

УДК 519.86: 004.032.26

В.Р. Петренко, К.О. Куделіна

## **КЛАСТЕРИЗАЦІЯ ЧАСОВИХ ПОСЛІДОВНОСТЕЙ З ВИКОРИСТАННЯМ НЕЙРОННОЇ МЕРЕЖІ КОХОНЕНА**

### **Вступ**

При обробці результатів експериментальних досліджень часто доводиться мати справу з великою кількістю випадкових часових послідовностей (ВЧП). Завдання дослідження при цьому, як правило, полягає в моделюванні цих ВЧП, наприклад, з метою прогнозування їхніх майбутніх значень.

Процедури синтезу моделей ВЧП [1, 2] є досить трудомісткими, і у випадку, коли потужність множини досліджуваних послідовностей велика, актуальною стає задача скорочення загальної трудомісткості досліджень. Найбільш природним підходом до її рішення є попереднє зменшення потужності множини досліджуваних ВЧП, що можливо в тих випадках, якщо деякі з досліджуваних ВЧП мають майже однакові значення імовірнісних характеристик, що допускає можливість опису цих послідовностей однією й тією ж моделлю. При цьому рішення задачі моделювання здійснюється шляхом послідовної реалізації наступних етапів:

- розділення вихідної множини послідовностей на деякі непересічні підмножини послідовностей із близькими значеннями характеристик;
- визначення усередненого вектора характеристик для кожної підмножини;
- синтез моделей для всіх підмножин шляхом використання усереднених векторів характеристик.

У даній роботі розглядається реалізація першого етапу, який можна інтерпретувати як рішення задачі кластеризації, тобто процес об'єднання в групи об'єктів, що мають схожі ознаки, коли ні групи, ні їхня кількість заздалегідь не визначені й формуються в процесі роботи системи виходячи з певної міри близькості об'єктів.

Завдання кластеризації часових послідовностей досліджувалися у деяких роботах [5,6]. Автори роботи [5] використовували в якості

вектора класифікаційних ознак вектор значень самих послідовностей, тобто розмірність вектора ознак дорівнювала довжині часових послідовностей. У роботі [6] нами запропоновано в якості вектору класифікаційних ознак використовувати вектор статистичних оцінок основних імовірнісних характеристик послідовностей (середнє, дисперсія, вибіркові статистично значимі автокореляції), що дозволяє значно скоротити розмірність класифікаційного простору при аналізі стаціонарних часових послідовностей, оскільки їх автокореляційні функції швидко загасають.

### Постановка задачі

Виходячи з того, що використання нейромережевих методів у задачах класифікації одержало в останній час широке поширення [3], в даній роботі для кластеризації множини часових послідовностей пропонується використовувати нейронну мережу Кохонена. Мета роботи – дослідження можливості використання мережі Кохонена для вирішення зазначеної задачі.

### Основна частина

#### Нейронна мережа Кохонена

Штучна нейронна мережа Кохонена або карта ознак, що самоорганізується (SOM) була запропонована Тойво Кохоненом на початку 80-х рр. [7]. Мережа розпізнає кластери в навчальних даних і розподіляє дані по відповідних кластерах. Якщо в майбутньому мережа зустрічається з набором даних, несхожих ні з одним із відомих зразків, вона відносить його до нового кластеру.

Мережа Кохонена являє собою двошарову мережу (рис.1). Кожний нейрон першого (розподільного) шару з'єднаний з усіма нейронами другого (вихідного) шару, які розташовані у вигляді двовимірної решітки.

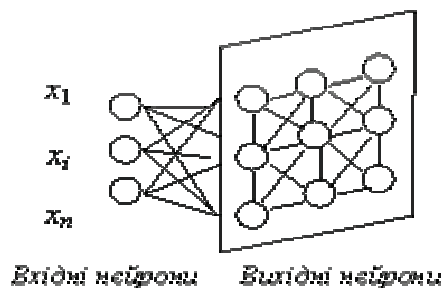


Рисунок 1 – Топологія нейронної мережі Кохонена

Нейрони вихідного шару називаються кластерними елементами, їх кількість визначає максимальну кількість груп, на які система може розділити вхідні дані. Збільшуючи кількість нейронів другого шару можна збільшувати деталізацію результатів процесу кластеризації.

Система працює за принципом змагання [3] – нейрони другого шару змагаються один з одним, перемагає той елемент-нейрон, чий вектор вагів ближче всього до вхідного вектору сигналів. За міру близькості двох векторів зазвичай беруть евклідову відстань між ними. Таким чином, кожний вхідний вектор відноситься до деякого кластерного елемента.

Основний ітераційний алгоритм Кохонена послідовно проходить ряд епох, на кожній епосі опрацьовується один навчальний приклад. Вхідні сигнали – вектори дійсних чисел – послідовно пред'являються мережі. Бажані вихідні сигнали не визначаються. Після пред'явлення достатнього числа вхідних векторів, синаптичні ваги мережі визначають кластери. Крім того, ваги організуються так, що топологічно близькі вузли чуттєві до схожих вхідних сигналів.

Для вихідного нейрона-переможця  $i$  для його сусідів з околу обновлюються вектори вагів за правилом:

$$w_{ij}(t+1) = w_{ij}(t) + e(t)h(t, j, m)(x_i - w_{ij}(t)) \quad , \quad (1)$$

де

$w_{ij}$  – значення вагового коефіцієнту зв'язку вхідного нейрона  $i$  та вихідного нейрона  $j$  в момент часу  $t$ ;

$h(t, j, m)$  – значення функції околу (наприклад, функція Гауса) з центральним нейроном вихідного шару  $m$  для нейрона вихідного шару  $j$  в момент часу  $t$ ;

$e(t)$  – коефіцієнт швидкості навчання в момент часу  $t$ ;

$x_i$  – вихід нейрона першого шару номер  $i$ .

В алгоритмі навчання мережі 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.

Властивість топологічної впорядкованості досягається в алгоритмі за допомогою використання поняття околу. Окіл – це декілька нейронів, що оточують нейрон-переможець. Розмір околу

поступово зменшується, так, що спочатку до нього належить досить велике число нейронів (можливо вся карта), на самих останніх етапах окіл стає нульовим і складається лише з нейрона-переможця. Мережа формує грубу структуру топологічного порядку, при якій схожі приклади активують групи нейронів, що близько знаходяться на топологічній карті. З кожною новою епохою швидкість навчання і розмір околу зменшуються, тим самим всередині ділянок карти виявляються більш тонкі розходження, що зрештою призводить до точнішого налаштування кожного нейрона.

Після того, як мережа навчена розпізнаванню структури даних, її можна використовувати як засіб візуалізації при аналізі даних.

### Використання мережі Кохонена для кластеризації множини ВЧП

Формування навчальних даних було здійснено шляхом генерації ВЧП з використанням моделі АРКС  $(p, q)$  з трьома різними наборами значень  $p, q, F_i, i = \overline{1, p}, \theta_j, j = \overline{1, q}, \theta_0, y_a^2$ .

$$x_t = F_1 \cdot x_{t-1} + F_2 \cdot x_{t-2} + \dots + F_p \cdot x_{t-p} + \theta_1 \cdot a_{t-1} + \dots + \theta_q \cdot a_{t-q} + \theta_0 + a_t, \quad (2)$$

де  $a_t \sim N(0, y_a^2)$ .

Конкретні значення параметрів моделі, що генерується, разом з початковими умовами та вибірковими оцінками імовірнісних характеристик 12 згенерованих часових послідовностей представлені в таблиці 1.

Таблиця 1

#### Параметри генерації та оцінки характеристик послідовностей

№ посл	Параметри моделі					Початк. умови			Оцінки характеристик					
	$F_1$	$F_2$	$\theta_1$	$\theta_0$	$\sigma_a^2$	$x_0$	$x_1$	$a_0$	$\bar{x}$	$\sigma_x^2$	$r(1)$	$r(2)$	$r(3)$	$r(4)$
1	0.6	0.35	—	0.05	0.5	1.15	1.05	0	1.122	1.713	0.908	0.865	0.796	0.729
2	-0.87	—	—	1.87	0.8	1.0	—	0	0.897	1.941	-0.748	0.646	-0.531	0.432
3	—	—	0.2	1.0	1.0	—	—	0	1.283	1.453	0.089	-0.049	-0.192	-0.122
4	—	—	0.2	1.0	1.0	—	—	0	1.033	1.420	-0.082	-0.010	0.271	-0.085
5	-0.87	—	—	1.87	0.8	1.15	—	0	0.894	1.907	-0.742	0.639	-0.522	0.423
6	0.6	0.35	—	0.05	0.5	1.1	1.2	0	1.161	1.734	0.910	0.866	0.798	0.732
7	-0.87	—	—	1.87	0.8	1.25	—	0	0.909	1.798	-0.705	0.620	-0.533	0.427
8	0.6	0.35	—	0.05	0.5	1.25	1.15	0	1.164	1.735	0.910	0.866	0.798	0.732
9	—	—	0.2	1.0	1.0	—	—	0	0.736	0.857	0.130	0.204	0.051	-0.096
10	—	—	0.2	1.0	1.0	—	—	0	0.989	1.457	0.254	0.007	-0.134	-0.082
11	0.6	0.35	—	0.05	0.5	1.2	1.18	0	1.222	1.605	0.823	0.820	0.781	0.707
12	-0.87	—	—	1.87	0.8	1.1	—	0	0.861	3.020	-0.743	0.641	-0.524	0.432

Мережа, що розробляється, має 6 входів, що відповідає кількості вибірових оцінок імовірнісних характеристик. Топологічна карта була обрана розміром 6x6. Таким чином кількість елементів вихідного шару дорівнює 36, а його ширина – 6.

Для побудови та навчання нейронної мережі Кохонена використовувався програмний пакет STATISTICA Neural Networks [4].

Навчальна множина складала 12 згенерованих послідовностей, з трьома різними наборами значень моделі АРКС ( $p, q$ ).

В результаті навчання побудованої мережі були ідентифіковані три кластери згідно з елементами-переможцями:  $G1 = \{1, 6, 8, 11\}$ ,  $G2 = \{2, 5, 7, 12\}$ ,  $G3 = \{3, 4, 9, 10\}$ . На рис. 2 приведено приклад топологічної карти з елементом-переможцем кластеру  $G3$ , де Case No=3 вказує на приналежність третьої часової послідовності кластеру  $G3$ .

Отримані результати у вигляді кластерів, кількість яких дорівнює кількості наборів параметрів, відповідає правдивій структурі множини досліджуваних послідовностей.

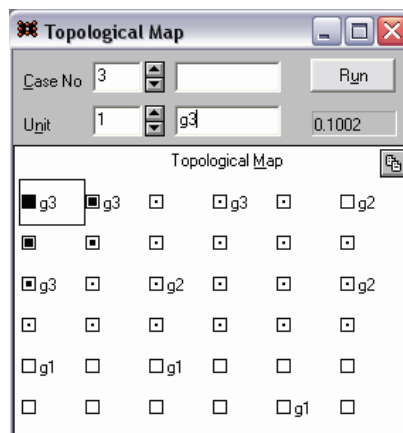


Рисунок 2 – Топологічна карта

### Висновки

Отримані результати підтверджують можливість використання нейронної мережі Кохонена для кластеризації множини ВЧП при вирішенні задачі їх моделювання. Запропонований підхід може бути використаний в складі спеціального програмного забезпечення автоматизованої системи оперативного управління ходом виробництва напівпровідникових матеріалів для визначення груп марок з ідентичною динамікою коефіцієнта виходу в кондиційну продукцію.

### СПИСОК ЛИТЕРАТУРЫ

1. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. Выпуск 1. пер. с англ. Левшина А.Л. – М.: Мир, 1974. – 406 с.
2. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования. – М.: Статистика, 1979. – 254 с
3. Головки В.А. Нейронные сети: обучение, организация и применение. Под общ. ред. А. И. Галушкина. – М: ИПРЖР, 2001. – 256 с
4. Нейронные сети. STATISTICA Neural Networks. Пер. с англ. – М.: Горячая линия-Телеком, 2000.– 182 с.
5. Вавилов Б.Т., Лунин Ю.В. О кластеризации временных рядов производственно-экономических показателей. //Вопросы радиоэлектроники. Серия АСУ. – 1979. – вып.3. – с. 71-80.
6. В.Р. Петренко, К.А. Куделина, В.А. Самоляк. Применение кластеризации в задаче моделирования стационарных временных последовательностей. //Новые технологии. № 4 (18). – 2007. – с. 90-94.
7. Т. Kohonen. “Self-organizing Maps”, Series in Information Sciences. – Springer, 1995.

Получено 20.03.2008 г.