

В цьому випадку, необхідно прийняти до уваги, що окремі параметри, що приймаються складовими для $f(x)$ повинні бути незалежними. Це означає, що в кожній $F_i x^i$, змінні аргументи повинні бути зв'язані відповідними предикатом F_i . Практично, досягнути виконання цієї умови є складно, оскільки в рамках єдиного технологічного процесу виробництва книги, в більшості випадків параметри, що його характеризують і відповідно, визначають параметри продукту, які мають власну фізичну інтерпретацію, зв'язані між собою. Тим не менше, задовольнити сформульовану вимогу можна в процесі формування частини рекурентної моделі, що представляє собою опис інтерпретації елементів предметної області. Інші деталі та особливості використання рекурентних моделей доцільно аналізувати при їх побудові.

1. *Піх І.В.* Параметричне прогнозування якості друкованої продукції . Квалілогія книги. 36. наук. прощ. Львів, 2003, УАД вип. 6. с.200-213
2. *Боровков А.А.* Теория вероятностей. М.: Наука, 1976, -352с.
3. *Шуригин А.М.* Прикладная статистика: робастность, оценивания, прогноз. М.: Финансы и статистика, 2005. – 224 с.
4. *Шуригин А.М.* Робастность и устойчивость статистических оценок. / Статистика, вероятность, экономика. Сб., науч. ст. М.: Наука, с. 90-98
5. *Бродский В.З.* Модели факторных планов. – Тезисы докладов IV Всесоюзной конференции по планированию и автоматизации эксперимента в научных исследованиях, часть 1, М.: 1973, с. 36-38
6. *Бочаров П.П., Печинкин А.В.* Теория вероятностей. Математическая статистика. М.: Гардарика, 1998. – 328 ст.
7. *Ландо С.К.* Лекции о производящих функциях. М.: МЦНМО, 2004. – 144с.

Поступила 16.09.2010р.

УДК 681.322

Ю.В.Стех, к.т.н., доц. каф. САП, НУ «Львівська політехніка»,
Файсал М.Е. Сардіх, аспірант каф. САП НУ «Львівська політехніка»,
М.В.Лобур, д.т.н., професор, зав. каф. САП, НУ «Львівська політехніка»,
М.С. Домброва, аспірант каф. САП НУ «Львівська політехніка»,
В.Є. Арцибасов, аспірант каф. САП НУ «Львівська політехніка»

АДАПТИВНИЙ МЕТОД І АЛГОРИТМ ПОШУКУ ЦЕНТРІВ КЛАСТЕРІВ ЗА ДОПОМОГОЮ НЕЙРОННОЇ МЕРЕЖІ

Developed an adaptive method and algorithm for finding the cluster centers by using neural network. The developed method requires a set threshold. Proposed heuristic computational formula for determining the optimal value of threshold. An

experimental study of the algorithm is given. Developed an algorithm for modifying the value of a threshold, which allows you to determine the best centers of the clusters.

Keywords – method, algorithm, cluster

Розроблено адаптивний метод і алгоритм пошуку центрів кластерів за допомогою нейронної мережі. Розроблений метод вимагає задання порогової величини. Запропоновано евристичну розрахункову формулу визначення оптимального значення порогової величини. Приведені результати експериментального дослідження розробленого алгоритму. Розроблено алгоритм модифікації значення порогової величини, який дозволяє визначити оптимальні центри кластерів.

Ключові слова – метод, алгоритм, кластер

Вступ

Важливим параметром сучасних алгоритмів неієрархічної кластеризації є кількість заданих (очікуваних) кластерів. Цей параметр використовується в таких алгоритмах, як k-means, k-medoids, ISODATA та їх модифікаціях. Некоректне задання цих параметрів приводить до отримання кластерів, які не відповідають дійсному розподілу образів по класах. У випадку n-вимірного простору оцінити достовірність результатів кластеризації досить важко. Тому актуальною є задача визначення оптимальної початкової кількості кластерів

Розробка адаптивного методу і алгоритму пошуку центрів кластерів

Ефективним методом пошуку початкових центрів кластерів є запропонований у [2,4,6] алгоритм пошуку центрів кластерів за допомогою нейронної мережі. В цьому алгоритмі досліджувана множина точок образів представляється за допомогою нейронної мережі, де кожному образу ставиться у відповідність нейрон. Таким чином досліджувана множина точок образів у d-вимірному просторі перетворюється в певний неорієнтований зважений граф нейронної мережі. Схема роботи алгоритму представлена на рис. 1.

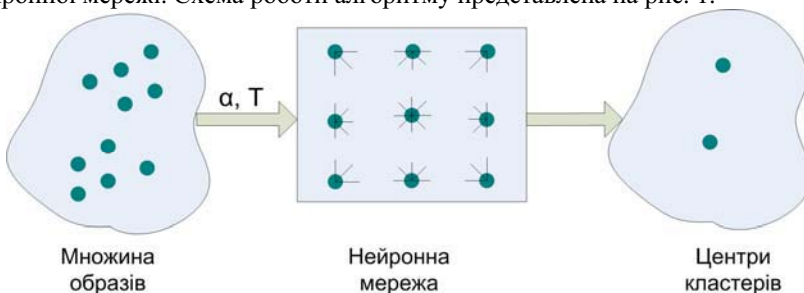


Рис. 1. Схема роботи алгоритму пошуку центрів кластерів за допомогою нейронної мережі

На початку для досліджуваної множини точок образів у d-вимірному просторі $\mathbf{X}_i = (x_{i1}, x_{i2}, \dots, x_{id}) \in \mathbf{R}^d$ обчислюється квадратна матриця евклідових відстаней між ними:

$$L = \begin{pmatrix} l_{11} & l_{12} & \dots & l_{1N} \\ l_{21} & l_{22} & \dots & l_{2N} \\ \dots & \dots & \dots & \dots \\ l_{N1} & l_{N2} & l_{N3} & l_{NN} \end{pmatrix}, \quad (1)$$

де $l_{ij} = \|\mathbf{X}_i - \mathbf{X}_j\|_2$, N – кількість точок образів у множині.

Кожна точка образу вважається нейроном з початковою активністю $S_i(0)$.

Робота алгоритму полягає у виконанні наступних кроків:

Крок 1. Для фіксованого наперед заданого значення порогової величини $T > 0$ визначаються валові коефіцієнти зв'язків між нейронами

$$w_{ij} = \begin{cases} \frac{T^2}{l_{ij} + T^2}, & \text{якщо } l_{ij} \leq T \\ 0, & \text{якщо } l_{ij} > T \end{cases} \quad (2)$$

Після виконання першого кроку множина точок образів представляється у вигляді повного неорієнтованого зваженого графа. Приклад структури такого графа для множини із шести точок образів наведений на рис. 2.

Крок 2. Обчислюються початкові активності нейронів.

$$S_i(0) = \sum_{j=1}^N w_{ij} \quad (3)$$

Крок 3. На подальших кроках здійснюється ітеративний перерахунок активностей нейронів:

$$S_i(t+1) = S_i(t) + \alpha \sum_{j=1}^N w_{ij} \cdot (S_i(t) - S_j(t)), \quad (4)$$

де α - параметр, який визначає швидкість навчання нейронної мережі, $\alpha \in (0,1)$.

Крок 4. Якщо в процесі зміни активності $S_i(t) < 0$, тоді $S_i(t) = 0$.

Алгоритм працює таким чином, що нейрони, які знаходяться на межі кластерних областей, передають свої активності нейронам, які знаходяться всередині областей кластерів. Процес навчання нейронної мережі сходиться до такого результату, коли в кожній області кластера залишається лише один активний нейрон – центр кластера. Однак, можливий такий варіант роботи алгоритму, коли виникають згустки центрів кластерів (декілька центрів кластерів у одній області). При цьому число нейронів з додатною активністю не змінюється. В такому випадку виконується корекція значення порогової величини (Рис.3):

$$T = T + \Delta T, \quad (5)$$

де $\Delta T = 0,01 \cdot T$.

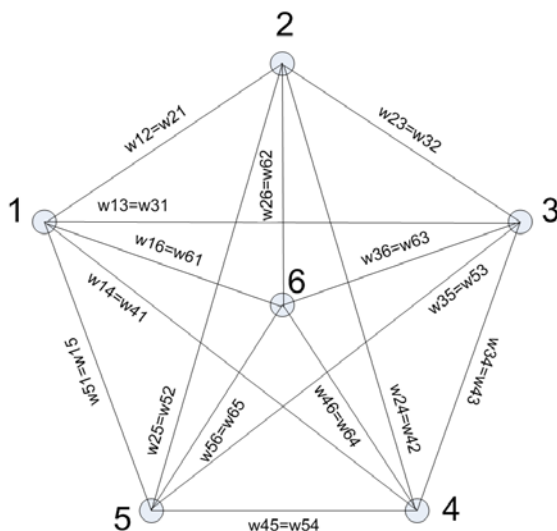
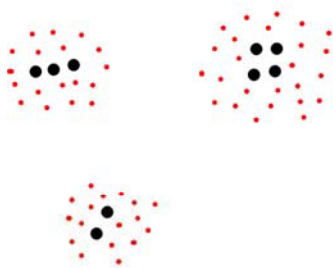


Рис. 2. Повний неорієнтований зважений граф нейронної мережі для шести образів

Умови закінчення роботи алгоритму і необхідності корекції значення порогової величини перевіряє користувач у діалоговому режимі. При цьому він контролює кількість нейронів з додатними значеннями активностей і величини елементів матриці евклідових відстаней між ними.



$$D_{ij} = \|\bar{x}_i - \bar{x}_j\|$$

$$D_{\max} = \max_{i,j} (D_{ij}), i \neq j$$

$$M_l = \sum_{i=1}^{K_l} m_i$$

$$M_{l+1} = \sum_{i=1}^{K_{l+1}} m_i$$

$$M_{l+1} = M_l$$

$$D_{ij} = \|\bar{x}_i - \bar{x}_j\|, i \neq j; i, j \in L$$

L – множина індексів отриманих центрів

$$MAX = \max_{i,j} \{D_{ij}\}, i \neq j$$

$$MAX > 0,5D_{\max}$$

$$T = T + 0,01\Delta T$$

Рис. 3 Алгоритм корекції значення порогової величини

Результати експериментального дослідження

На рис.4 показано тестову двовимірну область яка містить 50 точок розподілених по 5 кластерах. На рис.5 наведений графік залежності $K(T)$ для заданої тестової області.

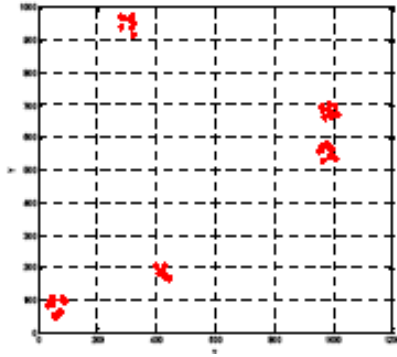


Рис.4 Двовимірна тестова область з п'ятьма кластерами

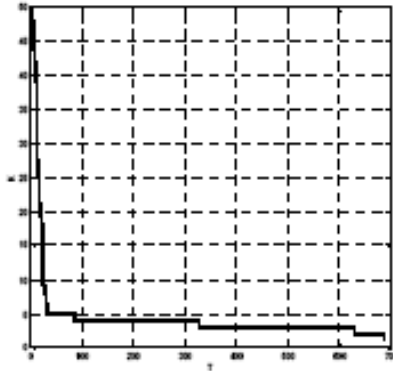


Рис.5 Графік залежності $K(T)$ для двовимірної тестової області з п'ятьма кластерами

Із даного графіка видно, що на початковій стадії зміни T кількість кластерів змінюється дуже швидко. Із зростанням T кількість кластерів стабілізується на рівні $K = 5$ і в подальшому майже не змінюється.

На рис.6 показано тестову двовимірну область, яка містить 10 кластерів.

На рис.7 показано графік залежності $K(T)$ для заданої тестової області.

Характер графічної залежності $K(T)$ аналогічний попередньому випадку. Із зростанням T кількість кластерів стабілізується при $K = 10$. Експериментальні результати показують, що доцільно вибирати T за наступною формулою:

$$T = \frac{0,1 * \max(l_{ij})}{2}$$

де l_{ij} - елемент матриці евклідових відстаней між точками образів (1).

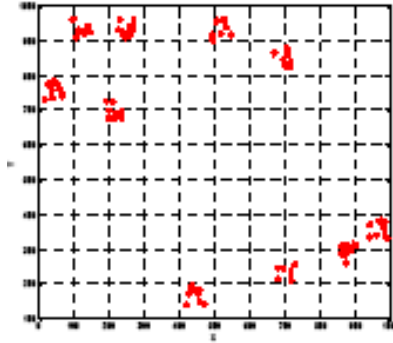


Рис.6 Двовимірна тестова область із десятима кластерами

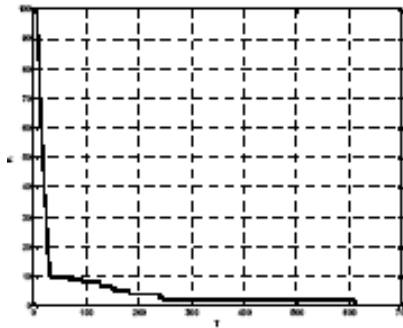


Рис.7 Графік залежності $K(T)$ для двовимірної тестової області з десятима кластерами

Висновки

Розроблений адаптивний метод і алгоритм пошуку центрів кластерів за допомогою нейронної мережі. Критичним значенням в розробленому алгоритмі є початкове значення порогової величини T . На основі експериментальних досліджень запропонована евристична формула розрахунку початкового значення порогової величини. Розроблено алгоритм адаптації, котрий дозволяє коректувати значенням порогової величини в діалоговому режимі для отримання в кожній кластерній області одного центра кластера.

1. Барсегян А.А., Курпянов М.С., Степаненко В.В., Холод И.И. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP.- СПб.: БХВ – Петербург, 2007.-384с.
2. Ким Дж.-О., Мьюллер Ч.У., Клекка У.Р. Факторный, дискриминантный и кластерный анализ. Пер. с англ. – М.: Финансы и статистика, 1989.-215с.
3. Ту Дж., Гонсалес Р. Принципы распознавания образов. Пер. с англ. – М.: Мир, 1978. – 411с.
4. Классификация и кластер/ под ред. Дж. Вэн Райзина. Пер. с англ.-М.: Мир, - 389с.

5. *A. J. Jain, M. N. Murty, P.J. Flynn* Data clustering: a review // *ACM Computing Surveys*, - 1999 - v..31, № 3 – P. 264-323.
6. *Стех Ю.В., Файсал М.Е. Сардіх, Лобур М.В., Керницький А.Б.* Алгоритм пошуку оптимальної кількості кластерів // *Вісник Національного університету «Львівська політехніка»*. – 2009.- №651 – С. 129-132.

Поступила 2.08.2010р.