

НЕЙРОІНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ

NEUROINFORMATICS AND INTELLIGENT SYSTEMS

УДК 004.8:004.032.26

ШВИДКА НЕЙРОННА МЕРЕЖА ТА ЇЇ АДАПТИВНЕ НАВЧАННЯ В ЗАДАЧАХ КЛАСИФІКАЦІЇ

Бодяньський Є. В. – д-р техн. наук, професор, професор кафедри штучного інтелекту, Харківський національний університет радіоелектроніки, Харків, Україна.

Шафроненко Є. О. – асистент кафедри медіаінженерії та інформаційних радіоелектронних систем, Харківський національний університет радіоелектроніки, Харків, Україна.

Бродецький Ф. А. – старший викладач кафедри інформатики, Харківський національний університет радіоелектроніки, Харків, Україна.

Таняньський О. С. – аспірант кафедри інформатики, Харківський національний університет радіоелектроніки, Харків, Україна.

АНОТАЦІЯ

Актуальність. Для вирішення широкого класу задач обробки інформації і, перш за все, розпізнавання образів за умов суттєвої нелінійності широке розповсюдження одержали штучні нейронні мережі, завдяки своїм універсальним апроксимуючим властивостям та здатності до навчання на основі тренувальних навчальних вибірок. Найбільшого розповсюдження отримали глибокі нейронні мережі, які дійсно демонструють дуже високу якість розпізнавання, але потребують надвеликих обсягів навчальних даних, які не завжди є доступними. За цих умов ефективними можуть бути, так звані, машини опорних векторів найменших квадратів, які не потребують великих обсягів навчальних вибірок, однак можуть навчатися лише у пакетному режимі і є достатньо громіздкими у чисельній реалізації. Тому достатньо актуальною є задача навчання LS-SVM у послідовному режимі за умов суттєвої нестационарності даних, що послідовно у онлайн режимі надходять на опрацювання у нейронну мережу.

Мета. Мета роботи полягає у запровадженні підходу до адаптивного навчання LS-SVM, що дозволяє відмовитися від перетворення зображень у векторні сигнали.

Метод. Запропоновано підхід для розпізнавання образів-зображень за допомогою машини опорних векторів найменших квадратів (LS-SVM) за умов, коли дані на обробку надходять у послідовному онлайн режимі. Перевагою запропонованого підходу є скорочення часу вирішення задачі розпізнавання образів-зображень, а також дозволяє реалізувати процес навчання на нестационарних тренувальних вибірках. Особливістю запропонованого методу є обчислювальна простота і висока швидкість, пов'язана з тим, що кількість нейронів у мережі не змінюється з часом, тобто архітектура залишається фіксованою у процесі налаштування.

Результати. Запропонований підхід до адаптивного навчання LS-SVM спрощує чисельну реалізацію нейронної мережі та дозволяє підвищити швидкість обробки інформації і, перш за все, налаштування її синаптичних ваг.

Висновки. Розглянута задача розпізнавання образів-зображень за допомогою машини опорних векторів найменших квадратів (LS-SVM) за умов, коли дані на обробку надходять у послідовному онлайн режимі. Процес навчання реалізується на ковзному вікні, що веде до того, що кількість нейронів у мережі не змінюється з часом, тобто архітектура залишається фіксованою у процесі налаштування. Такий підхід спрощує чисельну реалізацію системи та дозволяє реалізувати процес навчання на нестационарних тренувальних вибірках. Розглянута можливість навчання у ситуаціях, коли навчальні образи задані не лише у векторній формі, а й матричній, що дозволяє відмовитися від перетворення зображень у векторні сигнали.

КЛЮЧОВІ СЛОВА: адаптивне навчання, класифікація, швидка нейронна мережа, машина опорних векторів.

АБРЕВІАТУРИ

SVM – машина опорних векторів;

LS-SVM – машина опорних векторів найменших квадратів.

НОМЕНКЛАТУРА

w – вектор синаптичних ваг;

k – номер вектору-спостереження;

$x(k)$ – вектор-спостереження;

$\hat{y}(x)$ – вихідний сигнал мережі;

$\phi(x)$ – вектор, утворений ядерними дзвонуватими активаційними функціями;

$\phi_j(x)$ – значення j -ї ядерної дзвонуватої активаційної функції;

w_0 – пороговий параметр вибірки;

w_k – значення k -го вектора-спостереження;

$x(l)$ – вхідні вектори;

$(n \times 1)$ – вектор вхідних сигналів в мережу;

$\lambda(l)$ – множники Лагранжа;

E – цільова функція;

L^{SV} – квадратичний критерій;

$e(l)$ – середня похибка;

$y(l)$ – зовнішній навчальний сигнал;

$(\bullet)^+$ – символ псевдообернення за Муром – Пенроузом;

γ – параметр регуляризації;

I_{kk} – одинична $(k \times k)$ матриця;

I_k – $(k \times 1)$ – вектор, утворений одиницями;

$\Lambda(k)$ – вектор, утворений множниками Лагранжа;

$Y(k)$ – вектор, утворений вихідними сигналами мережі;

$\Omega(k)$ – активаційна ядерна функція;

σ^2 – параметр рецепторного поля активаційної функції;

$Sp(\bullet)$ – символ сліду матриці;

K – традиційний гауссіан;

$P(k)$ – обернення блочних матриць.

ВСТУП

Проблема класифікації – розпізнавання образів, у тому числі зображень, є однією з найважливіших проблем Data Mining, Data Stream Mining, Big Data Mining. В даний час існує багато підходів для її вирішення. На сьогодні найбільш ефективними з них є Deep Neural Networks, які демонструють дійсно вражаючі результати саме в задачах обробки зображень довільної природи. Водночас ці мережі мають ряд істотних недоліків, які обмежують їх використання, особливо в задачах, коли дані надходять на обробку послідовно, спостереження за спостереженням, можливо, в реальному часі. Це пояснюється тим, що глибокі мережі є досить повільними системами, які навчаються за допомогою зворотного поширення помилок протягом багатьох епох у пакетному (мініпакетному) режимі, що займає багато часу та вимагає великих обсягів навчальних даних, що не завжди підходить для реальних програм.

За цих умов ефективними можуть бути, так звані, машини опорних векторів найменших квадратів [1–3], які не потребують великих обсягів навчальних вибірок, однак можуть навчатися лише у пакетному режимі і є достатньо громіздкими у чисельній реалізації. Тому достатньо актуальною є задача навчання LS-SVM у послідовному режимі за умов суттєвої неста-

ціонарності даних, що послідовно у онлайн режимі надходять на опрацювання у нейронну мережу.

Об'єкт дослідження. Розпізнавання образів-зображень за допомогою машини опорних векторів найменших квадратів (LS-SVM) за умов, коли дані на обробку надходять у послідовному онлайн режимі.

Предмет дослідження. Підхід до адаптивного навчання LS-SVM.

Мета роботи полягає у запровадженні підходу до адаптивного навчання LS-SVM, що дозволяє відмовитися від перетворення зображень у векторні сигнали.

1 ПОСТАНОВКА ЗАВДАННЯ

Як вже відзначалося, навчання традиційних машин опорних векторів є достатньо громіздким з обчислювальної точки зору і пов'язане з вирішенням задачі нелінійного програмування з обмеженнями, а розмірність цієї задачі визначається обсягом навчальної вибірки. Ще раз підкреслимо, що навчання відбувається у пакетному офлайн режимі.

Нелінійне перетворення, що реалізуються нейронною мережею опорних векторів має вигляд

$$\hat{y}(x) = w^T \phi(x) + w_0,$$

де $w = (w_1, \dots, w_l, \dots, w_k)^T$ – вектор синаптичних ваг, розрахований на основі навчальної вибірки $x(1), \dots, x(l), \dots, x(k)$; $x(l) = (x_1(l), \dots, x_i(l), \dots, x_n(l))^T$ – $(n \times 1)$ – вектор вхідних сигналів у мережу; $\phi(x) = (\phi_1(x), \dots, \phi_l(x), \dots, \phi_k(x))^T$ – вектор, утворений ядерними дзвонуватими активаційними функціями, центри яких визначаються вхідними векторами $x(l), l = 1, \dots, k$.

2 ОГЛЯД ЛІТЕРАТУРИ

Для вирішення широкого класу задач обробки інформації і, перш за все, розпізнавання образів за умов суттєвої нелінійності широке розповсюдження одержали штучні нейронні мережі [4], завдяки своїм універсальним апроксимуючим властивостям та здатності до навчання на основі тренувальних навчальних вибірок. Тут найбільше розповсюдження отримали глибокі нейронні мережі і, перш за все, згорткові мережі [5], які дійсно демонструють дуже високу якість розпізнавання, але потребують надвеликих обсягів навчальних даних, які не завжди є доступними при вирішенні реальних конкретних задач. Для розглянутої задачі ефективним може бути використання опорної векторної машини (SVM) [6–8], яка оптимізує емпіричний критерій навчання ризику та коригує його параметри як на основі традиційного навчання під керівництвом, так і на основі «нейронів у точках даних» поняття [9]. Навчання SVM можна значно прискорити та звести до розв'язування систем лінійних рівнянь за допомогою так званих опорних векторних машин найменших квадратів (LS-SVM) [10].

У загальному випадку машини опорних векторів є окремим класом нейронних мереж, заснованих на мінімізації, так званого, емпіричного ризику [6] та налаштовуються у режимі контрольованого навчання, при цьому центри їх активаційних функцій розташовуються за принципом «Нейрони в точках даних». Ключовим моментом тут є, так звані, опорні (крайні) вектори, що формують достатньо компакту множину найбільш інформативних спостережень з навчальних даних. Слід також відмітити, що «класичні» SVM навчаються у пакетному режимі, при цьому навчальна вибірка повинна бути сформована заздалегідь.

3 МАТЕРІАЛИ І МЕТОДИ

При навчанні штучних нейронних мереж найчастіше використовується традиційний квадратичний критерій

$$E^{LS}(k) = \sum_{l=1}^k e^2(l),$$

де $e(l) = y(l) - \hat{y}(x(l))$, $l = 1, 2, \dots, k$.

Мінімізація цього критерія веде до стандартної оцінки найменших квадратів

$$L^{SV} = (w, w_0, e, \lambda) = E^{SV}(k) + \sum_{l=1}^k \lambda(l) (y(l) - w^T \varphi(x(l)) - w_0 - e(l)) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{l=1}^k \lambda(l) (y(l) - w^T \varphi(x(l)) - w_0 - e(l)),$$

при цьому в процесі оптимізації повинні бути знайдені не лише синаптичні ваги w, w_0 , але й невизначені множники Лагранжа $\lambda(l)$, $l = 1, 2, \dots, k$.

Система рівнянь Куна-Такера для введеної функції Лагранжа має вигляд

$$\begin{cases} \nabla_w L^{SV}(w, w_0, e, \lambda) = w - \sum_{l=1}^k \lambda(l) \varphi(x(l)) = \bar{0}, \\ \frac{\partial L^{SV}(w, w_0, e, \lambda)}{\partial w_0} = -\sum_{l=1}^k \lambda(l) = 0, \\ \frac{\partial L^{SV}(w, w_0, e, \lambda)}{\partial e(l)} = \gamma e(l) - \lambda(l) = 0, \\ \frac{\partial L^{SV}(w, w_0, e, \lambda)}{\partial \lambda(l)} = y(l) - w^T \varphi(x(l)) - w_0 - e(l) = 0 \end{cases}$$

(тут $\bar{0} = (k \times 1)$ – вектор утворений нулями), або

$$\begin{cases} w = \sum_{l=1}^k \lambda(l) \varphi(x(l)), \\ \sum_{l=1}^k \lambda(l) = 0, \\ \lambda(l) = \gamma e(l), \\ y(l) = w^T \varphi(x(l)) - w_0 - e(l) = 0. \end{cases}$$

$$w^{LS} = \left(\sum_{l=1}^k \varphi(x(l)) \varphi^T(x(l)) \right)^+ \varphi(x(l)) y(l).$$

Значно кращі результати можуть бути отримані на основі критерія емпіричного ризику

$$E^{SV}(k) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{l=1}^k e^2(l)$$

за наявності k лінійних обмежень-рівностей

$$\begin{cases} y(1) = w^T \varphi(x(1)) + w_0 + e(1), \\ \dots \\ y(l) = w^T \varphi(x(l)) + w_0 + e(l), \\ \dots \\ y(k) = w^T \varphi(x(k)) + w_0 + e(k). \end{cases}$$

При навчанні у пакетному режимі оцінювання синаптичних ваг машини опорних векторів найменших квадратів пов'язане із знаходженням сідлової точки функції Лагранжа

Нескладно помітити, що синаптичні ваги повністю визначаються невизначеними множниками Лагранжа (перше рівняння системи), а сама ця система може бути переписана у компактній векторно-матричній формі

$$\begin{pmatrix} 0 & I_k^T \\ I_k & \Omega(k) + \gamma^{-1} I_{kk} \end{pmatrix} \begin{pmatrix} w_0 \\ \Lambda(k) \end{pmatrix} = \begin{pmatrix} 0 \\ Y(k) \end{pmatrix},$$

де $\Lambda(k) = (\lambda(1), \dots, \lambda(l), \dots, \lambda(k))^T$; $Y(k) = (y(1), \dots, y(l), \dots, y(k))^T$, $\Omega(k) = \{\Omega_{ij} = \varphi^T(x(i)) \varphi(x(j)) = K(x(i), x(j))\}$ – активаційна ядерна функція, найчастіше традиційний гауссіан у формі

$$K(x(i), x(j)) = e^{-\frac{\|x(i) - x(j)\|^2}{2\sigma^2}}.$$

Тут цікаво відмітити, що вхідним сигналом у мережу традиційно є $(n \times 1)$ – вектор $x(l)$, при цьому, якщо вирішується задача розпізнавання образів-зображень, це зображення що має форму $(n_1 \times n_2)$ – матриці, попередньо повинно бути перетворено у $(n \times 1)$ – вектор за допомогою операцій згортки та субдискретизації.

Використання гауссіанів в якості активаційних функцій у LS-SVM дозволяє використовувати в якості

вхідного сигналу безпосередньо матрицю зображення, при цьому

$$K(x(i), x(j)) = e^{-\frac{Sp(x(i)-x(j))(x(i)-x(j))^T}{2\sigma^2}},$$

при цьому в якості відстані використовується не традиційна евклідова відстань, а матрична метрика Фробеніуса, що є узагальненням евклідової метрики на матричний випадок.

Таким чином, навчання LS-SVM пов'язане з вирішенням систем рівнянь

$$\begin{pmatrix} w_0 \\ \Lambda(k) \end{pmatrix} = \begin{pmatrix} 0 & I_k^T \\ I_k & \Omega(k) + \gamma^{-1} I_{kk} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ Y(k) \end{pmatrix} = P(k) \begin{pmatrix} 0 \\ Y(k) \end{pmatrix}.$$

Ця система описує процес навчання у пакетному офлайн режимі, коли вся навчальна вибірка сформована заздалегідь і не змінюється з часом.

$$P(k+1) = \begin{pmatrix} P(k) + P(k)\bar{K}(x(i)x(k+1)) \bullet & \bullet \bar{K}^T(x(i)x(k+1))P(k) \bullet & -(P(k)\bar{K}(x(i)x(k+1))) \bullet \\ \bullet (1 + \gamma^{-1} - \bar{K}^T(x(i)x(k+1))) \bullet & \bullet (1 + \gamma^{-1} - \bar{K}^T(x(i)x(k+1))) \bullet & \bullet (1 + \gamma^{-1} - \bar{K}^T(x(i)x(k+1))) \bullet \\ \bullet P(k)\bar{K}(x(i)x(k+1))^{-1} \bullet & \bullet P(k)\bar{K}(x(i)x(k+1))^{-1} \bullet & \bullet P(k)\bar{K}(x(i)x(k+1))^{-1} \bullet \\ -(\bar{K}^T(x(i)x(k+1))P(k)) \bullet & (1 + \gamma^{-1} - \bar{K}^T(x(i)x(k+1))) \bullet & \bullet P(k)\bar{K}(x(i)x(k+1))^{-1} \bullet \\ \bullet (1 + \gamma^{-1} - \bar{K}^T(x(i)x(k+1))) \bullet & \bullet P(k)\bar{K}(x(i)x(k+1))^{-1} \bullet & \bullet P(k)\bar{K}(x(i)x(k+1))^{-1} \bullet \end{pmatrix}.$$

Із зростанням навчальної вибірки при великих k використання цієї формули, яка хоча і видається дещо громіздкою, спрощує процес обернення матриць великої розмірності.

Із зростанням обсягів навчальної вибірки при великих k процес навчання LS-SVM стає громіздким, тому може потребувати досить багато часу. Крім того, у нестационарних ситуаціях, коли характеристики навчальної вибірки змінюються з часом, доцільно організувати «забування» застарілої інформації. При цьому можна організувати процес навчання на «ковзному вікні», коли з надходженням нового спостереження виключається одне застаріле. При цьому при надходженні нового $(k+1)$ -го спостереження із нейронної мережі вилучається $(k-s)$ -та ядерна активаційна функція (тут s – розмір ковзного вікна, що включає в себе тільки значущі спостереження). Таким чином, у системі фіксується кількість активаційних функцій, яка визначається розміром вікна s .

Введемо далі у розгляд $(s \times s)$ – матрицю ядерних активаційних функцій

Нескладно також організувати цей процес у онлайн режимі, коли у вже навчену на k спостереженнях систему надходить $(k+1)$ -ше спостереження $y(k+1)$.

Тоді для $(k+1)$ -го відліку можна записати [11]:

$$\begin{pmatrix} w_0 \\ \Lambda(k+1) \end{pmatrix} = \begin{pmatrix} 0 & I_{k+1}^T \\ I_k & \Omega(k+1) + \gamma^{-1} I_{k+1,k+1} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ Y(k+1) \end{pmatrix},$$

Звідки

$$\begin{pmatrix} 0 \\ Y(k) \\ y(k+1) \end{pmatrix} = \begin{pmatrix} P^{-1}(k) & \bar{K}(x(i), x(k+1)) \\ \bar{K}^T(x(i), x(k+1)) & 1 + \gamma^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \bar{Y}(k) \\ y(k+1) \end{pmatrix},$$

де

$$\bar{K}^T(x(i), x(k+1)) = (1, K(x(1), x(k+1)), \dots, K(x(k), x(k+1)))^T,$$

$\bar{Y}(k) = (0, Y^T(k))$, після чого, використовуючи формулу обернення блочних матриць, отримуємо

$\Omega(k, s) = \{\Omega_{ij} = \varphi^T(x(i)\varphi(x(j))) = K(x(i), x(j), s)\}$,
 $i = k-s+1, k-s+1, \dots, k$; $j = k-s+1, k-s+2, \dots, k$; на основі яких формується вихідний сигнал мережі

$$\hat{y}(x, k, s) = \sum_{l=k-s+1}^k \lambda(l, s) K(x, x(l), s) + w_0(k, s).$$

Параметри $\lambda(l, s)$, $w_0(k, s)$ можуть бути знайдені шляхом вирішення матричного рівняння

$$\begin{pmatrix} w_0 \\ \Lambda(k, s) \end{pmatrix} = \begin{pmatrix} 0 & I_s^T \\ I_s & \Omega(k, s) + \gamma^{-1} I_{s,s} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ Y(k, s) \end{pmatrix} = P(k, s) \begin{pmatrix} 0 \\ Y(k, s) \end{pmatrix},$$

де $\Lambda(k, s) = (\lambda(k-s+1, s), \lambda(k-s+2, s), \dots, \lambda(k, s))^T$,

$Y(k, s) = (y(k-s+1), y(k-s+2), \dots, y(k))^T$.

З знаходженням $(k+1)$ -го спостереження, його необхідно включити в матрицю $\Omega(k+1, s)$, одночасно з тим, виключаючи спостереження, що відповідає $(k-s)$ -му моменту часу.

Нескладно бачити, що оновлені параметри LS-SVM визначаються за допомогою співвідношенням

$$\begin{pmatrix} w_0(k+1, s) \\ \Lambda(k+1, s) \end{pmatrix} = \begin{pmatrix} 0 & I_s^T \\ I_s & \Omega(k+1, s) + \gamma^{-1} I_{s,s} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ Y(k+1, s) \end{pmatrix} = P(k+1, s) \begin{pmatrix} 0 \\ Y(k+1, s) \end{pmatrix},$$

де

$$\Lambda(k+1, s) = (\lambda(k-s+1, s), \lambda(k-s+2, s), \dots, \lambda(k+1, s))^T,$$

$$Y(k+1, s) = (y(k-s+1), y(k-s+2), \dots, y(k+1))^T.$$

Тут важливо відзначити, що така організація процесу навчання LS-SVM, дозволяє не тільки «придусувати» застарілу інформацію, але й зберігає архітектуру нейронної мережі з часом, що суттєво спрощує її чисельну реалізацію.

4 ЕКСПЕРИМЕНТИ

Для перевірки запропонованого підходу було використано набір даних «Fashion-MNIST» [12] – це набір даних із зображеннями статей Zalando, що складається з навчального набору з 60000 прикладів і тестового набору з 10000 прикладів. Кожен приклад – це зображення у відтинках сірого 28x28, пов'язане з міткою з 10 класів. Приклади спостережень наведено на рисунку 1.

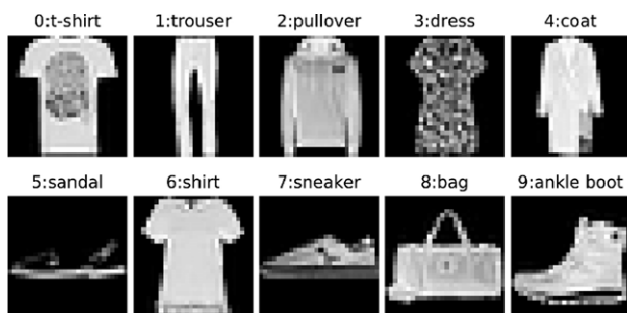


Рисунок 1 – Екземпляри вибірки «Fashion MNIST»

Для експериментальних досліджень та порівняльного аналізу запропонованого підходу були обрані методи K-NN та RCNN з високою швидкістю навчання, оскільки специфікою онлайн-обробки є їх апіорна та поточна невизначеність, тому швидкість навчання системи повинна бути високою.

Для аналізу швидкості та точності підходів, вибірку даних «Fashion-MNIST» було розбито на декілька

підмасивів: 5000, 10000, 15000 спостережень відповідно. Результати наведені на рисунку 2.

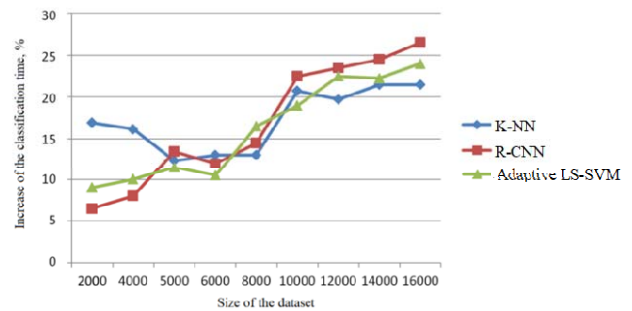


Рисунок 2 – Графік залежності часу класифікації від розміру вхідних даних

Якість класифікації вибірки «Fashion MNIST» був проведений в декілька етапів на 5000, 10000 та 15000 спостережень. Результати наведені в таблиці 1.

Таблиця 1 – Оцінка якості класифікації вибірки «FMNIST»

Методи	SI	CHI	DBI
5000 спостережень			
K-NN	0,3324	928,01	1,3
R-CNN	0,3335	974,42	1,10
Adaptive LS-SVM	0,3665	1420,28	1,15
10000 спостережень			
K-NN	0,6045	1460,65	1,5
R-CNN	0,6122	1945,5	1,25
Adaptive LS-SVM	0,7222	2800,05	1,32
15000 спостережень			
K-NN	0,8324	1560,00	1,8
R-CNN	0,8735	2000,28	1,15
Adaptive LS-SVM	0,9885	2985,2	1,28

Індекс силуету (SI) вимірює, наскільки кожен об'єкт в класі подібний до інших об'єктів у тому ж класі порівняно з об'єктами інших класів. За значенням від 0 до 1, вищі значення SI вказують на кращу якість класифікації.

Індекс Калінські-Харабаса (CHI) вимірює, наскільки добре класи відокремлені один від одного. Високі значення CHI вказують на кращу якість класифікації.

Індекс Девіса-Болдуїна (DBI) вимірює середню відстань між класами і внутрішньокласову відстань. Низькі значення DBI свідчать про кращу якість класифікації.

6 ОБГОВОРЕННЯ

Аналізуючи результати отриманих експериментальних досліджень та порівняльного аналізу роботи запропонованого підходу при вирішенні завдань розпізнавання образів з хорошою точністю та швидкістю в умовах обмеженого навчального набору даних, обробки зображень в онлайн-режимі.

В таблиці 1 метод адаптивної LS-SVM має найвищі значення SI для всіх трьох обсягів даних, що свідчить про його здатність ефективно відокремлювати класи навіть у великих обсягах даних, а також має

найвищі значення СНІ для всіх обсягів даних, що означає, що класи, сформовані Adaptive LS-SVM, добре відокремлені один від одного.

Цей підхід є особливо корисним при роботі з об'ємними даними, забезпечуючи високу точність класифікації та ефективність в умовах зміни типів вхідних даних. Його гнучкість та здатність до адаптації роблять з нього потужний інструмент для різноманітних завдань у сфері автоматичного розпізнавання образів-зображень. Запропонований підхід призначений для вирішення достатньо великого класу проблем у загальних рамках Data Stream Mining і Big Data Mining.

ВИСНОВКИ

Розглянута задача розпізнавання образів-зображень за допомогою машини опорних векторів найменших квадратів (LS-SVM) за умов, коли дані на обробку надходять у послідовному онлайн режимі. Процес навчання реалізується на ковзному вікні, що веде до того, що кількість нейронів у мережі не змінюється з часом, тобто архітектура залишається фіксованою у процесі налаштування. Такий підхід спрощує чисельну реалізацію системи та дозволяє реалізувати процес навчання на нестационарних тренувальних вибірках. Розглянута можливість навчання у ситуаціях, коли навчальні образи задані не лише у векторній формі, а й матричній, що дозволяє відмовитися від перетворення зображень у векторні сигнали.

Запропонований підхід до адаптивного навчання LS-SVM спрощує чисельну реалізацію нейронної мережі та дозволяє підвищити швидкість обробки інформації і, перш за все, налаштування її синаптичних ваг.

Наукова новизна: вперше запропоновано підхід до адаптивного навчання LS-SVM за умов, коли дані на обробку надходять у послідовному онлайн режимі.

Практичне значення: результати експерименту дозволяють рекомендувати запропонований підхід для використання на практиці для вирішення проблем автоматичного розпізнавання образів-зображень.

Перспективи подальших досліджень швидкі нейронні мережі розпізнавання образів-зображень для широкого класу практичних задач Data Stream Mining і Big Data Mining.

ПОДЯКА

Робота виконана в рамках науково-дослідного проєкту державного бюджету Харківського національного університету радіоелектроніки «Адаптивний бегінг гібридних систем обчислювального інтелекту на основі оптимального за швидкодією онлайн навчання» (ДР №0124U000363).

ЛІТЕРАТУРА

1. Goodfellow I. Deep learning / I. Goodfellow, J. Begin and A. Courville. – The MIT Press, 2016.
2. Graupe D. Deep learning neural networks: design and case studies / D. Graupe. – World Scientific Publishing Company, 2016. <https://doi.org/10.1142/10190>
3. Neural networks and deep learning / C. C. Aggarwal et al. – Cham : Springer, 2018. – Т. 10. № 978. <https://doi.org/10.1007/978-3-319-94463-0>
4. Poggio T. Networks for approximation and learning / T. Poggio, F. Girosi // Proceedings of the IEEE. – 1990. – Т. 78, № 9. – P. 1481–1497.
5. Haykin S. Neural networks: a comprehensive foundation. / S. Haykin. – Prentice Hall PTR, 2004. – Т. 2. – 1994.
6. Vapnik V. N. The Nature of Statistical Learning Theory / V. N. Vapnik. – New York : Springer, 1995.
7. Cortes C. Support-vector networks / C. Cortes and V. Vapnik // Machine Learning. – Sep. 1995. – Vol. 20, No. – P. 273–297, <https://doi.org/10.1007/bf00994018>.
8. Steinwart I. Support Vector Machines. / I. Steinwart and A. Christmann. – New York : Springer, 2008.
9. Pattern recognition using radial basis function network / [D. R. Zahirniak, R. Chapman, S. K. Rogers et al.] // Aerospace Application of Artificial Intelligence. – 1990. – P. 249–260.
10. Least Squares Support Vector Machines / [J. Vandewalle, B. D. Moor, T. V. Gestel et al.]. – World Scientific Publishing Company, 2003.
11. Adaptive least-squares support vector machine and its online learning / [Y. Bodyanskiy, A. Deineko, F. Brodetskyi, and D. Kosmin] // CEUR Workshop Proceedings. – Nov. 2020. – Vol. 2762, Art. no. 3. <http://ceur-ws.org/Vol-2762/paper3.pdf>
12. Xiao H. Fashion-mist: a novel image dataset for benchmarking machine learning algorithms / H. Xiao, K. Rasul, R. Vollgraf. – Available: <https://arxiv.org/abs/1708.07747>.

Стаття надійшла до редакції 02.04.2025.

Після доробки 12.06.2025.

FAST NEURAL NETWORK AND ITS ADAPTIVE LEARNING IN CLASSIFICATION PROBLEMS

Bodyanskiy Ye. V. – Dr. Sc., Professor at the Department of Artificial Intelligence, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Shafronenko Ye. O. – Assistant at the Department of Media Engineering and Information Radio Electronic Systems, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Brodetskyi F. A. – PhD Student, Assistant at the Department of Informatics, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

Tanianskyi O. S. – Postgraduate student at the Department of Informatics, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

ABSTRACT

Context. To solve a wide class of information processing tasks and, above all, pattern recognition under conditions of significant nonlinearity, artificial neural networks have become widely used, due to their universal approximating properties and ability to learn based on training samples. Deep neural networks have become the most widespread, which indeed demonstrate very high recognition quality, but require extremely large amounts of training data, which are not always available. Under these conditions, the so-called least squares support vector machines can be effective. They do not require large amounts of training samples but can be trained only in batch mode and are quite cumbersome in numerical implementation. Therefore, the problem of training LS-SVM in sequential mode under conditions of significant non-stationarity of data that are sequentially fed online to the neural network for processing is quite relevant.

Objective. The aim of the work is to introduce an approach to adaptive learning of LS-SVM, which allows us to abandon the conversion of images into vector signals.

Method. An approach for image recognition using a least squares support vector machine (LS-SVM) is proposed under conditions when data for processing is received in a sequential online mode. The advantage of the proposed approach is that reduces the time to solve the image recognition problem and allows the implementation of the learning process on non-stationary training samples. A feature of the proposed method is computational simplicity and high speed since the number of neurons in the network does not change over time, i.e., the architecture remains fixed during the tuning process.

Results. The proposed approach to adaptive learning of LS-SVM simplifies the numerical implementation of the neural network and allows for an increase in the speed of information processing and, above all, the tuning of its synaptic weights.

Conclusions. The problem of pattern recognition using the least squares support vector machine (LS-SVM) is considered under conditions when data for processing is received in a sequential online mode. The training process is implemented on a sliding window, which leads to the fact that the number of neurons in the network does not change over time, i.e. the architecture remains fixed during the tuning process. This approach simplifies the numerical implementation of the system and allows the training process to be implemented on non-stationary training samples. The possibility of training in situations where training images are given not only in vector form but also in matrix form allows us to abandon the conversion of images into vector signals.

KEYWORDS: Adaptive learning, classification, fast neural network, support vector machine.

REFERENCES

1. Goodfellow I., Bengio J. and Courville A. Deep learning. The MIT Press, 2016.
2. Graupe D. Deep learning neural networks: design and case studies. World Scientific Publishing Company, 2016. <https://doi.org/10.1142/10190>
3. C. C. Aggarwal et al. *Neural networks and deep learning*. Cham, Springer, 2018, T. 10, № 978. <https://doi.org/10.1007/978-3-319-94463-0>
4. Poggio, T. Girosi F. Networks for approximation and learning, *Proceedings of the IEEE*, 1990, T. 78, №. 9. pp. 1481–1497.
5. Haykin S. Neural networks: a comprehensive foundation. Prentice Hall PTR, 2004, T. 2, 1994.
6. Vapnik V. N. The Nature of Statistical Learning Theory. New York, Springer, 1995.
7. Cortes C. and Vapnik V. Support-vector networks, *Machine Learning*, Sep. 1995, Vol. 20, No. 3, pp. 273–297, <https://doi.org/10.1007/bf00994018>.
8. Steinwart I. and Christmann A. Support Vector Machines. New York, Springer, 2008.
9. Zahirniak D. R., Chapman R., Rogers S. K., Suter B. W., Kabrisky M., and Pyatti V. Pattern recognition using radial basis function network, *Aerospace Application of Artificial Intelligence*, 1990, pp. 249–260.
10. Vandewalle J., Moor B. D., Gestel T. V., Brabanter J. D., and Suykens J. A. K. Least Squares Support Vector Machines. World Scientific Publishing Company, 2003.
11. Bodyanskiy Y., Deineko, A. Brodetskyi F., and Kosmin D. Adaptive least-squares support vector machine and its online learning, *CEUR Workshop Proceedings*, Nov. 2020, vol. 2762, Art. no. 3. <http://ceur-ws.org/Vol-2762/paper3.pdf>
12. Xiao H., Rasul K., Vollgraf R. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. Available: <https://arxiv.org/abs/1708.07747>.