

УДК 004.93.1:656.073.5

Б.І. МОРОЗ, д-р техн. наук, проф., зав. каф., Академія митної служби України, Дніпропетровськ,
С.М. КОНОВАЛЕНКО, аспірант, зав. лаб., Академія митної служби України, Дніпропетровськ

ОПТИМІЗІЦІЯ ПАРАМЕТРІВ НАВЧАННЯ НЕЙРОМЕРЕЖЕВОЇ СИСТЕМИ ОБРОБКИ ІНФОРМАЦІЇ МИТНОГО КОНТРОЛЮ

Розглянуто аспекти підвищення якості навчання нейронних мереж типу багат шаровий перцептрон за рахунок аналізу та корегування таких параметрів, як крок навчання градієнтного методу, об'єм вхідного вектору, пераметр регуляризації. Графічно показано можливі залежності цільової функції від цих параметрів та приведено заходи, щодо зменшення негативного впливу певної ситуації. Іл.: 3. Бібліогр.: 9 назв.

Ключові слова: нейронна мережа, цільова функція, градієнтний метод.

Постановка проблеми. Розробка методів та засобів переробки інформації для потреб митної служби України є задача важлива та актуальна [1 – 3]. Оскільки досить часто трапляються випадки подання суб'єктами зовнішньоекономічної діяльності недостовірних відомостей про характеристики товарів з метою заниження (завищення) митної вартості, перевезення контрабанди, то існує потреба застосування методів і засобів бізнес-аналітики. До таких моделей обробки інформації висуваються такі вимоги, як здатність до навчання, якісного розпізнавання чи класифікації. Враховуючі можливості інформаційних технологій система аналізу та управління ризиками потребує імплементації потужних методів інтелектуальної обробки інформації.

Аналіз літератури. Публікації в галузі впровадження сучасних інформаційних технологій в автоматизовані системи митної служби носять досить концептуальний характер [2, 3], виділяючи проблематику та формалізуючи загальні аспекти функціонування. В роботах [4, 5] висвітлено такий важливий аспект, як підготовка "сирих" даних (інформація митного контролю), що подаються на вхід аналітичної системи. У працях [6, 7] було запропоновано використати математичну модель нейронної мережі типу багат шаровий перцептрон, що за своєю суттю є універсальними апроксиматором, для ідентифікації ризиків порушення митного законодавства. В цих роботах були отримані досить хороші результати розпізнавання ризиків, що інтерпретуються як

"Високий", "Помірний", "Низький", проте мало приділено уваги проблематиці якісного машинного навчання. Тобто, яким чином можна вплинути на якість навчання нейрокласифікатора та шляхи подолання проблем що виникають.

Метою статті є продовження циклу робіт щодо формалізації та розробки моделі ідентифікації ризиків порушення митного законодавства на основі математичного апарату нейронних мереж типу багатошаровий перцептрон. Мета роботи описати такий важливий аспект розробки інтелектуальної системи, як організація та керування процесом навчання з метою запобігання втрати нею такої властивості, як узагальнення.

Інтелектуальна система обробки інформації (ICOI). Процес розробки ICOI загалом передбачає собою наступні етапи (див. рис. 1):

1. Перцепція інформації з різноманітних джерел (як правило це різнотиповий вектор вхідних даних).
2. Підготовка або нормалізація вхідного вектору.
3. Обробка інформації.
4. Інтерпретація отриманих результатів.

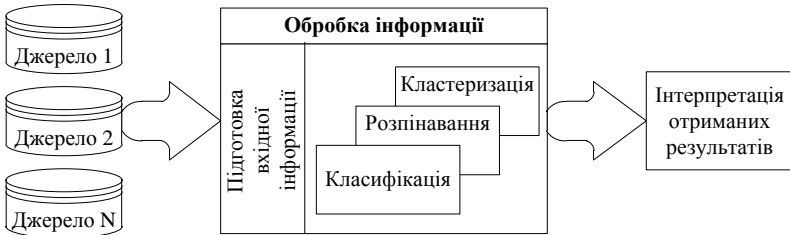


Рис. 1. Інтелектуальна система обробки інформації (ICOI)

Аналіз предметної галузі проблеми, що розглядається, передбачає собою обробку досить великих масивів даних, що носять переважно різнотиповий характер та відповідно виділення таких факторів та ознак, що мають найбільш важливе значення [5]. В роботі [7] було визначено вектор ознак інформації митного контролю та відповідним чином оброблено [4] для адекватного сприйняття нейрокласифікатором. Необхідно зауважити, що сформована для навчання системи вибірка повинна отримати такі властивості, як репрезентативність (для адекватного відображення предметної галузі) та несуперечливість.

Процес інтелектуальної обробки інформації залежить від вибору методів. Напевно, передусім це методи, що використовують алгоритми

машинного навчання. У нашому випадку було застосовано математичну модель штучних нейронних мереж типу багатощаровий персеptron [7].

І нарешті інтерпретація отриманих результатів ставить собою на меті передусім донесення у зрозумілій формулі для користувача про результат розпізнавання або класифікації. В контексті аналізу ризиків порушення митного законодавства це рівень ризику – "Високий", "Помірний" або "Низький".

Налаштування параметрів навчання ICOI. Якісне навчання ICOI є, напевно, найважливішою задачею, що необхідно вирішити розробнику, адже це безпосередньо впливає на результат [8]. Тому перш за все, необхідно сформовану множину вхідних даних розділити на:

1. Навчальну множину $\{x^{(1)}, y^{(1)}\}, \dots, \{x^{(m)}, y^{(m)}\}$.
2. Перевірочну – $\{x_{cv}^{(1)}, y_{cv}^{(1)}\}, \dots, \{x_{cv}^{(m_{cv})}, y_{cv}^{(m_{cv})}\}$.
3. Тестову – $\{x_{test}^{(1)}, y_{test}^{(1)}\}, \dots, \{x_{test}^{(m_{test})}, y_{test}^{(m_{test})}\}$.

Розподіл необхідно зробити у такій пропорції – відповідно 60 %, 20 % та 20% від загальної кількості. Це дозволить нам використовувати отримані похибки $J_{train}(\Theta)$ (*Training error*), $J_{cv}(\Theta)$ (*Cross Validation error*) та $J_{test}(\Theta)$ (*Test error*). Загалом функція похибки (цільова функція) має наступний вигляд:

$$J(\Theta) = \frac{1}{2m} \left[\sum_{i=1}^m (h_{\Theta}(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=2}^n \Theta_j^2 \right], \quad (1)$$

де $h_{\Theta}(x) = g(\Theta^T x)$ – модель, що налаштовується, або гіпотеза; Θ – вектор параметрів, що налаштовується; λ – параметр регуляризаційного члену.

Зауважимо, що цільова функція (1) записана в загальному випадку, без врахування шарів та кількості нейронів у персеptronі, що дозволяє враховувати такі моделі як лінійна, нелінійна та логістична регресія.

Використовуючи градієнтні методи оптимізації ми можемо знайти мінімум цільової функції, проте для адекватної роботи, як згадувалось вище, необхідно провести нормалізацію вхідного простору. Наприклад привести до діапазону $\{0 \dots 1\}$ (2), або відповідним кодуванням:

$$x_i = \frac{x_i}{\max(x)}; \quad x_i = \frac{x_i - \mu}{S}, \quad (2)$$

де μ – середнє значення вектора x у навчальній вибірці, а S – максимінне або стандартне відхилення.

Масштабування вхідних ознак допоможе пришвидшити збіжність оптимізаційного процесу. Так наприклад на рис. 2. а зображено поверхні рівня цільової функції $J(\Theta)$ для двох Θ . Тож можемо уявити як це полегшить оптимізацію для функції багатьох змінних.

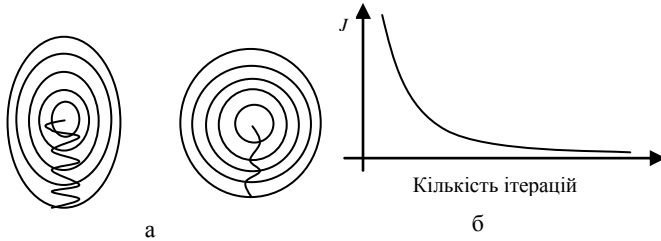


Рис. 2. а – порівняння процесу мінімізації до та після масштабування;
б – залежність функції похибки від кількості ітерацій (epoch)

Процес навчання персептрона градієнтними методами передбачає собою ітераційний алгоритм мінімізації цільової функції $J(\Theta)$ шляхом налаштування параметрів Θ (3).

$$\Theta_j = \Theta_j - \alpha \left(\frac{1}{m} \sum_{i=1}^m (h_{\Theta}(x^{(i)}) - y^{(i)}) x_j^{(i)} + \frac{\lambda}{m} \Theta_j \right), \quad (3)$$

де α – крок навчання.

Швидкість (крок) навчання α (*learning rate*) необхідно вибирати таким чином щоб уникнути наступних проблем [9]:

1. Якщо α занадто мала, то градієнтні методи доволі повільно збігаються.

2. Якщо α занадто велика, то можлива ситуація, коли оптимізаційний процес багаторазово "перестрибує" мінімум цільової функції, або навіть зовсім розбігається.

Під час оптимізації доцільно побудувати залежність цільової функції $J(\Theta)$ від кількості ітерацій та зупиняти процес навчання коли зменшення значення функції за одну ітерацію не буде менше ніж встановлений поріг, наприклад 10^{-3} . В ідеалі вона зображує релаксаційний процес показаний на рис. 2.б. В процесі навчання ICOI доцільно, та мабуть необхідно, обчислювати вищевказані похибки ($J_{train}(\Theta)$, $J_{cv}(\Theta)$ та $J_{test}(\Theta)$) на просторі відповідних множин. Аналізуючи ці похибки можна виявити декілька проблем, які впливають

на здатність класифікатора до узагальнення. Ці проблеми можна описати наступними термінами:

1. Недонавчання (*underfit*) – λ має велике значення (рис. 3. а)).
2. Перенавчання (*overfit*) – $\lambda = 0$ (рис. 3. а)).
3. Тільки правильні (*just right*) – λ має середнє значення (рис. 3. а)).

Побудова залежностей функцій похибки від регуляризаційного параметру λ та об'єму вибірки m (рис. 3) дозволяє виявити підступну проблему перенавчання, тобто ситуацію, коли класифікатор коректно працює лише на навчальній множині.

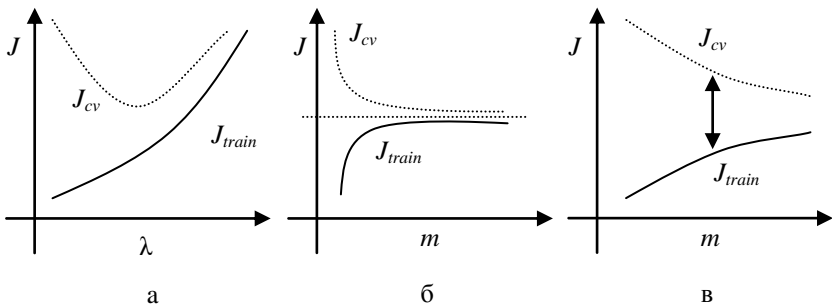


Рис. 3. а – залежність $J(\lambda)$; б – та в – залежність $J(m)$, де m – об'єм вибірки.

Якщо під час навчання та кросвалідації ми отримуємо велику помилку (рис. 3. б)), збільшення числа m не вирішить проблему. У ситуації (рис. 3. в)), навпаки – збільшення дещо покращить якість навчання в цілому [9]. Окрім того, до цього можна додати процедуру зменшення розмірності вхідного простору (якщо є така необхідність).

На останок додамо ще одну рекомендацію: перед навчанням нейронної мережі градієнтними методами варто включити в оптимізаційний алгоритм перевірку градієнта за відомою геометричною формулою наближеного обрахування градієнту. Зауважимо, що перевірку необхідно робити до навчання класифікатора алгоритмом зворотного розповсюдження помилки. Не врахувавши цього, ми отримаємо надлишковий код та збільшення обрахунків на кожній ітерації.

Висновки. В результаті роботи виділені методики аналізу якості процесу навчання ICOI та шляхи подолання такої проблеми нейронних мереж, як втрата здатності до узагальнення. Врахування цих рекомендацій покращує навчальний алгоритм та відповідно впливає на якість розпізнавання ризиків порушення митного законодавства. В

подальшому перспективним для дослідження є удосконалення архітектури нейрокласифікатора.

Список літератури: 1. Основи митної справи в Україні: Підручник / За ред. П.В. Паука. – К.: Знання, 2008. – 652 с. 2. Семенко О.М.-М. Підвищення ефективності роботи по протидії контрабанді та порушенням митних правил, митного контролю за рахунок впровадження автоматизованої системи аналізу та управління митними ризиками на основі Fuzzy-технології / О.М.-М. Семенко. – К.: LAT & K, 2008. – 238 с. 3. Митні інформаційні технології: навч. посібник / О.Ф. Волик, О.В. Кащесва, І.В. Дорда та ін. / За ред. П.В. Паука. – К.: Знання, 2011. – 391 с. 4. Konovalenko Sergii Preprocessing "raw" data sets as an important aspect of intelligent information processing / Sergii Konovalenko // International Book Series "Information Science and Computing". Artificial Intelligence Driven Solutions to Business and Engineering Problems. – Rzeszow and Sofia: ITNEA, 2012. – Book 27. – P. 133-139. 5. SPSS: Искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей: Пер. с нем. / Ахим Бююль, Петер Цёфель. – СПб.: ООО "ДиаСофтЮП", 2005. – 608 с. 6. Коноваленко С.М. Деякі аспекти розвитку системи аналізу ризиків порушення митного законодавства / С.М. Коноваленко, Б.І. Мороз // Вісник Національного технічного університету "Харківський політехнічний інститут". – Харків: НТУ "ХПИ", 2011. – № 17. – С. 109-114. 7. Коноваленко С. Применение нейронных сетей в рамках концепции "электронная таможня" / С. Коноваленко, Б. Мороз // International Book Series "Information Science and Computing". Applicable Information Models. – Sofia: ITNEA, 2011. – Book 22. – P. 104-110. 8. Christopher M. Bishop Pattern Recognition and Machine Learning (Information Science and Statistics). – Springer, 2007. – P. 738. 9. Andrew Ng Machine Learning Online Course Режим доступу: <http://www.coursera.org>.

Надійшла до редакції 25.03.2013

УДК 004.93.1:656.073.5

Оптимизация параметров обучения нейросетевой системы обработки информации таможенного контроля / Мороз Б.И., Коноваленко С.Н. // Вестник НТУ "ХПИ". Серия: Информатика и моделирование. – Харьков: НТУ "ХПИ". – 2013. – № 39 (1012). – С. 134 – 139.

Рассмотрены аспекты повышения качества обучения нейронных сетей типа многослойный перцептрон, за счет анализа и корректировки таких параметров, как шаг обучения градиентного метода, объем входного вектора, параметр регуляризации. Графически показаны возможные зависимости целевой функции от этих параметров и приведены средства уменьшения негативного влияния определенной ситуации. Ил.: 3. Библиогр.: 9 назв.

Ключевые слова: нейронная сеть, целевая функция, градиентный метод.

UDC 004.93.1:656.073.5

Optimization of parameters training a neural network of the system information processing of customs control / Moroz B.I., Konovalenko S.N. // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modelling. – Kharkov: NTU "KhPI". – 2013. – № 39 (1012). – P. 134 – 139.

Considered the aspects of improving the quality of training neural networks such as multilayer perceptron, by analyzing and adjusting the parameters such as: the learning step of gradient method, capacity of the input vector parameters of regularization. Graphically shown the possible dependence of the target function of these parameters and given the means of reducing the negative influence of a particular situation. Figs: 3. Refs: 9 titles.

Keywords: neural network, target function, gradient method.