

УДК 519.7

ЗАСТОСУВАННЯ САМООРГАНІЗОВАНИХ КАРТ В УМОВАХ КЛАСТЕРИЗАЦІЇ

Студ. Т. М. Максимчук, гр. БМО1 - 17

Науковий керівник доц. О.Л. Блохін

Київський національний університет технологій та дизайну

Метою цієї публікації є вивчення методу аналізу з використання самоорганізованих карт Кохонена, що дозволяє автоматизувати дані у пошуку закономірностей.

Самоорганізаційна карта Кохонена – нейронна мережа з нескерованим навчанням, що виконує завдання кластеризації. Є методом проектування багатовимірного простору в простір з нижчою розмірністю (найчастіше, двовимірний), застосовується також для вирішення завдань моделювання, прогнозування та інших.

Кластеризація – це розбиття множини даних на кластери. Кластери – підмножини однорідних одиниць сукупності, параметри яких заздалегідь невідомі.

Досить часто потрібно вирішити задачу аналізу даних, при умові, коли важко визначити чіткі дані об'єкта. У таких випадках використовується метод самоорганізованих карт Кохонена.

Карти Кохонена мають набір вхідних елементів, кількість яких збігається з розмірністю - подаються на вхід векторів, і набір вихідних елементів, кожен з яких відповідає одному кластеру (групі).

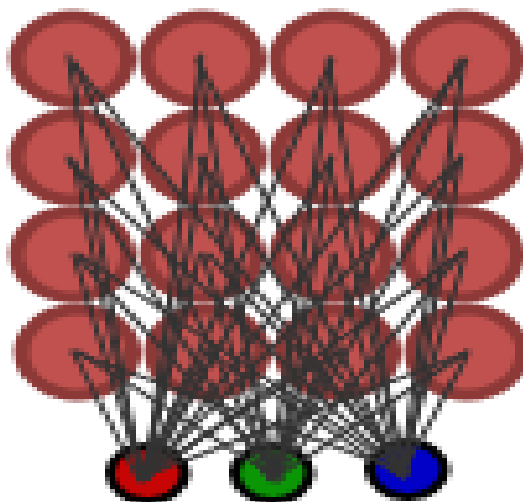


Рисунок 1 – Найпростіша мережа Кохонена

Зазвичай намагаються задавати кількість вихідних елементів меншу, ніж кількість вхідних, в такому випадку мережа дозволяє отримати спрощену характеристику об'єктів для подальшої роботи з ними. При подачі будь-якого вектора на вхід мережа повинна визначити, до якого з кластерів цей вектор найближче. В якості критерію близькості може бути обраний критерій мінімальності квадрата евклідова відстані. Розглянемо вхідний вектор як точку в n-вимірному просторі (n - кількість координат вектора, це число дорівнює числу вхідних нейронів).

Подавши вектор на вхід мережі ми отримаємо один кластер-переможець, відповідно, цей вектор буде належати саме до цього кластеру (групі). Саме так карти Кохонена вирішують задачу класифікації.



На вхід подається вектор з навчальної вибірки. Для цього вектора визначається кластер-переможець, і для цього кластера виробляють корекцію вагових коефіцієнтів таким чином, щоб він виявився ще ближче до вхідного вектору.

Найчастіше також коректують ваги декількох сусідів кластера-переможця. Вихідні (кластерні) елементи мережі Кохонена зазвичай представляють розташованими тим чи іншим чином в двовимірному просторі. Розмістимо, наприклад, вихідні елементи у вигляді квадратної сітки і поставимо початковий радіус навчання рівним

Подаємо на вхід мережі вектор і елементом-переможцем виявляється нейрон. За алгоритмом необхідно оновити значення ваг для цього нейрона, а також для тих, які потрапляють в коло заданого радіуса/ Ближче до кінця процесу навчання радіус зменшується. Нехай він став рівним одиниці, тоді оновлюватися будуть ваги наступних елементів.

Алгоритм застосування самоорганізованих карт в умовах кластеризації:

- беремо навчальний вектор і обчислюємо квадрат евклідова відстані від нього до кожного з кластерних елементів мережі
- знаходимо мінімальне з отриманих значень і визначаємо елемент-переможець для нейрона-переможця, а також для тих нейронів, які потрапили в заданий радіус,
- виконуємо коригування ваг зв'язків оновлюємо значення норми навчання і радіусу
- продовжуємо навчання, якщо не виконана умова зупинки навчання

При цьому описана технологія є універсальним методом аналізу. З її допомогою можна аналізувати різні стратегії діяльності, проводити аналіз результатів маркетингових досліджень, перевіряти кредитоспроможність клієнтів і т.д.

Таким чином, маючи перед собою карту і знаючи інформацію про деяку з частини досліджуваних об'єктів, ми можемо досить достовірно судити про об'єкти, з якими ми мало знайомі. Потрібно дізнатися, що з себе представляє новий партнер? Відобразимо його на карті і подивимося на сусідів. В результаті, можна отримувати інформацію з бази даних, ґрунтуючись на нечітких характеристиках.

Головна перевага самоорганізованих карт Кохонена є їх здатність "проектування" багатовимірних просторів на двовимірне. Важливою особливістю карт Кохонена є те, що об'єкти зі схожими характеристиками (за всіма компонентами) розташовуються поруч, утворюючи кластери.

Візуалізація даних у вигляді двовимірної карти дозволяє проводити кластеризацію даних і значно спрощує кореляційний аналіз.

ЛИТЕРАТУРА

1. Воронцов К. Лекции по искусственным нейронным сетям// <http://www.ccas.ru/voron/download/NeuralNets.pdf>
2. Мандель И. Кластерный анализ / Мандель И.Д. – М. : Финансы и статистика, 1988.
3. Нейский И. Классификация и сравнение методов кластеризации// http://it-claim.ru/Persons/Neyskiy/Article2_Neyskiy.pdf