

УДК 004.8

ИССЛЕДОВАНИЕ ПОСЕЩАЕМОСТИ ЭЛЕКТРОННЫХ КУРСОВ ДЛЯ ПОСТРОЕНИЯ ПРОФИЛЕЙ АКТИВНОСТИ ПОЛЬЗОВАТЕЛЕЙ С ИСПОЛЬЗОВАНИЕМ НЕЙРОННОЙ СЕТИ КОХОНЕНА

Родионов А.В.*ФГБОУ ВО «Байкальский государственный университет», Иркутск, e-mail: avr-v@yandex.ru*

На современном этапе развития образования в образовательных учреждениях достаточно четко прослеживается тенденция включения в учебный процесс электронных онлайн-курсов, которые реализуются с использованием различных систем управления образовательным контентом, в частности Moodle. При этом подчеркивается важность обеспечения достаточного уровня качества этих курсов для поддержания эффективности обучения. Для оценки качества электронных курсов используются различные квалиметрические модели, предполагающие формирование определенного набора показателей качества курса, которые могут быть оценены и впоследствии агрегированы. В такие наборы обычно входит множество показателей, связанных с качеством учебного материала, формами его подачи, техническими и финансовыми параметрами. В статье обосновывается значимость включения в такой набор показателя «интерес обучающегося», так как востребованность учащимися на протяжении всего периода обучения материалов электронного курса является важным критерием качества опубликованного материала. Такую востребованность можно оценить посредством анализа посещаемости электронного курса, при этом важно учитывать не только количество посещений курса, но и то, как посещаемость курса распределяется в течение всего периода работы с электронным курсом. В работе проводится исследование зависимости посещаемости электронных курсов от их структуры и содержания. Проведена кластеризация курсов с использованием нейронной сети Кохонена. В результате проведенного анализа определены три основных профиля посещаемости, которые соответствуют трем группам (кластерам) курсов, определены основные характеристические особенности этих групп. Построенные профили посещаемости могут быть использованы в качестве основы для классификатора курсов и их последующей оценки.

Ключевые слова: электронный курс, посещаемость курса, профиль посещаемости, нейронная сеть, качество курса

RESEARCH ON THE TRAFFIC OF ELECTRONIC COURSES TO BUILD USER ACTIVITY PROFILES USING A KOHONEN NETWORK

Rodionov A.V*Baikal State University, Irkutsk, e-mail: avr-v@yandex.ru*

At the present stage of development of education in educational institutions, there is a fairly clear tendency to include electronic online courses in the educational process, which are implemented using various educational content management systems, in particular, Moodle. This highlights the importance of ensuring that these courses are of sufficient quality to maintain learning effectiveness. To assess the quality of electronic courses, various qualimetric models are used, which involve the formation of a certain set of course quality indicators that can be evaluated and subsequently aggregated. Such sets usually include many indicators related to the quality of educational material, the forms of its presentation, technical and financial parameters. The article substantiates the importance of including the «student interest» indicator in such a set, since demand by students throughout the entire period of study of e-course materials is an important criterion for the quality of the published material. This demand can be assessed by analyzing the e-course attendance, and it is important to consider not only the number of course visits, but also how the course attendance is distributed throughout the entire period of work with the e-course. The paper investigates the dependence of the attendance of electronic courses on their structure and content. Courses were clustered using the Kohonen network. As a result of the analysis, three main attendance profiles have been identified that correspond to three groups (clusters) of courses, and the main characteristic features of these groups have been identified. The constructed attendance profiles can be used as the basis for a classifier of courses and their subsequent evaluation.

Keywords: e-course, course attendance, attendance profile, neural network, course quality

В настоящее время образовательные учреждения активно внедряют в учебный процесс электронные онлайн-курсы, используя различные системы управления учебным контентом, например Moodle. Причем часто речь идет не только о дистанционном, но и об очном и смешанном обучении, когда электронные курсы обеспечивают информационную, методическую и технологическую поддержку образовательного процесса. Для того, чтобы не потерять эффективность обучения и представлять интерес для обучающихся, электронные курсы должны быть достаточно высокого качества. При этом само понятие «качество» одно-

значно не определено и существуют разные подходы к его оценке [1]. В рамках квалиметрического подхода, который нашел свое отражение в ряде работ [2–4], процедура оценивания предполагает формирование некоторого набора показателей качества курса, которые можно каким-либо образом оценить с последующей «сверткой» оценок этих показателей для формирования итоговой (интегральной) оценки электронного курса [4, 5].

В числе показателей, характеризующих качество электронного курса, важное место занимает показатель, который можно назвать «интерес обучающегося». Работа

с электронным онлайн-курсом объективно предполагает ограничение контактов учащегося с преподавателем, вследствие чего курс необходимо разрабатывать и использовать таким образом, чтобы минимизировать возникающие дидактические потери. Это предполагает активную работу с электронным курсом как студента, так и преподавателя, преподаватель, в частности, должен использовать различные мультимедиа и интерактивные элементы, позволяющие обеспечить двухстороннее взаимодействие с обучающимся, выставление оценок, публикацию отзывов и рецензий, а студент, в свою очередь, может обращаться с вопросами по материалам курса и пр.). Все это имеет вполне объективное отражение в виде количества обращений пользователей к материалам курса (посещаемость курса), причем посещаемость должна быть относительно равномерно распределена по всему периоду обучения. В случае если преподаватели рассматривают электронный курс только как инструмент предоставления учебного материала и не используют в полной мере интерактивные элементы электронных курсов, то это будет проявляться в низкой и неоднородной посещаемости электронных курсов [6]. Таким образом, анализируя посещаемость электронного курса, можно попытаться оценить востребованность курса в зависимости от контента, что косвенно оценивает и качество самого учебного контента. При этом важно рассмотреть не просто количество посещений курса, а как посещаемость курса распределяется на всем периоде обучения. Это позволит построить типовые профили активности пользователей курса – распределение количества обращений к курсу по временным периодам, что можно использовать как основу для классификатора.

В работе ставится задача провести анализ посещаемости электронных курсов и выделить типовые профили активности по неделям в зависимости от содержания электронных курсов. Актуальность работы заключается в определении профилей активности, изучения зависимости профиля от содержания и условий использования электронных курсов. В работе использованы методы data mining и нейронные сети, что в современных условиях развития цифровых информационных сред позволяет эффективно исследовать скрытые взаимосвязи и минимизировать возможную субъективность получаемых оценок.

Материалы и методы исследования

В качестве инструмента для сегментации электронных курсов по посещаемости

и построения соответствующих профилей в работе использована одна из технологий интеллектуального анализа данных (data mining) – кластеризация. При работе с веб-контентом (электронные курсы в подавляющем большинстве случаев реализуются с применением веб-технологий) можно выделить два основных направления применения data mining: интеллектуальный анализ контента и интеллектуальный анализ использования сайтов [7, 8]. Данное исследование относится ко второму направлению, которое предполагает выявление закономерностей в действиях пользователей системы.

На этапе предварительной обработки осуществляется «добыча» данных из баз данных или файлов журнала системы с целью формирования датасета для следующего этапа обработки. На этом же этапе могут осуществляться операции нормализации данных, заполнения пропусков и пр. Источником данных для анализа в работе является система «Электронный университет» на основе E-LMS Moodle (<https://exam.bgu.ru/edu-portal>). Информация о деятельности пользователей системы хранится в базе данных MS Sql в таблице, описание полей которой приведено в табл. 1.

Таким образом, в данной таблице хранится вся необходимая для анализа информация, которую можно отфильтровать по ряду параметров: по дате, контексту, типу операций, пользователям и пр.

На следующем этапе решается задача кластеризации, суть которой состоит в том, чтобы провести сегментацию (распределение) электронных курсов по частоте обращений пользователей к ресурсам и элементам электронного курса. При выборе алгоритма следует учитывать достаточно большой размер, который будет иметь датасет, так как количество обращений к электронным курсам даже за семестровый курс обучения может исчисляться тысячами, что в случае массового использования курсов приводит к более чем миллиону записей. Применение нейронных сетей для решения задачи классификации позволяет избежать сложных и громоздких расчетов при оперировании большими объемами многомерных данных [9]. Самоорганизующаяся карта Кохонена (самоорганизующаяся карта признаков) представляет собой соревновательную нейронную сеть с обучением без учителя. Идея сети предложена финским ученым Т. Кохоненом в начале 1980-х гг. [10]. Сеть Кохонена представляет собой два слоя – входной и выходной, причем все нейроны входного слоя связаны с каждым из нейронов выходного (рис. 1).

Таблица 1

Поля таблицы с журналом действий пользователей

Поле	Описание
eventname	Описание события
component	Модуль, который вызвал событие
action	Тип действия
target	Цель, над которой выполняется действие
objecttable	Имя таблицы базы данных, которое представляет объект события
objectid	Идентификатор записи объекта из objecttable.
crud	Типы операции: «создать», «редактировать», «обновить» или «удалить»
contextlevel	«Уровень» совершенной операции (контекста): курс, модуль, категория курса и т.д.
contextinstanceid	В зависимости от уровня контекста это может быть идентификатор курса, идентификатор модуля курса, категория курса
userid	Идентификатор пользователя
courseid	Идентификатор курса. Используется только для контекстов на уровне курса и ниже
timecreated	Время, когда произошло событие

Таблица 2

Фрагмент датасета

#	courseid	Номер недели										
		36	37	38	39	45	...	47	48	49	50	51
5	4	0,09	0,73	0,45	0,36	0,36	...	0,18	0,27	1,00	0,64	0,27
6	5	0,25	0,50	0,63	0,38	0,88	...	0,50	0,25	0,75	0,38	0,75
7	6	0,27	0,36	0,55	0,45	0,18	...	0,45	0,27	0,45	0,91	0,55
8	7	0,00	0,19	0,50	0,19	0,38	...	1,00	0,63	0,56	0,44	0,19
9	8	0,20	0,40	0,70	0,20	0,20	...	0,20	0,30	0,30	0,40	0,30
10	9	0,00	0,44	0,56	0,44	0,33	...	0,33	0,44	0,56	1,00	0,22
11

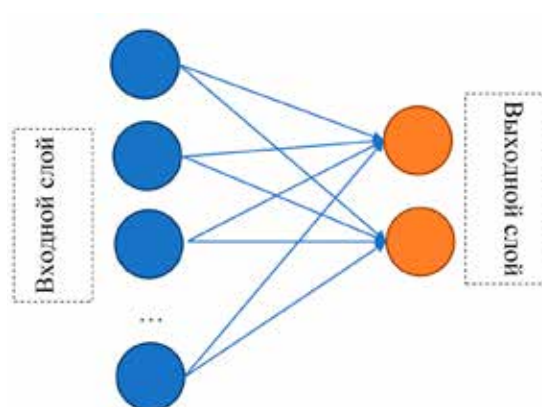


Рис. 1. Структура сети

Число входных нейронов равно количеству признаков объектов, а количество выходных нейронов – числу кластеров. Таким образом, задание структуры сети осуществ-

ляется путем выбора количества нейронов выходного слоя (слоя Кохонена). Нормализация входных переменных выполняется в пределах $[-1, 1]$ или $[0, 1]$. В качестве метрики обычно используется евклидово расстояние:

$$d_E(A, B) = \sqrt{\sum_{j=1}^J (a_j - b_j)^2} = \|A - B\|.$$

Для исследования была взята статистика посещаемости электронных курсов системы «Электронный университет» Байкальского государственного университета за первый учебный семестр 2022 учебного года. За этот период активными были 1728 учебных курсов. Количество обращений ко всем электронным курсам за данный период – 2470054. На этапе первоначальной

обработки был сформирован датасет в виде кросс-таблицы с группировкой запросов пользователей по неделям, выполнены операции замены пропусков (в случае, если в течение учебной недели не было обращений, ставился 0) и нормализации исходных данных в диапазоне $[0, 1]$ (максимум обращений к курсу в течение семестра брались за единицу, остальные значения по неделям рассчитывались относительно максимума). Фрагмент датасета представлен в табл. 2.

В рамках проводимого исследования для решения задачи сегментации курсов по посещаемости были построены несколько сетей: количество входных нейронов соответствовало учебным неделям выбранного семестра, а выходных нейронов равнялось числу идентифицируемых профилей посещаемости. Так как нормализация датасета была проведена в пределах $[0, 1]$, то случайная инициализация весовых коэффициентов с использованием генератора случайных чисел осуществлялась в диапазоне:

$$0,5 - \frac{1}{\sqrt{M}} \leq w_{ij} \leq 0,5 + \frac{1}{\sqrt{M}},$$

где M – количество входных переменных сети (количество нейронов во входном слое),

w_{ij} – весовой коэффициент, $i = \overline{1, M}$, $j = \overline{1, N}$, N – количество нейронов в выходном слое.

Процесс обучения состоит в циклической подаче обучающих примеров (входных векторов) текущей эпохи обучения и расчета евклидовых расстояний от них до центров всех кластеров [11]:

$$d_j = \sqrt{\sum_{i=1}^M (\tilde{x}_i - w_{i,j})^2},$$

где \tilde{x}_i – значения элементов входного вектора.

Для нейрона-победителя w_{ij} осуществляется коррекция весового коэффициента:

$$w_{i,j}^{l+1} = w_{i,j}^l + v(\tilde{x}_i - w_{i,j}^l),$$

где v – скорость обучения, равномерно убывающая от 1 до 0 по мере обучения сети. Процесс обучения проходит до того момента, пока изменения весовых коэффициентов не перестают быть значимыми. В качестве инструмента для работы с данными, построения сети и последующего анализа были использованы модули языка Python, в частности Pandas, Numpy и Sompy,

построение графиков осуществлялось с использованием программы Excel.

Рассмотрим сегментацию курсов по посещаемости при выборе трех выходных нейронов. В результате обучения сети разделение курсов произошло на три кластера, центры кластеров представлены в табл. 3, а график – на рис. 2.

На основе контекстного анализа курсов (проверялось соответствие рабочей программы дисциплины, наличие интерактивных и мультимедиаэлементов и пр.), попавших в первый кластер («кластер 0» в табл. 3 и на рис. 2), можно сделать вывод, что в этот кластер попали курсы, которые по сути представляют собой электронные учебники – простое переложение учебных пособий и конспектов лекций в «цифровую» форму, зачастую с использованием самых простых инструментов. Типовая структура – опубликованные учебные материалы в виде файлов лекций, презентаций и пр. при минимуме или полном отсутствии применения интерактивных элементов. График посещаемости на рис. 2 достаточно четко характеризует интерес обучающихся – видно явное снижение по мере обучения (синий график).

Второй кластер («кластер 1» в табл. 3 и на рис. 2) составляют курсы, в которых преподаватели сделали упор на промежуточную аттестацию, при этом в этих курсах может содержаться и теоретическая составляющая, но в большинстве случаев будет отсутствовать интерактивная деятельность. При этом на графике (рис. 2, оранжевый график) посещаемость растет именно к сессии, достигая там максимума.

Наконец третий кластер («кластер 2» в табл. 3 и на рис. 2) составляют курсы, в подавляющем большинстве содержащие как минимум теоретическую и практическую части, а также аттестации в виде тестов и заданий – текущую и промежуточную. Преподаватели используют курсы из этого кластера не только как учебник или систему тестирования, но и осуществляют взаимодействие со студентом посредством доступного им инструментария, принимают и комментируют лабораторные работы, используют средства коммуникации и пр. Посещаемость таких курсов достаточно стабильна, но видны достаточно выраженные недели роста и спада посещаемости. Сравнение посещаемости с учебными планами показало, что пики посещаемости совпадают с периодами сессии заочного обучения, что вполне объясняет их появление – электронные курсы часто используют не только для студентов очного, но и заочного обучения. Количество курсов, попавших в первый кластер – 22%, во второй кластер – 25%, в третий кластер – 53%.

Таблица 3

Центры кластеров

Номер кластера	Номер недели															
	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
Кластер 0	.14	.31	.16	.13	.21	.14	.26	.18	.23	.07	.11	.07	.05	.06	.05	.02
Кластер 1	.01	.06	.06	.05	.07	.08	.08	.16	.14	.15	.16	.23	.25	.40	.45	.21
Кластер 2	.10	.49	.40	.43	.58	.60	.38	.65	.69	.34	.70	.62	.58	.56	.63	.29

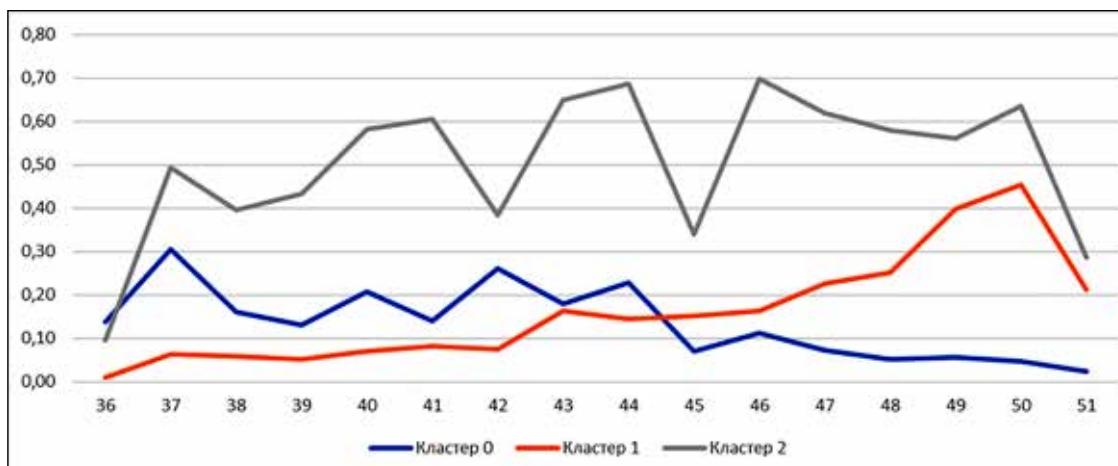


Рис. 2. Профили кластеров

Таблица 4

Центры кластеров

Номер кластера	Номера недель															
	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
Кластер 0	.13	.63	.35	.52	.67	.54	.49	.71	.82	.33	.79	.58	.55	.55	.69	.26
Кластер 1	.07	.35	.40	.33	.45	.58	.29	.58	.50	.34	.59	.58	.60	.51	.58	.27
Кластер 2	.01	.05	.04	.04	.07	.06	.08	.11	.11	.08	.11	.19	.31	.45	.51	.23
Кластер 3	.13	.26	.18	.14	.20	.18	.19	.17	.19	.13	.13	.12	.07	.07	.07	.02

Таким образом можно сделать вывод, что только примерно в половине из числа всех курсов ведется активная и равномерная деятельность.

Попробуем увеличить размерность сети до 4. В этом случае сеть разделит курсы на 4 кластера, центры кластеров представлены в табл. 4, а график – на рис. 3.

Наблюдаемая картина в целом схожа с предыдущими настройками кластеризации: четко выражены кластеры курсов с убывающей посещаемостью и возрастающей посещаемостью к концу семестрового обучения (началу сессии). При этом кластер с интерактивными курсами разделится

на два подкластера – в зависимости от того, в какие периоды выпадают сессии для заочного обучения. В кластер 0 попали 27% электронных курсов, в кластер 1 – 28%, в кластер 2 – 20% и в кластер 3 – 25%.

При увеличении размерности выходного слоя до 5 происходит выделение кластера, в котором собираются курсы, не используемые для заочного обучения (серая линия на рис. 4). Посещаемость курсов, которые используются как для очного, так и для заочного обучения, резко повышается в периоды заочных сессий и падает с их окончанием. Центры кластеров приведены в табл. 5, а профили – на рис. 4.

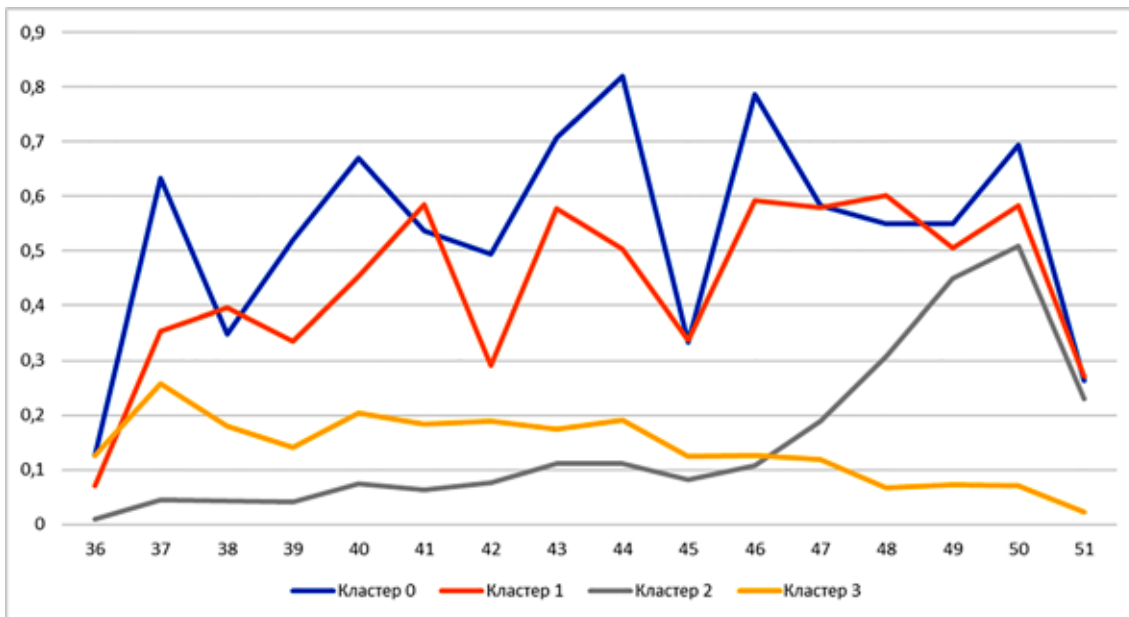


Рис. 3. Профили кластеров

Таблица 5

Центры кластеров

Номер кластера	Номера недель															
	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
Кластер 0	.13	.62	.26	.53	.66	.46	.53	.73	.82	.29	.74	.56	.50	.54	.70	.21
Кластер 1	.09	.49	.55	.41	.63	.73	.30	.71	.71	.38	.74	.60	.58	.54	.63	.37
Кластер 2	.05	.24	.26	.32	.28	.40	.28	.42	.36	.37	.44	.55	.52	.52	.55	.17
Кластер 3	.01	.03	.03	.02	.04	.03	.05	.08	.07	.07	.09	.17	.23	.44	.56	.22
Кластер 4	.11	.25	.19	.10	.19	.11	.21	.19	.24	.09	.16	.08	.06	.04	.05	.03

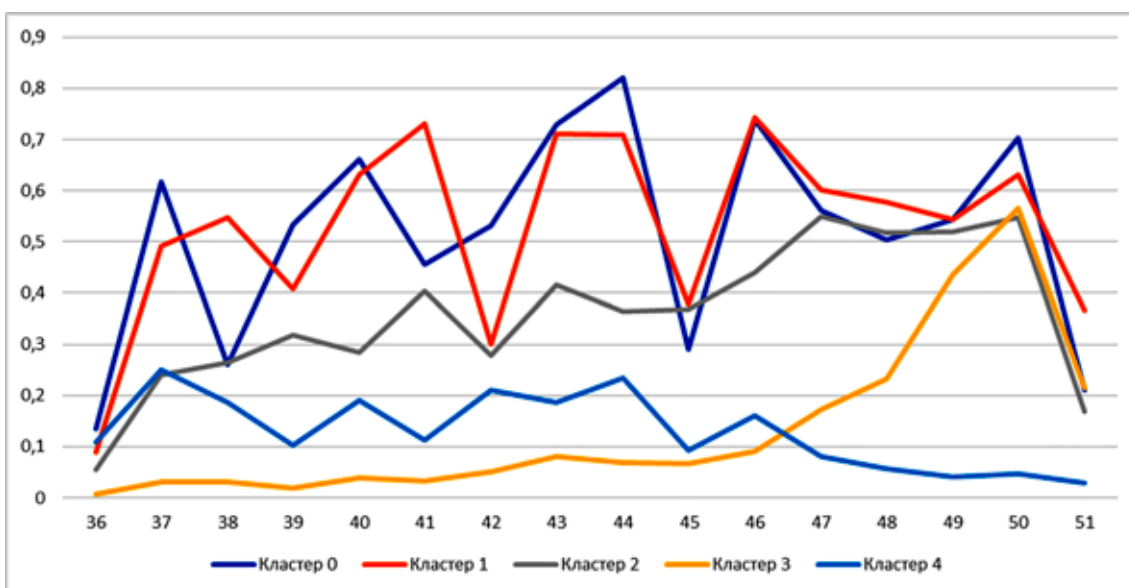


Рис. 4. Профили кластеров

В рамках проведенного исследования были рассмотрены и другие настройки сети, однако дальнейшее увеличение размерности сети еще больше «размывает» третий кластер курсов, практически не затрагивая первый и второй кластеры, которые получились при использовании сети размерности 3.

Результаты исследования и их обсуждение

Полученные результаты кластеризации, безусловно, зависимы от множества факторов: от процесса сбора и предварительной обработки исходных данных; метода кластеризации; выбранных параметров метода. Однако полученные результаты вполне адекватно отражают зависимость интереса участников обучения от структуры, содержания и применения электронного курса. Проведенное исследование показало, что в зависимости от контента и принципов использования электронных курсов можно четко выделить три профиля активности пользователей курсов (рис. 2): с сильным всплеском посещаемости в начале обучения и постепенным спадом интереса к завершению, обратный профиль, когда всплеск посещаемости идет на этапе завершения (аттестации), и профиль, когда посещаемость относительно равномерно распределена на всем промежутке обучения. В ряде случаев видится целесообразным выделять пять профилей, если важно учитывать условия применения курсов (рис. 4).

Используя данные профили, можно построить классификатор курсов и на его основе определять численную оценку показателя активности. В этом случае необходимо каждому кластеру сопоставить некоторую числовую оценку. Информация о посещаемости исследуемого курса подается на вход сети и относится к одному из существующих кластеров, либо делается вывод о невозможности такого отнесения. В случае отнесения курса к какому-либо кластеру он получает соответствующую, заранее определенную числовую оценку, которую можно использовать для построения интегральной оценки. При определении числового значения оценки следует учитывать, что оптимальными характеристиками с точки зрения качества образовательного процесса обладают курсы, попавшие в третий кластер (рис. 2). Курсы, попадающие во второй кластер, занимают свою нишу в качестве тестовых систем и могут быть использованы по назначению (например, для проведения промежуточной аттестации). Наибольшие риски несут курсы, попавшие в первый кластер – падение интереса в процессе обучения чревато не только потерей эффективности обучения по данным

дисциплинам, но и в целом образовательно-го процесса.

Однако использование показателей только на основе активности пользователей электронных курсов может привести к ошибочным выводам: например, далеко не всегда только курсы с интерактивным содержанием попадали в третий кластер – при анализе контента были выявлены курсы, не содержащие интерактивных элементов, не соответствующие рабочим программам либо имеющие иные, достаточно серьезные замечания (таких курсов около 20% из числа курсов, вошедших в этот кластер). Но даже в таком виде курсы были востребованы учащимися на протяжении всего периода обучения, что является косвенным подтверждением качества опубликованного материала. Таким образом, построенный по профилям активности пользователей курсов классификатор может служить важным инструментом для построения целостной оценки качества электронного курса, но его применение возможно только совместно с другими показателями.

Список литературы

1. Андреев А.А. Онлайн-обучение и его качество // Электронное обучение в непрерывном образовании. 2015. № 1–2. С. 82–91.
2. Шишлина Н.В. Квалиметрическая оценка электронных образовательных ресурсов вуза // Информационные технологии. Проблемы и решения: материалы международной научно-практической конференции. 2018. № 1 (5). С. 3–7.
3. Войцеховская М.Ф. Критерии оценки электронных курсов как средство управления качеством электронной информационно-образовательной среды вуза // Открытое и дистанционное образование. 2019. № 4 (76). С. 34–39. DOI: 10.17223/16095944/76/5.
4. Шалкина Т.Н. Метод деревьев решений в задаче синтеза алгоритма оценки качества электронных учебных курсов // Образовательные технологии и общество. 2016. Т. 19. № 4. С. 322–333.
5. Родионов А.В. Модель скоринговой оценки электронного курса как инструмент управления качеством образовательного процесса // Современные наукоемкие технологии. 2022. № 10–1. С. 37–41.
6. Шалкина Т.Н. Показатели и критерии качества электронного учебного курса // Образовательные технологии и общество. 2015. № 18–3. С. 608–619.
7. Миркин Б.Г. Методы кластер-анализа для поддержки принятия решений: обзор: препринт WP7/2011/03. М.: Издательский дом Национального исследовательского университета «Высшая школа экономики», 2011. 88 с.
8. Прохоров И.В., Канаков Ф.Л., Кустов Д.С. Разработка модели кластеризации пользователей сайта университета и определение их интересов на основе статистики посещений веб-сайта // Образование и наука в России и за рубежом. 2018. № 6 (41). С. 120–125.
9. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных. М.: ДМК Пресс, 2015. 400 с.
10. Кохонен Т. Самоорганизующиеся карты. М.: БИНОМ. Лаборатория знаний, 2008. 655 с.
11. Дударов С.П. Модели самоорганизующихся искусственных нейронных сетей для идентификации стационарных промышленных источников загрязнения воздуха // Математическое моделирование. 2017. Т. 29. № 1. С. 33–44.