

## **Определение причин, обуславливающих поступление в магистратуру, с применением методов кластеризации данных**

*Авторы:  
Ст. преподаватель каф. «НТПиПК» Пудеян Л.А.,  
Студентка группы ВМПИ11 Подколзина Л.А.*

Высшее образование – является одним из важнейших институтов общества. Решением одной из задач структурно-организационного характера стал переход к двухуровневой системе высшего профессионального образования в России. С 2011 года квалификации бакалавра и магистра стали основными для абитуриентов отечественных вузов [1-3]. Бакалавриат является обязательным для всех студентов первым уровнем подготовки. Магистратура – более высокая ступень образования, ориентированная на подготовку выпускников к научной деятельности, движущей силе прогресса. Подготовка специалистов высшей квалификации, обучение студентов, готовых выбрать себе научную сферу деятельности и приступить к профессиональному труду – продвижению науки – важная задача современного образования.

**Целью исследования** является поиск факторов, влияющих на решение выпускников продолжить обучение в магистратуре, ведь в образовательном процессе главную роль играет осознанность студента в получении новых знаний, повышении уровня профессиональной компетенции. Для успешной реализации поставленной цели был использован иерархический алгоритм кластерного анализа, позволяющий провести кластеризацию выпускников и их предпочтений на кластеры, соответствующие определенной группе.

*На первом этапе* были получены данные в результате проведенного анкетирования среди абитуриентов гуманитарных направлений подготовки в 2015 г. Основной акцент в анкетах был сделан на вопросы, касающиеся продолжения получения образования в рамках магистерской программы подготовки. Также содержались вопросы, затрагивающие проблемы процесса обучения и степень их влияния на дальнейшее желание продолжить обучение. Был сформирован csv-файл, представляющий собой нормированные показатели, полученные в ходе проведенного опроса.

*Вторым этапом работ* стал выбор основного метода обработки данных. Был использован кластерный анализ, позволяющий на основе множества показателей, характеризующих объект исследования, выделять его естественную структуру на основе кластеров. Задачей кластеризации является разбиение множества объектов на  $m$  кластеров, чтобы один объект принадлежал лишь одной группе разбиения, а объекты, находящиеся в пределах одной группы, были однородными.

**Теоретические данные.** Полученное множество  $I = \{I_1, I_2, \dots, I_n\}$  обозначим как  $n$  объектов. Результат измерения  $i$ -ой характеристики  $I_j$  объекта обозначим  $x_{ij}$ . Для множества  $I$  объектов мы располагаем множеством векторов измерений  $X = \{X_1, X_2, \dots, X_k\}$ , описывающих множество  $I$ . Множество

$X$  представлено как  $n$  точек в  $p$ -мерном евклидовом пространстве  $E_p$ . Произведем разбиение объектов  $I$  на  $m$  кластеров  $q_1, q_2, \dots, q_k$  так, что каждый  $I_j$  объект принадлежит одному и только одному кластеру и объекты, принадлежащие разным кластерам, являются разнородными (при этом  $m < k; m, k \in N$ ).  $K$  измерений  $X_1, X_2, \dots, X_k$  представлены в виде матрицы (1):

$$X = [X_1, X_2, \dots, X_k] = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots \\ x_{k1} & x_{k2} & \dots & x_{kk} \end{bmatrix} \quad (1)$$

Определение евклидова расстояния между парами векторов  $d(X_i, X_j)$  (2):

$$d(X_i, X_j) = \left[ \sum_{u=1}^N (x_{ui} - x_{uj})^2 \right]^{\frac{1}{2}} \quad (2)$$

Далее находим пару самых близких кластеров. Расчет расстояния и поиск пар методом полных связей повторяются до тех пор, пока все объекты не будут соединены в один кластер или до достижения заданного «порога» близости.

**Практическая реализация.** Процесс кластеризации был реализован на языке python с применением библиотек SciPy, numpy (выполнение анализа) и matplotlib (построение графика). Из исходного набора данных (файл data.csv) были получены списки идентификаторов объектов и матрица значений, проведено нормирование данных, кластеризация и произведен расчет евклидова расстояния между  $m$ -наблюдениями (анкетами респондента) в  $p$ -мерном пространстве (10)).

В ходе проведенного исследования были получены следующие результаты. Сравнительный анализ анкет респондентов показывает, что в среднем, решение о поступлении в магистратуру окончательно формируется к началу 4 курса (7 семестр). 1/5 часть опрошенных не считают поступление в магистратуру необходимым фактором для успешного построения карьеры. В качестве основных аргументов была названа разочарованность в выбранном направлении подготовки. Оставшиеся 4/5 респондентов считают необходимым повышать свой профессиональный уровень подготовки и углублять знания предметной области в магистратуре. Группы, выделяемые кластерным анализом, отличаются различной оценкой факторов, но в каждой из них находится значительная доля студентов, отмечающих высокую значимость магистратуры для построения успешной карьеры. Результаты, полученные в ходе исследования могут быть использованы при распределении мест для абитуриентов уровня второй ступени высшего образования.

### *Список использованных источников*

1. Главные события в современном образовании 2004-2011. Кафедра в Европу. М.: Медиа-лайн, 2010. <http://ru.education.mon.gov.ru/site-static/main-events.pdf> Дата обращения 12.04.2016.
2. Семин, Н.В. Академическая мобильность: проблемы выдачи и легализации документов об образовании /Н.В. Семин, Н.И. Зверев, А.Л. Демчук, М.Н. Житникова, Г.Ф. Ткач. – М.: Изд-во МГУ, 2007. – 148 с.
3. Концепция модернизации Российского образования на период до 2010 года / Распоряжение Правительства РФ от 29 декабря 2001 года № 1756-р. Доступ из справ.-правовой системы «КонсультантПлюс»