різноманітних припущень та виконання доведень.

Псевдотранзитивність рівня довіри.

Дана аксіома виводу *F*-залежностей не має прямої альтернативи для ІПЗ, за умови накладення лише одного обмеження на вихідні залежності. Доведемо дане твердження.

Розглянемо залежності $Conf\left(s\in S\to t\in T\right)=p$ і $Conf\left(t\in T\wedge q\in Q\to w\in W\right)=1$, наклавши обмеження істинності на $Conf\left(t\in T\wedge q\in Q\to w\in W\right)$.

Позначимо $X=\sigma_{s\in S}(R)$, $Y=\sigma_{t\in T}(R)$, $Z=\sigma_{q\in Q}(R)$, $V=\sigma_{w\in W}(R)$. Тоді

$$Conf\left(t\in T\wedge q\in Q\to w\in W\right)=\frac{\left|Y\cap Z\cap V\right|}{\left|Y\cap Z\right|}=1;$$

$$\left|X\cap Z\cap V\right|=\left|X\cap Z\right|;$$

$$Conf\left(s\in S\to t\in T\right)=\frac{\left|X\cap Y\right|}{\left|X\right|}=p;$$

$$Conf\left(s\in S\wedge q\in Q\to w\in W\right)=\frac{\left|X\cap Z\cap V\right|}{\left|X\cap Z\right|}=$$

$$=\frac{\left|\left((X\cap Y)\cup(X\setminus Y)\right)\cap Z\cap V\right|}{\left|\left((X\cap Y)\cup(X\setminus Y)\right)\cap Z\right|}=$$

$$=\frac{\left|(X\cap Y\cap Z\cap V)\cup\left((X\setminus Y)\cap Z\cap V\right)\right|}{\left|(X\cap Y\cap Z)\cup\left((X\setminus Y)\cap Z\right)\right|}=$$

$$=\frac{\left|(X\cap Y\cap Z)\cup\left((X\setminus Y)\cap Z\cap V\right)\right|}{\left|(X\cap Y\cap Z)\cup\left((X\setminus Y)\cap Z\right)\right|}=$$

$$= \frac{|X \cap Y \cap Z| + |(X \setminus Y) \cap Z \cap V| - |X \cap Y \cap Z \cap (X \setminus Y) \cap Z \cap V|}{|X \cap Y \cap Z| + |(X \setminus Y) \cap Z|} =$$

$$= \frac{|X \cap Y \cap Z| + |(X \setminus Y) \cap Z \cap V|}{|X \cap Y \cap Z| + |(X \setminus Y) \cap Z|} = (1)$$

$$X \cap Y \cap Z \subset X \cap Z;$$

$$(X \setminus Y) \cap Z \cap V \subset X \cap Z;$$

$$(X \setminus Y) \cap Z \subset X \cap Z.$$

Таким чином, $X \cap Z$ є універсальною множиною U даних виразів і результат обчислення (1) не зміниться, якщо розглядати лише кортежі з $X \cap Z$. Позначимо $Y' = Y \cap (X \cap Z), V' = V \cap (X \cap Z)$. Тоді

$$(1) = \frac{|Y'| + |\neg Y' \cap V'|}{|Y'| + |\neg Y'|} = \frac{|Y'| + |\neg Y' \cap V'|}{|U|};$$

$$|\neg Y' \cap V'| \in [0; |\neg Y'|];$$

$$\frac{|Y'| + |\neg Y' \cap V'|}{|U|} \in \left[\frac{|Y'|}{|U|}; 1\right].$$

Повернемось до введених позначень: $Y' = Y \cap (X \cap Z)$. Початкові умови не накладають обмежень на значення даного виразу, тому $|Y'| \in [0; |U|]$ і відповідно

$$\frac{|Y'| + |\neg Y' \cap V'|}{|U|} \in [0;1];$$

$$Conf(s \in S \wedge q \in Q \to w \in W) \in [0;1].$$

Тобто, залежності $Conf(s \in S \to t \in T) = p$ і $Conf(t \in T \wedge q \in Q \to w \in W) = 1$ не роблять ніякого впливу на залежність $Conf(s \in S \wedge q \in Q \to w \in W)$.

Розглянемо тепер обмеження іншої залежності, комбінуючи $Conf(s \in S \to t \in T) = 1$ і $Conf(t \in T \wedge q \in Q \to w \in W) = p$. Використовуючи вищенаведені позначення, отримаємо:

$$Conf(t \in T \wedge q \in Q \to w \in W) = \frac{|Y \cap Z \cap V|}{|Y \cap Z|} = p;$$

$$Conf(s \in S \to t \in T) = \frac{|X \cap Y|}{|X|} = 1;$$

$$|X \cap Y| = |X|.$$

В даному випадку початкові залежності ніяк не обмежують отриманий вираз і це очевидно вже на першому кроці: множина V може не мати спільних кортежів з X, одразу перетворивши $Conf(s \in S \wedge q \in Q \to w \in W)$ в нуль. З іншої сторони, можливий варіант і $V \subset X \cap Z$ – тоді $Conf(s \in S \wedge q \in Q \to w \in W) = 1$. Функція $Conf(s \in S \wedge q \in Q \to w \in W)$ лінійно залежить від $|X \cap Z \cap V|$, маючи змінною множину V, таким чином маючи область значень $[0;1]$. Отже, залежності $Conf(s \in S \to t \in T) = 1$ і $Conf(t \in T \wedge q \in Q \to w \in W) = p$ не обмежують область значень $Conf(t \in T \wedge q \in Q \to w \in W)$.

Таким чином, доведено, що псевдотранзитивність функціональних залежностей не має прямого аналогу серед ІПЗ.

Як бачимо з вищеописаних викладень, більшість аксіом виведення F-залежностей можуть бути трансформовані для ІПЗ лише зі значними обмеженнями умовної частини однієї з залежностей. До того ж деякі аналоги не дають чіткої формули обчислення рівня довіри нової залежності, а тільки накладають на нього обмеження. Таким чином, наведена множина правил виведення ІПЗ є не повною. Для забезпечення повноти правил виводу необхідно ввести додаткові правила виведення, специфічні для ІПЗ.

ОПЕРАЦІЇ НАД ІПЗ

Факторизація.

Назвемо розклад залежності F_1: $\pi_{A_{i_1} A_{i_2} \ldots A_{i_k}}(s) \in \{s_1, s_2, \ldots s_m\} \to \pi_{A_{j_1} A_{j_2} \ldots A_{j_l}}(s) \in \{t_1, t_2, \ldots t_n\}$ на множину залежностей $\left\{\pi_{A_{i_1} A_{i_2} \ldots A_{i_k}}(s) = s_i \to \pi_{A_{j_1} A_{j_2} \ldots A_{j_l}}(s) = t_j\right\}$, $i = \overline{1..m}, j = \overline{1..n}$ факторизацією і позначатимемо $F_I[Fact]$.

$$F_I = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\pi_{A_{i_1} A_{i_2} \ldots A_{i_k}}(s) = s_i \to \pi_{A_{j_1} A_{j_2} \ldots A_{j_l}}(s) = t_j\right). \quad (12)$$

Об'єднання.

Об'єднанням ІПЗ $s \in S_1 \to t \in T_1$ і $s \in S_2 \to t \in T_2$ назвемо нову ІПЗ $s \in S_1 \cup S_2 \to t \in T_1 \cup T_2$ (використовуються позначення ІПЗ в термінах реляційної алгебри (4)).

$$(s \in S_1 \to t \in T_1) + (s \in S_2 \to t \in T_2) = s \in S_1 \cup S_2 \to t \in T_1 \cup T_2. \quad (13)$$

Розглянемо властивості операції об'єднання ІПЗ.

Комутативність.

Операція об'єднання ІПЗ володіє властивістю комутативності.

$$(s \in S_1 \to t \in T_1) + (s \in S_2 \to t \in T_2) = (s \in S_2 \to t \in T_2) + (s \in S_1 \to t \in T_1). \quad (14)$$

Доведення:

$$(s \in S_1 \to t \in T_1) + (s \in S_2 \to t \in T_2) = (s \in S_1 \cup S_2 \to t \in T_1 \cup T_2) = (s \in S_2 \cup S_1 \to t \in T_2 \cup T_1) = (s \in S_2 \to t \in T_2) + (s \in S_1 \to t \in T_1).$$

Асоціативність.

Операція об'єднання ІПЗ володіє властивістю асоціативності.

$$(s \in S_1 \to t \in T_1) + ((s \in S_2 \to t \in T_2) + (s \in S_3 \to t \in T_3)) = ((s \in S_1 \to t \in T_1) + (s \in S_2 \to t \in T_2)) + (s \in S_3 \to t \in T_3). \quad (15)$$

Доведення:

$$(s \in S_1 \to t \in T_1) + ((s \in S_2 \to t \in T_2) + (s \in S_3 \to t \in T_3)) = (s \in S_1 \to t \in T_1) + (s \in S_2 \cup S_3 \to t \in T_2 \cup T_3) = s \in S_1 \cup S_2 \cup S_3 \to t \in T_1 \cup T_2 \cup T_3.$$

Розкладемо праву частину виразу асоціативності:

$$((s \in S_1 \to t \in T_1) + (s \in S_2 \to t \in T_2)) + (s \in S_3 \to t \in T_3) = (s \in S_1 \cup S_2 \to t \in T_1 \cup T_2) + (s \in S_3 \to t \in T_3) = s \in S_1 \cup S_2 \cup S_3 \to t \in T_1 \cup T_2 \cup T_3 =$$

$$(s \in S_1 \to t \in T_1) + ((s \in S_2 \to t \in T_2) + (s \in S_3 \to t \in T_3)).$$

ПРАВИЛА ВИВЕДЕННЯ ІПЗ

Як було показано, застосування трансформованих правил виведення функціональних залежностей недостатньо для забезпечення повноти множини правил виведення ІПЗ. Розглянемо правила виведення, специфічні для ІПЗ, що дозволять побудувати ефективні алгоритми пошуку цих залежностей в наборах даних.

Агрегування області визначення. Якщо наявні ІПЗ $s \in S_1 \to t \in T_1$ і $s \in S_2 \to t \in T_2$ та значення $\sigma_{s=s_i}(r(R))$, $\sigma_{t=t_j}(r(R))$, $\sigma_{s=s_i \wedge t=t_j}(r(R))$ такі, що $\bigcup_i s_i = S_1 \cup S_2$, $\bigcup_j s_j = T_1 \cup T_2$ то для ІПЗ

a) $h > 1$ $s \in S_1 \cup S_2 \to t \in T_1 \cup T_2$;

b) $s \in S_1 \cup S_2 \to t \in T_1 \cap T_2$;

c) $s \in S_1 \cap S_2 \to t \in T_1 \cup T_2$;

d) $s \in S_1 \cap S_2 \to t \in T_1 \cap T_2$;

e) $s \in S_1 \to t \in T_1 \cup T_2$;

f) $s \in S_1 \to t \in T_1 \cap T_2$;

g) $s \in S_2 \to t \in T_1 \cup T_2$;

h) $s \in S_2 \to t \in T_1 \cap T_2$;

i) $s \in S_1 \cup S_2 \to t \in T_1$;

j) $s \in S_1 \cap S_2 \to t \in T_1$;

k) $s \in S_1 \cup S_2 \to t \in T_2$;

l) $s \in S_1 \cap S_2 \to t \in T_2$.

можна обчислити усі параметри з формул (5)–(11).

Доведення:

$$\sigma_{x \in X}(r) = \sum_{z \in X} \sigma_{x=z}(r). \quad (16)$$

Розглянемо доведення найскладнішого (першого) з наведених наслідків. Інші обчислюються аналогічно і не подаються тут для лаконічності.

$$Sup(s \in S_1 \cup S_2 \to t \in T_1 \cup T_2) = Sup(s \in S_1 \cup S_2 \wedge t \in T_1 \cup T_2) = \frac{|\sigma_{s \in S_1 \cup S_2 \wedge t \in T_1 \cup T_2}(r)|}{|r|} = \frac{|\sigma_{s \in S_1 \wedge t \in T_1 \cup T_2}(r)| + |\sigma_{s \in S_2 \wedge t \in T_1 \cup T_2}(r)| - |\sigma_{s \in S_1 \cap S_2 \wedge t \in T_1 \cup T_2}(r)|}{|r|} =$$

$$= \frac{1}{|r|}(|\sigma_{s \in S_1 \wedge t \in T_1}(r)| + |\sigma_{s \in S_1 \wedge t \in T_2}(r)| - |\sigma_{s \in S_1 \wedge t \in T_1 \cap T_2}(r)| + |\sigma_{s \in S_2 \wedge t \in T_1}(r)| + |\sigma_{s \in S_2 \wedge t \in T_2}(r)| - |\sigma_{s \in S_2 \wedge t \in T_1 \cap T_2}(r)| - |\sigma_{s \in S_1 \cap S_2 \wedge t \in T_1}(r)| - |\sigma_{s \in S_1 \cap S_2 \wedge t \in T_2}(r)| + |\sigma_{s \in S_1 \cap S_2 \wedge t \in T_1 \cap T_2}(r)|). \quad (17)$$

$$Conf(s \in S_1 \cup S_2 \to t \in T_1 \cup T_2) = \frac{|\sigma_{s \in S_1 \cup S_2 \wedge t \in T_1 \cup T_2}(r)|}{|\sigma_{s \in S_1 \cup S_2}(r)|} = \frac{|r| \cdot Sup(s \in S_1 \cup S_2 \to t \in T_1 \cup T_2)}{|\sigma_{s \in S_1}(r)| + |\sigma_{s \in S_2}(r)| - |\sigma_{s \in S_1 \cap S_2}(r)|}. \quad (18)$$

$$Imp(s \in S_1 \cup S_2 \to t \in T_1 \cup T_2) = \frac{Conf(s \in S_1 \cup S_2 \to t \in T_1 \cup T_2)}{Sup(t \in T_1 \cup T_2)} = \frac{|r| \cdot Conf(s \in S_1 \cup S_2 \to t \in T_1 \cup T_2)}{|\sigma_{t \in T_1}(r)| + |\sigma_{t \in T_2}(r)| - |\sigma_{t \in T_1 \cap T_2}(r)|}. \quad (19)$$

$$I_{s\in S_1\cup S_2\leftrightarrow t\in T_1\cup T_2} = \sum_{x\in S_1\cup S_2}\sum_{y\in T_1\cup T_2} Sup\left(s=x\to t=y\right)\times$$

$$\times \log_2 Imp\left(s=x\to t=y\right). \qquad (20)$$

Складність обчислення формул (16)–(20) прямо лінійно залежить лише від потужності множин S_1 та S_2. До того ж існують структури даних (Fibonacci heap, binomial heap [7]), що дозволяють обчислювати об'єднання та перетин множин за сублінійний час.

Позначатимемо операцію агрегування ІПЗ $\overset{x}{\oplus}$, де x – тип агрегування. Оскільки найчастіше використовуваним агрегуванням буде форма, то для її позначення використовуватимемо спрощене позначення $\oplus$ (без вказування форми агрегації).

Реалізація алгоритмів виведення ІПЗ для правила агрегування може використовувати підтипи та часткові випадки правила агрегування. Наприклад, якщо $S_1\cap S_2=\varnothing$, формули (16)–(20) стають зовсім простими і обчислюються з асимптотичною складністю $O(1)$.

Дані правила (наведене правило агрегування ІПЗ включає 12 підправил, поданих пунктами наслідків) виведення особливо ефективне для даних з невеликими множинами значень, наприклад даних соціологічних та психологічних опитувань, спостережень погоди, досліджень транспортних потоків та ін.

ПОВНОТА ПРАВИЛ ВИВЕДЕННЯ ІПЗ

Розглянемо довільну ІПЗ $F_1: \bigvee_{k=1}^{h} P_k^e \to \bigvee_{l=1}^{g} Q_l^e$. Її можна отримати агрегуванням ІПЗ $P_h^e\to Q_g^e$ з $\bigvee_{k=1}^{h-1} P_k^e \to \bigvee_{l=1}^{g-1} Q_l^e$ якщо $h>1$ і $g>1$, з $P_h^e \to \bigvee_{l=1}^{g-1} Q_l^e$, якщо $h=1$ і $g>1$, з $\bigvee_{k=1}^{h-1} P_k^e \to Q_g^e$, якщо $h>1$ і $g=1$.

Таким чином, отримано розклад

$$\bigvee_{k=1}^{h} P_k^e \to \bigvee_{l=1}^{g} Q_l^e = \bigoplus_{i=0..\max(h-1,g-1)} \left(P^e_{\max(h-i,1)} \to Q^e_{\max(g-i,1)}\right). \qquad (21)$$

З формули (1)

$$P_k^e = A_{i_1}\in\left\{a_{(i_1)(j_1)}\right\}\wedge A_{i_2}\in\left\{a_{(i_2)(j_2)}\right\}\wedge A_{i_k}\in\left\{a_{(i_k)(j_k)}\right\}.$$

Отже, параметри довільної ІПЗ можна обчислити, маючи статистику $\sigma_{x=x_i}\left(r(R)\right), \sigma_{x=x_i\wedge y=y_j}\left(r(R)\right)$, $\sigma_{x=x_i\wedge y=y_j\wedge z=z_k}\left(r(R)\right)$ і т.д. Тобто для представлення усіх ІПЗ відношення, в яких є не більше k частин умов предикатів, необхідно

$$O\left(\max_{(i_1,i_2,\ldots,i_k)\in Z^k}\left(\left|class\left(A_{i_1}\right)\right|\cdot\left|class\left(A_{i_2}\right)\right|\cdot\ldots\cdot\left|class\left(A_{i_k}\right)\right|\right)\right)$$

пам'яті, де $Z\subset R$ – множина атрибутів, між якими шукаються залежності, $\left|class\left(A_{i_j}\right)\right|$ – кількість областей класифікації за атрибутом A_{i_j}. Найпростіший варіант: $class\left(A_{i_j}\right)=dom\left(A_{i_j}\right)$, проте для числових чи вимірювальних даних часто зручно розбивати їх на під області. Це збільшує інформативність знайдених залежностей та спрощує їх пошук.

Ведення повної статистики довільної глибини вкладення звичайно є неможливим через обмеження наявної пам'яті обчислювальної системи, але в реальності практично не використовуються залежності з більш, ніж 3–4 частинами умовного предикату. Відповідно, представлення усіх необхідних даних цілком можливе навіть для дуже великих масивів даних.

ВИСНОВКИ

Дана стаття описує результати досліджень властивостей та правил виведення ймовірнісних продукційних залежностей – класу асоціативних залежностей, що широко застосовується у аналізі даних комп'ютерних наук, енергетики, фізики, соціології та ін.

Основним результатом досліджень є правила виведення нових ІПЗ з деякої їх множини. Це дозволяє зберігати лише мінімальне покриття набору даних обраним класом залежностей, а не усі наявні залежності у відношенні. Така форма представлення даних таких дозволяє їх легко модифікувати (видаляти чи додавати кортежі, а також змінювати значення атрибутів існуючих кортежів). Виявлення цієї властивості ІПЗ дає важливу перевагу над багатьма методами аналізу статичних даних – при зміні даних не потрібно повністю перераховувати всю статистику даних, а лише оновити необхідні параметри.

У статті доводиться можливість ефективного обчислення таких характеристик, як рівень підтримки, рівень довіри, рівень покращення та повна взаємна інформація ІПЗ. Проте це далеко не всі параметри, що можуть бути ефективно обчислені з використанням вивчених властивостей та знайдених правил виведення ІПЗ. У наступних роботах планується глибше розглянути необхідні умови, яким повинен відповідати критерій якості, щоб його можна було ефективно обчислювати, застосовуючи правила виведення ІПЗ.

Застосування правил виведення ІПЗ, описаних у даній статті дозволяє зменшити необхідний обсяг дискового простору обчислювальної системи до

$$O\left(\max_{(i_1,i_2,\ldots,i_k)\in Z^k}\left(\left|class\left(A_{i_1}\right)\right|\cdot\left|class\left(A_{i_2}\right)\right|\cdot\ldots\cdot\left|class\left(A_{i_k}\right)\right|\right)\right),$$

де k – максимальна кількість атрибутів, що фігурує в умовній та результуючій частині шуканих ІПЗ. Зазвичай немає потреби в значеннях k>3, а якщо й виникає, то лише для деяких специфічних значень фіксованих атрибутів і тоді стає можливим зберігати окрему статистику для таких атрибутів.

Отже, правила виведення, описані у даній статті, дозволяють ефективно зберігати та знаходити ІПЗ у великих наборах даних, ґрунтуючись на мінімальному покритті ІПЗ.

СПИСОК ЛІТЕРАТУРИ

1. *Чесноков, С. В.* Детерминационный анализ социально-экономических даннях / С. В. Чесноков. – М. : Наука, 1982. – 168 с.
2. *Тітова, О. В.* Методи побудови та оцінки агрегованих асоціативних правил в інтелектуальних базах даних. Харків – 2006.
3. Головний сайт департаменту патології, UT Southwestern Medical Center [Електронний ресурс] – режим доступу http://pathcuric1.swmed.edu/pathdb/classifi.html
4. Опис утиліти BiNGO, сайт університету Гент, [Електронний ресурс] – режим http://www.psb.ugent.be/cbd/papers/BiNGO/Home.html.
5. Офіційний сайт National Institute of Allergy and Infectious Diseases (NIAID), NIH [Електронний ресурс] – режим доступу http://david.abcc.ncifcrf.gov/content.jsp?file=/ease/ease1.htm&type=1
6. Thun S. Laplacian smoothing / Norwig P., Thun S. Online lecture «Machine Learning», USA, Stanford University – 2011. https://www.ai-class.com/course/video/quizquestion/97.
7. Heaps: Heapsort, Binary Heap, Smoothsort, Soft Heap, Fibonacci Heap, Treap, Binomial Heap, Pairing Heap, Leftist Tree, Skew Heap. Memphis, Tennessee, Llc Books , General Books LLC – 2010, 74 p.

Стаття надійшла до редакції 12.03.2012.

Пшеничный А. Ю.

СВОЙСТВА АССОЦИАТИВНЫХ ЗАВИСИМОСТЕЙ В АНАЛИЗЕ ДАННЫХ

В данной работе поданы результаты исследований свойств ассоциативных зависимостей и возможностей их эффективного агрегирования. Разработан метод поиска ассоциативных зависимостей широкого класса в больших наборах данных.

Ключевые слова: ассоциативная зависимость, функциональная зависимость, зависимости данных, анализ данных.

Pshenychnyi O. Y.

ASSOCIATIVE DEPENDENCIES PROPERTIES IN DATA ANALYSIS

This paper describes the results of research in the field of associative dependencies properties and effective aggregation possibilities. Also it briefly describes the developed method of special class of associative dependencies detection in large data volumes. The main idea of this research is aggregation of elementary associative dependencies into more complicated once. This approach gives good performance results and allows processing data volumes with millions records. Current paper shows how it is possible to define algebra of associative dependencies with few main operations and rules of inference, taking place in such algebra. The rule set completeness is also proven here to be sure that no rules are lost during inference. The outcome of described theory is highly effective data analysis method, capable to detect wide range of associative dependencies in relational data.

Key words: associative dependency, functional dependency, data dependency, data analysis.

REFERENCES

1. Chesnokov S.V. Determinatsyonnyi analiz sotsyalno-ekonomicheskikh dannyh. Moskva, Nauka, 1982, 168 p.
2. Titova O.V. Metody pobudovy ta otsinky ahrehovanyh asotsiatyvnykh pravyl v intelektualnykh bazah danykh. Kharkiv, 2006.
3. The main site of the Department of Pathology, UT Southwestern Medical Center. http://pathcuric1.swmed.edu/pathdb/classifi.html
4. BiNGO utility description, Ghent university site, http://www.psb.ugent.be/cbd/papers/BiNGO/Home.html.
5. Site of National Institute of Allergy and Infectious Diseases (NIAID), NIH, http://david.abcc.ncifcrf.gov/content.jsp?file=/ease/ease1.htm&type=1.
6. Thun S. Laplacian smoothing / Norwig P., Thun S. Online lecture «Machine Learning», USA, Stanford University – 2011. https://www.ai-class.com/course/video/quizquestion/97.
7. Heaps: Heapsort, Binary Heap, Smoothsort, Soft Heap, Fibonacci Heap, Treap, Binomial Heap, Pairing Heap, Leftist Tree, Skew Heap. Memphis, Tennessee, Llc Books , General Books LLC, 2010, 74 p.

УДК 519.6

Чопоров С. В.[1], Гоменюк С. И.[2], Лисняк А. А.[3], Панасенко Е. В.[4]

[1]Канд. техн. наук, старший преподаватель Запорожского национального университета
[2]Д-р техн. наук, старший преподаватель профессор Запорожского национального университета
[3, 4] Канд. физ.-мат. наук, старший преподаватель Запорожского национального университета

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ НЕКОТОРЫХ КРЕПЕЖНЫХ СОЕДИНЕНИЙ НА БАЗЕ ТЕОРИИ R-ФУНКЦИЙ

В статье рассмотрена проблема математического моделирования сложных геометрических объектов на базе теории R-функций. Предложены новые математические модели наиболее распространенных гаечных и болтовых соединений.

Ключевые слова: математическая модель, R-функция, гайка, болт.

ПОСТАНОВКА ПРОБЛЕМЫ

Одним из наиболее динамично развивающихся направлений современной науки и техники является компьютерное моделирование сложных технических объектов и процессов, позволяющее заменить дорогостоящее и продолжительное исследование испытательного образца вычислительным экспериментом. При этом для практического применения многих вычислительных методов, как правило, необходимо построение математических