

УДК 004.65

МОДЕЛЬ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНЫХ КОМПЕТЕНЦИЙ С ИСПОЛЬЗОВАНИЕМ ИНСТРУМЕНТАРИЯ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

Ю. Б. Щенёва, ассистент кафедры ВПМ РГРТУ, Рязань, Россия;
orcid.org/0000-0001-6351-0399, e-mail: Shenyova@yandex.ru

А. Н. Пылькин, д.т.н., профессор кафедры ВПМ РГРТУ, Рязань, Россия;
orcid.org/0000-0001-9925-2870, e-mail: pylkin.a.n@rsreu.ru

Е. С. Щенёв, магистрант РГРТУ, Рязань, Россия;
orcid.org/0009-0005-8438-7445, e-mail: Jheka1235317@gmail.com

О. А. Бодров, к.т.н., доцент кафедры КТ РГРТУ, Рязань, Россия;
orcid.org/0009-0005-7225-6704, e-mail: bodrov.o.a@rsreu.ru

Рассматривается задача построения модели освоения образовательных компетенций с применением методов интеллектуального анализа данных. Целью работы является разработка комплексного подхода, включающая технологию извлечения и анализа данных с применением концепции многомерного представления и методов кластерного анализа. Рассматриваются методы классификации и упорядочивания частных показателей, определяющих уровень освоения образовательных компетенций. Обосновывается выбор метрики при получении обобщенного показателя в n-мерном пространстве. На основе результатов анализируемых показателей объясняются подход построения траекторий освоения образовательных компетенций, способы выбора приближений, обосновываются алгоритм и выбор метода кластеризации для эффективной оценки качества образовательного процесса.

Ключевые слова: образовательные компетенции, комплексный подход, многомерное пространство, частные и обобщенные показатели, метрика, траектории освоения компетенций, кластерный анализ, интеллектуальный анализ данных.

DOI: 10.21667/1995-4565-2023-84-119-132

Введение

Одним из важнейших условий эффективного формирования компетенций студентов в системе высшего образования является организация контроля качества обучения. Поэтому оценка успеваемости и уровня сформированности компетенций должна объективно показывать степень освоения полученных знаний, умений и практических навыков в решении задач из различных областей будущей профессиональной деятельности.

В настоящее время в большинстве случаев оценка качества результата обучения студентов осуществляется с помощью вычисления среднего значения некоторого числа показателей [1]. В ряде случаев такой переход от многокритериального оценивания к обобщенному критерию не всегда оправдан и зачастую приводит к потере информативных признаков. В связи с этим целесообразно рассмотреть другие методики оценивания эффективности процесса подготовки специалистов в вузе. Для этого необходимо разработать комплексный подход, включающий технологию извлечения анализа информации в больших базах данных с применением концепции многомерного представления и методов кластерного анализа.

Предлагаемый подход основан на анализе показателей групп студентов, включающий в себя их классификацию, упорядочение, а также построение модели, определяемой траекторией в n-мерном пространстве. Использование различных инструментов моделирования, выбор метрики для получения обобщенного показателя в n-мерном пространстве, получение траекторий освоения образовательных компетенций, способы выбора приближений, постро-

ение алгоритма и выбор методов кластеризации являются основными задачами разрабатываемого комплексного подхода.

Стоит отметить, что применение технологии интеллектуального анализа данных (Data Mining) позволит не только избавиться от узкого локального оценивания сформированности профессиональных компетенций у студентов, но и создать новые программные решения, ориентированные на повышение эффективности управления организационным процессом, что является актуальной научно-технической задачей.

Постановка задачи

Несмотря на наличие разнообразных способов оценки уровня сформированности компетенций, существует необходимость разработки качественно нового комплексного подхода, который позволит устранить недостатки существующих решений благодаря разработке методов анализа, классификации и упорядочивания показателей исследуемых групп студентов, основанных на концепции интеллектуального анализа данных.

Одним из возможных способов реализации такого подхода при оценке качества обучения можно считать построение модели, определяющей некоторую траекторию в n -мерном пространстве. Одной из основных проблем, которую требуется решить, является установление размерности пространства n , выбор и определение частных показателей w_i , ($i = 1, 2, \dots, n$), которые максимально характеризуют динамику освоения профессиональных компетенций, предусмотренных образовательным стандартом направления подготовки обучающихся. Частные показатели w_i , ($i = 1, 2, \dots, n$), имеющие смысл для некоторой точки в пространстве R_n , объединяют обобщенным показателем эффективности освоения учебных компетенций. В качестве обобщенного выбирается показатель на основе расстояния между точками пространства. При этом можно выделить 2 подхода:

- первый подход подразумевает нахождение расстояния от точки пространства, которая соответствует конкретному студенту, до точки, которая соответствует «идеальному» студенту. Очевидно, для повышения эффективности освоения образовательных компетенций необходимо уменьшать расстояние между этими точками;

- второй подход основывается на определении расстояния от точки пространства, которая соответствует конкретному студенту, до точки, которая соответствует «нулевому» студенту. Очевидно, для повышения эффективности освоения образовательных компетенций необходимо увеличение расстояния между этими точками.

Для повышения обоснованности и качества исследования необходимо обосновать выбор различных подходов к определению расстояния в n -мерном пространстве, которое задает обобщенный показатель. Новизна исследования заключается в том, что алгоритм вычисления обобщенного показателя использует сравнения множества разнообразных метрик: евклидово расстояние, квадрат евклидова расстояния, хаусдорфово расстояние, расстояние городских кварталов (манхэттенское расстояние), расстояние Чебышева, степенное расстояние и др. Проведя сравнительный анализ полученных показателей в рассматриваемых метриках, выбирается наиболее подходящий вариант для получения качественного результата.

На основании расчетов нормированных значений частных показателей и получения обобщенных показателей могут быть получены траектории освоения образовательных компетенций, на основе которых в дальнейшем будет разработана технология извлечения анализа информации в больших базах данных с применением концепции многомерного представления и методов кластерного анализа.

Теоретическая часть

Для оценки эффективности процесса обучения приходится обрабатывать и анализировать большие объемы информации, что приводит зачастую к огромным временным затратам. Поэтому возникает необходимость в разработке решений, направленных на создание математических, программных и аппаратных средств, которые предназначены для получения быстрого и качественного результата обработки и анализа данных. Наиболее перспективным

направлением развития в данной сфере является разработка программных средств интеллектуального анализа данных. С их помощью может производиться обработка многомерных структур данных большого объема [7].

Для получения новой технологии, основанной на концепции многомерного представления, и методов кластерного анализа необходимо разработать методы классификации и упорядочивания показателей, определяющих уровень освоения образовательных компетенций. Следует отметить, что не существует универсальных алгоритмов и методов классификации показателей анализируемых групп студентов [6]. Более того, применение различных метрик при получении обобщенного показателя может приводить к различным результатам на одном и том же наборе данных. Различные инструменты моделирования, выбор метрики для получения обобщенного показателя в n -мерном пространстве, получение траекторий образовательных компетенций, способы выбора приближений, построение алгоритма и выбор методов кластеризации – основные инструменты для получения комплексного подхода при решении поставленной задачи.

Процесс обучения студентов в вузе можно определить в виде последовательности некоторых событий, реализуемых в течение периода обучения. К таким событиям можно отнести: вступительные испытания, обучение и освоение компетенций, предусмотренных учебным процессом, сдача сессии и т.д. [1]. Выполнив анализ рассматриваемых событий, можно выделить и определить частные показатели, которые позволяют определить размерность n -мерного пространства, используемого в дальнейшем. Задавая конкретные значения частным показателям той или иной группы, проводя нормализацию полученных значений, можно получить величину обобщенного показателя, которая определяется как расстояние между точками в n -мерном пространстве. Выбор подхода при нахождении обобщенного показателя определяется, исходя из требований к задаче: для каждого случая рассчитывается расстояние от исследуемого до «идеального» объекта или от исследуемого до «нулевого» объекта. Для корректного вычисления значения обобщенного показателя исследуется множество разнообразных метрик: евклидово расстояние, квадрат евклидова расстояния, хаусдорфово расстояние и др. Проведя сравнительный анализ полученных значений обобщенных показателей и построив траектории освоения образовательных компетенций в рассматриваемых метриках, выбирается наиболее приемлемый вариант для получения качественного результата.

Положим число исследуемых событий равным n , где $n > 0$. Данное значение задает количество частных показателей, которое в дальнейшем используется при проведении исследования. Стоит отметить, что использование определения количества частных показателей носит экспертный характер и влияет на выбор размерности пространства, в котором происходит обработка многомерных сложноорганизованных данных больших объемов. Увеличивая значение n , можно более детально представить факторы, влияющие на процесс обучения. При свертке показателей происходит уменьшение размерности пространства. Стоит учитывать, что увеличение n может существенно уменьшить вес одного показателя, а уменьшение n может привести к потере важной информации, касающейся эффективности освоения компетенции обучающегося.

В соответствии с рассуждениями, проведенными выше, процесс обучения характеризуется некоторым числом частных показателей $w_1^*, w_2^*, \dots, w_n^*$. Каждый из частных показателей ограничен минимальным $w_{i\min}^*$ и максимальным $w_{i\max}^*$ значениями, которые могут быть установлены нормативными актами и, исходя из логики рассуждений [2].

Для дальнейшего исследования ограничимся рассмотрением укрупненного направления групп специальностей 09.00.00 «Информатика и вычислительная техника». С учетом этого выбора, выполнив анализ исследуемой информации, проведена классификация частных показателей, которые условно были разделены на группы, представленные в таблице 1 раздела «Экспериментальные исследования».

Для корректного использования частного показателя в обобщенном необходима предварительная нормализация значений частных показателей исследуемого объекта. Были рас-

смотрены различные способы нормировки, из которых был выбран способ, преобразующий числовое значение частного показателя w_i исследуемого объекта в отрезок $[0;1]$:

$$w_i = \frac{w_i^* - w_{i\min}^*}{w_{i\max}^* - w_{i\min}^*}, \quad (1)$$

где w_i^* – текущее значение соответствующего показателя; $w_{i\max}^*$, $w_{i\min}^*$ – соответственно максимальное и минимальное значения рассматриваемого показателя [3].

Как было сказано выше, показатель w_i принимает значения от 0 до 1, т.е. $0 \leq w_i \leq 1$, $i = 1, 2, \dots, n$. В этом случае степень освоения студентом образовательных компетенций может быть определена совокупностью из n вещественных чисел: $W = (w_1, w_2, \dots, w_n)$. Относительные показатели w_i не имеют размерности, что позволяет проводить процесс суммирования (и совместной обработки) разных переменных [4].

Выполнив нормализацию полученных значений, необходимо получить величину обобщенного показателя, которую определим как расстояние между точками в n -мерном пространстве. Показатели $w_1, w_2, \dots, w_n$ можно представить в виде некоторой точки в n -мерном пространстве. Для вычисления значения обобщенного показателя необходимо выбрать метрику, которая позволит получить наиболее оптимальный принцип моделирования траекторий в n -мерном пространстве.

При проведении анализа данных рассмотрим простейшее многомерное евклидово пространство, в котором n может быть любым натуральным числом. В этом случае совокупность частных показателей есть точка в n -мерном евклидовом пространстве R^n , т.е. $W = (w_1, w_2, \dots, w_n)$.

Если эту точку соотносить с показателями студента, имеющего в рассматриваемом списке позицию j , тогда следует обозначить $w^{(j)} = w_1^{(j)}, w_2^{(j)}, \dots, w_n^{(j)}$, где $j = 1..m$, m – количество студентов в исследуемой группе.

Для вычисления значения обобщенного показателя найдем расстояние между точкой $w^{(j)}$ и точкой $w^{(k)}$ (соответствующими j -у и k -у студентам). Это простое Евклидово расстояние, которое определяется по формуле:

$$d_{jk} = \left(\sum_{i=1}^n (w_i^{(j)} - w_i^{(k)})^2 \right)^{\frac{1}{2}}. \quad (2)$$

Для дальнейшего использования более удобным будет следующее обозначение: w_{ji} – частный i -й показатель j -го студента. Тогда формула (2) будет записана в следующем виде:

$$d_{jk} = \left(\sum_{i=1}^n (w_{ji} - w_{ki})^2 \right)^{\frac{1}{2}} = \sqrt{\sum_{i=1}^n (w_{ji} - w_{ki})^2}. \quad (3)$$

Для корректного вычисления значения обобщенного показателя рассмотрим и другие метрики. Определим расстояние, используя метрику взвешенного евклидова расстояния. Этот случай рассматривается, если переменные w_{ji} , $i = 1, 2, \dots, n$ ранжированы между собой по степени важности, т.е. им присвоены веса c_i . Вес частного показателя определяет важность учета при классификации данного признака [9]. При расчете меры сходства учитывается вес показателя. Вес показателя оценивается, как правило, путем экспертных оценок.

Взвешенное евклидово расстояние будем определять по формуле:

$$d_{jk} = \sqrt{\sum_{i=1}^k (w_{ji} - w_{ki})^2 \cdot c_i}, \quad (4)$$

где c_i – вес i -го показателя; $c_i > 0$, $\sum_{i=1}^n c_i = 1$.

Однако могут возникнуть ситуации, при которых веса $c_1, c_2, \dots, c_n$ трудно определить, в этом случае рекомендовано использовать в качестве метрики простое евклидово расстояние. Если все веса факторов одинаковы, а одному из них надо повысить его вес, то это определяется опросом экспертов или директивными документами. Например, при организации учебного процесса следует с большим весом учитывать показатели, связанные с профильными дисциплинами или практическим освоением соответствующего материала.

Еще одним вариантом расчета значения обобщенного показателя является использование в качестве меры квадрата евклидова расстояния:

$$d_{jk} = \sum_{i=1}^k (w_{ji} - w_{ki})^2. \quad (5)$$

Данная мера расстояния рекомендуется к использованию в случаях, когда большие веса присваиваются наиболее удаленным друг от друга объектам. Считается, что использование данной меры особенно важно для расчета относительных значений показателей.

В тех случаях, когда требуется различать два обобщенных показателя w_j и w_k по какому-либо частному показателю, необходимо использовать метрику (расстояние) Чебышева. Расстояние Чебышева определяется как максимальное расстояние между соответствующими частными показателями:

$$d_{jk} = \max_i |w_{ji} - w_{ki}|. \quad (6)$$

Основной недостаток данной метрики состоит в том, что расстояние Чебышева является грубой мерой различия из-за игнорирования значительного количества имеющейся информации.

Обобщением различных подходов к определению расстояния в n -мерном пространстве является расстояние Минковского:

$$d_{jk} = \left(\sum_{i=1}^n |w_{ji} - w_{ki}|^p \right)^{\frac{1}{p}}. \quad (7)$$

При $p = 2$ формула расстояния Минковского соответствует простому евклидовому расстоянию [см. формулы (2), (3), и (7)]. При $p = 1$ получается расстояние Хэмминга (расстояние городских кварталов, манхэттенское расстояние). В большинстве случаев расстояние Хэмминга приводит к таким же результатам, что и в случае использования простого евклидова расстояния.

Особое внимание заслуживает расстояние Хаусдорфа, которое позволяет измерить расстояние между подмножествами метрического пространства, т.е. вычислить наибольшее из всех возможных расстояний от каждой точки одного множества до ближайшей точки к ней другого множества [5].

Предположим (X, d) – полное метрическое пространство. Выделим из него пространство $H(X)$, состоящее из компактных подмножеств X . Для полученного пространства $H(X)$ определим метрику:

1) определим расстояние между $x \in X$ и $T \in H(X)$: $d(x, T) = \min \{d(x, y) : y \in T\}$;

2) определим расстояние между двумя компактными множествами A и T : $d(A, T) = \max \{d(x, T) : x \in A\}$.

Необходимо заметить, что непосредственное определение расстояния Хаусдорфа требует значительных вычислений, однако при проведении процедуры кластеризации огромную роль играет функция расстояния, которая основывается на использовании данной метрики.

Подводя итог вышесказанному, заключаем, что ввиду универсальности, а также для сокращения вычислительных затрат целесообразно использование простого евклидова расстояния для получения значения обобщенного показателя. Для выбора функции расстояния при проведении кластерного анализа предпочтительна метрика Хаусдорфа.

Кроме рассмотренных выше метрик с различными видами расстояний, однородность многокритериальных показателей может быть определена комбинацией методов корреляционного и кластерного анализа [8].

Для проведения следующего этапа исследования был выбран алгоритм кластеризации k -средних (в дальнейшем k -means). Его суть можно представить следующим образом: на начальном этапе множество элементов пространства разбивается на заранее известное число кластеров k . Действие алгоритма основано на минимизации среднего квадрата расстояния на точках каждого кластера. Основная идея заключается в том, что на каждой итерации вычисляется центр масс (центроид) для исследуемого кластера, полученного на предыдущем шаге, затем полученные значения разбиваются на кластеры вновь в зависимости от того, какой из новых центров оказался ближе по выбранной метрике. Алгоритм завершается, когда на какой-то итерации не происходит изменения кластеров.

Алгоритм кластеризации k -means может быть представлен следующей последовательностью шагов:

Шаг 1. Получить выборку данных, произвести анализ полученных значений, упорядочить их по исследуемым признакам, определить метрику, выбрать число кластеров k .

Шаг 2. Выбрать количество случайных центров в пространстве заданных классифицирующих признаков.

Шаг 3. Каждый объект выборки отнести к кластеру, к центру которого объект оказался ближе. Для этого необходимо найти расстояние между каждым значением исследуемой выборки до каждого из k -центроидов, а затем добавить его к наиболее близкому кластеру.

Шаг 4. Вычислить новое значение центроида, центр масс кластера, как среднее арифметическое признаков всех объектов этого кластера.

Шаг 5. Оценить точность каждого кластера, т.е. рассчитать средний квадрат расстояния внутри кластеров.

Шаг 6. Если отклонение внутри кластера минимально, то выбрать этот кластер для нового перераспределения объектов.

Шаг 7. Уточнить положение центров.

Шаг 8. Повторять шаги 3 – 7 до тех пор, пока центры кластеров перестанут изменять положение.

Шаг 9. Получить результаты кластеризации, оценить точность полученных значений и время, затраченное на его реализацию.

Проанализировав выбранный метод кластеризации, следует отметить, что он не обеспечивает в полной мере получение приемлемого решения ввиду высокой чувствительности к выбору начальных центроидов, а также поиск эффективных решений приводит к значительным временным затратам из-за необходимости многократных итераций. В связи с этим необходимо модифицировать данный метод или разработать новый, который обеспечит высокий уровень результатов.

Экспериментальные исследования

Для проведения экспериментального исследования была взята выборка студентов РГРТУ укрупненного направления групп специальностей 09.00.00 «Информатика и вычислительная техника». С учетом этого выбора, проанализировав исследуемую информацию, проведена классификация частных показателей, которые условно разделены на группы, представленные в таблице 1.

Для каждого студента определим элемент пространства R^n : $W = (w_1, w_2, \dots, w_n)$. Если необходимо явно указать принадлежность показателей некоторому конкретному студенту с номером j из некоторого массива данных (группа студентов; студенческий поток; студент, обучающийся по укрупненному направлению 09.00.00), тогда будем использовать обозначение:

$$(W_j = (w_{j1}, w_{j2}, \dots, w_{jn})).$$

Таблица 1 – Классификация групп частных показателей

Table 1 – Classification of private indicators groups

Показатель	Группа	Содержание (определение) показателя	Принимаемые значения
1	2	3	4
w_1	1	ЕГЭ по математике	0..1
w_2	1	ЕГЭ по информатике (физике)	0..1
w_3	1	ЕГЭ по русскому языку	0..1
w_4	1	Индивидуальные достижения	0..1
w_5, w_6, w_7	1	Текущая аттестация в 1-м семестре	{0; 0,5; 0,75; 1}
w_8	1	Промежуточная аттестация (сессия) в 1-м семестре	{0; 0,333; 0,667; 1}
w_9	1	Уровень получаемой стипендии	{0; 0,7; 0,75; 0,8; 0,85; 0,9; 0,95; 1}
w_{10}	1	Степень участия в профессиональных мероприятиях	{0; 0,5; 1}
w_{11}	1	Уровень научно-методических публикаций	{0; 0,75; 1}
w_{12}	1	Практическая подготовка (курсы партнеров)	{0; 1}
$w_{13}, \dots, w_{20}$	Структура показателей повторяет структуру показателей 1-го семестра $w_5, \dots, w_{13}$		
$w_{21}, \dots, w_{28}$	Структура показателей повторяет структуру показателей 1-го семестра $w_5, \dots, w_{13}$		
$w_{29}, \dots, w_{36}$	Структура показателей повторяет структуру показателей 1-го семестра $w_5, \dots, w_{13}$		
w_{37}	3	Промежуточная аттестация (сессия) в 5-м семестре	{0; 0,333; 0,667; 1}
w_{38}	4	Стипендия	{0; 0,7; 0,75; 0,8; 0,85; 0,9; 0,95; 1}
w_{39}	5	Участие в профессиональных мероприятиях	{0; 0,5; 1}
w_{40}	6	Научно-методические публикации	{0; 0,75; 1}
w_{41}	7	Практическая подготовка (курсы партнеров)	{0; 1}
w_{42}	8	Работа на предприятиях ИТ-сферы	{0; 1}
$w_{43}, \dots, w_{48}$	Структура показателей повторяет структуру показателей 5-го семестра		
$w_{49}, \dots, w_{54}$	Структура показателей повторяет структуру показателей 5-го семестра		
$w_{55}, \dots, w_{60}$	Структура показателей повторяет структуру показателей 5-го семестра		
w_{61}		Обучение на военной кафедре	{0; 1}
w_{62}		Оценка за ВКР	{0,5; 0,75; 1}

Для проведения анализа классифицируемых данных введем в рассмотрение понятия «нулевого» и «идеального студентов» (в дальнейшем нулевой и идеальный). Нулевой студент является гипотетическим студентом, у которого все частные показатели имеют минимально возможные значения $w_{i\min}$, $i = 1, 2, \dots, n$. Назовем нулевого студента «неуспевающим» (B – bad) и поставим ему в соответствие точку пространства $B = (w_{b1}, w_{b2}, \dots, w_{bn})$.

Очевидно, для нормированных значений:

$$B = (w_{b1}, w_{b2}, \dots, w_{bn}) = B(0, 0, \dots, 0),$$

что следует из применения формулы (1).

Идеальный студент (G – good) может быть представлен точкой в пространстве R^n :

$$G = (w_{g1}, w_{g2}, \dots, w_{gn}) = G(1, 1, \dots, 1).$$

Частные показатели $(w_1, w_2, \dots, w_n)$, имеющие смысл для некоторой точки в пространстве R^n , можно объединить некоторым обобщенным показателем эффективности освоения учебных компетенций. Наиболее очевидным показателем эффективности можно считать расстояние в пространстве R^n между точками $B = B(0, 0, \dots, 0)$ и $W = (w_1, w_2, \dots, w_n)$. Это расстояние показывает степень приближения некоторого студента к нулевому студенту. Естественно, чем значение этого обобщенного показателя больше, тем выше эффективность освоения учебных компетенций (и наоборот).

Тогда обобщенный показатель эффективности

$$d_{BW} = d(B, W) = \sqrt{\sum_{i=1}^n (b_i - w_i)^2} = \sqrt{\sum_{i=1}^n w_i^2}.$$

Если учитывать весовые коэффициенты $c_i, i = 1, 2, \dots, n$ для каждого частного показателя, тогда $d(B, W) = d_{BW} = \sqrt{\sum_{i=1}^n c_i w_i^2}$.

В качестве альтернативного (или добавочного) обобщенного показателя эффективности освоения учебных компетенций можно использовать расстояние между конкретным студентом и идеальным студентом. Этот показатель дает оценку близости траектории студента к траектории идеального студента. Для данного критерия

$$d_{WG} = d(W, G) = \sqrt{\sum_{i=1}^n (w_i - g_i)^2} = \sqrt{\sum_{i=1}^n (w_i - 1)^2}.$$

Получим зону возможных значений обобщенного показателя относительно нулевого студента (рисунок 1).

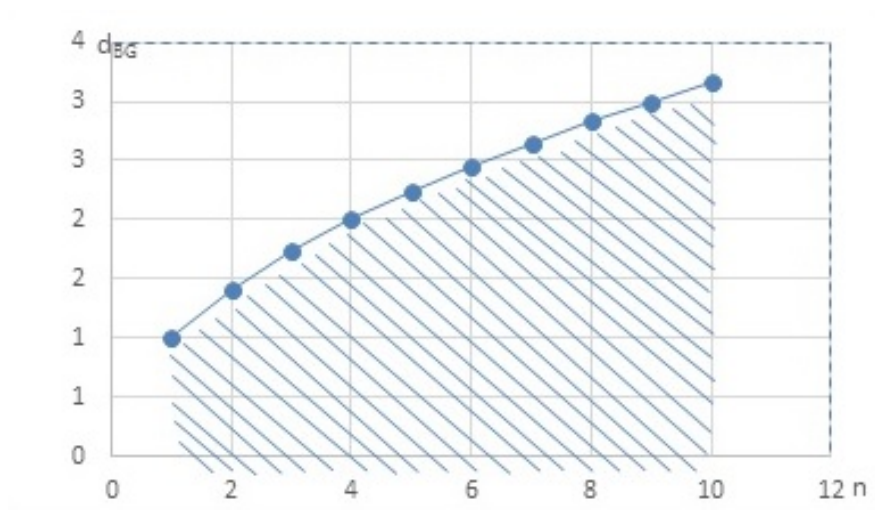


Рисунок 1—Зона возможных значений обобщенного показателя относительно нулевого студента

Figure 1 – Generalized indicator possible values relative to zero student

Расположение точек нулевого и идеального студентов для первых трех частных показателей может быть наглядно представлено в пространстве R^3 (рисунок 2).

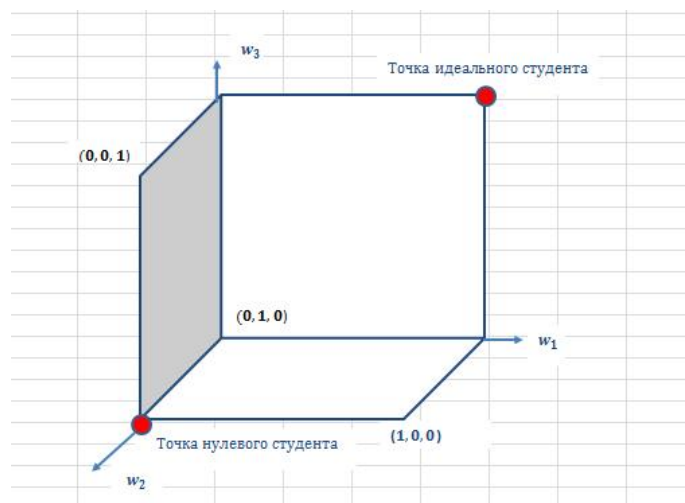


Рисунок 2–Расположение точек нулевого и идеального студентов в трехмерном пространстве
Figure 2 – Zero and perfect students points location id 3D space

В определенный момент времени значения некоторых частных показателей неизвестны. Будем считать их равными нулю. Пусть m – число известных частных показателей. Тогда

$$d(G, W) = \sqrt{\sum_{i=1}^n (1 - w_i)^2} = \sqrt{\sum_{i=1}^m (1 - w_i)^2 + \sum_{i=m+1}^n (1 - w_i)^2}.$$

Так как при $i = m+1, m+2, \dots, n$ имеем $w_i = 0$, тогда

$$d(G, W) = \sqrt{\sum_{i=1}^m (1 - w_i)^2 + \sum_{i=m+1}^n 1} \text{ или } d(G, W) = \sqrt{(n - m) + \sum_{i=1}^m (1 - w_i)^2}.$$

Каждый момент времени характеризуется реализацией того или иного события (частного показателя). В то же время некоторое событие происходит при безусловном наступлении предыдущих событий. Поэтому можно утверждать, что в некоторый момент времени известно некоторое количество m частных показателей: $w_1, w_2, \dots, w_m$ ($m \leq n$). Остальные показатели неизвестны (не реализованы) и равны нулю.

$$d(B, G) = \sqrt{\sum_{i=1}^n (w_i - g_i)^2} = \sqrt{\sum_{i=1}^n (0 - 1)^2} = \sqrt{n}.$$

Проанализировав частные показатели первой группы, проведя нормализацию полученных значений, можно представить результаты вступительных испытаний, например, для группы из 22 студентов в 3-мерном пространстве. Добавим к ним результаты «гипотетических» студентов, с максимальными и минимальными баллами соответственно. Каждая точка пространства соответствует результатам сдачи вступительных испытаний отдельного студента (рисунок 3).

На рисунке 3 выделяются три основные группы исследуемых студентов. К первой группе можно отнести студентов, у которых за все 3 экзамена получены баллы ЕГЭ, соответствующие «отличной» оценке. Вторую группу представляют студенты, баллы которых соответствуют оценкам «5» и «4». К третьей группе принадлежат студенты, результаты которых соответствуют оценке «3». Стоит отметить, что точки 24 и 25 определяют значения «идеального» и «нулевого» студентов соответственно.

Для проведения анализа освоения учебных компетенций в первом семестре с использованием траекторий многомерного пространства проведем выборку частных показателей 2-й и 3-й групп. Далее, проанализировав частные показатели этих групп, проведя нормировку полученных значений, получаем величину обобщенного показателя, которую определяем как расстояние между точками в n -мерном пространстве. Для наглядного отображения результатов освоения образовательных компетенций по результатам текущей аттестации за сентябрь, октябрь, ноябрь

(w_5, w_6, w_7) и промежуточной аттестации (w_8) построим траектории многомерного пространства. Построение выполним, используя, например, данные 4-х студентов из анализируемой группы и 2-х «гипотетических» студентов. Проведя расчет обобщенного показателя в евклидовой метрике для исследуемых объектов, получили начало траектории освоения учебных компетенций относительно «нулевого» и «идеального» студентов (рисунки 4, 5). На данном этапе обучения остальные показатели не реализованы и равны нулю, а размерность пространства равна 12.

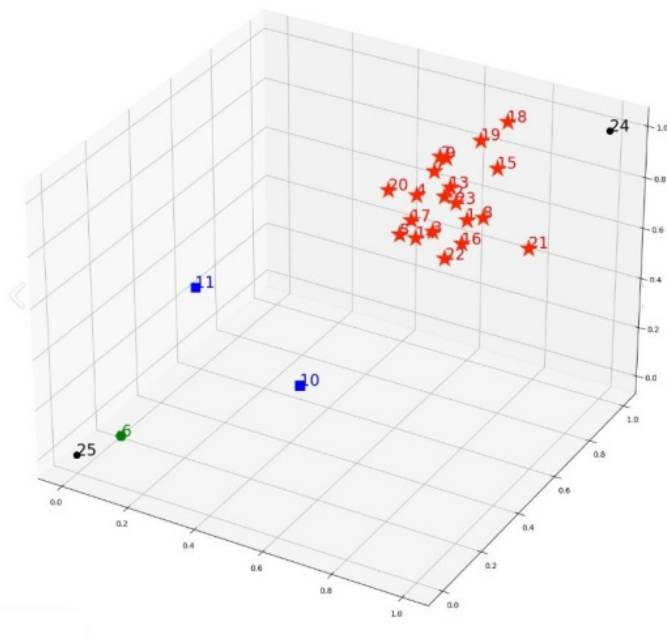


Рисунок 3—Результаты вступительных испытаний исследуемой группы студентов в 3-мерном пространстве

Figure 3 – Entrance test results of students group in 3D space

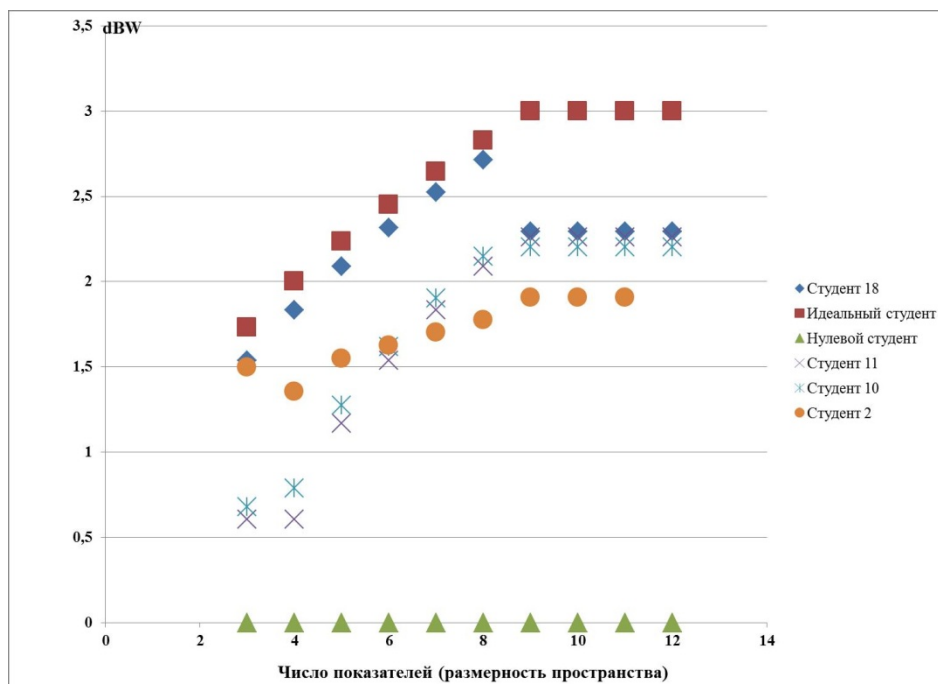


Рисунок 4—Траектория освоения учебных компетенций относительно «нулевого» студента

Figure 4 – Trajectory of educational competence development relative to «zero» student

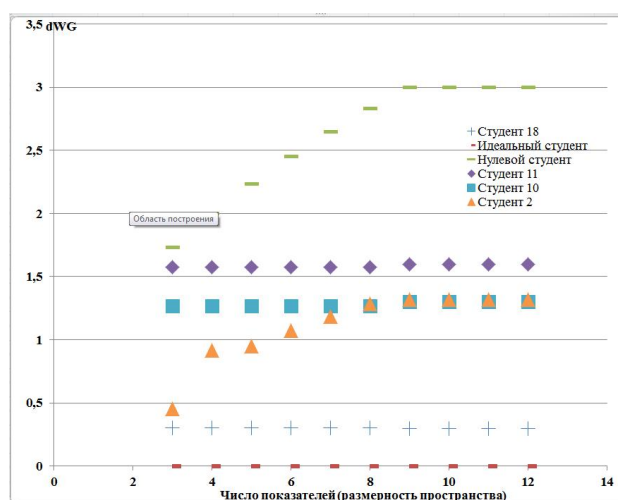


Рисунок 5—Траектория освоения учебных компетенций относительно «идеального» студента
Figure 5 – Trajectory of educational competence development relative to «perfect» student

Приведенные зависимости дают возможность получить дополнительную информацию по успеваемости конкретного студента в течение первого семестра обучения. Так, чем ближе расстояние от рассматриваемого студента к «идеальной» траектории, тем выше успеваемость обучающегося, т.е. значение обобщенного показателя эффективности на анализируемом участке не увеличивается относительно «идеального» студента. И чем дальше расстояние от траектории «нулевого» студента до рассматриваемой траектории, тем выше успеваемость студента, т.е. значение обобщенного показателя эффективности на анализируемом участке увеличивается относительно «нулевого» студента.

Можно предложить еще один способ сравнительного анализа результатов, достигнутых студентами. Для этих целей удобно использовать звездообразную диаграмму, которая применяется с целью группировки факторов, которые влияют на достижение глобальной цели, и в дальнейшем применить системный подход к принятию управленческих решений. На рисунке 6 приведено сечение множества критериев в определенный момент времени (звездообразная диаграмма) для рассматриваемого примера при $n = 8$.

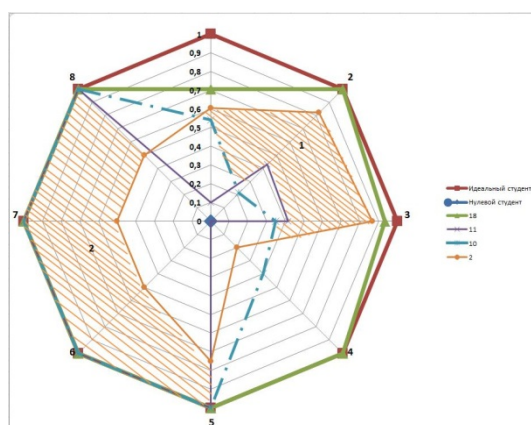


Рисунок 6—Сравнительный анализ результатов при помощи звездообразной диаграммы
Figure 6 – Comparative analysis of the results using star-shaped diagram

На полученной диаграмме по осям звездообразной диаграммы отложены значения частных показателей $w_{j1}, w_{j2}, \dots, w_{jn}$. В результате этого получается срез результатов освоения компетенций некоторого студента с номером j . Для сравнительного анализа отображены результаты «идеального» студента $j = 24$, нулевого студента $j = 25$, 4-х студентов, рассматриваемых выше ($j = 2; 10; 11; 18$). На диаграмме наглядно представлено преимущество студента ($j = 2$) при сдаче вступительных испытаний по сравнению с двумя студентами ($j = 10; 11$) (см.

заштрихованную зону 1). Зона 2 показывает снижение показателей освоения учебных компетенций данного студента в процессе обучения в первом семестре.

На следующем этапе выполнено исследование алгоритма кластеризации *k-means* данных, в ходе которого определена его высокая способность к обобщению. При проведении анализа исходные данные разбивались на выборки. Тестовая (контрольная) выборка в качестве исходных данных включала в себя 22 студента одной учебной группы и применялась для оценки точности метода кластеризации (рисунок 7).

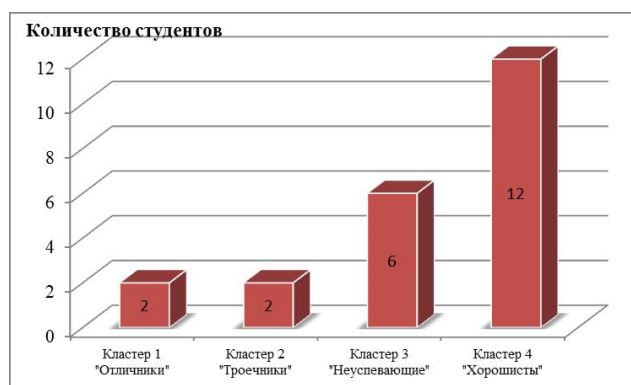


Рисунок 7—Результат кластеризации данных
Figure 7 – Figure7 – Data clustering result

Установлено, что в результате применения данного метода кластеризации выделяются 4 кластера, векторы данных которых можно сравнивать между собой, измеряя расстояние между ними. Показано, что полученные результаты кластеризации данных согласуются с траекториями освоения учебных компетенций, представленными на рисунках 4, 5, а именно, траектория студента 18 максимально приближается к траектории идеального студента и по полученным результатам кластеризации данный студент принадлежит кластеру «Отличники». Студенты 10 и 11 были отнесены к кластеру «Троечники», что также наглядно отражается по результатам траекторий освоения образовательных компетенций. Результаты полученных данных студента 2 иллюстрируют переход студента из кластера «Отличники» по результатам промежуточной и текущей аттестаций 1-го семестра в кластер «Хорошисты», а по результатам 2-го семестра переход в кластер «Троечники». Выявлено, что для получения адекватных результатов классификации с помощью метода кластеризации *k-means* данных необходима его модернизация для сокращения временных затрат на поиск эффективных решений. Дальнейшее исследование проводится при увеличении числа частных и обобщенных показателей и объемов выборки студентов.

Заключение

Подводя итог вышесказанному, можно заключить следующее: одним из возможных подходов при проведении анализа освоения образовательных компетенций является модель, определяемая траекторией в *n*-мерном пространстве. Значения частных и обобщенных показателей позволяют построить траектории освоения образовательных компетенций, анализ которых необходим для управления организационным процессом обучения студентов. В процессе увеличения количества показателей будет изменяться траектория освоения компетенций, на основе которой в дальнейшем может быть разработана технология анализа информации в базах данных по частным и обобщенным показателям с использованием методов кластерного анализа. Полученный алгоритм кластеризации данных использует множество метрик для определения меры сходства объектов. Однородность многокритериальных показателей определена методами корреляционного и кластерного анализа. В статье используются подходы, основанные на комбинации этих методов.

Библиографический список

1. **Щенёва Ю. Б.** Контроль освоения образовательных компетенций в процессе обучения IT-специалистов на основе анализа траекторий многомерного пространства // Специальный сборник трудов Всероссийской НПК с международным участием «Информационный обмен в междисциплинарных исследованиях II. Взгляд начинающих ученых». Рязань. 2023. С.87-91.
2. **Щенёва Ю. Б., Пылькин А. Н., Щенёв Е. С.** Формирование показателей эффективности управления организационным процессом подготовки IT-специалистов в вузе // Актуальные проблемы естественных, математических, технических наук и их преподавания: сб. науч. тр. Липецк, 2022. С. 207-212.
3. **Горленко О. А., Борбац Н. М.** Статистические методы в управлении качеством: учебник и практикум. М.: Юрайт, 2020. 306 с.
4. **Щенёва Ю. Б., Пылькин А. Н.** Повышение эффективности управления организационным процессом подготовки IT-специалистов в вузе // Математическое и программное обеспечение вычислительных систем: межв. сб. науч. трудов. Рязанский государственный радиотехнический университет. Рязань, 2022. С. 10-13.
5. **Russ Gordon**, Real Analysis: A First Course. Walla Walla, Washington, 1st Edition, 2011. 306 p.
6. **Ахо Альфред В.** Структуры данных и алгоритмы: Вильямс / пер. с английского и ред. Минько А. А., Ахо Альфред В., Хопкрофт Джон Э., Ульман Джеффри Д. – М. и др.: Вильямс, 2019. 382 с.
7. **Демидова Л. А. Соколова Ю. С.** Разработка двухуровневого классификатора сложноорганизованных многомерных данных больших объемов // Вестник Рязанского государственного радиотехнического университета. 2016. № 56. С. 71-82.
8. **Демидова Л. А., Тишкин Р. В., Юдаков А. А.** Разработка ансамбля алгоритмов кластеризации на основе матриц подобия меток кластеров и алгоритма спектральной факторизации // Вестник Рязанского государственного радиотехнического университета. 2013. № 46. С. 9-17.
9. **Майков К. А., Пылькин А. Н., Кузьменко С. Н., Теплов А. А.** Алгебраические особенности композиции алгоритмов вычисления фрактальных структур // Вестник Рязанского государственного радиотехнического университета. 2020. № 76. С. 117-127.

UDC 004.65

A MODEL FOR DEVELOPMENT OF EDUCATIONAL COMPETENCIES USING DATA MINING TOOLS

Y. B. Shcheneva, assistant of CAM department, RSREU, Ryazan, Russia;

orcid.org/0000-0001-6351-0399, e-mail:Shenyova@yandex.ru

A. N. Pylkin, Dr. Sc. (Tech.), professor of CAM department, RSREU, Ryazan, Russia;

orcid.org/0000-0001-9925-2870, e-mail:pylkin.a.n@rsreu.ru

E. S. Shchenev, post-graduate student, RSREU, Ryazan, Russia;

orcid.org/0009-0005-8438-7445, e-mail:Jheka1235317@gmail.com

O. A. Bodrov, Ph.D. (Tech.), docent of ST department, RSREU, Ryazan, Russia;

orcid.org/0009-0005-7225-6704, e-mail:bodrov.o.a@rsreu.ru

The article considers the problem of constructing a model to develop educational competencies using methods of data mining technology. The aim of the article is to develop integrated approach that includes data analysis technology using multidimensional representation concept and cluster analysis methods. Classification and particular indicators ordering methods determining the level of educational competencies development are considered. Metric choice is justified by obtaining generalized indicator in n-dimensional space. The approach to construct educational competence development trajectories and methods to choose approximations are explained, the algorithm and the choice of clustering method for effective assessment of educational process quality are substantiated based on analyzed indicators results.

Key words: educational competencies, integrated approach, multidimensional space, private and generalized indicators, metric, competence development trajectories, cluster analysis, intellectual data analysis.

DOI: 10.21667/1995-4565-2023-84-119-132

References

1. **Shcheneva Y. B.** Kontrol' osvoeniya obrazovatel'nyh kompetencij v processe obucheniya IT-specialistov na osnove analiza traektorij mnogomernogo prostranstva. *Special'nyj sbornik trudov Vserossijskoj NPK s mezhdunarodnym uchastiem «Informacionnyj obmen v mezhdisciplinarnyh issledovaniyah II. Vzglyad nachinayushchih uchenyh»*. Ryazan. 2023, pp.87-91. (in Russian).
2. **Shcheneva Y. B., Pylkin A. N., Shchenev E. S.** Formirovanie pokazatelej effektivnosti upravleniya organizacionnym processom podgotovki IT-specialistov v vuze. *Aktual'nye problemy estestvennyh, matematicheskikh, tekhnicheskikh nauk i ih prepodavaniya. Sbornik nauchnyh trudov*. Lipetsk, 2022, pp. 207-212. (in Russian).
3. **Gorlenko O. A., Borbats N. M.** *Statisticheskie metody v upravlenii kachestvom. Uchebnik i praktikum (Statistical methods in quality management. Textbook and workshop)*. Moscow: Yurayt, 2020. 306 p. (in Russian).
4. **Shcheneva Y. B., Pylkin A. N.** ovyschenie effektivnosti upravleniya organizacionnym processom podgotovki IT-specialistov v vuze. *Matematicheskoe i programmnoe obespechenie vychislitel'nyh sistem. Mezhvuzovskij sbornik nauchnyh trudov*. Ryazanskij gosudarstvennyj radiotekhnicheskij universitet. Ryazan, 2022, pp. 10-13. (in Russian).
5. **Russ Gordon.** *Real Analysis: A First Course*. Walla Walla, Washington, 1st Edition, 2011. 306 p.
6. **Aho Alfred V.** *Data structures and algorithms: Williams* / translated from English and ed. **Min-ko A. A., Aho Alfred V., Hopcroft John E., Ulman Jeffrey D. M.** et al.: Williams, 2019. 382 p.
7. **Demidova L. A., Sokolova Yu. S.** Razrabotka dvuhurovneвого klassifikatora slozhnoorganizovannyh mnogomernyh dannyh bol'shih ob'emov. *Vestnik Ryazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*. 2016, no. 56, pp. 71-82. (in Russian).
8. **Demidova L. A., Tishkin R. V., Yudakov A. A.** Razrabotka ansamblya algoritmov klasterizacii na osnove matric podobiya metok klasterov i algoritma spektral'noj faktorizacii. *Vestnik Ryazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*. 2013, no. 46, pp. 9-17. (in Russian).
9. **Majkov K. A., Pyl'kin A. N., Kuz'menko S. N., Teplov A. A.** Algebraicheskie osobennosti kompozicii algoritmov vychisleniya fraktal'nyh struktur. *Vestnik Ryazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*. 2020, no. 76, pp. 117-127. (in Russian).