

ПРИМЕНЕНИЕ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ ДЛЯ ПОВЫШЕНИЯ КАЧЕСТВА ПРИНЯТИЯ РЕШЕНИЙ ПРИ УПРАВЛЕНИИ РЕСУРСАМИ ИНФОРМАЦИОННОЙ МЕДИАСИСТЕМЫ

К. Б. Шабанов, В. В. Алексеев

*Кафедра «Информационные системы и защита информации»,
kotouchiha1@mail.ru; ФГБОУ ВО «ТГТУ», г. Тамбов, Россия*

Ключевые слова: генетический алгоритм; интеллектуальные системы; интеллектуальный анализ данных; нечеткая логика.

Аннотация: В связи с актуальностью автоматизации анализа данных рассмотрены базовые методы Data mining, такие как нейронные сети, генетические алгоритмы, методы нечеткой логики. Показана сущность данных методов и их применимость на практике, в частности, методов нечеткой логики для задачи повышения качества принятия решения при управлении ресурсами информационной медиасистемы.

Введение

Актуальность автоматизации анализа данных сложно переоценить, так как с каждым годом, объем данных на различных носителях растет [1, 2]. Исследования и обнаружение алгоритмами, средствами искусственного интеллекта в сырых данных скрытых знаний, которые ранее не были известны, нетривиальны, практически полезны и доступны для интерпретации человеком, то есть методы интеллектуального анализа данных (*Data mining*) на практике применяются в различных сферах: торговле, телекоммуникациях, промышленности, медицине, экономике, интернет-технологиях, страховом бизнесе и т.д.

Постановка задачи

Необходимо найти наиболее эффективный метод для решения задачи продвижения программного продукта на рынке. Под продвижением будем понимать любую форму сообщений, используемую разработчиком программного продукта (ПП) путем проведения маркетинговых коммуникаций в целях донесения до целевой аудитории основного конкурентного преимущества продукта, которое в свою очередь влияет на выбор и использование данного ПП потребителем [3]. Программный продукт – любое мобильное приложение, независимо от модели его распространения. Информационная медиасистема – медиаресурс, используемый для проведения маркетинговых коммуникаций. В данном случае рассмотрим медиаресурс сети Интернет *YouTube*, который является ведущим видеохостингом в мире согласно рейтинговой системе *Alexa Rank*. В свою очередь медиаресурс включает в себя множество маркетинговых каналов для продвижения. Ресурсы – это маркетинговый бюджет, определенный для маркетинговых коммуникаций на конкретном медиаресурсе. Эффективность – точность анализа данных маркетинговых каналов информационной медиасистемы и ПП для принятия решения о вложении маркетингового бюджета в конкретный маркетинговый канал.

Цель статьи – рассмотреть базовые методы *Data mining* в разрезе повышения эффективности управления ресурсами для продвижения ПП в информационной медиасистеме.

Базовые методы интеллектуального анализа данных и решаемые ими задачи

Нейронные сети. Класс таких моделей основан на биологической аналогии с мозгом человека и предназначен, после прохождения этапа так называемого обучения на имеющихся данных, для решения разнообразных задач анализа данных [4].

Нейронные сети являются обучаемыми, и после этапа выбора архитектуры сети, который в свою очередь также является важным, так как от него зависит итоговая эффективность работы сети, происходит процесс обучения.

Обучение может быть двух видов: с учителем и без него.

В первом случае сети предоставляется обучающая выборка (целевой вектор), выходной результат работы сети (выходной вектор) сравнивается с целевым вектором, и в соответствии с алгоритмом обучения изменяются веса. Спектр задач, решаемых нейронными сетями, достаточно широк: классификация, кластеризация, аппроксимация, прогноз, оптимизация, управление [4].

Генетические алгоритмы. Данные системы появились в процессе поиска новых методов математических расчетов, поскольку возможности как детерминированных, так и статистических методов оказались недостаточными для описания вновь появляющихся классов систем.

В методах эффективного поиска сформировалось понятие «эволюционные вычисления». В связи с этим оказалось полезным сочетать детерминированную составляющую алгоритмов со случайной [5]. Данные методы относятся к универсальным методам оптимизации, позволяют решать задачи разных типов и различной сложности, отличаются возможностью реализации как однокритериального поиска, так и многокритериального.

Нечеткая логика. Суть систем, основанных на нечеткой логике, лежит в понятии нечетких множеств.

Пусть E – универсальное множество, x – элемент E , а G – некоторое свойство. Обычное (четкое) подмножество A универсального множества E , элементы которого удовлетворяют свойству G , определяется как множество упорядоченных пар

$$A = \left\{ \begin{array}{c} \mu_A(x) \\ x \end{array} \right\}, \quad (1)$$

где $\mu_A(x)$ – характеристическая функция, принимающая значение 1, если x удовлетворяет свойству G , и 0 – в противном случае.

Нечеткое подмножество отличается от обычного тем, что для элементов x из E нет однозначного ответа «да – нет» относительно свойства G . В связи с этим нечеткое подмножество A определяется по формуле (1), но в данном случае $\mu_A(x)$ – характеристическая функция принадлежности, принимающая значение в некотором вполне упорядоченном множестве $M = [0, 1]$ [4].

В 1992 г. Ванг доказал следующую теорему: для каждой вещественной непрерывной функции $G(x)$, на компактном (замкнутом) множестве U , и для произвольного $\varepsilon > 0$ существует нечеткая система, формирующая выходную функцию $f(x)$, такую, что

$$\sup \|G(x) - F(x)\| \leq \varepsilon, \quad (2)$$

где $\|*\|$ – символ принятого расстояния между функциями [6].

Данный метод целесообразно применять при моделировании организационно-технических систем, когда неотъемлемым фактором их функционирования является человек. В основе этого высказывания лежит принцип несовместимости Заде, который утверждает, что по мере усложнения системы способность дать точные и в то же время имеющие практическое значение суждения уменьшается.

Данный метод используется в поддержке принятия решений, когда информации не хватает и она невысокого качества. В этом случае лицо принимающее решение (ЛПР) отходит от точечных числовых оценок и заменяет их качественными лингвистическими переменными на естественном языке (высокий\низкий, много\мало\nормально и т.д.).

Для решения рассматриваемой задачи ЛПР необходимо проанализировать следующие данные:

- заработок на одного пользователя за все время пользования продуктом в разбивке по странам;
- целевая аудитория ПП и маркетингового канала;
- средняя конверсия в установку для ПП;
- тематика ПП и маркетингового канала;
- среднее количество просмотров маркетинговой коммуникации;
- кликабельность для конкретного формата маркетинговой коммуникации.

На основе анализа вышеперечисленных данных ЛПР прогнозирует эффективность вложений ресурсов в данный маркетинговый канал с применением таких критериев, как окупаемость инвестиций; совпадение тематик и др.

Каждый из данных факторов является сложным, а оценка совпадения тематик – нечетким фактором. Исходя из данной специфики, выбор метода нечеткой логики является обусловленным.

Применение нечеткой логики для повышения эффективности управления ресурсами информационной медиасистемы

Для построения модели нечеткого логического вывода определим лингвистические переменные (табл. 1).

Граничные значения каждой из функций принадлежности для каждого термина определены экспертным путем.

В качестве алгоритма нечеткого логического вывода использовался алгоритм Мамдани, как самый доступный для моделирования. Далее опишем функции принадлежности для термов лингвистических переменных. Покажем, например, функции принадлежности для термов переменной «Окупаемость»:

– «Не окупится»:

$$\mu(x, 50, 100) \equiv \begin{cases} 1, & x \leq 50; \\ 1 - \frac{2(x-50)^2}{2500}, & 50 < x < 75; \\ \frac{2(100-x)^2}{2500}, & 75 < x < 100; \\ 0, & x \geq 100. \end{cases} \quad (3)$$

– «Окупится»:

$$\mu(x, 5, 100) \equiv \begin{cases} e^{-\left(\frac{x-100}{5}\right)^2}; \\ 0 \leq x < \infty. \end{cases} \quad (4)$$

Лингвистические переменные и их термы

Лингвистическая переменная	Область определения значений входных значений	Термы		
		Окупаемость	[0; ∞]	Не окупится
Пол	[0; 100]	Плохой	Удовлетворительный	Хороший
Возраст				
Тематика				

– «Принесет прибыль»:

$$\mu(x, 100, 120) \equiv \begin{cases} 1, & x \leq 100; \\ \frac{2(x-100)^2}{400}, & 100 < x \leq 110; \\ 1 - \frac{2(120-x)^2}{400}, & 110 < x < 120; \\ 1, & 120 \leq x. \end{cases} \quad (5)$$

Смоделируем данную систему нечеткого логического вывода с помощью пакет прикладных программ для решения задач технических вычислений MATLAB *Fuzzy Logic Toolbox*.

Для того чтобы проверить адекватность применяемого метода, рассчитаем реальные значения характеристик маркетинговых каналов, с которыми уже проводились маркетинговые коммуникации для конкретного ПП.

Задача проверки адекватности разработанных моделей будет определяться разницей между значениями процента ошибок эксперта ω и процента ошибок системы ω'

$$\Delta W = \omega - \omega', \quad (6)$$

где $\omega = \frac{e}{E} \cdot 100\%$, e – число ошибок, E – общее число решений.

В случае если $\Delta W \geq 0$, то можно сделать вывод об эффективности использования данного метода. Если $\Delta W \leq 0$ – есть необходимость в дальнейшей настройке системы, либо выборе другого метода.

Далее рассчитаем значения процента ошибок ω для эксперта

$$\omega = \frac{5}{20} \cdot 100\% = 25\%$$

и системы

$$\omega' = \frac{2}{20} \cdot 100\% = 10\%.$$

Найдем разность ΔW

$$\Delta W = 25 - 10 = 15\%.$$

Заключение

Исходя из полученных данных, сделан вывод, что метод нечеткой логики применим для задачи повышения качества принятия решений при управлении ресурсами информационной медиасистемы и позволяет снизить процент ошибок при принятии решений на 15 %.

Список литературы

1. Hilbert, M. The World's Technological Capacity to Store, Communicate, and Compute Information / M. Hilbert, P. Lopez // *Science*. – 2011. – Vol. 332. – P. 60 – 65.
2. The Zettabyte Era: Trends and Analysis. – Текст : электронный // Cisco. – 2015. – 29 p. – URL : https://files.ifi.uzh.ch/hilty/t/Literature_by_RQs/RQ%20102/2015_Cisco_Zettabyte_Era.pdf (дата обращения: 11.12.2020).
3. Ехлаков, Ю. П. Основные положения по разработке программы продвижения программных продуктов в сети Интернет / Ю. П. Ехлаков, Д. Бараксанов // *Бизнес-информатика*. – 2012. – № 4 (22). – С. 33 – 39.
4. Методы и модели анализа данных: OLAP и Data Mining / А. А. Барсегян, М. С. Куприянов, В. В. Степаненко, И. И. Холод. – СПб. : БХВ-Петербург, 2004. – 336 с.
5. Советов, Б. Я. Интеллектуальные системы и технологии : учеб. / Б. Я. Советов, В. В. Цехановский, В. Д. Чертовской. – М. : Академия, 2013. – 320 с.
6. Круглов, В. В. Искусственные нейронные сети. Теория и практика / В. В. Круглов, В. В. Борисов. – 2-е изд. – М. : Горячая линия – Телеком, 2002. – 382 с.

Application of Intellectual Methods of Data Analysis to Improve the Quality of Decision Making in Management of Resources for the Information Media System

K. V. Shabanov, V. V. Alekseev

*Department of Information Systems and Information Security,
kotouchihal@mail.ru; TSTU, Tambov, Russia*

Keywords: genetic algorithm; intelligent systems; data mining; fuzzy logic.

Abstract: In connection with the relevance of data analysis automation, basic data mining methods are considered, such as neural networks, genetic algorithms, and fuzzy logic methods. The essence of these methods and their practical applicability are shown, in particular, methods of fuzzy logic for the problem of improving the quality of decision-making when managing the resources of the information media system.

References

1. Hilbert M., Lopez P. The World's Technological Capacity to Store, Communicate, and Compute Information, *Science*, 2011, vol. 332, pp. 60-65.
2. https://files.ifi.uzh.ch/hilty/t/Literature_by_RQs/RQ%20102/2015_Cisco_Zettabyte_Era.pdf (accessed 11 December 2020).
3. Yekhlakov Yu.P., Baraksanov D. [Basic provisions for the development of a program for the promotion of software products on the Internet], *Biznes-informatika* [Business Informatics], 2012, no. 4 (22), pp. 33-39. (In Russ., abstract in Eng.)

4. Barsegyan A.A., Kupriyanov M.S., Stepanenko V.V., Kholod I.I. *Metody i modeli analiza dannykh: OLAP i Data Mining* [Methods and models for data analysis: OLAP and Data Mining], St. Petersburg: BKHV-Peterburg, 2004, 336 p. (In Russ.)

5. Sovetov B.Ya., Tsekhanovskiy V.V., Chertovskoy V.D. *Intellektual'nyye sistemy i tekhnologii: uchebnik* [Intelligent systems and technologies: textbook], Moscow: Akademiya, 2013, 320 p. (In Russ.)

6. Kruglov V.V., Borisov V.V. *Iskusstvennyye neyronnyye seti. Teoriya i praktika* [Artificial neural networks. Theory and practice], Moscow: Goryachaya liniya - Telekom, 2002, 382 p. (In Russ.)

Anwendung von Miningmethoden der Daten zur Verbesserung der Qualität der Entscheidungsfindung bei der Verwaltung der Ressourcen des Informationsmediensystems

Zusammenfassung: Im Zusammenhang mit der Relevanz der Automatisierung der Datenanalyse werden grundlegende Data-Mining-Methoden wie neuronale Netze, genetische Algorithmen und Fuzzy-Logik-Methoden untersucht. Es ist das Wesen dieser Methoden und ihre Anwendbarkeit in der Praxis, insbesondere die Methoden der Fuzzy-Logik für das Problem der Verbesserung der Qualität der Entscheidungsfindung bei der Verwaltung der Ressourcen des Informationsmediensystems, gezeigt.

Utilisation des méthodes d'analyse des données pour l'élévation de la qualité de la prise des décisions dans la gestion des ressources du système de médias

Résumé: Compte tenue de l'actualité de l'automatisation de l'analyse des données, sont examinées les méthodes de base Data mining telles que les réseaux neuronaux, les algorithmes génétiques et les méthodes de logique floue. Il est montrée l'essence de ces méthodes et leur applicabilité dans la pratique, en particulier des méthodes de logique floue pour améliorer la qualité de la prise des décisions dans la gestion des ressources du système de médias d'information.

Авторы: *Шабанов Константин Борисович* – аспирант кафедры «Информационные системы и защита информации»; *Алексеев Владимир Витальевич* – доктор технических наук, профессор, заведующий кафедрой «Информационные системы и защита информации», ФГБОУ ВО «ТГТУ», г. Тамбов, Россия.

Рецензент: *Тютюнник Вячеслав Михайлович* – доктор технических наук, профессор кафедры «Конструирование радиоэлектронных и микропроцессорных систем», ФГБОУ ВО «ТГТУ», г. Тамбов, Россия.