

11. ПРИМЕНЕНИЕ КЛАСТЕРНОГО АНАЛИЗА КАК МЕТОДА МОДЕРНИЗАЦИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

Адоньева К.К., студент гр.173902

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Федосенко В.А. – канд. тех. наук, доцент кафедры ЭИ

Аннотация. Сегодня, благодаря цифровым технологиям, изменения и улучшения происходят практически в каждой сфере жизни человека. Что касается сферы образования – она также требует модернизации. Метод кластеризации мог бы отлично подойти для этого, а также помочь образовательному процессу перейти на новый уровень.

Кластерный анализ – это общее название множества вычислительных процедур, используемых при создании классификации. Главная цель кластерного анализа – нахождение групп схожих объектов в выборке данных. Эти группы удобно называть кластерами. Не существует общепринятого определения термина «кластер», однако считается, что кластеры обладают некоторыми свойствами, наиболее важными из которых являются плотность, дисперсия, размеры, форма, а также отделимость [1].

Кластерный анализ применяется для решения широкого спектра задач, но чаще всего речь идёт именно о задаче сегментации. Все исследования, посвященные проблеме сегментации, безотносительно того, какой используется метод, имеют целью идентифицировать, устойчивы группы, каждая из которых объединяет в себя объекты похожими характеристиками.

Цели и задачи кластеризации[2]:

1. Более глубокое понимание темы

Деление разрозненной информации на группы помогает аналитику понять, какие именно данные собраны. Потом их проще будет обрабатывать — например, применять к разным кластерам конкретные методы анализа.

2. Выявление аномалий

После кластеризации могут появиться отдельные данные, которые не относятся ни к одному из кластеров. Их нужно изучить, чтобы понять, ошибка это или же какой-то интересный феномен.

3. Расширение данных

Иногда при сборе информации у каких-то данных больше признаков, а у каких-то меньше. Кластеризация поможет предположить отсутствующие признаки у других элементов кластера.

4. Сжатие данных

Если данных слишком много, можно поделить их на кластеры, усреднить и оставить по одному объекту на каждый кластер. Это позволит в дальнейшем при анализе использовать меньше мощности.

Выделяют две группы методов кластерного анализа: иерархические и неиерархические. К основным методам иерархического кластерного анализа относятся метод ближнего соседа, метод полной связи и метод Варда.

Метод ближнего соседа состоит в том, что алгоритм находит расстояния между запросом и всеми примерами в данных, выбирая определенное количество примеров (k), наиболее близких к запросу, затем голосует за наиболее часто встречающуюся метку (в случае задачи классификации) или усредняет метки (в случае задачи регрессии).

Метод полной связи, который часто называют методом ближнего соседа, имеет схожий принцип. При использовании этого метода расстояние между кластерами вычисляется по принципу «дальнего соседа», то есть за расстояние между кластерами принимается расстояние между самыми удаленными элементами двух кластеров.

Метод Варда это принцип, при котором начала в обоих кластерах для всех имеющихся наблюдений производится расчёт средних значений отдельных переменных. Затем вычисляются квадраты евклидовых расстояний от отдельных наблюдений каждого кластера до этого кластерного среднего значения, а затем данные дистанции суммируются.

Неиерархических методов больше, хотя работают они на одних и тех же принципах. По сути, они представляют итеративные методы дробления исходной совокупности. В процессе деления формируются новые кластеры и так до тех пор, пока не будет выполнено правило остановки. Между собой методы различаются выбором начальной точки, правилом формирования новых кластеров, а также правилом остановки[3].

Суть метода, описанного здесь, состоит в применении интеллектуального анализа в образовательных данных, с целью выявления закономерностей и группирования учащихся на основе поданных результатов централизованного тестирования или экзаменов. Группировка проводится для того, чтобы разделить студентов на две группы. Разница между этими группами в скорости усвоения материала. Подразумевается использование метода Варда.

Цель этого метода состоит в том, чтобы дать абсолютно всем студентам работать в своём темпе. Те студенты, которые попадают в группу с высокой скоростью усваивания материала, могут учиться и при этом у них не стоит задача «подождать» остальных студентов, у которых могут возникать разного рода проблемы. При этом, студенты с низкой скоростью усваивания материала не будут чувствовать ощущение неполноценности.

При проведении кластерного анализа важно соблюдать его этапы, так как их правильная последовательность максимизирует точность конечных результатов. При анализе темы образовательного процесса было принято решение выделить пять этапов.

Первым этапом можно назвать отбор выборки для кластеризации. В нашем случае выборка – набор результатов централизованного тестирования или экзаменов тех, кто уже точно принят в университет.

Второй этап – определение переменных, по которым будут оцениваться объекты в выборке. Здесь результаты тестирования оцениваются по признаку большего усвоения предметов.

Третий этап – непосредственное вычисление значений той или иной степени различия. Можно охарактеризовать анализом поданных данных для дальнейшего деления на кластеры. Для большей объективности определяется сложность задания и количество информации нужной для решения той или иной задачи, чтобы минимизировать субъективность оценки задания сложность, которой может иметь вес, но оцениваться одинаково.

Четвёртый этап – создание групп сходных объектов. На этом этапе подразумевается деление всех студентов потока на две группы: те, кто усваивает материал лучше, а также те, у кого есть с этим трудности.

Пятый этап – интерпретация полученных результатов.

Таким образом, после проведения кластерного анализа остаются две группы с различными скоростями усвоения материала. Как только они определены, преподаватели могут адаптировать свои методы обучения и учебный план, чтобы лучше удовлетворить потребности каждой группы, для достижения более эффективного и действенного образовательного процесса.

Одним из основных преимуществ использования кластерного анализа в образовании является то, что он позволяет преподавателям выявлять учащихся, которым может потребоваться дополнительная поддержка или ресурсы. Например, если кластерный анализ показывает, что у определенной группы учащихся возникают трудности с изучением определенного предмета или стиля обучения, учителя могут обеспечить целенаправленные вмешательства, чтобы помочь этим учащимся преодолеть свои трудности.

Еще одним преимуществом использования кластерного анализа является то, что он может помочь преподавателям определить учащихся, которые могут извлечь выгоду из расширенных или

ускоренных возможностей обучения. Группируя учащихся в зависимости от их способностей и интересов, преподаватели могут предлагать более сложные учебные задания тем, кто к ним готов, при этом гарантируя, что другие учащиеся получат соответствующую поддержку и ресурсы.

Кроме того, кластерный анализ может помочь преподавателям персонализировать процесс обучения для каждого учащегося. Группируя учащихся на основе их стилей обучения и предпочтений, преподаватели могут предоставлять индивидуальные инструкции и занятия, адаптированные к уникальным потребностям каждого учащегося.

Однако можно указать и некоторые ограничения использования кластерного анализа в образовании. Например, метод в значительной степени зависит от качества и количества доступных данных, и преподавателям может потребоваться собрать и проанализировать значительный объем данных для получения значимых результатов. Кроме того, кластерный анализ может не подходить для всех типов образовательных данных или контекстов, и преподаватели должны тщательно рассмотреть вопрос о соответствии метода их конкретным потребностям.

Подводя итоги, можно сказать, что кластерный анализ является мощным методом, который можно использовать для модернизации образовательного процесса. Выявляя закономерности и группируя учащихся на основе их сходства, преподаватели могут проводить целенаправленные вмешательства, персонализированное обучение и более эффективное и действенное обучение. Тем не менее, преподаватели также должны знать об ограничениях метода и тщательно учитывать его пригодность для своих конкретных потребностей.

Список использованных источников:

1. Кластерный анализ в задачах классификации, оптимизации и прогнозирования//Л.Х. Гитис // Издательство Московского государственного горного университета, 2004. – с. 253.

2. Что такое кластерный анализ и как с помощью него группируют данные// Практимум. [Электронный ресурс]. Режим доступа: <https://practicum.yandex.ru/blog/chto-takoe-klasterizaciya-i-klasternyi-analiz/>. – Дата доступа: 28.03.2023.

3. Кластерный анализ: сущность, преимущества и недостатки // Библиотека. [Электронный ресурс]. Режим доступа: <https://cyberleninka.ru/article/n/klasternyi-analiz-suschnost-preimuschestva-i-nedostatki/viewer/>. – Дата доступа: 29.03.2023.