

УДК 51-77

DOI 10.51691/2541-8327\_2023\_6\_23

***ОСНОВНЫЕ ПОНЯТИЯ КЛАСТЕРНОГО АНАЛИЗА И ЕГО  
ПРИМЕНЕНИЕ В ПРИКЛАДНЫХ НАУКАХ***

***Каратаев А.А.,***

*студент,*

*Северо-Восточный федеральный университет им.М.К. Аммосова,*

*Якутск, Россия*

**Аннотация**

Анализ данных сегодня является основой для принятия управленческих решений. Статистические методы исследования являются базой для освоения указанной сквозной технологии. Огромную роль в статистическом анализе играет кластеризация. В статье рассматриваются базовые понятия кластеризации, как основной методики выявления схожих групп. Основной акцент сделан на применении методов кластерного анализа в различных прикладных областях.

**Ключевые слова:** кластерный анализ, методы и алгоритмы кластеризации

***BASIC CONCEPTS OF CLUSTER ANALYSIS AND ITS APPLICATION IN  
APPLIED SCIENCES***

***Karataev A.A.,***

*student,*

*North-Eastern Federal University,*

*Yakutsk, Russia*

**Abstract**

Data analysis today is the basis for making managerial decisions. Statistical research methods are the basis for the development of this end-to-end technology. Clustering plays an important role in statistical analysis. The article discusses the basic concepts of clustering as the main technique for identifying similar groups. The main emphasis is placed on the application of cluster analysis methods in various applied areas.

**Keywords:** cluster analysis, clustering methods and algorithms

В настоящее время развитие информационно-коммуникационных технологий позволяет ежесекундно генерировать огромные объемы информации. Сегодня данные называют «нефтью» цифровой экономики, а сквозная технология «анализ данных» является одной из базовых технологий, которую необходимо осваивать большинству специалистов, работающих в новых форматах производства.

Кластерный анализ – это общий логический процесс, сформулированный как процедура группировки объектов в группы на основе их сходства и различия. Группа методик, объединенных этим названием, широко используется во многих прикладных сферах деятельности человека, в которых необходимо выявление определенных общих черт и закономерностей: обработка баз данных, сегментация изображений, аналитическая обработка интернет-документов и т.д. [1].

Есть два типа кластерного анализа: иерархический и неиерархический.

Методы иерархического кластерного анализа включали анализируемые объекты в иерархическую систему кластеров. Эта система определяется как система взаимно различных непустых подмножеств исходного множества объектов. Основной характеристикой иерархических методов кластерного анализа является создание декомпозиции исходного множества объектов, при

котором каждая частичная декомпозиция уточняет следующую или предыдущую декомпозицию[2].

По способу создания декомпозиций методы иерархической кластеризации делятся на несколько групп:

1. Агломеративная кластеризация — в начале кластеризации отдельные объекты рассматриваются как самостоятельные группы. Далее по определенным алгоритмам наиболее похожие отдельные точки массива объединяются более крупные кластеры до тех пор, пока не будут выполнены заданные критерии качественной декомпозиции.
2. Дивизиональная кластеризация — в начале процесса кластеризации все объекты находятся в одном массиве. Затем весь набор данных по определенным методикам разбивается на более мелкие кластеры. Существует несколько методов кластерного анализа [3].

Методы иерархического кластерного анализа [4]:

1. Простой метод связи
2. Метод средней связи
3. Полный метод связывания
4. Метод центроида
5. Медианный метод
6. Метод Уорда

Для неиерархических методов кластерного анализа характерна классификация объектов на заранее определенное количество дизъюнктивных кластеров. Эти методы кластеризации можно разделить на 2 группы [5]:

- методы жесткой кластеризации — четкое отнесение объекта к кластеру;
- нечеткий кластерный анализ — вычисляет степень релевантности объектов кластерам.

Области применения:

1. В сфере маркетинга - для сегментирования рынка (выявление закономерностей в покупках, совершаемых клиентами; выделение групп потребителей со схожими стереотипами поведения и т.п.).
2. В экономике и финансах - в экономике и финансах кластерный анализ применяется для анализа рынков, выявления групп компаний с похожими финансовыми показателями, моделирования рисков и прогнозирования финансовых временных рядов. Например, кластерный анализ может помочь идентифицировать сходные группы компаний по сектору, размеру или финансовым показателям, что может быть полезно для принятия инвестиционных решений и управления портфелем.
3. В страховом бизнесе - для получения профилей клиентов (с целью определения услуг страхования, обеспечивающих наименьшие для компании риски).
4. В медицине - кластерный анализ широко используется в биологии и медицине для классификации биологических организмов, выявления генетических паттернов, идентификации подтипов заболеваний и прогнозирования результата лечения. Например, кластерный анализ может помочь идентифицировать группы пациентов с похожими клиническими характеристиками или генетическими мутациями, что может привести к более индивидуализированному подходу к лечению и диагностике.
5. В телекоммуникационном бизнесе - для поиска родственных групп клиентов с похожими типами пользования услугами (с целью разработки привлекательных наборов цен и услуг).
6. В социологии - кластерный анализ применяется для выявления типологий поведения, классификации социальных групп, анализа социальных сетей и моделирования предпочтений потребителей. Например, кластерный анализ может помочь идентифицировать группы потребителей с похожими предпочтениями и поведением, что может быть полезно для разработки маркетинговых стратегий и персонализации продуктов и услуг.

Так же кластерный анализ широко используется в сфере IT, а в частности в сегменте машинного обучения. Эта методология является основой для сегментации пользователей, выявления аномалий и анализа текстовых данных. Например, кластерный анализ может помочь разделить пользователей на группы с похожими интересами или поведением, что может быть полезно для рекомендательных систем и персонализации контента.

Подводя итог, можно сказать, что в наше время значение методов кластеризации для прикладной науки только возрастает, так как процесс накопления данных носит взрывной характер, поэтому важно уметь ориентироваться в «море» данных и с помощью анализа, извлекать из этого пользу.

### **Библиографический список:**

1. Бериков В.Б., Лбов Г.С. Современные тенденции в кластерном анализе [Электронный ресурс] // сайт «Сравнительный анализ в биологии (и не только)». — URL: [https://biocomparison.ucoz.ru/\\_ld/0/49\\_berikov\\_lbov.pdf](https://biocomparison.ucoz.ru/_ld/0/49_berikov_lbov.pdf) (дата обращения: 19.06.2023)
2. Tom Tullis, Bill Albert, Collecting, Analyzing, and Presenting Usability Metrics. [Электронный ресурс] — URL: <https://www.sciencedirect.com/science/article/abs/pii/B9780124157811000091> (дата обращения: 19.06.2023)
3. Maurice Roux. A Comparative Study of Divisive and Agglomerative Hierarchical Clustering Algorithms. [Электронный ресурс] — URL: <https://hal.science/hal-02085844/document> (дата обращения: 19.06.2023)
4. Maurizio Vichi, Carlo Cavicchia. Hierarchical Means Clustering. [Электронный ресурс] — URL: <https://link.springer.com/article/10.1007/s00357-022-09419-7> (дата обращения: 19.06.2023)
5. Ian Davidson. Understanding K-Means Non-hierarchical Clustering. [Электронный ресурс] — URL:

[https://www.researchgate.net/publication/228574607\\_Understanding\\_K-means\\_non-hierarchical\\_clustering](https://www.researchgate.net/publication/228574607_Understanding_K-means_non-hierarchical_clustering) (дата обращения: 19.06.2023)

*Оригинальность 79%*