

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ

Национальный исследовательский Томский государственный университет
Томский государственный университет систем управления и радиоэлектроники
Болгарская Академия наук
ООО «Научно исследовательское предприятие «Лазерные технологии»

ИННОВАТИКА-2019

СБОРНИК МАТЕРИАЛОВ

**XV Международной школы-конференции студентов,
аспирантов и молодых ученых
25–27 апреля 2019 г.
г. Томск, Россия**

Под редакцией А.Н. Солдатов, С.Л. Минькова

Scientific & Technical Translations



ИЗДАТЕЛЬСТВО

Томск – 2019

АЛГОРИТМ ПРЕДСКАЗАНИЯ ПРОДАЖ С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИЙ МАШИННОГО ОБУЧЕНИЯ

К.А. Гайдук, Д.В. Шашев

*Национальный исследовательский Томский государственный университет
asablindme@gmail.com*

THE SALES PREDICTION ALGORITHM WITH USE OF MACHINE LEARNING TECHNOLOGIES

K.A. Gaiduk, D.V. Shashev

National Research Tomsk State University

The object of the study is the dependence of sales of goods of a certain type on calendar holidays. Determining the dependence of sales of a certain type of goods on calendar holidays and creating a sales forecasting algorithm using machine learning technologies.

Keywords: sales of goods, machine learning technologies, dependence of sales, forecasting algorithm, prediction algorithm, supervised learning, unsupervised learning.

Машинное обучение - класс методов искусственного интеллекта, характерной чертой которых является не прямое решение задачи, а обучение в процессе применения решений множества сходных задач. Для построения таких методов используются средства математической статистики, численных методов, методов оптимизации, теории вероятностей, теории графов, различные техники работы с данными в цифровой форме [1].

Различают два типа обучения: обучение по прецедентам, или индуктивное обучение, и дедуктивное обучение. Обучение по прецедентам основано на выявлении эмпирических закономерностей в данных. Дедуктивное обучение предполагает формализацию знаний экспертов и их перенос в компьютер в виде базы знаний. Дедуктивное обучение принято относить к области экспертных систем, поэтому термины машинное обучение и обучение по прецедентам можно считать синонимами.

Многие методы индуктивного обучения разрабатывались как альтернатива классическим статистическим подходам.

Для реализации алгоритма предсказания продаж в зависимости от календарных праздников используется интеллектуальный анализ данных [1].

Интеллектуальный анализ данных – это исследование и обнаружение с помощью алгоритмических средств в сырых данных скрытых знаний,

являющихся ранее не известными, нетривиальными, практически полезными и доступными для интерпретации человеком [2].

Нейронные сети имитируют структуру головного мозга: каждый искусственный нейрон соединяется с несколькими другими нейронами. Нейросети имеют многослойную структуру: нейроны на одном слое передают данные нескольким нейронам на следующем и т. д. В конечном счете данные достигают выходного слоя, где сеть выдает предположение о том, как решить задачу, классифицировать объект и т. п. Нейросети применяются в целом ряде отраслей. В здравоохранении их используют при анализе медицинских снимков с целью ускорения диагностических процедур и поиска лекарств. В телекоммуникационной отрасли и медиаиндустрии нейросети можно применять для машинного перевода, распознавания мошенничеств и предоставления услуг виртуальных ассистентов. В финансовой отрасли их используют для распознавания мошенничеств, управления портфелями и анализа риска. В розничной торговле — для избавления от очередей в кассу и для персонализации обслуживания покупателей.

Частным случаем задачи классификации является задача прогнозирования временных рядов. Прогнозирование временных рядов осуществляется в три этапа [3].

Сначала строится модель, которая характеризует временной ряд. Для этого применяются различные методы статистики и классификации. Затем построенная модель оценивается. Для оценки модели имеющиеся данные разбиваются на две выборки: обучающую и тестовую. Построение модели выполняется на обучающем множестве, а затем с ее помощью строят прогноз на тестовом множестве. Спрогнозированные результаты сравнивают с реальными данными и по коэффициенту ошибки оценивают модель. И на последнем этапе построенная модель используется для прогноза будущих событий. Это происходит, если модель получила удовлетворительную оценку [4].

Также необходимо решить задачу по минимизации количества фиктивных переменных без потери информационной составляющей. В ходе этой процедуры выделяются основные производители и основные страны по исходной выборке, остальные переменные с около нулевой дисперсией удаляются. Данная процедура повышает предсказательную способность и устойчивость модели на тестовой выборке [5].

Литература

1. Майер-Шенбергер В., Кузьер К. Большие данные. Революция, которая изменит то, как мы живём, работаем и мыслим. М.: Манн, Иванов и Фербер, 2014. 240с.
2. Анализ данных и процессов. / А. А. Барсегян, М. С. Куприянов и др. СПб.: БХВ-Петербург, 2009. 512 с.
3. Шаль С. А. Развитие рынка технологий «Big Data»: современные тенденции и сферы применения. / I Международная научная конференция «Конвергентные когнитивно-информационные технологии» и XI Международная научно-практическая конференция «Современные информационные технологии и ИТ-образование», Москва, 25-26 ноября 2016 г. / Московский государственный университет имени М.В. Ломоносова, Москва, 2016, С 148 – 151.
4. Велевич Р.П., Давыдова Г. А. Экономика торговой организации. Учебное пособие. – Минск: Вышэйш. шк., 2008. – 371 с.
5. Kuhn M., Johnson K. Applied Predictive Modeling. Springer, 2013. 600 p. — 203 illus., 153 illus. in color.