

УДК 004.89

Агаев Ф.Т., Мамедова Г.А.

Институт Информационных Технологий НАНА, Баку, Азербайджан
depart10@iit.ab.az

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ОБРАЗОВАТЕЛЬНЫХ ПРОГРАММ ИКТ-СПЕЦИАЛЬНОСТЕЙ УНИВЕРСИТЕТОВ МИРА С ПРИМЕНЕНИЕМ МЕТОДОВ ПОИСКА АССОЦИАТИВНЫХ ПРАВИЛ

В статье рассматриваются вопросы сопоставительного анализа образовательных программ ИКТ-специальностей ведущих университетов мира с использованием методов извлечения нужной информации в больших объемах данных (Data Mining). Использование методов поиска ассоциативных правил в настоящее время позволяет ставить вопрос об информационной поддержке процедуры сопоставительного анализа университетских образовательных программ.

Ключевые слова: образовательные программы, перечень дисциплин, множество транзакций, ассоциативные правила.

Введение

За довольно короткий период независимости Азербайджан добился значительного экономического роста во многих сферах деятельности. Одним из условий долгосрочного экономического благосостояния страны является развитие человеческого капитала, а именно его образовательного и интеллектуального потенциалов. Поэтому первоочередная задача системы профессионального образования – обеспечение национальной экономики высококвалифицированными кадрами, способными эффективно решать профессиональные задачи в динамичных современных условиях.

Переход на двухуровневую систему высшего профессионального образования стал основой реформирования образовательных систем не только в зарубежных странах, но и в Азербайджане. В реформировании европейской системы высшего образования большую роль сыграла Болонская декларация, одной из задач которой являлось создание системы обучения, основанной на двух основных циклах и третьем – научном уровне. После Болонского соглашения намечается тенденция унификации образовательных программ различных стран [1].

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 [2].

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

Сопоставление образовательных программ различных университетов мира

Возможность использования средств автоматизированного сравнения образовательных программ обусловлена достаточно хорошей структурированностью текста этих документов. Учебная программа относится к одному из направлений подготовки области ИКТ. На сегодняшний день образование в области ИКТ осуществляется по пяти крупным направлениям специализации: компьютерные науки,

информационные технологии, информационные системы, программная инженерия, компьютерная инженерия.

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 программы» размещаются данные из конкретной учебной программы вуза, т.е. название 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.

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 [3].

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 можно успешно применить и для сравнения множества образовательных программ ведущих университетов мира с целью поиска информации, общей для большинства учебных программ вузов. Вначале можно сравнить названия дисциплин университетов, изучаемых по конкретной специальности и уровню специализации (бакалавр или магистр), затем отдельно можно сравнить 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, обрабатываемыми или отменяемыми 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 i_j , обозначим следующим образом:

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

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

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

Такой набор называется k -элементным набором множества I .

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

$$D_F = \{T_r \mid F \subseteq T_r; r = 1..m\}.$$

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

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

Минимальное значение поддержки интересующих наборов *Suppmin* должен указать эксперт. Набор $F = \{i_j | i_j \in I; j = 1..n\}$ называется частым (large item set), если значение его поддержки больше минимального значения поддержки, заданного экспертом. Т.е. $Supp(F) > Suppmin$.

Задачей поиска ассоциативных правил является нахождение множества всех часто встречаемых наборов, таких, как: $L = \{F | Supp(F) > Suppmin\}$ [5].

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). Алгоритм Apriori описан в 1994 г. Рамакришнан Срикантом (Ramakrishnan Srikant) и Ракеш Агравалом (Rakesh Agrawal) [7]. Он гласит, что поддержка любого набора объектов не может превышать минимальной поддержки любого из его подмножеств:

$$Supp_F \leq Supp_E,$$

где $E \subset F$.

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 (рис. 1). 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 рисунок 1, 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.

На рис. 1 вся эта ветвь, начиная с $\{i_1, i_2\}$, выделена фоном. Использование свойства антимонотонности позволяет существенно сократить пространство поиска.

В нашем случае база данных учебных программ состоит из большого количества данных, и с увеличением количества сравниваемых университетов база данных разрастается. Для применения данного алгоритма необходимо изменить структуру данных, данные представить в бинарном виде и привести таблицы в нормализованный вид. Преобразованная база данных будет содержать несколько таблиц (количество таблиц равно количеству направлений специализации), в которых данные будут представлены в бинарном виде (таблица 1). Количество таблиц будет равно количеству направлений специализации (в нашем случае – 5). Каждая запись будет соответствовать уровню специализации (бакалавр или магистр), а также специальности.

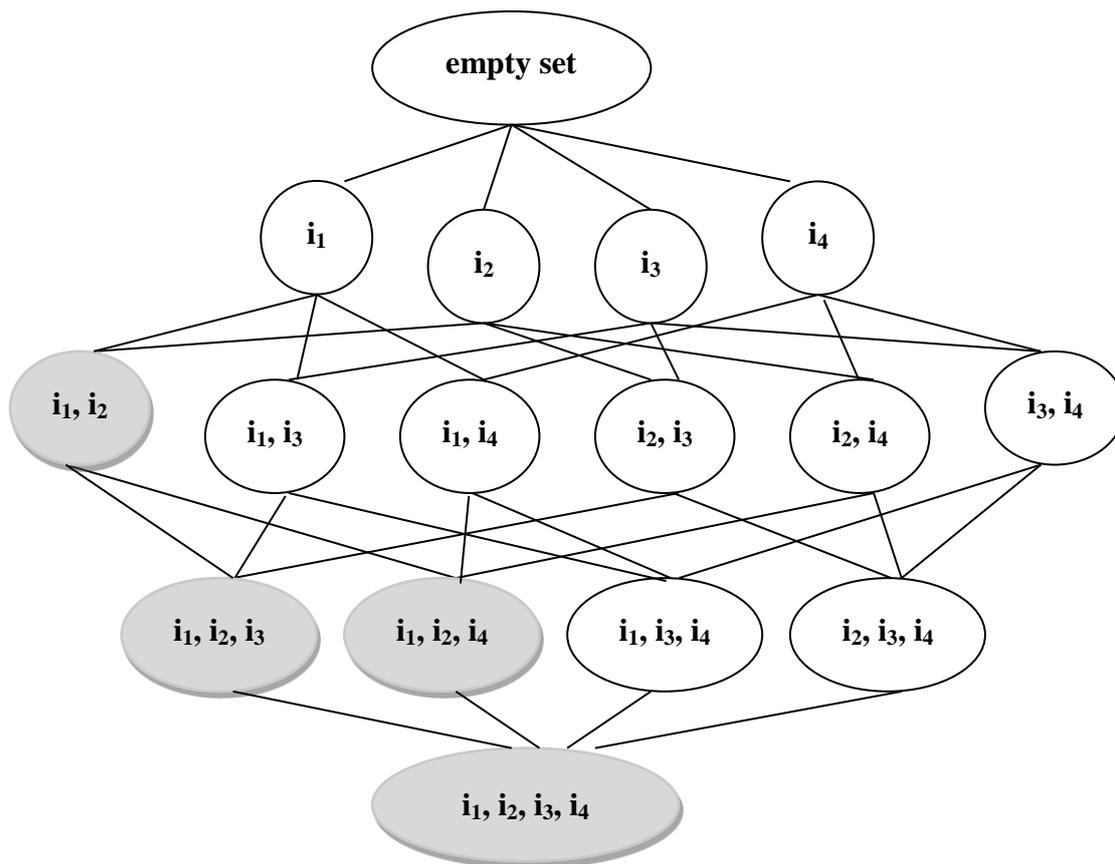


Рис. 1. Пример взаимосвязей между объектами 4-элементного набора

В столбце *KS* таблицы 1 «101, 102, ..., 201, 202...» обозначают условный код специальности, где первая цифра «1» обозначает уровень бакалавра, цифра «2» – уровень магистра, две последние цифры – порядковый номер специальности. Количество столбцов *n* в таблице равно общему количеству дисциплин, присутствующих в множестве транзакций *T*. Дисциплины в заголовке таблицы обозначены $i_1, i_2, i_3, \dots, i_n$.

Таблица 1

Взаимосвязи между специальностями и преподаваемыми дисциплинами

KS	i_1	i_2	i_3	i_4	i_5	i_6	...	i_n
101	1	0	0	1	0	0	...	0
102	0	0	1	0	1	0	...	1
103	1	1	1	0	0	0	...	1
...
201	0	1	0	1	1	1	...	0
202	1	1	0	0	1	0	...	1
...

В нашем случае каждая запись таблицы – это транзакция, где в соответствующем столбце стоит 1, если дисциплина присутствует в транзакции, и 0 – в противном случае. Необходимо заметить, что исходный вид таблиц отличается от приведенного в таблице 1, т.к. для каждого направления, уровня специализации и специальности существует свой набор дисциплин.

При реализации алгоритма Apriori на первом шаге подсчитываются 1-элементные часто встречающиеся наборы. Для этого необходимо пройти по всему набору дисциплин и подсчитать для них поддержку, т.е. сколько раз эти дисциплины встречаются в базе. Все последующие шаги алгоритма состоят из двух частей: объединение часто встречающихся наборов дисциплин с кандидатами из других наборов потенциально часто встречающихся наборов элементов (которые называют кандидатами) и подсчет поддержки для кандидатов. Каждый k -элементный кандидат будет формироваться путем добавления к $(k-1)$ -элементному частому набору элемента из другого $(k-1)$ -элементного набора. Для каждого построенного k -элементного набора множеств необходимо удалить кандидата, если хотя бы один из $(k-1)$ -наборов не является часто встречающимся. После генерации кандидатов следующей задачей является подсчет поддержки для каждого кандидата, чтобы значение поддержки было выше заданного экспертом Suppmin .

Заметим, что количество кандидатов может быть очень большим, поэтому необходимо сравнивать каждую транзакцию с каждым кандидатом. Эффективным и быстрым способом для хранения кандидатов является использование хэш-дерева [8]. В узлах хэш-дерева содержатся хэш-таблицы с указателями на потомков, а листья – указывают на кандидатов. Каждый раз, когда формируются кандидаты, строится хэш-дерево. Вначале дерево состоит только из корня и является пустым множеством. Когда появляется новый кандидат, он заносится в корень дерева. Как только количество кандидатов становится больше предельной величины, корень преобразуется в хэш-таблицу, т.е. становится узлом хэш-дерева с указателями на потомков. После построения хэш-дерева с кандидатами-наборами легко подсчитать поддержку для каждого кандидата и проверить, удовлетворяют ли значения поддержки кандидатов минимальному порогу.

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 в систему профессионального образования в азербайджанских вузах наиболее ее эффективных элементов.

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, www.mit.edu, www.stanford.edu, www.nus.edu.sg, www.ucsb.edu, www.polytechnique.edu, www.ethz.ch, www.cam.ac.uk, www.u-tokyo.ac.jp/en/, www.ust.hk.

3. Кузьменко В.И. Data Mining: поиск закономерностей.
http://www.cnews.ru/reviews/new/rynok_bi_v_rossii_2013/articles/data_mining_poisk_zakon_omernostej/
4. Дюк В.А., Самойленко А.П. Data Mining: учебный курс. СПб.: Питер, 2001, с.368.
5. Поиск ассоциативных правил в Data Mining. Лекции Новосибирского технического государственного университета, 2013. www.ami.nstu.ru.
6. Смит Д.М., Смит Д.К. Абстракции баз данных: агрегация и обобщение. Системы управления базами данных. 2006, №2, с.141–160.
7. Agrawal R., Srikant R. Fast algorithms for mining association rules. 20th VLDB Conference, Santiago, Chile, 1994, pp.487–499.
8. Park J.S., Chen M.S., and Yu P.S. Efficient parallel mining for association rules. In Proc. of the 4th Int. Conf. on Information and Knowledge Management, Baltimore, Maryland, 1995, pp.31–36.

UOT 004.89

Ağayev Firudin T., Məmmədova Gülarə A.

AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
depart10@rambler.ru

Assosiativ qaydaların axtarış metodlarının tətbiqi ilə dünya universitetlərinin İKT ixtisasları üzrə tədris proqramlarının müqayisəli analizi

Məqalədə böyük həcmli verilənlərdən lazım olan informasiyanın əldə olunması (Data Mining) üsullarından istifadə etməklə dünyanın aparıcı universitetlərinin İKT ixtisasları üzrə təhsil proqramlarının müqayisəli analizi göstərilir. Assosiativ qaydaların axtarış üsullarının istifadəsi universitetlərin təhsil proqramlarının müqayisəli analizində informasiya dəstəyi məsələlərinin həllinə imkan yaradır.

Açar sözlər: təhsil proqramları, fənlərin siyahısı, tranzaksiyalar çoxluğu, assosiyasiya qaydaları.

Agayev T. Firudin, Mammadova A. Gulara

Institute of Information Technology of ANAS, Baku, Azerbaijan
depart10@rambler.ru

Comparison analysis of association rules search to compare educational programs of ICT professions in world universities

The article discusses the comparative analysis of the ICT professions education programs of world leading universities using Data Mining methods (extraction of desired information in large volumes of data). Using the search methods of mining association rules now allows to raise the question of the informational support in comparative analysis procedure of university education programs.

Keywords: educational programs, disciplines list, multiple transactions, association rules.