in technology, management and optimization of the size of structures, optimal planning of complex technical systems such as information systems, computer networks, transport and telecommunications networks, etc .;

in mathematical economics: solving large macroeconomic models (such as input-output model), microeconomic models or models of entrepreneurship, decision theory and game theory.

As we see mathematics, including linear programming has wide application in various fields of applied science. So naturally, some problems arise when building an interactive electronic atlas Kyiv were consolidated to known linear programming problems. In particular task of finding favorite places was reduced to the traveling salesman problem.

Also the development of the atlas there were tasks associated with finding the minimum spanning tree. Recall the definition of minimal spanning tree. Suppose we have a graph: the set of vertexes and edges with weights that connects them. Spanning tree is called a subset of edges of the graph, you can go from any vertex to an arbitrary vertex, using the edge of this set. At a minimum spanning tree called the tree with minimum weight - the sum of the weights (length) of edges. To solve this problem we were using standard methods, such as Kruskal and Prim.

Keywords: linear programming, mathematical models, interactive atlas Kyiv, GIS technology.

Иксанов А. М., Полоцкий С. В., Голубцов А. Г. **Применение задач линейного программирования при разработке интерактивного атласа Киева**. В работе кратко рассказывается о математическом, в частности линейном, программировании. Несмотря на простоту теории, линейное программирование имеет широкое практическое применение. В работе рассматриваются некоторые важные задачи с линейными моделями и говорится какие именно использовались при построении интерактивного электронного атласа города Киева.

Ключевые слова: линейное программирование, математические модели, интерактивный атлас Киева, ГИС-технологии.

Надійшла до редколегії 24.11.2016

УДК 519.8:004.4

Іксанов О.М., Полоцький С.В.
Київський національний університет імені Тараса Шевченка

ЗАДАЧІ КЛАСТЕРНОЇ ОПТИМІЗАЦІЇ У ПРОЕКТУВАННІ СТРУКТУРИ ІНТЕРАКТИВНОГО АТЛАСУ КИЄВА

Ключові слова: кластерна оптимізація, методи кластеризації, електронний атлас Києва, ГІС-технології

Актуальність теми. Кожного дня нам доводиться працювати з інформацією. Вона можу бути різною: економічного, соціального, політичного характеру. Тому і за структурою може дуже відрізнятися. Важливо навчитися з нею працювати, а саме обробляти та робити висновки. На даний момент обробка інформації є дуже актуальною. Незважаючи на те, що ми живемо в еру сучасних комп'ютерів та обчислювальних систем, дуже часто їх потужностей не вистачає для розв'язання задачі в початковому вигляді (мається на увазі розв'язання задачі за обмежений час). Тому приходиться займатися попередньою обробкою інформації. В цьому пожуть нам допомогти задачі кластерної оптимізації. Цим задачам уже багато років, але зараз вони актуальні як ніколи раніше. У роботі [4] можна ознайомитися з основними методами кластеризації та умовами застосування кожного з них. У роботах [1,3] розглядаються два сучасних методи, які мають широке практичне застосування. Нижче в роботі про це пишеться.

Кластерний аналіз або кластеризація полягає у групуванні безлічі об'єктів таким чином, що об'єкти, що знаходяться в тій же самій групі (так званий кластер) більш схожі (в тому чи іншому сенсі) один до одного, ніж в інших групах (кластери) . Це основне завдання пошукового інтелектуального аналізу даних, а також загальна методика аналізу статистичних даних, використовуються в багатьох областях, в тому числі машинному навчанні, розпізнавання образів, аналізі зображень, пошуку інформації, біоінформатики, стиснення даних та комп'ютерній графіці.

Виклад основного матеріалу. Під кластерним аналізом потрібно розуміти не один конкретний алгоритм, але проблему, яку потрібно вирішити. Це може бути досягнуто за допомогою різних алгоритмів, у яких істотно відрізняється поняття кластера та ефективні способи їх пошуку. Популярні

поняття кластерів включають групи з невеликими відстанями між членами кластеру, щільні області простору даних, інтервалів або окремих статистичних розподілів. Тому кластеризація може бути сформульована як задача багато критеріальної оптимізації. Відповідний алгоритм кластеризації і параметрів настройки (включаючи значення, такі як функції відстані, щоб використовувати, порогове значення щільності або кількість очікуваних кластерів) залежать від індивідуального набору даних. Кластерний аналіз являє собою ітеративний процес виявлення знань або інтерактивної багатоцільової оптимізації, яка включає в себе метод проб і невдачі. Часто буває необхідно змінити дані попередньої обробки і параметрів моделі, щоб досягнути бажаного результату.

Крім терміна кластеризації, існує цілий ряд термінів зі схожими значеннями, в тому числі автоматичної класифікації, чисельної систематики, botryology (від грецького "βότρυς винограду") і типологічного аналізу.

Серед основних **методів** кластеризації можна виділити наступні.

Ієрархічна кластеризація. Основна ідея полягає в тому, що об'єкти мають більший зв'язок з об'єктами як лежать від них на меншій відстані, ніж з тими, що лежать далі. Ці алгоритми об'єднують "об'єкти" в "кластери" в залежності від їх відстані. Кластер може бути заданий такою величиною як діаметр. Суть полягає в тому, що до одного кластера відносяться об'єкти, відстань між якими не перевищує діаметру. При різних значеннях діаметру будуть різні кластери. Їх можна представити з використанням дендрограми, який пояснює походження назви «ієрархічна кластеризація». На ній видно як деякі кластери зливаються в один при збільшення діаметру.

Зв'язок на основі кластеризації — це ціле сімейство методів, які відрізняються способом обчислення відстані. Крім звичайного вибору функції відстані, користувач також повинен вирішити, за критерієм зчеплення (оскільки кластер складається з декількох об'єктів, є кілька кандидатів, щоб обчислити відстань до), щоб використовувати.

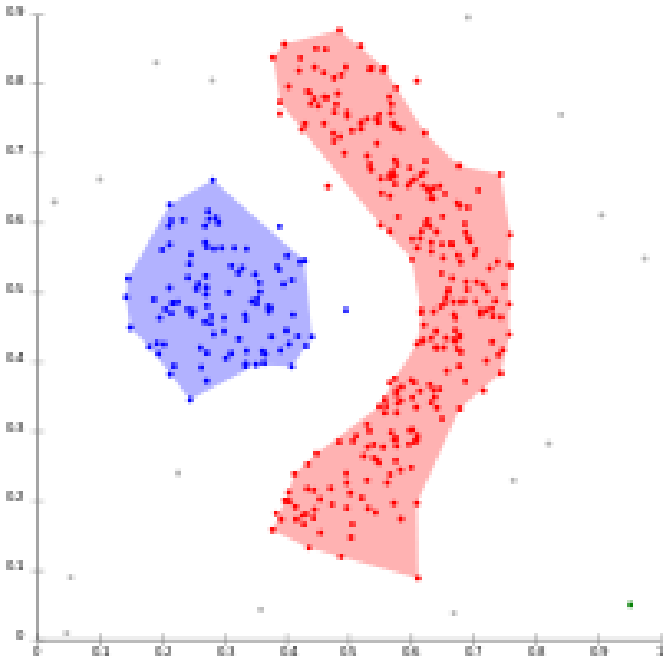
Кластеризація на базі центроїда. Кластери представлені центральним вектором, який не обов'язково належить вибірці даних. Коли число кластерів фіксується рівним K, то задача кластеризації зводиться до такої задачі оптимізації: знайти K центрів кластерів K та віднести об'єкти до найближчого центру кластера, таким чином, що квадрати відстаней від кластеру були мінімальними.

Сама задача оптимізації є NP-складною, і, таким чином, загальний підхід до розв'язку задачі полягає у пошуку наближених рішень. Добре себе зарекомендував відомий алгоритм Ллойда, який часто насправді називається метод "K-середніх". Правда він дял пошуку локального оптимуму, тому потрібно запускати кілька разів з різними випадковими початковими даними. Є багато варіацій методу K-середніх , але основним недоліком є те, що усі вони для початку своєї роботи вимагають кількість кластерів.

Розподіл на основі кластеризації. Модель кластеризації, пов'язана зі статистикою, заснована на моделях розподілу. Кластери можуть потім легко бути визначені як об'єкти, що належать, швидше за все, до того ж самого розподілу. Хороша властивістю цього підходу полягає в тому, що це дуже нагадує спосіб генерування штучних наборів даних: шляхом вибірки випадкових об'єктів з розподілу.

Теоретично ці методи працюють чудово, але у них виникає така проблема як overfitting (перенавчання).

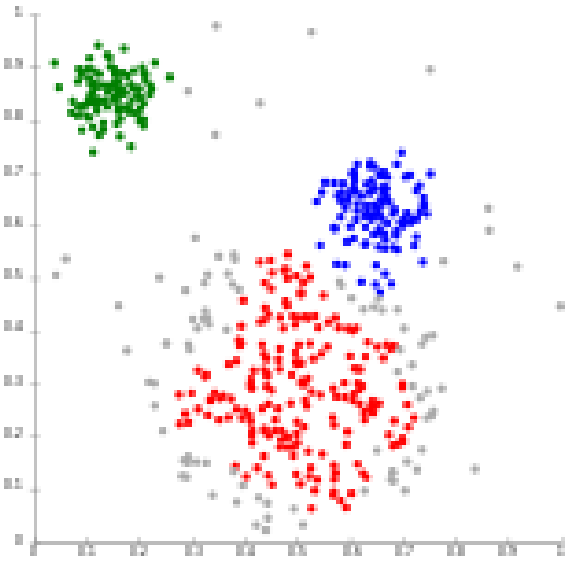
Розподіл на основі кластеризації виробляє складні моделі для кластерів, які можуть захопити і залежність кореляції між атрибутами. Проте, ці алгоритми дають додаткову складність користувачу: для багатьох наборів реальних даних, не може бути наперед визначена математична модель. Результат роботи можна побачити на наступному зображенні



Щільність на основі кластеризації

В щільності на основі кластеризації, кластери визначаються як області з більш високою щільністю. Найбільш популярним методом кластеризації на основі щільності є DBSCAN. На відміну від багатьох нових методів, він має цілком певну модель кластера під назвою "щільність-досяжності". Подібно кластеризації на основі зв'язків, в основі метода лежить з'єднання точок в межах певних порогів відстані. Проте, з'єднуються лише точки, які відповідають критеріям щільності, в первинному варіанті визначається як мінімальна кількість інших об'єктів в межах цього радіусу. Кластер складається з усіх підключених до щільності об'єктів (які можуть утворювати кластер довільної форми, на відміну від багатьох інших методів) плюс всі об'єкти, які знаходяться в межах дальності цих об'єктів. Ще одна цікава властивість DBSCAN є те, що його складність досить низька.

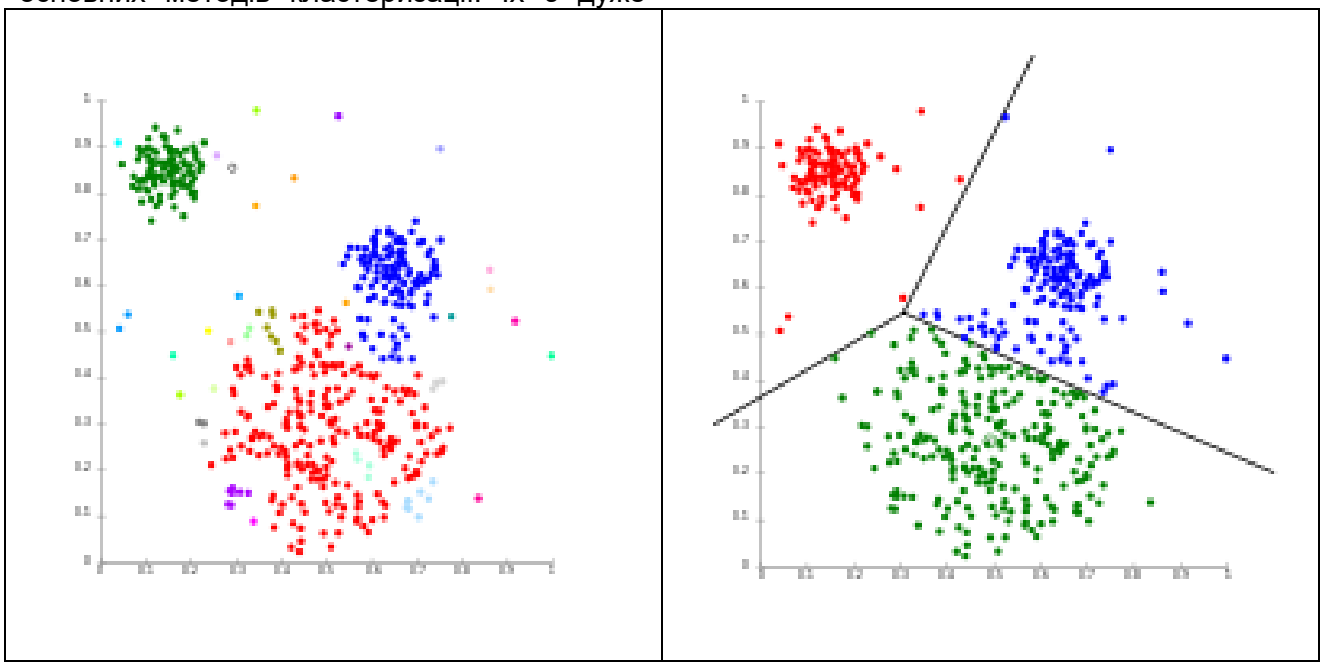
Основним недоліком DBSCAN є те, що він очікує свого роду падіння щільності для виявлення кластера кордонів. Крім того, вони не можуть виявити внутрішні кластерні структури, які поширені в більшості даних реальному житті. Робота даного методу ілюструється наступним зображенням



В останні роки значні зусилля були спрямовані на поліпшення ефективності існуючих алгоритмів. Це привело до розробки методів попередньої кластеризації, таких як навісна кластеризацію, який може обробляти великі масиви даних ефективно, але в результаті "кластери" є лише грубою розміткою з набору даних, які потім аналізуються існуючими методами, такими як k-засоби кластеризації. Різні інші підходи до кластеризації були випробувані такі як кластеризація на основі насіння.

Вище ми навели короткий перелік основних методів кластеризації. Їх є дуже багато і який саме підходить для вирішення тієї чи іншої задачі потрібно дивитися на практиці. При розробці електронного атласу Києва ми користувалися найпростішими методами, такими як метод К-середніх та кластеризація на базі центроіда. Ми їх використовували для знаходження областей з найбільшим розташуванням промислових обєктів, найбільш та найменш заселених частин міста з етнічним поділом та без нього.

Для кращого розуміння розглянемо кілька прикладів:



Ми бачимо на площині багато точок. Під кожною точкою потрібно розуміти об'єкт з певними властивостями. Наприклад точки з просторовими координатами. В один розфарбовані точки, віднесені до одного і того ж кластеру. Цей приклад демонструє як одна і там ж сама множина ділиться на кластери різними способами. На малюнку зліва застосовується ієрархічна кластеризація, а на малюнку справа кластеризація на в основі якої лежить метод К-середніх

Список літератури

1. Microsoft academic search: most cited data mining articles: DBSCAN. **2.** Lloyd, S. (1982). "Least squares quantization in PCM". IEEE Transactions on Information Theory. 28 (2): 129–137. **3.** Kriegel, Hans-Peter; Kröger, Peer; Sander, Jörg; Zimek, Arthur (2011). "Density-based Clustering". WIREs Data Mining and Knowledge Discovery. 1 (3): 231–240. **4.** Christopher M. Bishop. Pattern Recognition and Machine Learning, 2006 Springer Science+Business Media, LLC

Іксанов О.М., Полоцький С.В. **Задачі кластерної оптимізації у проектуванні структури інтерактивного атласу Києва.** У роботі розглядаються основні методи кластреної оптимізації. Наводяться приклади в якому випадку який метод краще використовувати. Також вказано, які методи для вирішення яких задач були використані при побудові електронного інтерактивного атласу міста Києва.

Ключові слова: кластерна оптимізація, методи кластеризації, електронний атлас Києва, ГІС-технології.

Iksanov O., Polotskyi S. **The objectives of the cluster structure to optimize the design of interactive atlas of Kyiv.** Clustering can be considered the most important *unsupervised learning* problem; so, as every other problem of this kind, it deals with finding a *structure* in a collection of unlabeled

data. A loose definition of clustering could be “the process of organizing objects into groups whose members are similar in some way”.

A *cluster* is therefore a collection of objects which are “similar” between them and are “dissimilar” to the objects belonging to other clusters.

In this case we easily identify the 4 clusters into which the data can be divided; the similarity criterion is *distance*: two or more objects belong to the same cluster if they are “close” according to a given distance (in this case geometrical distance). This is called *distance-based clustering*.

Another kind of clustering is *conceptual clustering*: two or more objects belong to the same cluster if this one defines a concept *common* to all that objects. In other words, objects are grouped according to their fit to descriptive concepts, not according to simple similarity measures.

So, the goal of clustering is to determine the intrinsic grouping in a set of unlabeled data. But how to decide what constitutes a good clustering? It can be shown that there is no absolute “best” criterion which would be independent of the final aim of the clustering. Consequently, it is the user which must supply this criterion, in such a way that the result of the clustering will suit their needs.

For instance, we could be interested in finding representatives for homogeneous groups (*data reduction*), in finding “natural clusters” and describe their unknown properties (*“natural” data types*), in finding useful and suitable groupings (*“useful” data classes*) or in finding unusual data objects (*outlier detection*).

Clustering algorithms may be classified like this Exclusive Clustering, Overlapping Clustering, Hierarchical Clustering, Probabilistic Clustering. In the first case data are grouped in an exclusive way, so that if a certain datum belongs to a definite cluster then it could not be included in another cluster.

On the contrary the second type, the overlapping clustering, uses fuzzy sets to cluster data, so that each point may belong to two or more clusters with different degrees of membership. In this case, data will be associated to an appropriate membership value. Instead, a hierarchical clustering algorithm is based on the union between the two nearest clusters. The beginning condition is realized by setting every datum as a cluster. After a few iterations it reaches the final clusters wanted.

Finally, the last kind of clustering use a completely probabilistic approach. Delevoping interactive atlas we have used K-means and Hierarchical clustering algorithms.

Keywords: cluster optimization, clustering methods, electronic atlas Kyiv, GIS technology.

Иксанов А.Н., Полоцкий С.В. **Задачи кластерной оптимизации в проектировании структуры интерактивного атласа Киева.** В работе рассмотрены основные методы кластерной оптимизации. Приводятся примеры в каких случаях какой метод лучше использовать. Также указано какие методы решения задач были использованы при построении электронного интерактивного атласа города Киева.

Ключевые слова: кластерная оптимизация, методы кластеризации, электронный атлас Киева, ГИС-технологии.

Надійшла до редколегії 01.12.2016