

# Применение методов Data Mining для разработки национальных образовательных стандартов ИКТ специальностей вузов

Фирудин Агаев<sup>1</sup>, Гюлара Мамедова<sup>2</sup>

<sup>1,2</sup> Институт Информационных Технологий НАНА, Баку, Азербайджан  
depart10@iit.ab.az

**Аннотация** – В статье рассматривается использование методов Data mining при сопоставительном анализе образовательных программ ИКТ специальностей ведущих университетов мира. Сопоставление программ позволит разработать образовательные стандарты для ИКТ специальностей вузов Азербайджана.

**Ключевые слова** – образовательные программы, перечень дисциплин, перечень тем и разделов образовательной программы, сопоставление программ, множество транзакций, поиск ассоциативных правил

## I. ВВЕДЕНИЕ

Построение информационного общества в Азербайджане – одна из приоритетных задач национального развития. В республике успешно реализуется стратегия построения «электронного государства», неотъемлемой и составной частью которого является система образования.

Развитие информационных технологий, глобализация информационных и технологических отношений требуют подготовки высокопрофессиональных специалистов в республике. Повышение конкурентоспособности национальных кадров на рынке труда и увеличение академической мобильности между странами требуют согласования форматов образовательных программ различных стран в рамках Болонского процесса.

В настоящее время в мире существуют десятки различных моделей высшего образования. Анализ зарубежного опыта, отраженного в университетских образовательных программах, с целью совершенствования национальной системы образования является важным и актуальным [1].

Сопоставительный анализ образовательных программ является сложной процедурой, поскольку они представлены в документах разной формы, часть информации, которая в них содержится, является неструктурированной, а процессы сопоставления – неформализованными [2]. Вместе с тем развитие методов искусственного интеллекта и возможностей вычислительной техники в настоящее время позволяет формализовать процедуры сопоставительного анализа университетских образовательных программ.

Для реализации процесса сопоставительного анализа учебные программы загружаются в реляционную базу данных. Структура базы данных состоит из трех взаимосвязанных таблиц: «Направления», «Специальности» и «Учебные программы». В первой таблице содержатся названия направлений специализации, во второй – название и код специальности. В таблице «Учебные планы» размещаются данные из конкретной учебной программы вуза, т.е. названия дисциплины, темы и раздела. В разных университетах мира по одной и той же специальности названия дисциплин, тем и разделов некоторым образом различаются. При сравнительном анализе образовательных программ основной задачей является выявление информации, общей для всех исследуемых вузов [6].

## II. ПРИМЕНЕНИЕ МЕТОДОВ ПОИСКА АССОЦИАТИВНЫХ ПРАВИЛ ПРИ СОПОСТАВЛЕНИИ ОБРАЗОВАТЕЛЬНЫХ ПРОГРАММ

Одной из наиболее распространенных задач искусственного интеллекта при сопоставлении данных является определение часто встречающихся наборов объектов в большом множестве наборов.

При решении этой задачи впервые был предложен метод поиска ассоциативных правил для нахождения типичных шаблонов покупок, совершаемых в супермаркетах [4].

По нашему мнению, этот метод успешно можно применить и для сравнения множества образовательных программ ведущих университетов мира с целью поиска информации, общей для большинства учебных программ вузов. Вначале можно сравнивать названия дисциплин университетов, изучаемых по конкретной специальности и уровню специализации (бакалавр или магистр), затем в отдельности названия тем и разделов изучаемой дисциплины. Эту информацию в последующем можно использовать для улучшения образовательных программ университетов Азербайджана и приведения этих программ к мировому уровню. Формальная постановка задачи следующая.

Пусть имеется база данных, состоящая из некоторого количества транзакций. Транзакция в информатике означает группу логически объединенных последовательных операций по работе с данными,

обрабатываемую или отменяемую целиком. В данном случае транзакция – это набор дисциплин, изучаемых в конкретном вузе по конкретному направлению и уровню специализации.

Пусть  $I = \{i_1, i_2, \dots, i_n\}$  – множество элементов (набор всевозможных дисциплин, тем или разделов) общим числом  $n$ .

Пусть  $T$  – множество транзакций,  $T = \{T_1, T_2, \dots, T_m\}$ , где каждая транзакция  $T_j$  – это набор элементов из  $I$ .

Множество транзакций, в которые входит объект  $ij$ , обозначим следующим образом:

$$T_{ij} = \{Tr \mid ij \in Tr; j=1, \dots, n; r=1, \dots, m\} \subseteq T.$$

Пусть  $F$  – некоторый произвольный набор, состоящий из  $k$  элементов (дисциплин, тем или разделов):

$$.F = \{ij \mid ij \in I; j=1..n\}.$$

Такой набор называется  $k$ -элементным набором множества  $I$ .

Множество транзакций, в которые входит набор  $F$ , обозначим следующим образом:

$$DF = \{Tr \mid F \subseteq Tr; r = 1..m\} \subseteq D.$$

Отношение количества транзакций, в которое входит набор  $F$ , к общему количеству транзакций называется поддержкой (support) набора  $F$  и обозначается  $Supp(F)$ :

$$Supp(F) = \frac{|D_F|}{|D|}.$$

Минимальное значение поддержки интересующих наборов  $Supp_{min}$  должен указать эксперт.

Набор  $.F = \{ij \mid ij \in I; j=1..n\}$  называется частым (large itemset), если значение его поддержки больше минимального значения поддержки, заданного экспертом. Т.е.  $Supp(F) > Supp_{min}$ .

Задачей поиска ассоциативных правил является нахождение множества всех часто встречаемых наборов, таких, как:

$$L = \{F \mid Supp(F) > Supp_{min}\}.$$

В нашем случае база данных учебных программ состоит из большого количества данных и с увеличением количества сравниваемых университетов разрастается. В этой связи для решения задачи требуются эффективные алгоритмы, одним из которых является алгоритм Apriori [3]. Для применения данного алгоритма необходимо изменить структуру данных, данные представить в бинарном виде и привести таблицу в нормализованный вид.

Задача поиска часто встречаемых наборов объектов в большом множестве объектов – очень трудоемкая, требующая перебора большого множества наборов. Для сокращения пространства поиска используется свойство антимонотонности (Apriori), которое гласит: поддержка

любого набора элементов не может превышать минимальной поддержки любого из его подмножеств [5].

Например, поддержка 4-элементного набора  $I(4) = \{i_1, i_2, i_3, i_4\}$  будет всегда меньше или равна поддержке любых 3-элементных наборов из множества  $I(4)$ . Т.к. любая транзакция, содержащая  $\{i_1, i_2, i_3, i_4\}$ , также должна содержать любые 3-элементные наборы множества  $I(4)$ , при этом обратное не верно.

Это свойство поиска ассоциативных правил называется антимонотонность и служит для снижения размерности пространства поиска.

Из свойства антимонотонности вытекает следующее правило:

любой  $k$ -элементный набор будет часто встречающимся тогда и только тогда, когда все его  $(k-1)$ -элементные подмножества будут часто встречающимися.

Наборы всех возможных элементов из множества  $I$  можно представить в виде решетки, начинающейся с пустого множества (empty set). На 1-м уровне располагаются одноэлементные наборы, на 2-м – двухэлементные, связанные с одноэлементными и т.д., на  $k$  уровне находятся  $k$ -элементные наборы, связанные со всеми своими  $(k-1)$ -элементными подмножествами.

Рассмотрим пример, иллюстрирующий набор элементов  $I(4) = \{i_1, i_2, i_3, i_4\}$ .

Предположим, что набор из элементов  $\{i_1, i_2\}$  имеет поддержку ниже заданного порога и, соответственно, не является часто встречающимся.

Тогда, согласно свойству антимонотонности, все множества, содержащие элементы  $i_1, i_2$ , также не являются часто встречающимися и отбрасываются.

Использование свойства антимонотонности позволяет существенно сократить пространство поиска.

### III. ЗАКЛЮЧЕНИЕ

По многим направлениям модернизации высшего профессионального образования в республике в качестве основы сегодня апробируется опыт развитых, прежде всего европейских, стран, утверждается невозможность их прямого копирования. Одновременно с этим признается, что сравнительный анализ зарубежной и национальной практики создает предпосылки для внедрения наиболее эффективных элементов в систему профессионального образования в азербайджанских вузах.

Решение задачи поиска ассоциативных правил, как и любой задачи, сводится к обработке исходных данных и получению результатов. Используя вышеуказанный метод, на основе сравнения образовательных программ различных университетов можно определить набор дисциплин (тем или разделов), который чаще остальных преподается на данном уровне подготовки (бакалавр или магистр) и по определенному направлению специализации.

Обеспечение учебного процесса в зарубежных вузах сопровождается наличием банков дисциплин по различным направлениям и специализациям. Эти дисциплины и их содержание постоянно обновляются, что позволяет корректировать процесс обучения в соответствии с развитием новых научных направлений и требованиями рынка труда. Азербайджанским вузам еще предстоит решить эту проблему и при разработке образовательных стандартов третьего поколения создать для каждого направления обучения свой банк дисциплин, предлагаемых студентам на выбор.

#### ЛИТЕРАТУРА

- [1] Образовательное законодательство и образовательные системы зарубежных стран. Под ред. проф. Козырина А.Н., М., Academia, 2007, 432 с.
- [2] [www.berkeley.edu](http://www.berkeley.edu), [www.mit.edu](http://www.mit.edu), [www.stanford.edu](http://www.stanford.edu), [www.nus.edu.sg](http://www.nus.edu.sg), [www.ucsb.edu](http://www.ucsb.edu), [www.polytechnique.edu](http://www.polytechnique.edu), [www.ethz.ch](http://www.ethz.ch), [www.cam.ac.uk](http://www.cam.ac.uk), [www.u-tokyo.ac.jp/en/](http://www.u-tokyo.ac.jp/en/), [www.ust.hk](http://www.ust.hk).
- [3] Кузьменко В.И. Data Mining: поиск закономерностей. CNews, 2013, №7.
- [4] Дюк В.А., Самойленко А.П. Data Mining: учебный курс. СПб.: Питер, 2001.
- [5] Поиск ассоциативных правил в Data Mining. Лекции Новосибирского технического государственного университета, [www.ami.nstu.ru](http://www.ami.nstu.ru), 2013.
- [6] Смит Д.М., Смит Д.К. Абстракции баз данных: агрегация и обобщение. Системы управления базами данных. 2006, №2, с. 141–160.
- [7] [www.it-edu.ru](http://www.it-edu.ru)