

## КЛАСИФІКАЦІЯ ОБ'ЄКТІВ З ВИКОРИСТАННЯМ МЕТОДІВ ШТУЧНОГО ІНТЕЛЕКТУ КЛАСТЕРНОГО АНАЛІЗУ

Періг В.М.

Тернопільський національний економічний університет

### I. Постановка проблеми

Одним із підходів до вирішення завдання класифікації об'єктів по сукупності ознак, які їх описують, є кластерний аналіз – багатомірна статистична процедура, яка класифікує об'єкти або спостереження в однорідні групи (кластери). Кластери формують таким чином, щоб до складу кожного з них входили близькі за ознаками об'єкти. Елементи різних кластерів, зазвичай, суттєво відрізняються.

Головною метою кластерного аналізу є визначення множин подібних об'єктів у вибірці. Відсутність жорсткого визначення вимог до об'єктів класифікації робить кластерний аналіз універсальним інструментом. Саме тому спектр його прикладних застосувань є надзвичайно широким. Однак, універсальність застосування призвела до появи несумісних методів і підходів, які в окремих випадках ускладнюють однозначне використання та коректну інтерпретацію результатів кластерного аналізу. Тому на сьогоднішній день актуальним є аналіз можливості застосування та адаптація методів кластерного аналізу до вирішення задачі класифікації інформаційно-телекомунікаційних систем.

### II. Мета роботи

Метою дослідження є аналіз різних підходів до вирішення основної задачі кластеризації – визначення відображення

$$\alpha: X \rightarrow Y, \quad (1)$$

за яким кожному об'єкту  $x_i \in X$  ставиться у відповідність кластер  $y_j \in Y$ . В окремих випадках множина  $Y$  може бути визначеною завчасно. Однак, зазвичай, її визначення проводиться при кластеризації. У цьому випадку необхідно знайти потужність множини кластерів і об'єкти (кластери), які до неї входять.

Розглянуто такі методи кластеризації: K-means (к-середніх) – є найбільш популярним при вирішенні широкого кола прикладних задач кластеризації інформації; графові – в їх основу покладено використання об'єктів біграфів для побудови кластерів; статистичні – передбачають використання методів статистичного аналізу початкових даних, які описують об'єкти кластеризації для встановлення спільних ознак і подальшого їх поділу на кластери; евристичні (FOREL, FOREL2) – в основі роботи відповідного алгоритму покладено використання принципу компактності: близьким за змістом об'єктам в просторі ознак відповідають окремі множини (кластери).

Усі проаналізовані методи кластеризації інформації зазвичай ґрунтуються на визначенні метрики в тому чи іншому вигляді. Відстань між об'єктами використовується для визначення областей близьких об'єктів (кластерів). Однак, в окремих задачах, зокрема в задачі класифікації інформаційно-телекомунікаційних систем, визначення метрики є складним і не завжди однозначним. Тому для кластеризації інформаційно-телекомунікаційних систем потрібно використати інший підхід, який не потребує явного визначення метрики.

Проведення кластеризації без визначення метрики на просторі ознак об'єктів є можливим у випадку використання методів штучного інтелекту, зокрема, нейронних мереж. При вирішенні таких задач традиційним є застосування нейронних мереж Кохонена. Вона принципово відрізняється від багатьох інших нейромереж. Основна її особливість полягає у використанні неконтрольованого навчання – навчальна множина складається лише із значень вхідних змінних. Основним елементом цих нейромереж є шар Кохонена, який складається з лінійних формальних нейронів. Ці нейрони представляють собою адаптивні лінійні суматори.

У нейромережі Кохонена використовується шар паралельно працюючих лінійних елементів (моделей нейронів). Усі вони мають однакову кількість входів ( $n$ ), на які отримують один і той самий вектор вхідних сигналів  $x=(x_1, \dots, x_n)$ . Проміжний результат на виході  $i$ -го нейрона обчислюється наступним чином:

$$y_i = v_{i0} + \sum_{j=1}^n v_{ij} \cdot x_j, \quad (2)$$

де  $u_i$  – проміжний сигнал на виході  $i$ -того нейрона,  $v_{ij}$  – вагові коефіцієнти,  $v_{i0}$  – коефіцієнт зсуву,  $x_i$  – вхідні сигнали.

Для вирішення задачі кластеризації інформаційно-телекомунікаційних систем (ІТС) розроблено програмне забезпечення, в якому змодельована робота нейромережі Кохонена. Для опису ІТС використано сукупність ознак, яка формується користувачем на попередньому етапі. Ознаки можуть включати множину фіксованих значень. У цьому випадку кожне з можливих значень описується текстовим рядком і кодується (автоматично) числовим значенням. Важливо при формуванні таких ознак задавати можливі значення у вірній послідовності, яка відповідає змісту ознаки. Інший тип ознаки – логічна, що може приймати лише 2 значення: так (кодується 1) або ні (кодується 0). Передбачена також можливість використання числових ознак. Слід відмітити, що оскільки в основі роботи нейромережі Кохонена покладена Евклідова метрика, надзвичайно важливим є адекватне кодування ознак. Більша різниця в числових значеннях має відповідати більшій відмінності характеристик, і навпаки. Це підтверджено при тестуванні програми.

Після формування сукупності ознак для побудови нейромережі Кохонена в програмі передбачено її навчання, яке у цій нейромережі полягає в обчисленні відповідних коефіцієнтів. Початковими даними для навчання нейромережі є центри кластерів – сукупності ознак, які найбільш точно відповідають окремим класам інформаційно-телекомунікаційних систем, за якими буде проведена наступна класифікація.

Після формування нейромережі існує можливість її збереження разом з описом ознак у файл для подальшого використання. Надалі програма надає можливість охарактеризувати певну ІТС за вказаними ознаками, визначити, до якого класу вона належить.

Дослідна експлуатація програми показала, що її можна застосовувати для вирішення задач класифікації ІТС.

### **III. Висновки**

Перевага використання нейромережі у порівнянні з класичним алгоритмічним підходом полягає у можливості швидкої паралельної обробки інформації. Проте, у випадку задач з порівняно невеликою кількістю ознак (параметрів) об'єктів кластеризації обчислювальної потужності сучасних ЕОМ більш, ніж достатньо для використання класичного підходу (обчислення евклідових відстаней і пошук мінімальної відстані). І лише у випадку значної кількості ознак доцільним є використання нейромережі Кохонена. При цьому існує можливість навіть при емуляції нейромережі на ПЕОМ досягти значного підвищення ефективності обробки інформації за рахунок паралельної емуляції роботи окремих нейронів на різних ядрах мікропроцесора.

Одним з недоліків існуючих підходів до використання нейромереж Кохонена є намагання понизити розмірність до одномірного (шар нейронів) і двомірного (мережа нейронів) варіанту для зручності інтерпретації результатів обробки і відображення на екрані монітора. Однак, внаслідок такого обмеження взаємозв'язків між нейронами до утворення лише одномірних або двомірних структур суттєво знижуються потенційні можливості таких нейромереж.

Для одношарових нейромереж існує проблема лінійної подільності. Нейромережа Кохонена аналогічна до лінійного поділу простору ознак на багатокутники, тобто кластери при використанні нейромережі Кохонена можуть мати лише форму багатокутника. Лінійна подільність обмежує нейромережі Кохонена лише тими задачами кластеризації, в яких множини точок в просторі ознак, які відповідають початковим значенням, можуть бути розділеними лінійно. Можуть виникати і задачі, в яких кластери мають більш складну форму, і звести її до лінійної неможливо.

У зв'язку з цим, розглянемо можливість застосування для кластеризації ІТС нейромереж з більш складною структурою.

### **Список використаних джерел**

1. Борисов А.Н. Обработка нечеткой информации в системах принятия решений / Борисов А.Н., Алексеев А.В., Меркурьева Г.В. — М. : Радио и связь, 1989. — 632 с.
2. Загоруйко Н.Г. Прикладные методы анализа данных и знаний / Загоруйко Н.Г. — Новосибирск, 1999. — 434 с.
3. Вишневыский В.М. Теоретические основы проектирования компьютерных сетей / Вишневыский В.М. — М. : Техносфера, 2003. — 512 с.
4. Основы В.Н. построения телекоммуникационных систем и сетей / [Гордиенко В.Н., Крухмалева А.Д. и др.] , под ред. Гордиенко В.Н. и Крухмалева А.Д. — М. : Горячая линия - Телеком, 2004. — 510 с.
5. Гольштейн А.Б. Интеллектуальные сети / Гольштейн А.Б., Ехриель И.М., Рерле Р.Д. — М. : Радио и связь, 2000. — 500 с.
6. Советов Б.Я. Построение сетей интегрального обслуживания / Советов Б.Я., Яковлев С.А. — Л. : Машиностроение, 1990. — 332 с.
7. Гольштейн А.Б., Гольштейн Б.С. Технология и протоколы MPLS / Гольштейн А.Б., Гольштейн Б.С. — С-Пб. : БХВ-Петербург, 2003. — 304 с.