

ПРИМЕНЕНИЕ КЛАСТЕРНЫХ МЕТОДОВ ДЛЯ АНАЛИЗА УСПЕВАЕМОСТИ СТУДЕНТОВ

*С.А. Ермоченко, С.В. Сергеевко
Витебск, ВГУ имени П.М. Машерова*

Задачи кластерного анализа относятся к области автоматической классификации объектов. Кластерные методы позволяют разбить некоторое множество объектов на непересекающиеся классы, используя различные числовые характеристики рассматриваемых объектов. При этом числовые характеристики могут иметь различные диапазоны значений и единицы измерения [1].

В данной работе предлагается применять методы автоматической классификации для результатов экзаменационной сессии в отдельно взятых студенческих группах. В этом случае характеристиками объектов будут являться экзаменационные оценки по всем дисциплинам в этой группе.

Такой подход позволит не только разбить студентов на различные классы по уровню успеваемости, основанной на средней оценке в экзаменационную сессию, но и учесть различную направленность дисциплин. При анализе результатов экзаменационной сессии все дисциплины были разбиты на три блока:

- социально-гуманитарные дисциплины;
- математические дисциплины;
- дисциплины, связанные с программированием.

Материал и методы. В качестве метода исследования применяется кластерный метод, ориентированный на построение дендрограмм на основе матрицы расстояний между объектами.

Для применения данного метода необходимо для произвольной пары объектов a и b ввести понятие некоторой метрики $\rho(a, b)$. Данная функция определяет степень близости двух объектов между собой, основанную на числовых характеристиках этих объектов. Рассмотрим в качестве объектов студентов некоторой группы x_i , $i = 1, 2, \dots, n$, где n – количество студентов в группе. Тогда матрица $A = \{a_{ij} = \rho(x_i, x_j) \mid i \in \{1, \dots, n\}, j \in \{1, \dots, n\}\}$ называется матрицей расстояний [1]. Полученная матрица расстояний позволяет поэтапно объединять наиболее близкие объекты (с минимальным расстоянием) в кластеры. Такой процесс объединения продолжается до тех пор, пока все объекты не будут объединены в один кластер.

В качестве функции расстояния $\rho(x_i, x_j)$ в данной работе использовались:

- среднее арифметическое всех экзаменационных оценок;
- средневзвешенное экзаменационных оценок, при этом веса каждой дисциплине назначаются исходя того, к какому блоку дисциплин они относятся;
- Евклидова метрика [2] аффинно-преобразованного пространства средневзвешенных успеваемости по блокам дисциплин. Эта метрика вычисляется за 5 шагов:

1. Каждая десятибалльная оценка, полученная на некотором экзамене, представляется вещественным числом из отрезка $[0; 1]$. При этом учитывается диапазон фактически выставленных оценок.
2. Каждому экзамену по каждому блоку дисциплины назначается вес, который характеризует влияние оценок по этому экзамену на средневзвешенную успеваемость по данному блоку дисциплин. Сумма весов всех экзаменов для каждого блока должна быть равна 1.
3. Для каждого студента вычисляются средневзвешенные успеваемости по каждому блоку дисциплин на основании полученных оценок с учетом соответствующих весов.

4. Все средневзвешенные успеваемости умножаются на коэффициент масштабирования соответствующего блока. Этот коэффициент характеризует влияние данного блока дисциплин на дальнейший анализ успеваемости.
5. В полученном многомерном пространстве расстояние между точками (студентами) вычисляется как корень квадратный из суммы квадратов разностей средневзвешенных успеваемости по соответствующим блокам дисциплин.

То есть, для анализа применяется евклидова метрика пространства, полученного на 4-ом шаге.

Результаты и их обсуждение. В качестве входного множества объектов рассматривались результаты зимней экзаменационной сессии студентов специальности «прикладная математика» второго, третьего и четвертого курсов.

При разбиении на кластеры с применением функции расстояния, заданной средним арифметическим оценок, результат, как и ожидалось, показал распределение студентов по 3 кластерам, между которыми существовали ярко выраженные отличия. Один из кластеров – это хорошо успевающие студенты со средним баллом от 8 и выше. Второй кластер – это студенты со средней успеваемостью со средним баллом от 6 до 7,5. Третий кластер – это студенты с низкой успеваемостью со средним баллом 5,5 и меньше. Стоит заметить, что в анализе принимали участие результаты только студентов, успешно сдавших сессию.

При использовании функции расстояния, основанной на средневзвешенной экзаменационной оценке, в зависимости от используемых весовых коэффициентов, результаты получались различными. Так при увеличении весовых коэффициентов на блоки дисциплин математических или связанных с программированием, существенных различий в кластерной структуре не замечено. Однако следует отметить, что существует небольшое количество студентов (3–4 студента, в зависимости от группы) которые в зависимости от весовых коэффициентов, назначенных рассматриваемым двум блокам, переходят из одного кластера в другой. Это объясняется тем, что данные студенты имеют большие успехи в дисциплинах математического направления или дисциплинах, связанных с программированием. Но при этом увеличение весового коэффициента для блока социально-гуманитарных дисциплин приводит к тому, что ярко выраженное различие между кластерами исчезает. Данный факт говорит о том, что большинство студентов группы усваивают дисциплины данного блока примерно на одинаковом уровне.

При использовании евклидовой метрики аффинно-преобразованного пространства средневзвешенных успеваемости по блокам дисциплин результаты во многом повторяют те, которые были получены при функции расстояния, основанной на средневзвешенной экзаменационной оценке. В частности, наблюдается снижение различия между кластерами при увеличении коэффициента масштабирования для блока социально-гуманитарных дисциплин, а также переход студентов между кластерами при изменении коэффициентов масштабирования. Однако следует отметить, что эти особенности выражены в меньшей степени.

Заключение. Прделанная работа не претендует на исчерпывающий анализ результатов успеваемости студентов, а лишь демонстрирует возможность применения кластерных методов для анализа успеваемости с несколько иной стороны.

В дальнейшем планируется продолжить анализ успеваемости студентов с применением кластерных методов, включить в число исследуемых результатов экзаменационных сессий результаты студентов других специальностей. Рассмотреть результаты за несколько лет и проанализировать динамику успеваемости по блокам дисциплин.

Список литературы

1. Олдендерфер М.С. Кластерный анализ: факторный, дискриминантный и кластерный анализ / М.С. Олдендерфер, Р.К. Блэшфилд. – Москва: Финансы и статистика, 1989 – 215 с.
2. Постников, М. М. Аналитическая геометрия. – Москва: Наука, 1973. – 754 с.