

## ФОРМИРОВАНИЕ УЧЕБНЫХ ПЛАНОВ МЕТОДОМ АССОЦИАЦИИ

Китаев Д.Ф., Макаров А.А., Смольников С.Д., Логвинова Е.А.

*ЧОУ ВПО «Международный институт рынка», г. Самара, Россия (443030, г. Самара, ул. Г.С. Аксакова, 21), e-mail: maa@imi-samara.ru*

Статья представляет собой дальнейшее развитие практики целенаправленного использования инструментов анализа данных применительно к задаче эффективного формирования учебных планов и программ в рамках федеральных государственных образовательных стандартов третьего поколения. Показано, каким образом с помощью понятия «дескриптор» дидактической единицы, коэффициентов, построенных из дескрипторов и понятия «ассоциация учебных элементов», можно формализовать понятие «близость дидактических элементов» и тем самым выстроить оптимальную последовательность вхождения учебных элементов в соответствующие планы и программы. Описанная методика поясняется на примере построения последовательности изучения учебных элементов фрагмента дисциплины «Анализ данных». Приводятся рассчитанные матрицы близости учебных элементов и мер ассоциаций.

Ключевые слова: учебный элемент, дидактическая единица, рабочий и учебный план, ассоциация, дескриптор, профессиональный стандарт, целочисленное программирование.

## THE CURRICULUM METHOD ASSOCIATION

Kitaev D.F., Makarov A.A., Smolnikov S.D., Logvinova E.A.

*CHOU VO «International market Institute», Samara, Russia, 443030, Samara, street of G.S. Aksakov, 21), e-mail: maa@imi-samara.ru*

The article is a further development of the practice of purposeful use of data analysis tools for the effective formation of educational plans and programs within the Federal state educational standards of the third generation. Shows how using the term "descriptor" didactic units, the ratios constructed from descriptors and the concept of "Association for educational items, we can formalize the notion of "proximity didactic elements" and thereby to build an optimal sequence of entering the educational elements into relevant plans and programmers. The described methodology is illustrated on the example of constructing a sequence of learning elements of a fragment of the discipline of "Data analysis". Shows the calculated proximity matrix of the training elements and measures of associations.

Keywords: educational element, didactic unit, work and curriculum, association, descriptor, professional standard, integer programming.

Статья представляет собой дальнейшее развитие практики использования математических инструментов анализа данных применительно к задачам формирования учебных планов и эффективного управления ВУЗами (4,5).

Сфера услуг в настоящее время стремительно развивается. Потребители нуждаются в качественных транспортных, медицинских, финансовых (3) и прочих услугах.

В образовании, которое также является специфической услугой, в настоящее время осуществляется переход к компетентностной парадигме, позволяющей повысить эффективность образовательного процесса и приблизить качество подготовки специалистов к возросшим требованиям работодателей /1/.

В федеральных образовательных стандартах третьего поколения предписано формирование рабочих программ и учебных планов осуществлять на основе соответствующих наборов профессиональных и общекультурных компетенций. В работе (4)

авторами была описана методика формирования набора дидактических единиц, реализующих ту или иную профессиональную компетенцию на основе некоторого набора начальных компетенций и базы дидактических единиц. Эта методика предполагает использование соответствующим образом сконструированной и обученной нейронной сети. Для обучения такой нейронной сети, как известно, необходим достаточно большой набор обучающей выборки, что не всегда возможно обеспечить для всех компетенций (особенно это касается профессиональных компетенций). Поэтому опишем другой возможный вариант решения той же задачи.

Предположим, нам нужно сформировать учебный план, реализующий ту или иную компетенцию и имеющий объем  $V$  зачетных единиц. У нас имеется база из  $N$  дидактических единиц, каждая из которых имеет объем  $v_i$  ( $i=1,..N$ ) и относительную полезность  $c_i$  применительно к рассматриваемой компетенции.

Введем  $N$  булевых переменных  $x_i$ :

$$x_i = \begin{cases} 1, & \text{если дидактическая единица входит в состав плана} \\ 0, & \text{в другом случае} \end{cases}$$

Заметим, что по смыслу дидактическая единица – это целостный учебный элемент, который нельзя декомпозировать на более простые элементы без потери смысла. Поэтому такие единицы либо входят в план целиком, либо не входят вообще.

Тогда целевая функция задачи формирования плана – это максимизация суммарной полезности:

$$\sum_{i=1}^N c_i x_i \rightarrow \max$$

с очевидными ограничениями:

$$\sum_{i=1}^N v_i x_i \leq V$$

В такой постановке эта задача решается методами целочисленного программирования, которые широко используются в моделировании управленческих и экономических систем (5).

После формирования оптимального набора единиц (учебных элементов) следующей по важности задачей является задача определения правильной (эффективной) последовательности их изучения. Для решения этой задачи введем понятие «близость дидактических единиц» на основе представления о дескрипторах. Дескриптором будем считать ключевое слово (понятие, термин и т.п.), имеющий вполне определенную семантику в контексте изучаемого учебного элемента (например, «база данных», «имитационная модель», «класс объектов» и т.д.).

Определим относительную близость дидактических единиц  $d_i$  и  $d_j$  в виде:

$$r_i = \frac{N_{ij}}{N_i}, \quad r_j = \frac{N_{ij}}{N_j}$$

где  $N_{ij}$  – число совпадающих дескрипторов в единицах  $d_i$  и  $d_j$ , а  $N_i$ ; и  $N_j$  – соответственно, общее число дескрипторов в них. Очевидно, что в общем случае  $r_i \neq r_j$ .

Отметим, что если  $r_i \sim r_j > 0,5$ , то соответствующие единицы имеют слишком много «пересечений» и правомерно рассматривать вопрос об их объединении. Наоборот, если  $r_i \sim r_j < 0,1$ , то единицы можно считать независимыми, в учебных планах ставить их параллельно (например, как базовые).

Введем понятие меры ассоциации  $S_{ij}$  единиц  $d_i$  и  $d_j$ :

$$S_{ij} = \frac{2N_{ij}}{(N_i + N_j)}$$

Теперь с помощью введенных величин можно проанализировать относительный вклад той или иной дидактической единицы в учебный план (рабочую программу), а также установить предположительный порядок их изучения. В самом деле, если  $r_i \ll r_j$ , то это значит, что смысловое поле единицы  $d_i$  много шире, чем у  $d_j$ . Тогда, скорее всего, первая единица является базовой и ее следует изучать первой, а вторая единица – уточняет и (или) развивает первую в определенном контексте или направлении.

Если  $r_i \sim r_j$  и при этом  $S_{ij}$  мало, то рассматриваемые единицы практически независимы, и их можно изучать в любой последовательности или параллельно. Наконец, если  $r_i \sim r_j$  и при этом  $S_{ij}$  не мало, то соответствующие учебные элементы имеют много общих дескрипторов, и при этом их смысловые поля сравнимы. В этом случае эти единицы должны в учебных планах и программах стоять рядом, но порядок изучения должен быть определен исходя из каких-то дополнительных критериев.

Заметим, что значения  $S_{ij}$ , также как  $r_i$  и  $r_j$  находятся в интервале (0, 1), и изменение этого параметра при соответствующих изменениях  $r_i$  и  $r_j$  небольшие. Поэтому на практике значительно удобнее пользоваться другой величиной – коэффициентом улучшения ассоциации  $K_{ij}$ :

$$K_{ij} = \frac{S_{ij}}{S_{ij} - r_i r_j}$$

Дальнейший алгоритм выглядит примерно так: определяется (исходя из некоторых соображений) базовая дидактическая единица  $d_1$ . Определяются все коэффициенты  $K_{1j}$  для всех  $j > 1$ . Из них находится максимальный (допустим  $d_{1m}$ ), и выстраивается следующая последовательность изучения дидактических единиц:  $d_1 \rightarrow d_m$ . Далее, заменяем  $m$  на 2 и рассматриваем все коэффициенты  $K_{2j}$  при  $j > 2$ , определяем максимальный и продолжаем

выстраивать цепочку последовательности вхождения учебных элементов в программу (план).

Понятно, что таким образом можно сформировать только линейную последовательность изучения учебных элементов, однако в ряде случаев (и нередко) при формировании учебных планов одна и та же дидактическая единица используется как «стартовая позиция» для двух и более дидактических единиц (см. рис.1). В этом случае можно предложить следующий алгоритм: после нахождения двух близких к  $d_1$  единиц ( $d_2$  и  $d_3$ ), определяется их относительная близость  $K_{23}$ . Если эта величина мала, но при этом  $r_2$  и  $r_3$  значительны, то это может означать, что единицы  $d_2$  и  $d_3$  используют разные общие с  $d_1$  наборы дескрипторов. Это и будет служить веским доводом в пользу того, что обе единицы базируются на  $d_1$  и могут изучаться параллельно, то есть, образуют дерево процесса обучения с двумя ветками. Таким же образом можно построить учебные планы (программы) с любым числом ветвей.

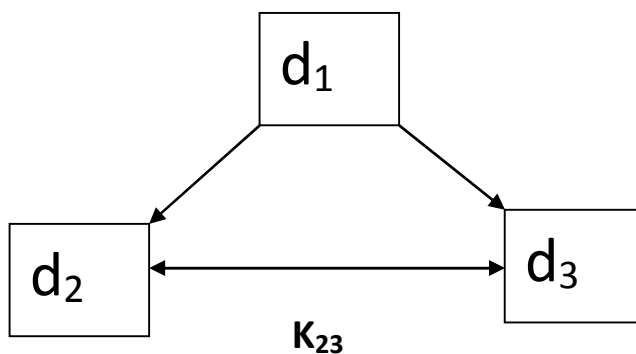


Рис.1. Разветвления в порядке следования учебных элементов

Поясним описанный выше алгоритм на простом примере. В таблице 1 приведен фрагмент текстового файла, содержащего пять учебных элементов с соответствующими дескрипторами для курса «Анализ данных».

**Таблица 1**

Учебные элементы и дескрипторы для курса «Анализ данных»

Номер элемента	Название элемента	Набор дескрипторов
EL01	Технология Data Mining	Хранилище данных, оптимизация, машинное обучение, кластеризация, классификация, ассоциации, шаблоны, деревья решений, анализ рисков, скоринг, профайлинг, прогнозирование, регрессионный анализ, корреляционный анализ, временной ряд, консолидация данных, Data Mining, нейронные сети, статистика
EL02	Введение в консолидацию данных	Консолидация данных, извлечение данных, преобразование данных, загрузка данных, очистка данных, обогащение данных, хранилище данных, витрины данных, оценка качества данных

EL03	Технология ETL	Извлечение данных, очищение данных, преобразование данных, загрузка данных, агрегирование данных, перевод значений, поток данных
EL04	Концепция хранилища данных	Хранилище данных, извлечение данных, очищение данных, преобразование данных, загрузка данных, витрины данных, метаданные, агрегирование данных, обогащение данных, оценка качества данных, интегрированные данные, детализация данных, предметно-ориентированные данные, хронология данных, неизменяемость данных
EL05	Введение в кластеризацию данных	Метрика, нейронная сеть, машинное обучение, профайлинг, сегментация, статистика, сжатие данных

В таблицах 2 и 3 приведены результаты расчетов матриц близостей дидактических единиц и мер ассоциаций.

**Таблица 2**

Матрица близости дидактических единиц

	EL01	EL02	EL03	EL04	EL05
EL01	1	0	0	0,05	0,2
EL02	0	1	0,44	0,78	0
EL03	0	0,57	1	0,57	0
EL04	0,07	0,47	0,27	1	0
EL05	0,57	0	0	0	1

**Таблица 3**

Матрица мер ассоциаций

	EL01	EL02	EL03	EL04
EL02	0			
EL03	0	0,62		
EL04	0,06	0,88	0,73	
EL05	0,3	0	0	0

Проанализируем эти результаты. Учебный элемент EL01 не имеет пересечений с элементами EL02 и EL03 и незначительное пересечение с элементом EL04 ( $r_{14} = 0,05$ ). Но с элементом EL05 этот коэффициент уже существенен:  $r_{15} = 0,2$ . Связь этих элементов подтверждает и мера близости ассоциаций:  $S_{15} = 0,3$ . Поскольку  $r_{51} = 0,57 > r_{15}$ , то элемент EL01 является базовым и должен изучаться первым, а элемент EL05 развивает и детализирует содержание первого элемента. Еще более тесной является связь между вторым

и четвертым учебными элементами. Аналогичный анализ соответствующих коэффициентов матриц близости элементов и мер ассоциаций указывает, что элемент EL04 должен изучаться первым. Что касается, например, учебных элементов EL01 и EL03 или EL02 и EL05, то эти элементы могут изучаться независимо друг от друга.

В заключение отметим, что для нахождения ассоциаций и построения правильной последовательности изучения учебных элементов для многосеместровых курсов, содержащих большое число элементов, можно использовать пакеты прикладных программ, предназначенных для анализа больших массивов данных. Например, в пакете Deductor Academic имеются готовые инструменты под названием «Ассоциативные правила» и «Деревья решений». И хотя эти инструменты предназначены, прежде всего, для выявления бизнес-правил при анализе транзакций в больших супермаркетах, нетрудно приспособить их и для описанных выше задач. Для сравнительно небольших по объему дисциплин достаточно использовать электронные таблицы Excel.

### Список литературы

1. Ассоциация предприятий компьютерных и информационных технологий. Профессиональные стандарты в области информационных технологий – М.:2008. – 616с.
2. Грешнов М.В. Модель финансового взаимодействия государства и малого бизнеса // Современные проблемы науки и образования. – 2013. - №6; URL: <http://www.science-education.ru/113-11373>
3. Лазарев А.В., Павлович В.Е. Специфика финансовой услуги // Экономические науки. - 2011.- № 80. - С. 146-149.
4. Макаров А.А., Смольников С.Д., Китаев Д.Ф. Процессный подход в разработке учебных планов (электронный ресурс) // Современные проблемы науки и образования. – 2013. – № 6; URL: <http://www.science-education.ru/113-11253>
5. Хаймович И.Н., Дровяников В.И. Особенности интеграции экономико-математического инструментария в информационную систему управления ВУЗа // Вестник Самарского государственного экономического университета. – 2010. - №74. – С.17-20.

### Рецензенты:

Прохоров С.А., д.т.н., профессор, зав. кафедрой информационных систем и технологий ФГБОУ ВПО Самарского аэрокосмического им. академика С.П. Королева – Национального исследовательского университета, г. Самара;

Хаймович И.Н., д.т.н., кафедры информационных систем и компьютерных технологий ЧОУ ВО «Международный институт рынка», г. Самара.