Математические структуры и моделирование
2017. № 2(42). С. 49–58

УДК 004.93

ПРИМЕНЕНИЕ МЕТОДА ПОСТРОЕНИЯ АССОЦИАТИВНЫХ ПРАВИЛ К АНАЛИЗУ ДЕЯТЕЛЬНОСТИ ОБЩЕСТВЕННЫХ ОРГАНИЗАЦИЙ

С.В. Белим
д.ф.-м.н., профессор, заведующий кафедрой информационной безопасности, e-mail: sbelim@mail.ru
Т.Б. Смирнова
д.и.н., доцент, профессор кафедры этнологии, антропологии, археологии и музеологии, e-mail: SmirnovaTB@omsu.ru
А.Н. Мироненко
к.т.н., доцент кафедры информационной безопасности, e-mail: MironenkoAN@omsu.ru

Омский государственный университет им. Ф.М. Достоевского

Аннотация. В статье предложен математический метод анализа показателей деятельности общественных организаций, основанный на построении ассоциативных правил. В качестве транзакций использованы опросные анкеты. На основе кодирования ответов руководителей общественных организаций российских немцев построены ассоциативные правила. Выделены правила, обладающие высокой поддержкой и достоверностью. На основе ассоциативных правил построен ориентированный взвешенный граф взаимосвязей. На данном графе выполняется поиск сообществ вершин (community), связанных друг с другом сильнее, чем с остальными вершинами. Выделенное сообщество позволило выявить факторы деятельности общественных организаций российских немцев, наиболее тесно связанные друг с другом.

Ключевые слова: ассоциативные правила, общественные организации, сообщества на графах, Data Mining.

Введение

Метод формирования ассоциативных правил традиционно используется для выявления устойчивых связей между предпочтениями покупателей в торговых сетях [1, 2]. Информация, полученная после выявления ассоциативных правил, используется в дальнейшем для позиционирования товаров в торговых залах. Взаимосвязанные товары размещаются близко друг к другу для увеличения продаж. Однако в последнее время метод стал активно использоваться и в других областях исследований для выявления закономерностей на основе статистических данных, таких как обработка изображений [3, 4], анализ распространения биологических видов [5], поиск точек интереса горной промышленности по фотографиям поверхности [6] и т. д.

Метод построения ассоциативных правил в социологических исследованиях стал применяться относительно недавно. Европейское социологическое общество ESS использовало его для выявления связи между страной проживания респондента и устоявшимися стереотипами [7]. Наибольшее распространение метод построения ассоциативных правил получил при анализе социальных сетей. В статье [8] поиск ассоциативных правил используется для исследования влияния пользователей социальных сетей друг на друга. На основе анализа полученных зависимостей определяются наиболее влиятельные пользователи, через которых возможно наиболее быстрое и эффективное продвижение информации. В работах [9, 10] выявлялись ассоциативные правила, позволяющие установить связь между полом студента и выбираемыми им курсами на основе записей в Facebook. Аналогичные исследования проведены в работе [11], авторы которой исследуют влияние социальной сети Facebook на образовательный процесс студентов Турции. В работе [12] авторы на основе анализа социальных сетей выявляют факторы, влияющие на хобби пользователей. Возможности использования ассоциативных правил в криминалистике рассмотрены в работе [13]. Авторы исследуют связь между побудительными мотивами и видом преступления. В статье [14] с помощью ассоциативных правил конструируется социальная сеть на основе базы данных о террористических атаках. Работа [15] посвящена исследованию влияния характера людей на склонность к наркомании с помощью построения ассоциативных правил на базе данных из наркологических клиник. Применение метода поиска ассоциативных правил к данным переписи населения в рамках проекта SPIN представлена в статье [16].

Данная статья посвящена определению закономерностей в деятельности общественных организаций российских немцев на основе их базы данных, полученной с помощью анкетирования. База данных анкет используется для построения ассоциативных правил и выявления закономерностей между различными аспектами деятельности общественных организаций.

1. Постановка задачи и методы решения

Для использования метода построения ассоциативных правил необходимо сформировать множество возможных записей и список транзакций. В данной работе анализировалась деятельность общественных организаций российских немцев в различных регионах. Исходная информация бралась из анкет, заполнявшихся руководителями организаций. В качестве транзакции выбиралась одна анкета. Каждая анкета содержала вопросы различного формата. Первый тип вопросов предполагал два варианта ответа «Да» или «Нет». Второй тип вопросов допускал выбор одного из четырёх или пяти вариантов, причём допускался дополнительный ответ «прочее», который не кодировался, так как содержал неопределённость. Также в анкете присутствовали вопросы с недетерминированным ответом, но они не учитывались при формировании транзакций, так как все ответы были различными и не могли привести к выявлению ассоциативных правил со сколько-нибудь заметной поддержкой. Каждый вариант ответа кодировался своей записью. Вопрос обозначался идентификатором из

одной или двух латинских букв. Например, вопрос, обозначаемый «A», имел два варианта ответа, кодируемых «A1» и «A2». В вопросах, содержащих выбор из двух альтернатив, нельзя ограничиваться кодированием только одной из них, не смотря на возможность однозначного восстановления второй. Такое ограничение может приводить к потере ассоциативных правил.

Пусть I — множество всех ответов, которые могут присутствовать в транзакции. Каждая транзакция T — это набор элементов из I ($T \subseteq I$). D — множество всех транзакций. Говорят, что транзакция T содержит набор элементов X, если $X \subseteq T$ и $X \subseteq I$. Ассоциативным правилом называется импликация $X \Rightarrow Y$, где $X \subseteq I$, $Y \subseteq I$ и $X \cap Y = \emptyset$.

Каждое ассоциативное правило характеризуется некоторым набором параметров. Первый параметр, называемый поддержкой, показывает частоту встречаемости данного правила в имеющемся наборе транзакций. Поддержка правила $X \Rightarrow Y$ вычисляется как процент транзакций, содержащий множество $X \cup Y$:

$$supp(X \Rightarrow Y) = \frac{N(X \cup Y)}{|D|} \cdot 100\ \%,$$

где $N(X \cup Y)$ — количество транзакций, содержащих множество $X \cup Y$.

Достоверность правила показывает, с какой вероятностью из X следует Y. Достоверность ассоциативного правила $X \Rightarrow Y$ вычисляется как процент транзакций, содержащих как X, так и Y, в множестве транзакций, содержащих X:

$$conf(X \Rightarrow Y) = \frac{supp(X \Rightarrow Y)}{supp(X)}.$$

Задача поиска ассоциативных правил состоит в нахождении наборов элементов, поддержка которых не ниже чем minsupport. Из найденных наборов выделяются правила с достоверностью не ниже minconfidence.

2. Ассоциативные правила

В общей сложности было обработано 107 анкет, каждая из которых рассматривалась как независимая транзакция. После кодирования были получены транзакции с различным числом записей от 24 до 50. На основе данных транзакций был осуществлён поиск ассоциативных правил с поддержкой не менее 60 % и достоверностью не менее 80 %. Для поиска ассоциативных правил был использован алгоритм APriori [17]. Ассоциативные правила, удовлетворяющие данным свойствам, приведены в Таблице 1.

Таблица 1. Ассоциативные правила по деятельности общественных организаций российских немцев

N	Предпосылка	Следствие	$supp$	$conf$
1	Более 50 % посетителей центра являются российскими немцами	Интернет используется в работе центра несколько раз в день	61,68	80,72
2	Более 50 % посетителей центра являются российскими немцами	В организации используются языковые курсы для взрослых	60,75	80,72
3	Более 50 % посетителей центра являются российскими немцами	В организации знают о том, что Германия реализует специальные программы	62,62	80,72
4	Более 50 % посетителей центра являются российскими немцами	Знания немецкого языка за последние 10 лет улучшились	62,62	81,93
5	Интернет используется в работе центра несколько раз в день	Знания немецкого языка за последние 10 лет улучшились	62,62	80,00
6	В организации используются языковые курсы для взрослых	Более 50 % посетителей центра являются российскими немцами	63,55	82,72
7	В организации используются языковые курсы для взрослых	Интернет используется в работе центра несколько раз в день	62,62	82,72
8	В организации используются языковые курсы для взрослых	Знания немецкого языка за последние 10 лет улучшились	63,55	85,19
9	Знания немецкого языка за последние 10 лет улучшились	Более 50 % посетителей центра являются российскими немцами	62,62	80,95
10	Знания немецкого языка за последние 10 лет улучшились	Интернет используется в работе центра несколько раз в день	62,62	80,95
11	Знания немецкого языка за последние 10 лет улучшились	В организации используются языковые курсы для взрослых	64,49	82,14
12	Партнерами в работе являются организации в России	Интернет используется в работе центра несколько раз в день	62,62	83,54
13	Партнерами в работе являются организации в России	Знания немецкого языка за последние 10 лет улучшились	63,55	82,28

Таблица 1. Ассоциативные правила по деятельности общественных организаций российских немцев (продолжение)

N	Предпосылка	Следствие	*supp*	*conf*
14	Сотрудничество с МСНК осуществляется постоянно с высокой степенью эффективности	Интернет используется в работе центра несколько раз в день	63,55	88,16
15	В организации знают о том, что Германия реализует специальные программы	Более 50 % посетителей центра являются российскими немцами	64,49	80,72
16	Для изучения немецкого языка используются детские и молодёжные языковые клубы	Знания немецкого языка за последние 10 лет улучшились	50,47	90,00
17	Партнёрами в работе являются организации в России **и** Сотрудничество с МСНК осуществляется постоянно с высокой степенью эффективности	Интернет используется в работе центра несколько раз в день	51,40	90,16
18	Сотрудничество с МСНК осуществляется постоянно с высокой степенью эффективности **и** Предлагаются программы целевой направленности для всех возрастов	Интернет используется в работе центра несколько раз в день	51,40	96,49
19	Предлагаются программы целевой направленности для всех возрастов **и** Интернет используется в работе центра несколько раз в день	Сотрудничество с МСНК осуществляется постоянно с высокой степенью эффективности	51,40	90,16
20	Сотрудничество с МСНК осуществляется постоянно с высокой степенью эффективности **и** Знания немецкого языка за последние 10 лет улучшились	Интернет используется в работе центра несколько раз в день	52,34	90,32

Ассоциативные правила в таблице 1 необходимо интерпретировать в

формате:

«**Если** Предпосылка, **то** Следствие».

Анализ показывает, что поиск ассоциативных правил позволяет выявить как достаточно очевидные взаимосвязи между различными аспектами деятельности общественных организаций, так и достаточно неожиданные влияния факторов друг на друга. При этом правила с одним утверждением в предпосылке обладают большей поддержкой, но меньшей достоверностью. Ассоциативные правила, содержащие конъюнкцию двух утверждений в предпосылке, характеризуются меньшей поддержкой, но очень высокой достоверностью. Следует отметить, что в формировании ассоциативных правил участвует всего 9 утверждений из 173 возможных. Между остальными утверждениями ассоциативные правила с достаточно высокими поддержкой и достоверностью отсутствуют.

3. Теоретико-графовый анализ

Построим граф связей на основе выявленных ассоциативных правил. В таблице 2 приведены обозначения для утверждений, встречающихся в ассоциативных правилах.

Таблица 2. Обозначения для утверждений, встречающихся в ассоциативных правилах.

v_1	Более 50 % посетителей центра являются российскими немцами
v_2	Интернет используется в работе центра несколько раз в день
v_3	В организации используются языковые курсы для взрослых
v_4	Знания немецкого языка за последние 10 лет улучшились
v_5	Партнёрами в работе являются организации в России
v_6	Сотрудничество с МСНК осуществляется постоянно с высокой степенью эффективности
v_7	В организации знают о том, что Германия реализует специальные программы
v_8	Для изучения немецкого языка используются детские и молодёжные языковые клубы

На рисунке 1 приведён граф, построенный на основе ассоциативных правил с одной предпосылкой. В качестве веса рёбер использованы значения достоверности.

Матрица весов построенного графа приведена в Таблице 3. Вес отсутствующих дуг равен нулю.

Применим к данному графу алгоритм поиска сообществ (community) [18]. Для этого необходимо предварительно получить приведённый вид матрицы ве-

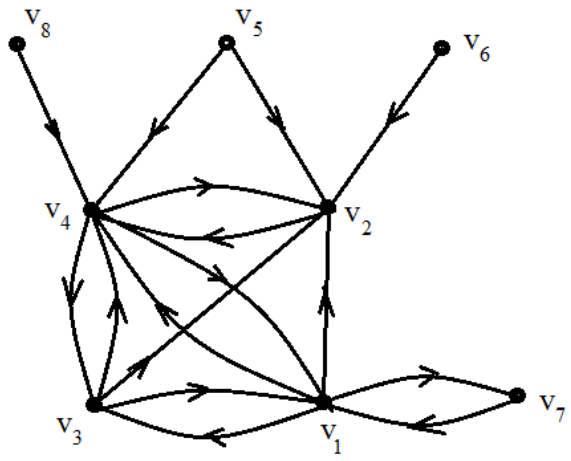

Рис. 1. Граф связи утверждений с ассоциативными правилами

Таблица 3. Матрица весов графа ассоциативных правил

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8
v_1	0	80,72	80,72	81,93	0	0	80,72	0
v_2	0	0	0	80,00	0	0	0	0
v_3	82,72	82,72	0	85,19	0	0	0	0
v_4	80,95	80,95	82,14	0	0	0	0	0
v_5	0	83,54	0	82,28	0	0	0	0
v_6	0	88,16	0	0	0	0	0	0
v_7	80,72	0	0	0	0	0	0	0
v_8	0	0	0	90,00	0	0	0	0

сов $e = E/m$, где

$$m = \sum_{i,j=1}^{N} E_{ij}.$$

N — количество вершин. В приведённой матрице весов элемент e_{ij} показывает долю веса заданного ребра в общем весе графа. В дальнейшем под матрицей весов будет пониматься именно приведённый вид. Легко увидеть, что

$$\sum_{i,j=1}^{N} e_{ij} = 1.$$

Для выявления сообществ используется функция модульности (modularity),

показывающая оптимальность разбиения графа на подграфы:

$$Q(e) = \sum_{i=1}^{N} e_{ii} - \sum_{i=1}^{N} p_{ii},$$

где p_{ii} — «ожидаемая связность». В канонической модели [18] p_{ii} определяется через исходящую степень вершины a_i, и входящую степень вершины b_i:

$$p_{ii} = a_i b_i.$$

В этом случае модульность записывается в виде:

$$Q(e) = \sum_{i=1}^{N} e_{ii} - \sum_{i=1}^{N} a_i b_i,$$

где

$$a_i = \sum_{j=1, j\neq i}^{N} e_{ij}, \quad b_i = \sum_{j=1, j\neq i}^{N} e_{ji}.$$

Для поиска сообществ на графах используется алгоритм образования стяжек. Выделим в графе G подграф G' и заменим все его вершины одной вершиной, при этом вершины подграфа $G\backslash G'$ остаются неизменными. Образованная вершина связана дугами с теми вершинами графа $G\backslash G'$, с которыми были связаны вершины, вошедшие в стяжку. Вес вершины, вошедшей в стяжку, равен сумме весов вершин и дуг, вошедших в стяжку.

Под сообществом будем понимать подграф исходного графа, который при образовании из него стяжки максимизирует функцию модульности графа $Q(e)$. Нашей задачей становится выявление сообществ на графе ассоциативных правил. В силу того что исходный граф имеет малое количество вершин, задача может быть решена полным перебором.

Функция модульности исходного графа равна $Q = -0{,}1497$. Объединение вершин v_1, v_3, v_4 в одно сообщество приводит к значению функции модульности $Q_{1,3,4} = -0{,}0361$, то есть такое объединение является выгодным и показывает тесную связь этих вершин. Объединение в сообщество вершин v_1, v_3, v_4 и v_2 приводит к значению модульности $Q_{1,2,3,4} = -0{,}0233$. Остальные варианты объединения не повышают функцию модульности, то есть не являются выгодными.

Из этого анализа можно сделать вывод о тесной связи таких аспектов деятельности общественных организаций российских немцев: «Более 50 % посетителей центра являются российскими немцами», «Интернет используется в работе центра несколько раз в день», «В организации используются языковые курсы для взрослых», «Знания немецкого языка за последние 10 лет улучшились». Эти четыре направления деятельности наиболее тесно взаимосвязаны между собой, и их надо рассматривать в совокупности.

4. Выводы

Предложенный в данной статье подход, основанный на поиске ассоциативных правил и дальнейшем представлении связей между вопросами анкеты в виде ориентированного графа, позволяет выявить закономерности, проявляющиеся в деятельности общественных организаций. Анализ графа взаимосвязей с помощью поиска сообществ вершин даёт возможность определять наиболее тесно взаимосвязанные аспекты деятельности общественных организаций.

Литература

1. Agrawal R., Imielinski T., Swami A. Mining Association Rules between Sets of Items in Large Databases // In Proc. of the 1993 ACM SIGMOD International Conference on Management of Data. Washington DC, USA. P. 207–216.
2. Shaheen M., Shahbaz M., Guergachi A. Context Based Positive and Negative Spatio Temporal Association Rule Mining // Elsevier Knowledge-Based Systems. 2013. P. 261–273.
3. Белим С.В., Майоров-Зильбернагель А.О. Восстановление изображений со статическими пропусками на основе метода ассоциативных правил // Вестник компьютерных и информационных технологий. 2014. № 12. С. 18–23.
4. Белим С.В., Майоров-Зильбернагель А.О. Алгоритм поиска повреждённых пикселей и удаления импульсного шума на изображениях с использованием метода ассоциативных правил // Наука и образование: электронное научно-техническое издание. 2014. № 12., URL: `http://technomag.bmstu.ru/doc/744983.html` (дата обращения 20.02.2017).
5. Атепалихин М.С., Кассал Б.Ю., Белим С.В. Использование ассоциативных правил для выявления взаимосвязи мест обитания биологических видов // Вестник Омского университета. 2014. № 2(72). С. 25–29.
6. Lee I., Cai G., Lee K. Mining Points-of-Interest Association Rules from Geo-tagged Photos // 46th Hawaii International Conference on System Sciences. 2013. P. 1580–1588.
7. European Social Survey (2012) Sampling for the European Social Survey Round VI: Principles and Requirements Mannheim, European Social Survey, GESIS.
8. Erlandsson F., Brodka P., Borg A., Johnson H. Finding Influential Users in Social Media Using Association Rule Learning // arXiv:1604.08075v2. 2016.
9. Erlandsson F., Borg A., Johnson H., Brodka P. Predicting User Participation in Social Media // In Advances in Network Science. Springer International Publishing : Cham, Switserland, 2016. P. 126–135.
10. Nancy P., Geetha Ramani R., Jacob S. Mining of Association Patterns in Social Network Data (Face Book 100 Universities) through Data Mining Techniques and Methods // In Advances in Computing and Information Technology. Springer : Berlin/Heidelberg, Germany, 2013. V. 178. P. 107–117.
11. Bozkir A.S., Mazman S.G., Sezer E.A. Identification of User Patterns in Social Networks by Data Mining Techniques: Facebook Case // IMCW 2010. P. 145–153.
12. Yu X., Liu H., Shi J., Hwang J.N., Wan W., Lu. J. Association Rule Mining of Personal Hobbies in Social Networks // In Proceedings of the 2014 IEEE International Congress on Big Data (BigData Congress), Anchorage, AK, USA, 2014. P. 310–314.

13. Pereira B.L., Brandao W.C. ARCA: Mining crime patterns ising association rules // IADIS International Conference Applied Computing 2014 (IADIS AC2014).
14. Gorecki J., Slaninova K. Building synthetic social network using association rules and clustering methods: case study on global terrorism database // Acta academica karviniensia. URL: `http://www.slu.cz/opf/cz/informace/acta-academica-karviniensia/casopisy-aak/aak-rocnik-2013/docs-3-2013/Gorecki_Slaninova.pdf` (дата обращения 20.02.2017).
15. Zahedi F., Zare-Mirakabad M.R. Employing data mining to explore association rules in drug addicts // Journal of AI and Data Mining. 2014. V. 2, No. 2. P. 135–139.
16. Malerba D., Esposito F., Lisi F.A. Mining Spatial Association Rules in Census Data // Specifying collective outputs from UN/ECE activities on statistical information technology. P. 541–550.
17. Agrawal R., Srikant R. Fast Discovery of Association Rules // In Proc. of the 20th International Conference on VLDB, Santiago, Chile, 1994. P. 487–499.
18. Newman M.E.J. Mixing patterns in networks // Phys. Rev. E. 2003. V. 67. P.026126-1–026126-13.

APPLICATION OF AN ASSOCIATIVE RULES METHOD TO THE ACTIVITIES ANALYSIS OF PUBLIC ORGANIZATIONS

S.V. Belim
Dr.Sc. (Phys.-Math.), Professor, e-mail: sbelim@mail.ru
T.B. Smirnova
Dr.Sc. (History), Associate Professor, e-mail: SmirnovaTB@omsu.ru
A.N. Mironenko
Ph.D. (Eng.), Associate Professor, e-mail: MironenkoAN@omsu.ru

Dostoevsky Omsk State University

Abstract. In the article the activities indicators analysis method of public organizations is developed. The method of the associative rules is used. Transactions are created of polling questionnaires. From coded responses of questionnaires the associative rules are obtained. The rules possessing high support and confidence are selected. On the basis of the associative rules the oriented weighted correlations graph is constructed. On this graph search of communities is executed. The selected community allowed to reveal activities factors of public organizations the most tightly connected with each other.

Keywords: associative rules, public organizations, community, Data Mining.

Дата поступления в редакцию: 22.02.2017