

8.5. Лабораторна робота №5 «Аналітичні рішення на основі самоорганізуючих карт Кохонена».

Завдання до лабораторної роботи.

1. Відповідно до заданого варіанту підготувати необхідні дані у вигляді таблиць MS Excel і зберегти їх як персональні файли. Для підготовки даних використовувати тематичні сайти Інтернет, результати проходження практик, довідники і каталоги.

2. Виконати завантаження даних з таблиць MS Excel за допомогою Майстра імпорту відповідно до методики, представленої в розділі «Порядок виконання роботи».

3. Отримати самоорганізуючі карти Кохонена відповідно до методики, представленої в розділі «Порядок виконання роботи».

4. Виконати економічний аналіз різних ситуацій, змінюючи основні дані на картах як у бік збільшення, так і у бік їх зменшення відповідно до методики, представленої в розділі «Порядок виконання роботи».

Варіанти завдань:

N варіанту	Галузь
1	Оцінка ефективності вкладень в рекламу
2	Оцінка ефективності вкладень в будівництво
3	Оцінка ефективності вкладень в курорти
4	Оцінка ефективності вкладень в промисловість
5	Аналіз ринку ЗМІ
6	Аналіз ринку нерухомості
7	Аналіз ринку книжкової продукції
8	Аналіз ринку медичних препаратів
9	Аналіз ринку комп'ютерної техніки
10	Аналіз ринку техніки зв'язку
11	Створення профілів клієнтів (розділення на групи) по суспільних інтересах
12	Створення профілів клієнтів (розділення на групи) по спортивних інтересах
13	Створення профілів клієнтів (розділення на групи) по наукових інтересах

Теоретична частина.

Самоорганізуючі карти можуть використовуватися для вирішення таких задач як моделювання, прогнозування, пошук закономірностей у великих масивах даних, виявлення наборів незалежних ознак і стискування інформації.

Алгоритм функціонування самоорганізуючих карт (Self Organizing Maps - SOM) є одним з варіантів кластеризації багатовимірних векторів - алгоритм проектування із збереженням топологічної подібності. Прикладом таких алгоритмів може служити алгоритм k -найближчих середніх (c-means). Важливою відмінністю алгоритму SOM є те, що в ньому всі нейрони (вузли, центри класів) впорядковані в деяку структуру (зазвичай двовимірну сітку). При цьому, в ході навчання модифікується не лише нейрон-переможець (нейрон карти, який найбільшою мірою відповідає вектору входів і визначає до якого класу відноситься приклад), але і його сусіди, хоча і у меншій мірі. За рахунок цього SOM можна вважати одним з методів проектування багатовимірного простору в простір з нижчою розмірністю. При використанні

цього алгоритму, вектора, близькі у вихідному просторі, виявляються поруч і на отриманій карті.

SOM здійснює використання впорядкованої структури нейронів. Зазвичай використовуються одно- і двовимірні сітки. При цьому кожен нейрон є n -мірним вектором-стовпцем, де n визначається розмірністю вхідного простору (розмірністю вхідних векторів). Вживання одно- і двовимірних сіток пов'язане з тим, що виникають проблеми при відображенні просторових структур більшої розмірності (при цьому знову виникають проблеми з пониженням розмірності до двовимірної, уявної на моніторі). Зазвичай, нейрони розташовуються у вузлах двовимірної сітки з прямокутними або шестикутними чарунками. При цьому, як було сказано вище, нейрони також взаємодіють один з одним. Величина цієї взаємодії визначається відстанню між нейронами на карті. При реалізації алгоритму SOM заздалегідь задається конфігурація сітки (прямокутна або шестикутна), а також кількість нейронів в мережі. Деякі джерела рекомендують використовувати максимально можливу кількість нейронів в карті. При цьому початковий радіус навчання (neighborhood в англійській літературі) в значній мірі впливає на здатність узагальнення за допомогою отриманої карти. У разі, коли кількість вузлів карти перевищує кількість прикладів в навчальній вибірці, успіх використання алгоритму у великій мірі залежить від відповідного вибору початкового радіусу навчання. Проте, у разі, коли розмір карти складає десятки тисяч нейронів, час, потрібний на навчання карти, зазвичай буває дуже велике для вирішення практичних задач. Таким чином, необхідно досягати допустимий компроміс при виборі кількості вузлів.

Підготовка навчальної вибірки. Відмінність в підготовці навчальної вибірки в цьому алгоритмі полягає в тому, що *вихідні поля в такій вибірці можуть бути відсутніми зовсім*. Навіть якщо в навчальній вибірці будуть присутні вихідні поля, вони не братимуть участь при навчанні нейромережі. Проте вони братимуть участь при відображенні карт. Нормалізація полів тут

така ж, як для дерев рішень. Навчальна вибірка налаштовується так само, як і для нейромережі і дерева рішень.

Навчання. Перед початком навчання карти необхідне проініціалізувати вагові коефіцієнти нейронів. Вдало вибраний спосіб ініціалізації може істотно прискорити навчання і привести до здобуття якісніших результатів. Існують три способи ініціалізації початкових вагів:

- ініціалізація випадковими значеннями, коли всім вагам даються малі випадкові величини;
- ініціалізація прикладами, коли в якості початкових значень задаються значення випадково вибраних прикладів з навчальної вибірки;
- лінійна ініціалізація. В цьому випадку ваги ініціюються значеннями векторів, лінійно впорядкованих уздовж лінійного підпростору, який проходить між двома головними власними векторами вхідного набору даних.

Візуалізація. Отриману в результаті навчання карту можна представити у вигляді прошаркового пирога, кожен шар якого є розфарбовуванням, породженим однією з компонент вхідних даних. Отриманий набір розфарбовувань може використовуватися для аналізу закономірностей, які є між компонентами набору даних. Після формування карти, отримується набір вузлів, який можна відображати у вигляді двовимірної картинки. При цьому кожному вузлу карти можна поставити у відповідність ділянку на малюнку (чотири або шестикутну), координати якої визначаються координатами відповідного вузла в ґратах. Тепер для візуалізації залишається лише визначити колір чарунок цієї картинки. Для цього і використовуються значення компонент. Найпростіший варіант - використання градацій сірого. В цьому випадку чарунки, відповідні вузлам карти, в які попали елементи з мінімальними значеннями компонента або не попало взагалі жодного запису, будуть змальовані чорним кольором, а чарунки, в які попали записи з максимальними значеннями такого компонента, відповідатимуть чарункам білого кольору. В принципі можна використовувати будь-яку градієнтну палітру для розфарбовування. Отримані розфарбовування в сукупності

утворюють атлас, що відображає розташування компонент, зв'язки між ними, а також відносне розташування різних значень компонент.

Порядок виконання роботи.

1. Продовжимо приклад з кредитуванням фізичних осіб. За допомогою карт Кохонена можна розглянути залежності між різними характеристиками кредиторів і виділити сегменти кредиторів, об'єднавши їх по схожих ознаках. Навчальна вибірка буде така ж, що і для нейромереж і дерев рішень. Але поле «Автомобіль» використовувати не будемо. Це недоцільно, оскільки дані, що подаються на вхід карт Кохонена мають бути такими, аби між ними можна було обчислити відстань або була можливість розташувати їх в порядку зростання або убутання. Значення ж «вітчизняна», «імпортна» і «немає автомобіля» порівнювати між собою не можна (рис. 8.24).

Сумма кредита	Возраст	Образование	Площадь	Срок проживания	Давать кредит
7000	37	Специальное	37	22	да
7500	38	Среднее	29	12	да
14500	60	Высшее	34	30	нет
15000	28	Специальное	14	21	да
32000	59	Специальное	53	29	да
11500	25	Специальное	28	9	да
5000	57	Специальное	18	34	да
61500	29	Высшее	26	18	нет
13500	37	Специальное	46	26	нет
25000	36	Специальное	20	21	нет
25500	66	Высшее	45	30	нет

Рис. 8.24. Навчальна вибірка для отримання карт Кохонена.

2. Застосуємо «Мастер обработки» і в першому вікні вибираємо «Карта Кохонена». Нормалізація полів буде така ж, як і для дерев рішень. Розбиття навчальної вибірки на дві множини залишимо за умовчанням, так само, як і останні налаштування.

3. Кроки 5 – 8 виконуються за умовчанням. На кроці 9 вибираємо режим «Карты Кохонена». Крок 10 дозволяє вибрати в лівому вікні допустимі відображення карт. Поставимо відмітки у вхідних і вихідному стовпцях.

4. Результати роботи алгоритму відображаються на картах. Кожному вхідному полю відповідає своя карта. Наприклад, в один і той же кластер були згруповані молоді кредитори з маленьким терміном мешкання в даній місцевості, спеціальною або середньою освітою, середньою житлоплощею, що беруть високі суми кредиту (рис. 8.25).

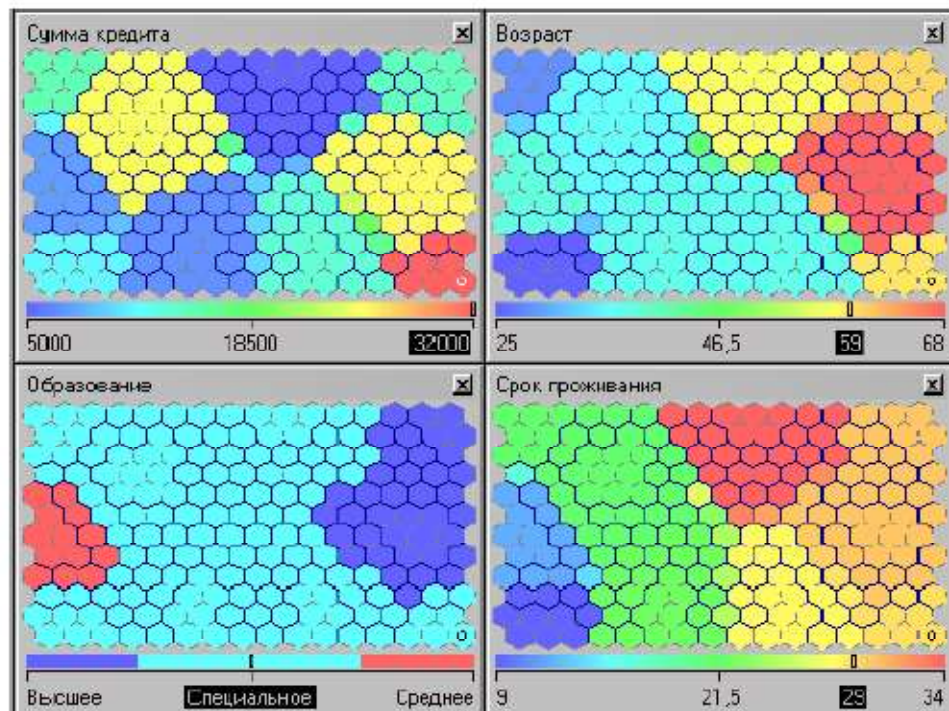


Рис. 8.25. Отримана карта Кохонена.