

Применение методов интеллектуального анализа данных при разработке системы классификации компетентностей студентов для web-сайта университета

А. Мансур, Ж. Мохаммад, В.В. Галушка

Донской государственной технической университет, Ростов-на-Дону

Аннотация: Статья посвящена проблеме применения искусственных нейронных сетей как одного из методов интеллектуального анализа данных для классификации компетентностей студентов университета при определении учебной специализации. В ней описываются сложности процесса выбора специализации и связанные с этим негативные последствия, а также предлагается подход к решению данной проблемы с применением программных средств классификации. В качестве основы для реализации предлагаемых средств используется система учёта успеваемости студентов на веб-сайте университета г. Тартус, Сирия. В статье приведены критерии отбора данных для формирования обучающей выборки, которая включает в себя успеваемость по некоторым предметам в качестве входных векторов. Значения выходного вектора зависят от окончательной специальности и попадают в обучающее множество только в случае её правильного выбора. На основе указанных данных сформирована структура многослойной искусственной нейронной сети и выбран алгоритм обучения, результаты работы которого отражены на веб-сайте университета в виде советов по выбору будущей специализации, что позволило повысить эффективность учебного процесса.

Ключевые слова: интеллектуальный анализ данных, искусственные нейронные сети, web-дизайн, машинное обучение, классификация.

Введение

Интеллектуальный анализ данных (англ. DataMining) это автоматическое исследование интересных и невидимых шаблонов в базе данных. Этот процесс направлен на анализ данных и извлечение скрытой в них информации, чтобы помочь учреждениям обнаружить важные сведения в своих базах данных (БД), а также разработать мощный инструмент для их интерпретации и извлечения практически полезных знаний, которые могли бы помочь в принятии решений [1, 2]. Использование методов интеллектуального анализа данных для извлечения полезной информации из веб-сайтов называется Web-mining [3]. При этом выделенные шаблоны используются для улучшения структуры веб-сайтов, повышения доступности представленной на них информации, а также для оптимизации процессов

автоматического поиска данных и представления информационных ресурсов в вебе [4].

В данной работе рассматривается применение методов DataMining на веб-сайте университета для помощи студентам в выборе будущей специализации на основе классификации их компетенций с целью повышения эффективности дальнейшего обучения.

1. Задача классификации компетентностей студентов

Описываемые далее результаты, получены в процессе исследований, проводившихся на базе факультета информационных и коммуникационных технологий в Университете Тартуса, Сирия. По итогам обучения на факультете выдаётся диплом по одной из специальностей: «Информационные технологии», «Коммуникационные технологии» или «Компьютерные системы».

Для автоматизации управления учебным процессом инженерного колледжа в Тартусском университете используется, помимо других средств, веб-сайт, предоставляющий студентам возможность самостоятельной регистрации на курсы (предметы/дисциплины) перед началом учебного года.

Общая продолжительность обучения по инженерным специальностям в Сирии составляет пять лет. Студенты университета в течение первых трех лет изучают одинаковые курсы, которые являются базовыми для всех перечисленных ранее специальностей. В конце третьего курса студент должен выбрать специализацию, которую он хочет изучать в течение последних двух лет специализации. Процесс определения специализации осуществляется через электронный веб-сайт университета во время регистрации студента на четвёртый год и заключается в выборе одной из специальностей, предлагаемых в списке. При этом, им часто приходится сталкиваться с трудностями при принятии этого решения, а зачастую этот выбор делается случайным образом, в результате чего, позже, некоторые

меняют своё мнение из-за неудач в выбранной сфере, что отрицательно сказывается на успеваемости в дополнение к возникающим административным проблемам.

Решение описанной проблемы возможно благодаря использованию методов интеллектуального анализа данных для определения наиболее подходящей специальности, на основе академических достижений за предыдущие три года. Полученная таким образом информация будет отражать интересы студентов и научные способности, а выдаваемые советы способны помочь ученику в правильном выборе, имеющем научное основание и соответствующем его склонностям и способностям, что в некоторой степени повышает вероятность его успеха при окончании университета.

2. Исходные данные

Подготовка и предобработка исходных данных является важным этапом в процессе интеллектуального анализа [5]. Сайт университета использует в качестве системы управления базами данных mysqlnd 5.0.11, а хранимая ей БД включает в себя 24 таблицы [6]. Для задачи, описанной выше, достаточно выбрать данные из части таблиц, схема связей которых показана на рисунке 1.

Представленный фрагмент базы данных включает в себя таблицу «Students» (ученики/студенты), которая связана с таблицей «Courses» (предметы) через отношение «Registration» (регистрация). В начале каждого семестра студенту автоматически назначается набор предметов. Таблица «Results» (результаты) содержит оценки работы студентов в течение каждого семестра и определяет, успешную или неуспешную сдачу того или иного предмета.

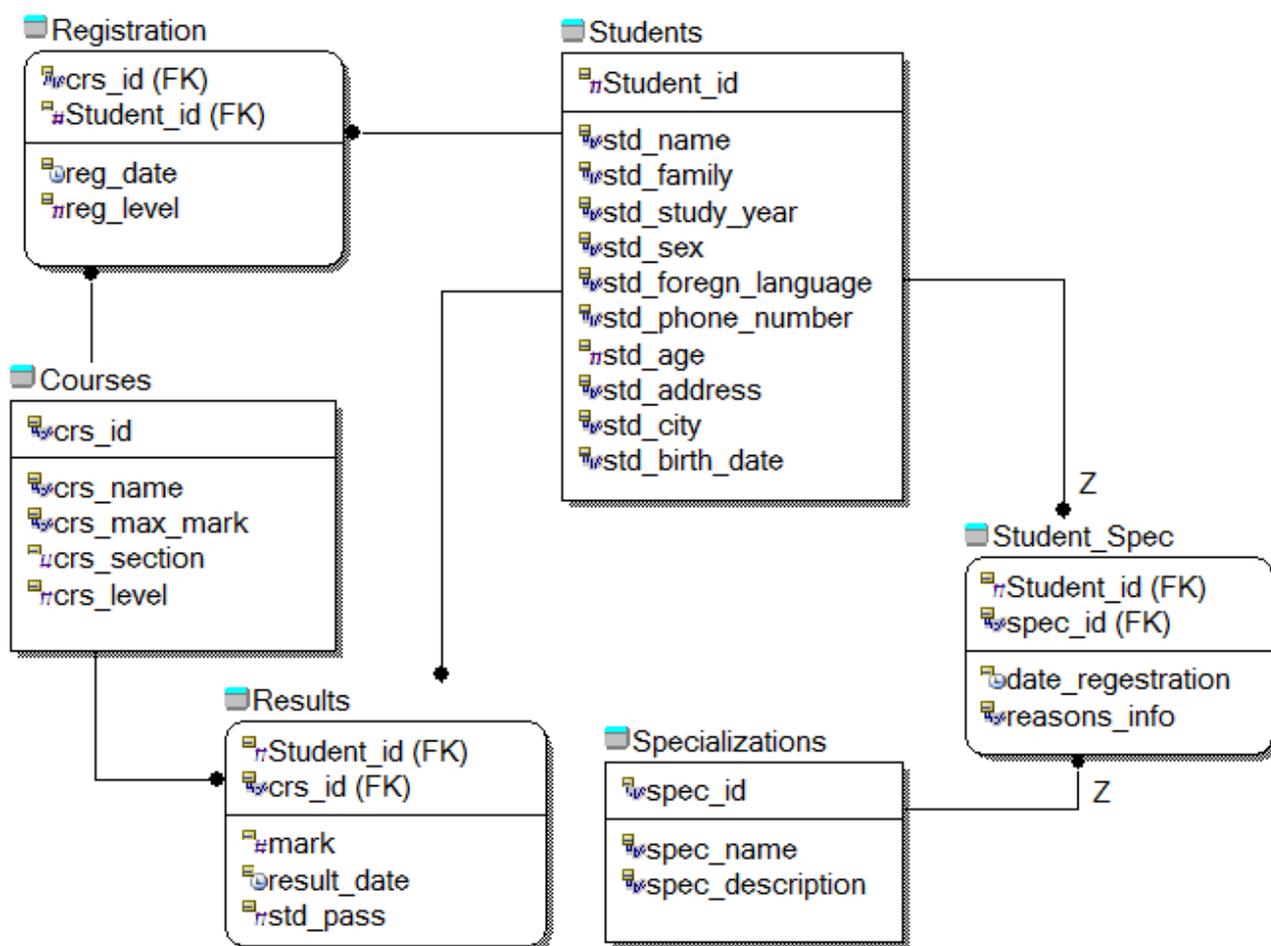


Рис. 1. – Фрагмент схемы базы данных сайта университета

Таблица «Specialization» (специализации) содержит уникальный номер, имя и подробную информацию о специализациях, имеющихся в университете. Она связана с таблицей «Students» через отношение «Student-Specialization», которое связывает каждого ученика с выбранной специализацией.

3. Применение искусственных нейронных сетей для решения задачи классификации

Классификация — это метод сопоставления целевых данных с predetermined группами или классами. Существует достаточно много методов, используемых для решения задачи классификации,

однако практика применения нейронных сетей показывает их большую эффективность при решении данного вида задач, по сравнению с традиционными методами, используемыми в этой области [7].

Основываясь на описанных ранее принципах программной реализации искусственных нейронных сетей и используя разработанный набор классов [8], можно создать нейронную сеть для классификации компетенций студентов. Она будет состоять из трех слоёв — входного, скрытого и выходного, количество нейронов, в каждом из которых необходимо определить [9].

Будем исходить из предположения, что число предметов, представляющих интерес, составляет 20 из 42, изучаемых учащимся в течение первых трех лет, исключая математику, языки, культуру и подобные предметы, не относящиеся к специальности. Оценки по этим предметам будут использоваться в качестве входных векторов нейронной сети, поэтому число нейронов во входном слое будет равно двадцати. Количество нейронов в выходном слое также определяется предметной областью, то есть исходя из необходимости выделения трёх классов, количество нейронов выходного слоя будет равно трём. Число нейронов скрытого слоя определяется эмпирически, исходя из опыта и предварительных практических результатов, и в данном случае равняется десяти.

4. Обучение нейронной сети

После создания нейронной сети её можно обучить выполнению классификации. Этот процесс состоит из двух этапов: обучения и тестирования. Для обучения, необходимо подготовить обучающие пары, каждая из которых состоит из входного вектора нейронной сети и соответствующего ему целевого вектора.

В качестве обучающего множества из базы данных была выбрана группа из 100 студентов, которые закончили университет с хорошей успеваемостью (более 80%), что означает, что их выбор был правильным и соответствовал компетенциям [10]. Сведения об их успеваемости приведены в таблице 1. В ней в 1-м столбце указана специальность, выбранная студентом на 4 курсе, а в остальных — оценки по предметам, изучаемым им ранее. Эти оценки составляют входной вектор нейронной сети, а целевой вектор формируется исходя из специальности. Пусть для каждой специальности значение 1 в целевом векторе находится на i -ом месте, где i -номер специальности в списке, а все остальные значения целевого вектора равны 0, то есть:

- Информационные технологии: [1, 0, 0],
- Коммуникационные технологии: [0, 1, 0],
- Компьютерные системы: [0, 0, 1].

Таблица № 1

Фрагмент исходных данных для обучения

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
ИТ	77	68	74	70	85	88	80	78	66	75	90	82	70	89	96	86	88	85	81	90
КТ	92	88	90	69	91	75	87	95	96	86	70	77	80	70	68	73	78	80	82	81
КС	77	68	74	99	71	93	96	77	79	70	80	91	98	82	77	80	75	95	80	74

Обучение нейронной сети происходит с использованием алгоритма обратного распространения ошибки и останавливается, когда ошибка достигает наименьшего допустимого значения. Этот процесс является вычислительно затратным, однако производится один раз, до ввода системы в эксплуатацию. По окончании обучения и тестирования, сеть может быть использована в автоматическом режиме при выборе студентом специализации для формирования советов по её выбору.

Вывод

Шаблоны и закономерности, полученные в процессе интеллектуального анализа данных, могут содержать очень важную информацию. Их использование для классификации компетентности студентов позволило получать информацию о студенческом научном потенциале и перспективах успеха в изучаемой области. Применение описанного метода на веб-сайте университета в виде руководства и советов по выбору будущей специализации позволило повысить эффективность учебного процесса за счёт более точного выбора студентами подходящего им направления.

Литература

1. Макленнен Дж., Танг Ч., Криват Б. Microsoft SQL Server 2008: Datamining — интеллектуальный анализ данных. Пер. с англ. СПб.: БХВ-Петербург, 2009. 720 с.
2. Larose, D. and Ch. Larose, 2014. Discovering knowledge in data: an introduction to data mining. Second Edition. New Jersey: John Wiley & Sons, Inc., 316 p.
3. Liu, B., 2011. Web Data Mining Exploring Hyperlinks, Contents, and Usage Data. Second Edition. Springer, 622 p.
4. Han, J. and M. Kamber, 2010. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, 310 p.
5. Репрезентативность данных // BaseGroup Labs. Технологии анализа данных URL: basegroup.ru/community/glossary/representativeness (дата обращения: 2.04.2017).
6. Фролов А.В., Фролов Г.В. Базы данных в Интернете: практическое руководство по созданию Web-приложений с базами данных. М.: Русская Редакция, 2011. 432 с.

7. Барский А.Б. Нейронные сети: распознавание, управление, принятие решений. М.: Финансы и статистика, 2004. 176 с.

8. Али М., Галушка В.В. Особенности применения средств объектно-ориентированного программирования для реализации многослойных искусственных нейронных сетей прямого распространения // Автоматизация: проблемы, идеи, решения. Уфа: Омега сайнс, 2017. с. 107-109.

9. Галушка В.В., Фатхи В.А. Формирование обучающей выборки при использовании искусственных нейронных сетей в задачах поиска ошибок баз данных // Инженерный вестник Дона, 2013, №2 URL: ivdon.ru/ru/magazine/archive/n2y2013/1597/.

10. Зятева О.А., Питухин Е.А., Пешкова И.В., Шабалина И.М. Интеллектуальный анализ данных при категоризации преподавателей вуза на основе наукометрических показателей // Инженерный вестник Дона, 2017, №4 URL: ivdon.ru/ru/magazine/archive/n4y2017/4580/.

References

1. Maklennen Dzh., Tang Ch., Krivat B. Microsoft SQL Server 2008: Datamining — intellektual'nyy analiz dannykh. Per. s angl. [Microsoft SQL Server 2008: Datamining] SPb.: BKhV-Peterburg, 2009. 720 p.

2. Larose, D. and Ch. Larose, 2014. Discovering knowledge in data: an introduction to data mining. Second Edition. New Jersey: John Wiley & Sons, Inc., 316 p.

3. Liu, B., 2011. Web Data Mining Exploring Hyperlinks, Contents, and Usage Data. Second Edition. Springer, 622 p.

4. Han, J. and M. Kamber, 2010. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, pp: 310.

5. Rerezentativnost' dannykh. BaseGroup Labs. Tekhnologii analiza dannykh [Data representativeness]. URL:



basegroup.ru/community/glossary/representativeness (data obrashcheniya: 2.04.2017)

6. Frolov A.V., Frolov G.V. Bazy dannykh v Internete: prakticheskoe rukovodstvo po sozdaniyu Web-prilozheniy s bazami dannykh. [Databases on the Internet: A Practical Guide to Creating Web Applications with Databases]. M.: Russkaya Redaktsiya, 2011. 432 p.

7. Barskiy A.B. Neyronnye seti: raspoznavanie, upravlenie, prinyatie resheniy. [Neural networks: recognition, management, decision making] M.: Finansy i statistika, 2004. 176 p.

8. Ali M., Galushka V.V. Osobennosti primeneniya sredstv ob"ektno-orientirovannogo programmirovaniya dlya realizatsii mnogosloynnykh iskusstvennykh neyronnykh setey pryamogo rasprostraneniya [Features of the application of object-oriented programming tools for the implementation of multilayered artificial neural networks of direct propagation] Avtomatizatsiya: problemy, idei, resheniya. Ufa: OMEGA SAYNS, 2017, pp. 107-109.

9. Galushka V.V., Fatkhi V.A. Inzhenernyj vestnik Dona (Rus), 2013, №2. URL: ivdon.ru/ru/magazine/archive/n2y2013/1597/.

10. Zyateva O.A., Pitukhin E.A., Peshkova I.V., Shabalina I.M. Inzhenernyj vestnik Dona (Rus), 2017, №4. URL: ivdon.ru/ru/magazine/archive/n4y2017/4580/.