



УДК 519.7

МАТЕМАТИЧНИЙ АПАРАТ САМООРГАНІЗОВАНИХ КАРТ КОХОНЕНА

Студ. А.В. Володько, гр. БМО-1-17

Науковий керівник доц. О.Л. Блохін

Київський національний університет технологій та дизайну

Самоорганізована карта Кохонена (англ. Self-organizing map - SOM) - нейронна мережа з навчанням без учителя, що виконує завдання візуалізації і кластеризації. Ідея мережі запропонована фінським вченим Т. Кохоненом. Є методом проектування багатовимірного простору в простір з більш низькою розмірністю (найчастіше, двовимірне), застосовується також для вирішення завдань моделювання, прогнозування, виявлення наборів незалежних ознак, пошуку закономірностей у великих масивах даних, розробці комп'ютерних ігор, квантування кольорів до їх обмеженого числа індексів в кольоровій палітрі: при друку на принтері і раніше на ПК або ж на приставках з дисплеєм зі знизеним числом кольорів, для архіваторів (загального призначення) або відео-кодеків, тощо. Є однією з версій нейронних мереж Кохонена.

Самоорганізована карта складається з компонентів, які називаються вузлами або нейронами. Їх кількість задається аналітиком. Кожен з вузлів описується двома векторами. Перший - т. зв. вектор ваги m , що має таку ж розмірність, що і вхідні дані. Другий - вектор r , що представляє собою координати вузла на карті.

Карта Кохонена візуально відображається за допомогою осередків прямокутної або шестикутної форми; остання застосовується частіше, оскільки в цьому випадку відстані між центрами суміжних осередків однакові, що підвищує коректність візуалізації карти.

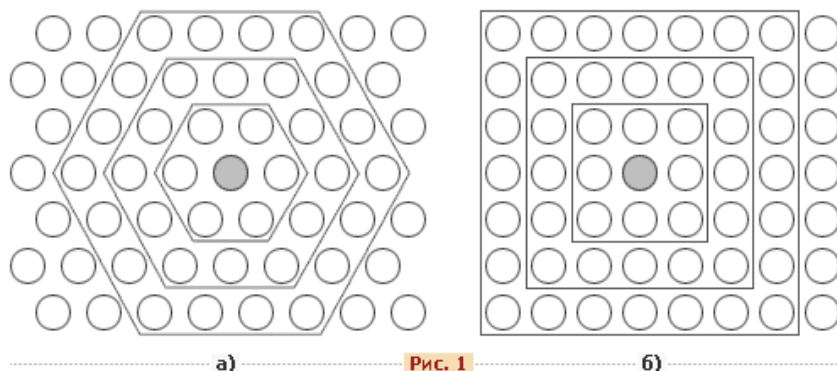
Спочатку відома розмірність вхідних даних, по ній певним чином будується початковий варіант карти. В процесі навчання вектори ваги вузлів наближаються до вхідних даних. Для кожного спостереження (семпли) вибирається найбільш схожий по вектору ваги вузол, і значення його вектора ваги наближається до спостереження. Також до спостереження наближаються вектори ваги декількох вузлів, розташованих поруч, таким чином якщо в безлічі вхідних даних два спостереження були схожі, на карті їм будуть відповідати близькі вузли.

Циклічний процес навчання, перебираючи вхідні дані, закінчується після досягнення картою допустимої (заздалегідь заданій аналітиком) похибки, або після здійснення заданої кількості ітерацій. Таким чином, в результаті навчання карта Кохонена класифікує вхідні дані на кластери і візуально відображає багатовимірні вхідні дані в двовимірній площині, розподіляючи вектори близьких ознак в сусідні осередки і розфарбовуючи їх в залежності від аналізованих параметрів нейронів.

В результаті роботи алгоритму виходять такі карти:

- карта входів нейронів - візуалізує внутрішню структуру вхідних даних шляхом підстроювання ваг нейронів карти. Зазвичай використовується кілька карт входів, кожна з яких відображає один з них і розфарбовується в залежності від ваги нейрона. На одній з карт певним кольором позначають область, в яку включаються приблизно однакові входи для аналізованих прикладів.
- карта виходів нейронів - візуалізує модель взаємного розташування вхідних прикладів. Окреслені області на карті представляють собою кластери, що складаються з нейронів зі схожими значеннями виходів.

- спеціальні карти - це карта кластерів, отриманих в результаті застосування алгоритму самоорганізуючої карти Кохонена, а також інші карти, які їх характеризують.



Кластером буде група векторів, відстань між якими всередині цієї групи менше, ніж відстань до сусідніх груп. Структура кластерів при використанні алгоритму SOM може бути відображена шляхом візуалізації відстані між опорними векторами (ваговими коефіцієнтами нейронів).

При використанні цього методу найчастіше використовується уніфікована матриця відстаней (u-matrix). При використанні цього методу обчислюється відстань між вектором ваг нейрона в сітці і його найближчими сусідами. Потім ці значення використовуються для визначення кольору, яким цей вузол буде нанесений.

Зазвичай використовують градації сірого, причому чим більше відстань, тим темніше нанесений вузол. При такому використанні вузлів з найбільшою відстанню між ними і сусідами відповідає чорний колір, а довколишнім вузлам - білий.

Особливостями моделі є стійкість до зашумлених даних, швидке і некероване навчання, можливість спрощення багатовимірних вхідних даних за допомогою візуалізації.

Самоорганізовані карти Кохонена можуть бути використані для кластерного аналізу тільки в тому випадку, якщо заздалегідь відомо число кластерів.

Важливим недоліком є те, що остаточний результат роботи нейронних мереж залежить від початкових установок мережі. З іншого боку, нейронні мережі теоретично можуть апроксимувати будь-яку безперервну функцію, що дозволяє досліднику не брати заздалегідь будь-які гіпотези щодо моделі.

ЛІТЕРАТУРА

1. Ф.Уоссермен, "Нейрокомп'ютерна техніка", М.: Мир
2. А.Ежов, С.Шумській, "Нейрокомп'ютеринг і його застосування в економіці і бізнесі".
3. Т.Кохонен, "Self-Organizing Maps", Springer.
4. Т.Кохонен, "Self-Organizing Maps" (2-nd edition), Springer.
5. Juho Vensano, "Data Mining Techniques Baseg on the Self Organized Map"