УДК: 004.832.32.

Кравець І.О., Шульженко Г.А.

РОЗРОБКА ТА ДОСЛІДЖЕННЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ НАДАННЯ БАНКІВСЬКИХ ПОСЛУГ

Було проведено дослідження нейронних мереж для розв'язання задач класифікації та кластеризації у банківської справі. Для задач класифікації розроблена мережа Кохонена з квантуванням навчального вектора. Для задач кластеризації було запропоновано модифікацію алгоритму кластеризації з використанням мережі Кохонена та оптимізацією числа вихідних нейронів або кластерів. Було побудовано 2-вимірну мапу Кохонена.

Ключові слова: *нейронні мережі, кластеризація, класифікація, карта Кохонена.*

Проведено исследование нейронных сетей для задач классификации и кластеризации в банковском деле. Для задач классификации разработана сеть Кохонена с квантованием обучающего вектора. Для задач кластеризации предложена модификация алгоритма кластеризации с использованием сети Кохонена и оптимизацией числа выходных нейронов или кластеров. Построена 2-мерная карта Кохонена.

Ключевые слова: *нейронные сети, кластеризация, классификация, карта Кохонена.*

The efficiency of neuron' networks, which are applied in bank's working, are researched. The neuron' networks and its learning algorithms for classification, the cluster analyses, are developed. The two sizes Kohonen's map was built.

Key words: *neuron networks, clusterization, classification, map of Kokhonena.*

ВСТУП

Банківська справа є класичним прикладом використання інтелектуальних технологій. Насамперед, це рішення проблеми про можливу некредитоздатність клієнтів банку. Позбавлені суб'єктивної упередженості, ці рішення опираються тільки на базу даних банку, де записується детальна інформація про кожного клієнта й в остаточному підсумку факт його кредитоспроможності. Класифікаційні алгоритми Data Mining обробляють ці дані, і отримані результати використаються далі для прийняття рішень. Необхідною задачею, яка виникає у банках, також є віднесення клієнтів та партнерів банків до певних груп по їх характеристикам (депозити юридичних осіб, активи банків, прибутковість банків то що), число якіх може бути достатньо великим.

З точки зору інтелектуального аналізу даних це є задачі класифікації та кластеризації. Однак, вони мають ряд специфічних ознак (великий обсяг інформації, наявність інформації у якісному та кількісному вигляді, необхідність автоматичного генерування гіпотез прихованих закономірностей) і потребують розробки специфічного алгоритмічного та програмного

Нейронні мережі дозволяють вирішувати таки неформалізовані задачі. Дослідник при цьому одержує дуже ефективну модель проблемної області і може дуже просто моделювати різні ситуації, візуально побачити результат.

На даний час ведуться інтенсивні роботи по розробці методів та алгоритмів для задач інтелектуального аналізу даних у тому числі у банківській справі. На українському ринку технології інтелектуальних обчислень роблять лише перші кроки.

З одного боку існують статистичні пакети SPSS, Statistica, Matlab, в яких використані деякі алгоритми інтелектуального аналізу даних, але вони не працюють в режимі on-line і вимагають від користувача введення інформації перед кожним аналізом даних, а також знання методів статистичного та інтелектуального аналізу даних. З іншого боку існують комерційні програмні продукти, такі як SoMine, NeuroShell, NeuroScalp, Deductor (додаток до Olap), які використовують інтелектуальний аналіз даних в режимі on-line та розв'язують перелічені вище задачі Але ці пакети є дуже коштовними та не кожен банк взмозі купувати його і навчити персонал. Тому розробка інтелектуально-аналітичної системи, яка б поєднувала базу даних та виконувала інтелектуальний аналіз даних в цьому режимі та була достатньо дешевою є актуальною задачею.

ЗАДАЧІ ДОСЛІДЖЕННЯ

Задачею дослідження було розробка структури, алгоритмів навчання та дослідження нейронних мереж для розв'язання задач класифікації та кластеризації у банківської справі. Наступною задачею було створення інформаційно-аналітичної системи, яка поєднує базу даних та аналітичний блок, і може слугувати прикладом використання методів інтелектуального аналізу даних для задач банківської справи.

ПІДГОТОВКА ДАНИХ

Процес кредитування пов'язаний з дією численних і багатоманітних чинників ризику, здатних спричинити за собою непогашення позики в обумовлений термін. Надаючи позики, комерційний банк повинен вивчати чинники, які можуть спричинити за собою їх непогашення. Таке вивчення іменують аналізом кредитоспроможності (credit analysis). Кредитний цикл за Програмою мікрокредитування починається із залучення клієнтів, консультації потенційного позичальника і закінчується повним погашенням кредитної заборгованості позичальника перед Банком. Договірні відносини між Банком і Клієнтом по кредитуванню закріплюються наступними документами: анкета Позичальника, пам'ятка клієнта, довідка про умови кредитування та орієнтовну сукупну вартість, довідка про умови кредитування та орієнтовну сукупну вартість споживчого кредиту. На основі одержаних від клієнта даних кредитним експертом здійснює обробку заявки банк(у даному випадку Приват-Банк) використовує власну форму обробки заявки, яка складається з 3 розділів:

1. Загальні дані.
2. Фінансові показники.
3. Характеристика кредиту.

Найбільш важливим розділом є другий розділ, так як він характеризує платоспроможність і кредитоспроможність позичальника. Питома вага розділу в загальній сумі складає 50 %. 30 % в загальній сумі складає третій розділ, в якому аналізується мета і строк кредитування, забезпечення, а також наявність поручителів. І 20 % складає перший розділ, який включає загальні дані про позичальника (вік, час проживання в даній місцевості, освіта, місце роботи, посада, стаж роботи на підприємстві, а також сімейний стан). В перелік даних, які необхідно знати про позичальника входять наступні показники:

- вік;
- сімейний стан;
- освіта;
- працевлаштування;
- посада;
- стаж роботи;
- розмір заробітної плати;
- загальний дохід сім'ї;
- розмір доходу з інших джерел;
- наявність автомобіля;
- володіння іншим майном;

- кредитна історія;
- позичаємо сума.

Ці дані, які підлягають кодуванню та нормуванню, є елементами вхідного вектора.

РОЗВ'ЯЗАННЯ ЗАДАЧІ ПРО ВИДАЧУ КРЕДИТУ

Задача про видачу кредиту клієнту є задачею класифікації. Для її розв'язання було досліджено нейронні мережі зворотного поширення похибки та мережу з квантуванням навчального вектора. Мережа зворотного поширення похибки складається з 3 прошарків: вхідного прошарку, у якого число нейронів дорівнює числу ознак об'єкту (13), прихованого прошарку з 5 нейронів, та вихідного прошарку з 2 нейронів. Ця мережа навчається з вчителем. Розмір навчальної вибірці 30 прикладів. Число ітераційних циклів для 1 приклада 10000.

Другою була досліджена мережа з квантуванням навчального вектора.Мережа складається з вхідного прошарку з13 нейронів та вихідного прошарку Кохонена з 2 нейронів. Переможцем є той нейрон, вектор ваг якого ω_c найближчий до вхідного образу x, тобто нейрон c визначається як

$$c = \arg\min_j(\|x - \omega_j\|^2). \tag{4.6}$$

Значення всіх ваг ω_j, що мінімізують помилку класифікації, обчислюються асимптотично. При цьому корекція ваг відбувається за правилом

$$\omega_c(k+1) = \begin{cases} \omega_c(k) + \alpha(k)[x(k) - \omega_c(k)], \text{ якщо } \omega_c \text{ й } x \text{ належать одному класу;} \\ \omega_c(t) + \alpha(k)[x(t) - \omega_c(k)] \text{ в іншому випадку;} \end{cases} \tag{4.7}$$

$$\omega_o(k+1) = \omega_j(k) \text{ для всіх } j \neq c,$$

де $\alpha(k) \in (0,]$. Параметр $\alpha(k)$ може залишатися постійним або монотонно зменшуватися. Таким чином, вектор ваг нейрона-переможця ω_c, що найближче розташований до поданого вхідного вектора, зміщується в напрямку останнього, якщо вхідний вектор відноситься до одного з ним класу, і віддаляється від нього в іншому разі. Блок-схема алгоритму наведена на рис. 1.

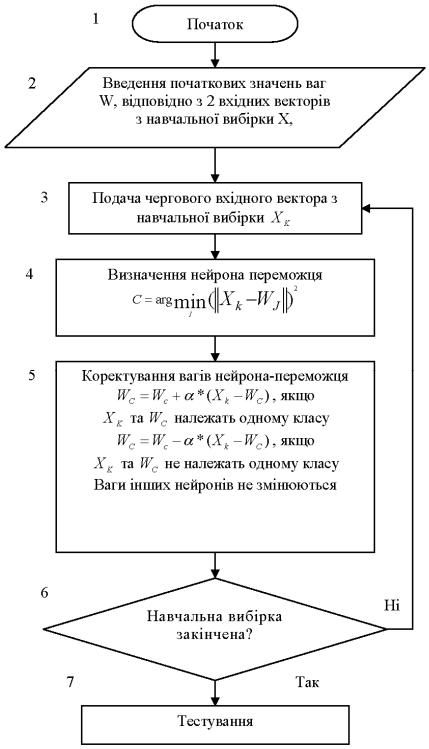
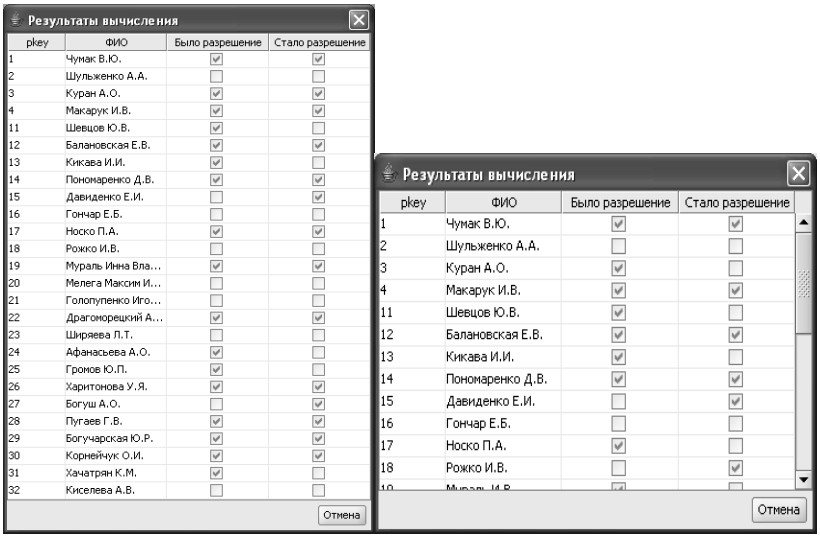


Рис. 1. Блок-схема алгоритму класифікації з використанням мережі Кохонен з квантуванням навчального вектора

Результати роботи мережі з квантуванням навчального вектора та мережі зворотного поширення похибки наведені на рис. 2 (прізвища клієнтів вигадані, але дані реальні).



Результаты вычисления

pkey	ФИО	Было разрешение	Стало разрешение
1	Чумак В.Ю.	☑	☑
2	Шульженко А.А.	☐	☐
3	Куран А.О.	☑	☑
4	Макарук И.В.	☑	☑
11	Шевцов Ю.В.	☑	☐
12	Балановская Е.В.	☑	☑
13	Кикава И.И.	☑	☐
14	Пономаренко Д.В.	☑	☑
15	Давиденко Е.И.	☐	☑
16	Гончар Е.Б.	☐	☐
17	Носко П.А.	☑	☑
18	Рожко И.В.	☐	☐
19	Мураль Инна Вла...	☑	☑
20	Мелега Максим И...	☐	☐
21	Голопупенко Иго...	☐	☐
22	Драгоморецкий А...	☑	☑
23	Ширяева Л.Т.	☐	☐
24	Афанасьева А.О.	☑	☐
25	Громов Ю.П.	☑	☐
26	Харитонова У.Я.	☑	☑
27	Богуш А.О.	☐	☑
28	Пугаев Г.В.	☑	☑
29	Богучарская Ю.Р.	☑	☑
30	Корнейчук О.И.	☑	☑
31	Хачатрян К.М.	☑	☐
32	Киселева А.В.	☐	☐

Отмена

Результаты вычисления

pkey	ФИО	Было разрешение	Стало разрешение
1	Чумак В.Ю.	☑	☑
2	Шульженко А.А.	☐	☐
3	Куран А.О.	☑	☐
4	Макарук И.В.	☑	☑
11	Шевцов Ю.В.	☑	☐
12	Балановская Е.В.	☑	☑
13	Кикава И.И.	☑	☐
14	Пономаренко Д.В.	☑	☑
15	Давиденко Е.И.	☐	☑
16	Гончар Е.Б.	☐	☐
17	Носко П.А.	☑	☐
18	Рожко И.В.	☐	☑

Отмена

Рис. 2. Порівняння результатів класифікації мережею з квантуванням навчального вектора (а) та мережею зворотного поширення похибки (б).

Як видно з рис. 2., мережа з квантуванням навчального вектора дає кращі результати.

РОЗВ'ЯЗАННЯ ЗАДАЧІ ВІДНЕСЕННЯ КЛІЄНТІВ ДО ПЕВНИХ ГРУП (КЛАСТЕРИЗАЦІЇ)

Для розв'язку задач кластеризації запропоновано та розроблено алгоритм кластеризації, який використовує мережу Кохонена, знаходить центри кластерів(ваги вихідних нейронів) та отримує оптимальне число вихідних нейронів [24].

Алгоритм складається із таких кроків:

1. Ініціалізація мережі. Число вихідних нейронів 3. Ваговим коефіцієнтам мережі надаються малі випадкові значення.
2. Пред'явлення мережі нового вхідного сигналу.
3. Обчислення відстані до всіх нейронів мережі:

відстані dj від вхідного сигналу до кожного нейрона j визначаються за формулою:

$$d_{i,j} = \sum_{i=1}^{N} (x_{i,j}(t) - w_{i,j}(t))^2,$$

де xi – i-ий елемент вхідного сигналу в момент часу t, wij(t) – вага зв'язку від i-го елемента вхідного сигналу до нейрона j у момент часу t.

4. Вибір нейрона з найменшою відстанню: вибирається нейрон-переможець j*, для якого відстань dj найменше.
5. Налаштування ваг нейрона j* і його сусідів:

робиться налаштування ваг для нейрона j* і всіх нейронів з його околу NE. Нові значення ваг:

$$wij(t+1)=wij(t)+r(t)(xi(t)-wij(t)), \quad (4.10)$$

де r(t) – швидкість навчання, що зменшується з часом (додатне число, менше одиниці).

6. Якщо навчальна вибірка не закінчена, то повернення до кроку 2. Якщо навчальна вибірка закінчена, то повторюємо цикл навчання знову. Число циклів навчання, кожен з котрих включає всю вибірку(епох) 10000.
7. Обчислення функції мети $$J(M,W) = \sum_{i=1}^{c} \sum_{j=1}^{d} d_A^2(m_j, w^{(i)}) \quad (4.11)$$

8. Порівняння функції мети із функцією мети попереднього кроку. Якщо функція мети зменшується, то додається ще один вихідний нейрон переходимо до пункту 2. Якщо функція мети не зменшується, то виводимо результати.

В алгоритмі використовується коефіцієнт швидкості навчання, який поступово зменшується, для тонкішої корекції на новій епосі. В результаті позиція центру встановлюється в певній позиції, яка задовільним чином кластеризує приклади, для яких даний нейрон є переможцем. Блок-схема алгоритму представлена на рис. 3.

На підставі даних о розбитті клієнтів на 3 кластера було побудовано 2-вимірну мапу Кохонена для задач кластеризації (розмір 10*10 нейронів).

Для побудови було запропоновано 2-етапний підхід: на першому етапі проводиться кластеризація та помічаються вхідні вектора. Потім навчається мережа з 2-вимірним вихідним прошарком, де нейрони-переможці помічаються міткою відповідного кластеру. Далі кожному непоміченому нейрону даються меткі нейрона-переможця, ваги якого найближче до ваг цього нейрону.

Приклад мапи Кохонена наведений на рис 4.

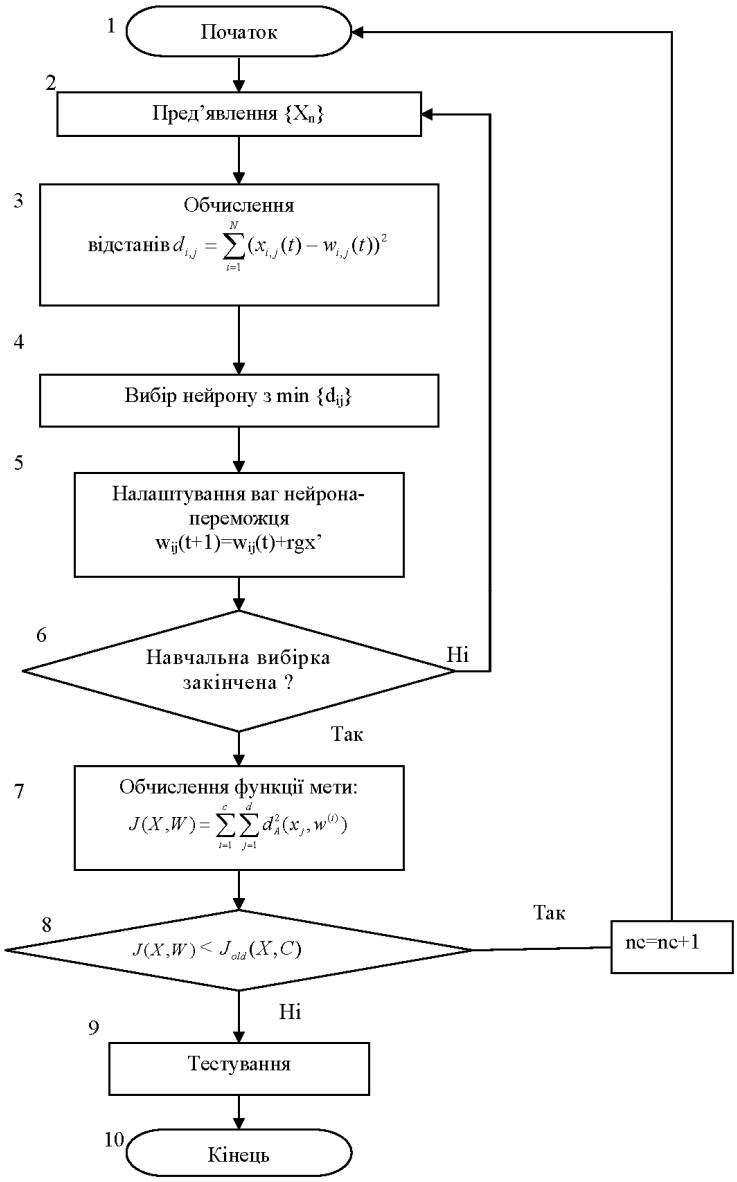


Рис. 3. Блок-схема алгоритму кластерізації з використанням мережі Кохонена

Рис. 4. Мапа Кохонена

Графічне відображення результатів виконання алгоритму кластеризації допомогає більш наглядно зрозуміти загальний розподіл клієнтів на три кластери (надійні клієнти, середні та ризиковані). На мапі надійні клієнти відображені сірим кольором – це люди, в основному вік яких належить до діапазону від 25 до 49 років, які мають або авто або інше майно, або у яких загальний дохід сім'є є достатньо високим і існують інші джерела доходів. До середнього класу надійності клієнтів віднесені студенти або пенсіонери, або виконавці зі стажем роботи від 1 до 3 років, і на карті вони відображені білим кольором. Чорним кольором відображені ризиковані клієнти, яким не слід видавати кредит, тому що вони або мають інші непогашені кредити та погану кредитну історію, або в них є діти до 18 років і низький рівень загального доходу сім'ї.

На карті немає однозначно чіткого розподілу на кластери, це пояснюється по-перше, похибкою, яка так чи інакше присутня в процесі навчання нейронної мережі та при визначенні нейрона-переможця, по-друге, можливо не зовсім коректною базою знань, яка використовувалася для навчання і тестування, тому що коректна робота алгоритму напряму залежить від правильно і однозначно чітко сформованих вхідних даних. Крім того, бувають такі неординарні клієнти, які можуть створювати похибки в вірному навчанні, наприклад син олігарха, за віком та за студентським станом може належати до ризикованого кластеру, але наявність надійного поручителя в ролі заможного батька дасть підстави для надання кредиту. І таких прикладів можна навести чимало.

ВИСНОВКИ

Було проведено дослідження нейронних мереж для розв'язання задач класифікації та кластеризації у банківської справі. В результаті досліджень було виявлено, що:

- для задач класифікації більш ефективним є алгоритм класифікації з використанням мережі Кохонена та та квантуванням навчального вектора ніж алгоритм з використанням мережі зворотнього поширення похибки, який виявив залежність кількості ітерації від початкових значень вагів нейронів та зупинявся у точках локальних мінімумів похибки.
- для задач кластеризації більш доцільним є алгоритм кластерізації з використанням мережі Кохонена.
- Було запропоновано власну модифікацію алгоритму кластерізації з використанням мережі Кохонена та оптимізацією числа вихідних нейронів або кластерів;
- було побудовано 2 вимірну мапу Кохонена для задач кластеризації. Для побудови було запропоновано 2 етапний підхід: на першому етапі проводиться кластеризація та помічаються вхідні вектора. Потім навчається мережа з 2 вимірним вихідним прошарком, де нейрони-переможці помічаються міткою відповідного кластеру. Далі кожному

непоміченому нейрону даються меткі нейрона-переможця, ваги якого найближче до ваг цього нейрону.

Створена інформаційно-аналітична система, яка поєднує базу даних та аналітичний блок, і може слугувати прикладом використання методів інтелектуального аналізу даних для задач банківської справи. Для розробки бази даних даного програмного забезпечення було обрано середовище MS Access, що забезпечує швидкий доступ і обробку великих масивів даних, не потребує спеціальної установки, бо є додатком MS Office При розробці програмного забезпечення було використано середовище Java, що відноситься до так званих систем швидкої розробки прикладних програм, і являє собою потужний генератор коду, візуальний дизайнер прикладних програм та засіб ведення баз даних. Java має можливість під'єднувати базу даних і дає можливість реалізувати різні складні додатки.

ЛІТЕРАТУРА

1. Шульга Н.П.Оцінка кредитоспроможності клієнта: Рекомендації банкіру при видачі кредиту. – К.: КІБ «Україна», 1995.
2. Kohonen T. 1984. Self–organization and associative memory. Series in Information Sciences, vol. 8. Berlin: Springer verlag.
3. Уоссермен З.Ф. Нейрокомпьютерная техника: теория и практика. – М.: Мир, 1992.
4. Саймон Хайкин Нейронные сети: полный курс, 2-е издание: Пер. с англ. – М.: Издательский дом «Вильямс», 2006. – 1104 с.
5. Дебок Г., Кохонен Т. Аналіз финансовых данных с помощью самоорганизующихся карт / Пер.с англ. – М.: Издательский Дом «АЛЬПИНА», 2001. – 317 с.
6. Руденко О.Г., Бодянський Є.В. Штучні нейронні мережі : Навчальний посібник. – Харків: ТОВ «Компанія СМІТ», 2006. – 404 с.

Рецензенти: д.т.н., проф. Данілов В.Я.
д.т.н., проф. Бідюк П.І.

© Кравець І.О., Шульженко Г.А., 2009 *Стаття надійшла до редколегії 16.01.09*