

Розроблена база даних забезпечить якісний та зручний облік виховної роботи куратора, також усіх необхідних відомостей про студентів певної групи, що здійснюється куратором.

Список використаних джерел

1. Пасічник В.В., Берко А.Ю., Верес О.М. Системи баз даних та знань. Книга 1. Організація баз даних та знань: Навч. посібник.- Львів: "Магнолія 2006", 2008.
2. Тимошенко Л.М., Ткач І.І., Біркова Н.М. Практикум з дисципліни «Організація баз даних». – Тернопіль: ТНЕУ, 2007.

УДК 004.832.32

ВИКОРИСТАННЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ РОЗВ'ЯЗАННЯ ЗАДАЧ КЛАСИФІКАЦІЇ ТА КЛАСТЕРИЗАЦІЇ

Гончар Л.І., Гончар Т.В., Козаревський В.В.

Тернопільський національний економічний університет

З точки зору інтелектуального аналізу даних задачі класифікації та кластеризації мають ряд специфічних ознак (великий обсяг інформації, наявність інформації у якісному та кількісному виглядах, необхідність автоматичного генерування гіпотез прихованих закономірностей) і потребують розробки специфічного алгоритмічного та програмного забезпечення. Нейронні мережі дозволяють вирішувати такі неформалізовані задачі.

З одного боку існують статистичні пакети, такі як 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 (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 надзвичайно дорогими. Тому розробка інтелектуально-аналітичної системи, яка б поєднувала базу даних та виконувала інтелектуальний аналіз даних в цьому режимі і була достатньо дешевою, є надзвичайно актуальною задачею [5].

На даний час ведуться інтенсивні роботи по розробці методів та алгоритмів для задач інтелектуального аналізу даних, у тому числі у банківській справі.

Задачею дослідження є розробка структури, алгоритмів навчання та вивчення можливостей нейронних мереж для розв'язання задач класифікації та кластеризації у банківській справі. Саме банківська справа є класичним прикладом використання інтелектуальних технологій. Необхідною задачею, яка виникає у банках, є віднесення клієнтів та партнерів банків до певних груп по їх характеристикам (депозити юридичних осіб, активи банків, прибутковість банків тощо), число яких може бути достатньо великим.

З точки зору інтелектуального аналізу даних це є задачі класифікації та кластеризації.

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 вибірки 30 прикладів. Другою була досліджена мережа з квантуванням навчального вектора. Мережа складається з вхідного прошарку із 13 нейронів та вихідного прошарку Кохонена з 2 нейронів. Переможцем є той нейрон, вектор ваг якого ω_c найближчий до вхідного образу x . Значення всіх ваг ω_j , що мінімізують помилку класифікації, обчислюються асимптотично.

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 протилежному випадку. Як показали результати досліджень, мережа з квантуванням навчального вектора дає кращі результати.

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 [2].

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 d_j від вхідного сигналу до кожного нейрона j визначаються за формулою:

$$d_j = \sum_{i=1}^N (x_i(t) \cdot w_{i,j}(t))^2 \quad (1)$$

де x_i – i -ий елемент вхідного сигналу в момент часу t , $w_{ij}(t)$ – 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 d_j є найменшою.

5. Налаштування ваг нейрона j^* і його сусідів: робиться налаштування ваг для нейрона j^* і всіх нейронів з його околу. Нові значення ваг:

$$w_{ij}(t+1) = w_{ij}(t) + r(t)(x_i(t) - w_{ij}(t)), \quad (2)$$

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

6. Якщо навчальна вибірка не закінчена, то відбувається повернення до кроку 2. В іншому випадку повторюємо цикл навчання знову. Число циклів навчання, кожен з яких включає всю вибірку (epoch), 10000.

7. Обчислення функції мети.

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

В алгоритмі використовується коефіцієнт швидкості навчання, який поступово зменшується, для точнішої корекції на новій епосі. У результаті позиція центру встановлюється в такій позиції, яка довільним чином кластеризує приклади, для яких даний нейрон є переможцем.

На підставі даних про розбиття клієнтів банку на 3 кластера було побудовано 2-вимірну мапу Кохонена для задач кластеризації (розмір 10×10 нейронів). При цьому 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 мітки нейрона-переможця, ваги якого найближче до ваг цього нейрону.

Графічне відображення результатів виконання алгоритму кластеризації допомагає більш наглядно зрозуміти загальний розподіл клієнтів на три кластери (надійні клієнти, середні та ризиковані). 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 є база знань, яка використовувалася для навчання і тестування, тому що робота алгоритму напряму залежить від правильних і однозначно чітко сформованих вхідних даних.

Таким чином, у результаті досліджень було виявлено, що:

- для задач класифікації більш ефективним є алгоритм класифікації з використанням мережі Кохонена та квантуванням навчального вектора, ніж алгоритм із використанням мережі зворотнього поширення похибки, який виявив залежність кількості ітерації від початкових значень ваг нейронів та зупинявся у точках локальних мінімумів похибки.

- для задач кластеризації більш доцільним є алгоритм кластеризації з використанням мережі Кохонена.

Було запропоновано власну модифікацію алгоритму кластеризації з використанням мережі Кохонена та оптимізацією числа вихідних нейронів або кластерів.

Створена інформаційно-аналітична система 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 введення баз даних.

Список використаних джерел

1. Дебок Г., Кохонен Т. Аналіз финансовых данных с помощью самоорганизующихся карт / Пер.с англ. – М.: Издательский Дом «АЛЬПИНА», 2001. – 317 с.
2. Руденко О.Г., Бодянский С.В. Штучні нейронні мережі : Навчальний посібник. – Харків: ТОВ «Компанія СМІТ», 2006. – 404 с.
3. Саймон Хайкин Нейронные сети: полный курс, 2-е издание: Пер. с англ.–М.: Издательский дом «Вильямс», 2006. – 1104с.
4. Уоссермен З.Ф. Нейрокомпьютерная техника: теория и практика. – М.: Мир, 1992.
5. Шульга Н.П.Оцінка кредитоспроможності клієнта: Рекомендації банкіру при видачі кредиту. – К.: КІБ «Україна», 1995.

УДК 681.3

ВИКОРИСТАННЯ БАНКІВ ДАНИХ В ІНФОРМАЦІЙНИХ СИСТЕМАХ

Гулова Г.В., Кінах Я.І.

Тернопільський національний економічний університет

I. Постановка проблеми

Банком даних (БНД) називають систему спеціальним чином організованих баз даних, програмних, технічних, мовних і організаційно-методичних засобів, призначених для забезпечення централізованого нагромадження і колективного багаточільового використання даних [1, 2].

II. Мета роботи

Метою дослідження є ґрунтовне знайомство з характерними головними рисами БНД

III. Характерні головні риси БНД

Бази даних створюються як правило для вирішення не однієї, декількох пов'язаних завдань, не одним, а групою користувачів.

У БНД є потужні засоби, що полегшують для користувачів роботу з даними (СУБД) [4]. Централізоване управління даними має переваги в порівнянні зі звичайною файловою системою:

- скорочення надмірності зберігання даних;
- скорочення трудомісткості розробки, експлуатації і модернізації ІС;
- забезпечення зручного доступу до даних як користувачам;
- професіоналам в області обробки даних, так і кінцевим користувачам.

IV. Головні вимоги, що пред'являються до БНД

- адекватність відображення предметної області;
- несуперечність даних, актуальність інформації;
- можливість взаємодії користувачів різних категорій, дуже висока ефективність доступу до даних;
- дружельюбність інтерфейсів, малий час на навчання;
- забезпечення секретності і розмежування доступу до даних для різних користувачів;
- надійність зберігання і захист даних.

V. База даних як ядро БНД

База даних - це поєднання сукупності взаємопов'язаних даних, що знаходяться під управлінням СУД [7, 10]. Метаінформація включає в себе опис структури БД (схеми і підсхеми), модель предметної області, інформацію про користувачів і їх права, опис форми вхідних і вихідних документів. Централізоване сховище метаінформації називається словником даних. Незвично невеличкі значення мають словники даних в системах автоматизованого проектування ІС.

VI. Програмні засоби СУБД

Програмні засоби СУБД підрозділяють на:

- ядро СУБД, яке забезпечує введення, висновок, обробку та зберігання даних у БД;
- транслятори, що забезпечують переклад мови СУБД на мову, що використовується ядром системи;