

DOI: 10.15593/2499-9873/2019.1.07

УДК 51-7

И.А. Задворная, О.М. Ромакина

Саратовский национальный исследовательский государственный университет имени Н.Г. Чернышевского, Саратов, Россия

СОЗДАНИЕ НЕЙРОННОЙ СЕТИ ДЛЯ ВЫЯВЛЕНИЯ ПОТЕНЦИАЛЬНЫХ КЛИЕНТОВ БАНКА

Рассмотрено применение алгоритма Data Mining «Нейронная сеть» для анализа имеющихся данных потенциальных клиентов банка. Рассматриваются возможности и особенности создания и использования нейронных сетей при проведении комплексного анализа данных и построения системы для дальнейшего прогнозирования. Авторами ставится задача выявления основных характеристик потенциальных клиентов, которые влияют на желание человека продолжать взаимодействие с банком после предложения им услуг. Исходя из поставленной задачи, на основе созданной многомерной структуры данных создается нейронная сеть для возможности прогнозирования перспективности взаимодействия с каждым клиентом. Построенная нейронная сеть позволяет на основе новых клиентских данных выявлять потенциальных клиентов для разработки дальнейших маркетинговых кампаний и оптимизации банковских стратегий, направленных на улучшение клиентского взаимодействия в долгосрочной перспективе.

Ключевые слова: интеллектуальный анализ данных, нейронная сеть, банковские клиентские данные, многомерная структура данных, хранилище данных, сигмоидальная функция, прогнозирование поведения клиентов.

I.A. Zadvornaya, O.M. Romakina

Saratov State University, Saratov, Russian Federation

CREATING NEURAL NETWORK FOR IDENTIFICATION OF POTENTIAL BANK CUSTOMERS

The article is devoted to the application of the Data Mining «Neural Network» algorithm for analyzing the available data of potential bank customers. The article discusses the possibilities and features of the creation and use of neural networks when conducting a comprehensive data analysis and building a system for further forecasting. The authors set the task of identifying the main characteristics of potential customers that affect a person's desire to continue interacting with the bank after offering services to them. Based on the task, the neural network is created on the basis of the created multidimensional data structure to be able to predict the prospects of interaction with each client. The constructed neural network allows identifying potential customers based on new customer data to develop further marketing campaigns and optimize banking strategies aimed at improving long-term customer interaction.

Keywords: Data Mining, neural network, bank client data, multidimensional data structure, data warehouse, sigmoid function, prediction of customer behavior.

В деловом мире корпоративные и клиентские данные все в большей мере становятся признанными в качестве стратегического актива. Возможность извлекать полезные знания, скрытые в этих данных, и воздействовать на них становится все более важной в современном конкурентном мире. Весь процесс применения компьютерной методологии, включая новые методы обнаружения знаний в данных, называется интеллектуальным анализом данных. Его целью является выявление достоверных новых потенциально полезных и понятных корреляций и шаблонов в существующих данных. Добыча данных – это междисциплинарная область, охватывающая целый ряд дисциплин: статистика, машинное обучение, базы данных, распознавание образов и другие области [1]. Методы интеллектуального анализа данных предсказывают значения данных, используя некоторые уже известные результаты, которые были найдены с применением другого набора данных.

Анализ клиентских данных в банковской сфере имеет первостепенное значение для данной области. В этой сфере существуют сложные взаимосвязи, для изучения которых одним из главных инструментов является Data Mining. Нейронные сети с успехом применяются для решения многих практических задач в банковской сфере. Например, предсказание платежеспособности, определение рисков вложений, финансовой эффективности различных банковских операций. В исследовании, представленном в сборнике статей [2], рассматривалась возможность применения Data Mining (в частности, нейронных сетей) к анализу данных владельцев кредитных карт банка для прогнозирования вероятности невозврата средств.

Банковская сфера как объект имеет сложное детерминированное квазислучайное поведение, для описания которого нейронные сети являются идеальным, по мнению авторов, инструментом. Нейросети успешно применяются для решения многих практических задач в банковской сфере: для оценки платежеспособности клиентов (на вход нейронной сети подаются данные о клиенте, на выходе – прогнозируемая степень его платежеспособности), предсказания банкротства, прогнозирования остатков на корреспондентских счетах, оценки риска страхования, предсказания изменения стоимости акций и экономической эффективности финансирования экономических и инновационных проектов и т.п.

Нейронная сеть моделируется по аналогии с работой человеческого мозга. Этот процесс является одним из способов распознавания образов и минимизации ошибок. Нейронную сеть можно также определить как способ извлечения полезных знаний из данных и обучения на основе данного опыта [3].

Нейронные сети состоят из узлов, расположенных в слоях. Схема конструкции зависит от типа и сложности нейронной сети. Перед началом процесса данные разделяются на учебные и тестовые наборы данных. Затем каждому из узлов в первом слое назначаются синаптические веса. Во время каждой итерации входы обрабатываются в результате прохождения через систему и сравниваются с фактическим значением. Полученная ошибка измеряется и возвращается в результате работы системы для корректировки синаптического веса каждого параметра сети. В большинстве случаев веса определяются лучше при прогнозировании фактических значений. Процесс завершается, когда достигается заданный минимальный уровень ошибки [4].

Для вычисления значений синаптических весов выходных сигналов искусственных нейронов используются функции активации, иначе называемые активационными функциями. В качестве аргумента они принимают входной сигнал от предыдущего слоя нейронов в модели. В данном исследовании используется сигмоидальная функция активации. Такая функция позволяет усиливать слабые сигналы в нейронной сети, при этом не перенасыщает модель в момент прохождения сильных сигналов [5].

Одним из преимуществ нейронной сети является ее способность выявлять нелинейные отношения в данных. Это позволяет анализировать сложные структуры. Однако одним из недостатков сети является склонность к переполнению данными. Это может привести к ухудшению качества модели применительно к новым данным [6]. Результаты работы нейронной сети также часто трудно интерпретировать.

Существует несколько определений и подходов к концепции построения нейронной сети, мы остановимся на одном из них. Искусственная нейронная сеть представляет собой массивный параллельный распределенный процесс, состоящий из простых подпроцессов. Она обладает способностью «учиться» на основе эмпирических знаний, выраженных через параметры межсетевое соединения, и может предоставлять такие знания для использования [7].

Нейронную сеть определим на основе четырех компонентов: U – множество узлов (нейронов), L^w – набор взвешенных связей между узлами, In – набор входных функций, и g – активационная функция [8]. Таким образом, набор элементов выглядит следующим образом:

$$\langle U, L^w, \text{In}, g \rangle.$$

Рассмотрим данные компоненты подробнее. Множество узлов U по местонахождению в модели можно разделить на три группы: U_{in} – набор входных нейронов, образующих входной слой; U_h – набор нейронов внутри модели, составляющих внутренний (скрытый) слой; U_{out} – набор выходных нейронов, представляющий собой выходной слой.

Набор связей между нейронами L^w состоит из двух компонентов: L – множество связей между узлами и W – набор числовых весов по всем связям между узлами.

Каждая связь задается как упорядоченная пара $\langle j, i \rangle$, где j и i – соответствующие индексы узлов U_j и U_i . Эта пара определяет направленную связь из U_j в U_i . А W_{ji} – вес связи между соответствующими узлами U_j и U_i .

Набор линейных входных функций In состоит из функций in_i (взвешенный вход для узла U_j) по каждому узлу U_i ; a_j – выход узла U_j , который служит в качестве входа для узла U_i ; a_i – множество всех входов в узел U_j . Общий вид функции

$$\text{in}_i(a_i) = \sum_j W_{ji} \cdot a_j.$$

Функция активации g преобразует взвешенный вход в конечное значение, которое служит активационным значением a_i . В качестве функции активации выступает следующая сигмоидальная функция

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}}.$$

Каждый блок обработки в нейронной сети выполняет простое вычисление. На основе входных сигналов от входных каналов блок вычисляет новый уровень активации для своего выходного канала. Рассмотрим этапы построения нейронной сети [9]:

1. Определяем соответствующие узлы (входной, выходной и скрытый слои) и их количество.
2. Определяем связи между узлами для формирования сети.
3. Инициализируем веса всех узлов.
4. Выбираем примеры для обучения.
5. Определяем результат работы сети для примеров.
6. Изменяем веса в соответствии с использованием алгоритма обучения, применяемого к набору примеров, чтобы минимизировать ошибку прогнозирования.

Свойством, которое имеет первостепенное значение для искусственной нейронной сети, является способность сети к обучению на реальных примерах и за счет этого повышение ее производительности. Нейронная сеть «узнает» о своей среде посредством интерактивного процесса корректировок, применяемых к весам его соединений. Сеть становится более осведомленной о своей среде после каждой итерации в процессе обучения.

Разработаем нейронную сеть для набора данных из открытого источника [10]. Данные в этом наборе основаны на информации о проведении телефонных маркетинговых кампаний банка. Создадим нейронную сеть, способную прогнозировать согласие/несогласие потенциального клиента банка на открытие депозитного счета после предложения данной услуги. Важно установить, какие характеристики больше всего влияют на выходной результат работы модели.

Хотя организация хранилища данных не является обязательной для интеллектуального анализа данных, на практике задача анализа значительно упрощается при использовании хранилища. Это особенно справедливо в случае анализа данных крупных компаний. Основная задача хранилища данных – увеличение скорости и качества извлечения из данных знаний, используемых впоследствии в процессе принятия решений.

Для реализации хранилища данных созданы следующие сущности: «Возрастная группа», «Образование», «Работа», «Семья», «Кредитная история», «Банковская история», «Время». В этих сущностях

хранятся данные о возрасте клиента, его образовании (начальное, среднее, высшее), типе занятости, семейном положении, кредитной истории и истории взаимодействия банка с данным клиентом. Они выступают в качестве таблиц измерений. Таблица фактов содержит соответствующие значения внешних ключей таблиц измерений и результата взаимодействия клиента и банка [11].

Для построения хранилища данных и нейронной сети воспользуемся готовым решением Microsoft SQL Server with Analysis Services [12]. Возможности этого программного средства соответствуют требованиям к программному обеспечению для разработки: наличие свободной или учебной лицензии, возможность визуализации данных хранилища, удобный интерфейс и возможность анализа данных с использованием технологий Data Mining. С помощью выбранного программного продукта данные организуются в многомерной базе данных и затем выполняется их анализ.

Алгоритмы, предоставляемые в Microsoft SQL Server Data Mining, являются наиболее популярными, хорошо изученными методами получения шаблонов из данных.

Современный алгоритм нейронной сети Microsoft представляет собой реализацию популярной и адаптируемой архитектуры нейронной сети для машинного обучения. Каждому состоянию входного атрибута алгоритм ставит в соответствие возможное состояние прогнозируемого атрибута.

При создании моделей интеллектуального анализа данных с использованием алгоритма Microsoft Neural Network количество сетей, содержащихся в одной модели интеллектуального анализа данных, зависит от количества состояний (значений атрибутов) во входных столбцах, а также от числа прогнозируемых столбцов, которые использует модель.

Данный алгоритм создает сеть, состоящую минимум из трех уровней узлов. В качестве параметра выходного слоя рассматривается результат проведенной маркетинговой кампании. Для входного слоя будем использовать описанные ранее показатели.

Значения в построенной нейронной сети рассматриваются как сигналы, передающиеся в следующий слой, ослабляясь или усиливаясь в зависимости от числовых значений (весов), приписываемых межнейронным связям. В результате на выходе нейрона самого верхнего слоя

формируется выходное значение. Для дальнейшего применения модели необходимо провести обучение сети на полученных ранее данных, для которых известны как входные, так и выходные значения. Тренировка состоит в подборе весов межнейронных связей, обеспечивающих наибольшую близость ответов сети к известным правильным ответам. Такая структура обеспечивает возможность определения набора значений характеристик, позволяющих отделить одну категорию данных от другой. Можно определить, какие характеристики оказывают влияние на предсказываемый параметр и какова степень этого влияния.

Посмотрим на взаимозависимость параметров. Подадим на вход нейронной сети заданные значения нескольких входных параметров: будем рассматривать клиента до 33 лет, студента, с семейным положением «не замужем/не женат». Таким образом, сразу задаем значения трех входных переменных.

В результате увидим, что подобное ограничение параметров оказывает большое влияние на положительный результат проведения маркетинговой кампании. Входные значения других переменных не влияют на полученное на выходе значение. В результате при заданных данных о клиенте маркетинговая кампания с большой вероятностью окажется успешной независимо от наличия другой информации.

Наиболее значимыми входными параметрами клиентов по результатам моделирования маркетинговых кампаний оказываются: длительность связи более 7 мин в течение данной маркетинговой кампании; параметр «Образование» – «необразованный»; клиент является студентом или пенсионером. Во всех перечисленных случаях велика вероятность получения согласия клиента на предложение банка.

Для наглядной оценки достоверности построенной модели воспользуемся графиком точности прогнозов [13]. Согласно графику (рисунок), полученная модель достаточно точна и отклонение от идеальной прогностической модели составляет не более 12 %.

Данные, описывающие характеристики клиентов той или иной компании, являются одной из тех областей, в которых интеллектуальный анализ данных может иметь важное практическое применение. Широкая доступность больших коллекций данных потенциальных клиентов позволяет провести анализ, который может дать банкам беспрецедентную возможность лучше понять природу и особенности взаимодействия с потенциальными клиентами.

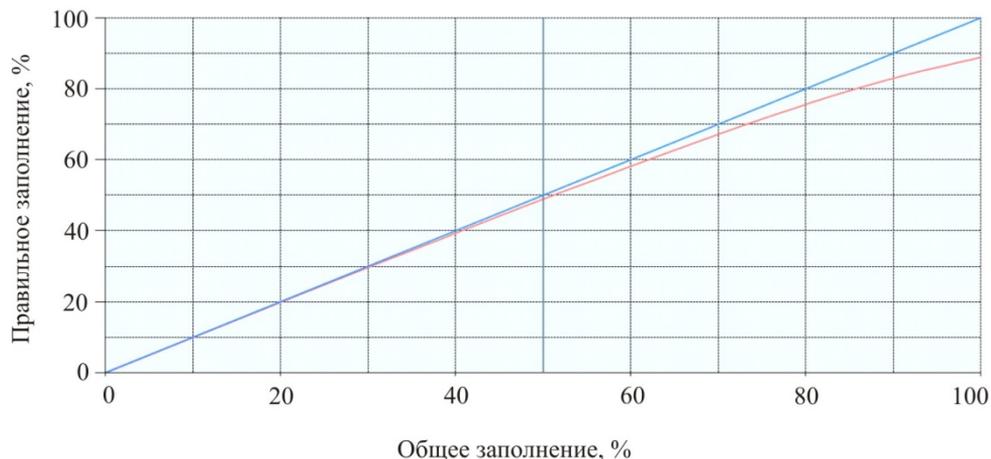


Рис. График точности прогнозов для нейронной сети

В настоящей работе была создана модель нейронной сети для исследования поведения клиентов по имеющейся базе персональных характеристик. Данное исследование может способствовать дальнейшему изучению влияния персональных характеристик клиентов на использование ими различных банковских продуктов с целью совершенствования существующей маркетинговой политики банка.

Список литературы

1. Pyle D. Business Modeling and Data Mining. – Waltham: Morgan Kaufmann Publishers, 2003. – 650 p.
2. Задворная И.А., Ромакина О.М. Применение алгоритмов Data Mining для анализа данных в сфере кредитования // Математическое и компьютерное моделирование в экономике, страховании и управлении рисками: материалы VII междунар. молодеж. науч.-практ. конф. – Саратов: Научная книга, 2018. – С. 61–66.
3. Nong Ye The Handbook of Data Mining. – London: Lawrence Erlbaum Associates, Inc., 2003. – 722 p.
4. Rud O.P. Data Mining Cookbook. – N.Y.: John Wiley and Sons, 2001. – 479 p.
5. Wang L., Fu X. Data Mining with Computational Intelligence. – Berlin: Springer, 2005. – 294 p.
6. Larose D.T. Discovering knowledge in data: an introduction to Data Mining. – N.Y.: John Wiley and Sons, 2005. – 241 p.
7. Kantardzic M. Data Mining: Concepts, Models, Methods, and Algorithms. – N.Y.: John Wiley and Sons, 2003. – 343 p.

8. Kovalerchuk B., Vityaev E. Data Mining in Finance: Advances in Relational and Hybrid Methods. – N.Y.: Kluwer Academic Publishers, 2002. – 322 p.
9. Wang J. Encyclopedia of data warehousing and mining. – Hershey: Idea Group Inc., 2005. – 1279 p.
10. Moro M., Cortez P., Rita P. UCI Machine Learning Repository. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 2014. – URL: <http://archive.ics.uci.edu/ml/datasets/Bank+Marketing> (accessed 03 Aug. 2018).
11. Inmon W.H. Building the Data Warehouse. – 3rd ed. – N.Y.: Wiley Computer Publishing, 2002. – 576 p.
12. MacLennan J., Crivat B. Data Mining with Microsoft SQL Server 2008. – N.Y.: John Wiley and Sons, 2009. – 672 p.
13. Lift Chart (Analysis Services – Data Mining). – URL: <https://docs.microsoft.com/en-us/sql/analysis-services/data-mining/lift-chart-analysis-services-data-mining?view=sql-server-2017> (accessed 09 Aug. 2018).

References

1. Pyle D. Business Modeling and Data Mining. Waltham: Morgan Kaufmann Publishers, 2003. 650 p.
2. Zadvornaia I.A. Romakina O.M. Primenenie algoritmov Data Mining dlia analiza dannykh v sfere kreditovaniia. Matematicheskoe i komp'iuternoe modelirovanie v ekonomike, strakhovanii i upravlenii riskami: Materialy VII Mezhdunarodnoi molodezhnoi nauchno-prakticheskoi konferentsii. Saratov: Izd-vo «Nauchnaia kniga», 2018. Pp. 61-66.
3. Nong Ye The Handbook of Data Mining. London: Lawrence Erlbaum Associates, Inc, 2003. 722 p.
4. Rud O.P. Data Mining Cookbook. N.Y.: John Wiley and Sons, 2001. 479 p.
5. Wang L., Fu X. Data Mining with Computational Intelligence. Berlin: Springer, 2005. 294 p.
6. Larose D.T. Discovering knowledge in data: an introduction to Data Mining. N.Y.: John Wiley and Sons, 2005. 241 p.
7. Kantardzic M. Data Mining: Concepts, Models, Methods, and Algorithms. N.Y.: John Wiley and Sons, 2003. 343 p.
8. Kovalerchuk B., Vityaev E. Data Mining in Finance: Advances in Relational and Hybrid Methods. N.Y.: Kluwer Academic Publishers, 2002. 322 p.
9. Wang J. Encyclopedia of data warehousing and mining. Hershey: Idea Group Inc, 2005. 1279 p.
10. Moro M., Cortez P., Rita P. UCI Machine Learning Repository. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support

Systems, Elsevier, 2014. URL: <http://archive.ics.uci.edu/ml/datasets/Bank+Marketing> (online; accessed: 03.08.2018).

11. Inmon W.H. Building the Data Warehouse, Third Edition. N.Y.: Wiley Computer Publishing, 2002. 576 p.

12. MacLennan J., Crivat B. Data Mining with Microsoft SQL Server 2008. N.Y.: John Wiley and Sons, 2009. 672 p.

13. Lift Chart (Analysis Services – Data Mining). URL: <https://docs.microsoft.com/en-us/sql/analysis-services/data-mining/lift-chart-analysis-services-data-mining?view=sql-server-2017> (online; accessed: 09.08.2018).

Получено 31.01.2019

Об авторах

Задворная Ирина Александровна (Саратов, Россия) – магистрант кафедры математического и компьютерного моделирования, Саратовский национальный исследовательский государственный университет имени Н.Г. Чернышевского (410062, г. Саратов, 2-й Московский проезд, 9, e-mail: zadvornayaia@gmail.com).

Ромакина Оксана Михайловна (Саратов, Россия) – кандидат физико-математических наук, доцент кафедры математического и компьютерного моделирования, Саратовский национальный исследовательский государственный университет имени Н.Г. Чернышевского (410005, г. Саратов, ул. Большая Горная, 289-53, e-mail: romakinaom@gmail.com).

About the authors

Irina A. Zadvornaya (Saratov, Russian Federation) – Master Student of the Department of Mathematical and Computer Modeling, Saratov State University (410062, 9, 2nd Moskovskii proezd, Saratov, Russian Federation, e-mail: zadvornayaia@gmail.com).

Oksana M. Romakina (Saratov, Russian Federation) – Ph.D. in Physics and Mathematics, Associate Professor of Department of mathematical and computer modeling, Saratov State University (410005, 289-53, Bolshaya Gornaya st., Saratov, Russian Federation, e-mail: romakinaom@gmail.com).