

ИСПОЛЬЗОВАНИЕ ТЕХНОЛОГИИ DATA MINING ДЛЯ АЛГОРИТМЗАЦИИ ПРОГНОЗИРОВАНИЯ ДАННЫХ НА ПОРТАЛЕ ГОСУСЛУГ

©2020 А. П. Преображенский, В. О. Логвина

Воронежский институт высоких технологий (Воронеж, Россия)

Информация, поступающая на портал государственных услуг, требует использования специальных методов ее обработки и последующего анализа. Технологии Data Mining используются для реализации сложных задач, посредством решения которых осуществляется прогнозирование показателей на портале госуслуг.

Ключевые слова: Data Mining, задача прогнозирования, нейронная сеть, период прогнозирования, горизонт прогнозирования, интервал прогнозирования, алгоритм нейросетевого прогнозирования.

Data Mining может рассматриваться в виде совокупности математического инструментария. В него входит большое число практически важных подходов, связанных со статистикой, кибернетическими методами и т. д.

Происходит объединение на основе соответствующих принципов различных методов, являющихся формализованными. Формализация может и отсутствовать.

Кроме того, могут осуществляться исследования в рамках количественного и качественного анализа данных. В рассматриваемой технологии различным разработчикам удалось осуществить объединение моделей, методик, которые являются полезными, с точки зрения принятия решений для самых разных, практически важных приложений в промышленной, технической, образовательной и других областях.

Какие ключевые подходы лежат в основе технологии Data Mining? Исследователи отмечают перспективы развития методов, связанных с классификацией, моделированием, а также прогнозированием. В ряде практических задач приходится детальным образом описывать применение статистических методов. Если говорить о простоте реализации на практике методик Data Mining, то их могут применять и не очень подготовленные специалисты. Связано это с тем, что результаты разработок

представляются достаточно наглядным образом. Но есть и специфика, определяемая необходимостью понимания главных механизмов анализа в ходе статистической обработки. В рассматриваемой технологии большая роль принадлежит процессам принятия решений. То есть, на первом этапе происходит формирование модели. В дальнейшем в ходе анализа конкретных областей эта модель уточняется, и ее применяют при решении соответствующих проблем. В технологии Data Mining могут быть следующие стадии:

1. Проведение свободного поиска.
2. Осуществление процессов прогностического моделирования.
3. Проведение анализа исключений.

На рисунке 1 представлен обобщенный алгоритм прогнозирования с помощью нейронной сети. На первом шаге происходит обработка исходных данных, а затем построение скользящего окна, которое содержит необходимое количество вариаций множеств данных. Глубина погружения (количество прошлых отсчетов, попадающих в окно) является одним из основных параметров при построении скользящего окна. На следующих этапах происходит разбиение исходного множества данных на обучающее и тестовое множества и уже построение самой многослойной нейронной сети, которая будет обучаться на основе метода обратного распространения ошибки. В данной работе будем рассматривать задачу долгосрочного и краткосрочного прогнозирования основных показателей Единого портала государственных услуг. В качестве этих показателей возьмем количество

Преображенский Андрей Петрович – Воронежский институт высоких технологий, профессор, app@vvt.ru.

Логвина Виктория Олеговна – Воронежский институт высоких технологий, студент, l0oggv45fg@yandex.ru.

подключенных услуг пользователями. Прогнозирование будет проводиться по общему количеству услуг, подключенных за определенный период, по определенному

виду услуг, по территориальным признакам (какое количество услуг было предоставлено в конкретной территории).



Рисунок 1. Обобщенный алгоритм нейросетевого прогнозирования

В качестве периода прогнозирования возьмем следующие временные интервалы: неделю, месяц и год. Построение искусственных нейронных сетей для прогнозирования показателей портала госуслуг осуществлялось с использованием системы STATISTICA.

Для этого были построены несколько моделей нейронных сетей (табл. 1).

По данным таблицы видим, что наиболее удачной и эффективной оказалась сеть MLP 4–4–1, которая представляет собой персептрон с одним скрытым слоем и состоящий из шести нейронов. В качестве функции активации выбрана логистическая.

Таблица 1

Выбор сети для решения задач прогнозирования

№	Тип сети	Точность			Функция активации скрытого слоя	Функция активации выходного слоя
		Обучающая выборка	Тестовая выборка	Контрольная выборка		
1	MLP 4–6–1	0,991344	0,993839	0,997658	Logistic	Logistic
2	MLP 4–3–1	0,990201	0,981627	0,975421	Logistic	Exponential
3	MLP 4–4–1	0,987641	0,975234	0,984532	Exponential	Tanh
4	MLP 4–5–1	0,991124	0,973261	0,985431	Exponential	Identity
5	MLP 4–3–1	0,973251	0,982718	0,974357	Exponential	Identity

Построим полученную модель, которая изображена на рисунке 2. По модели видно, что на вход нейросети подаются 4 входных сигнала (параметра).

Эти параметры соответствуют значениям основных показателей портала за предыдущие 4 периода.

Также по модели видно, что на выходе мы получаем спрогнозированное значение

исследуемого нами показателя на будущий период.

Полученная сеть может быть использована для решения задач и краткосрочного, и среднесрочного и долгосрочного прогнозирования.

Аналогично построим несколько моделей нейронных сетей, которые отображены в таблице 2.

Архитектура : МП 4-4-6-1:1 , N = 4
 Производительность обуч. = 0,991344 , Контр. производительность = 0,997658 , Тест. производительность = 0,993839

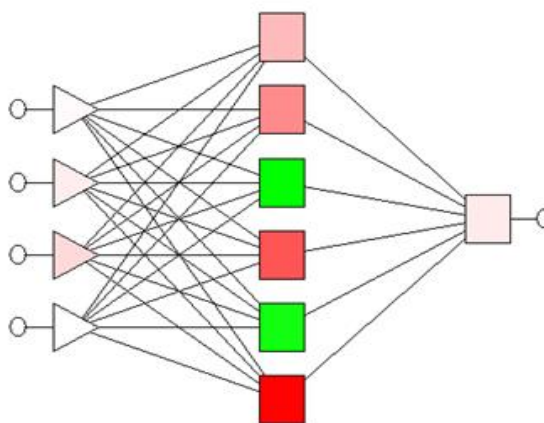


Рисунок 2. Архитектура сети для прогнозирования по четырем предшествующим периодам

Выбор сети для решения задач оперативного прогнозирования

№	Тип сети	Точность			Функция активации скрытого слоя	Функция активации выходного слоя
		Обучающая выборка	Тестовая выборка	Контрольная выборка		
1	MLP 2–3–1	0,954417	0,959695	0,956686	Logistic	Identity
2	MLP 2–6–1	0,940201	0,931824	0,921861	Identity	Identity
3	MLP 2–9–1	0,927531	0,905134	0,914532	Exponential	Identity

Также для решения задач краткосрочного прогнозирования, то есть для прогнозирования информации, которая быстро изменяется во времени можно использовать нейронную сеть с двумя входами.

Проанализировав результаты, можно сказать, что наиболее удачной и эффективной оказалась сеть MLP 2–3–1,

которая представляет собой персептрон с одним скрытым слоем и состоящий из шести нейронов. В качестве функции активации выбрана – логистическая для скрытого слоя и тождественная – для выходного слоя.

Архитектура сети показана на рисунке 3.

Архитектура : МП 2:2-3-1:1 , N = 1
Производительность обуч. = 0,954417 , Контр. производительность = 0,956686 , Тест. производительность = 0,959695

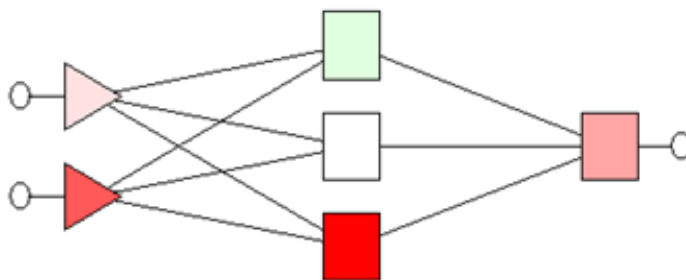


Рисунок 3. Архитектура сети для краткосрочного прогнозирования

Таким образом, можно спрогнозировать для пользователя общее количество услуг на будущие периоды на основе значений в предыдущих четырех месяцах, которые были указаны пользователем.

ЛИТЕРАТУРА

1. Луньков А. Д. Интеллектуальный анализ данных / А. Д. Луньков, А. Д. Галушкин, А. В. Харламов. – Учебно-методическое пособие. – Саратов: СГУ. – 96 с.

2. Галушкин А. И. Нейронные сети / А. И. Галушкин. – Горячая линия Телеком, 2010. – 496 с.

3. Рутковская Д. Нейронные сети, генетические алгоритмы и нечеткие системы / Д. Рутковская, М. Пилиньский, Л. Рутковский – М: Горячая линия Телеком, 2013. – 384 с.

4. Барсегян А. А. Технологии анализа данных: Data Mining, VisualMining, TextMining, OLAP./ А. А. Барсегян, М. С. Куприянов, В. В. Степаненко, И. И.Холод – СПб.: БХВ-Петербург. – 2007.

5. Львович И. Я. Основы информатики / И. Я. Львович, Ю. П. Преображенский, В. В. Ермолова // Воронеж. – 2014. – 339 с.

6. Цепковская Т. А. Характеристики платформ для ускорения вычислений / Т. А. Цепковская // Современные инструментальные системы, информационные технологии и инновации. Сборник научных трудов XV-ой Международной научно-практической конференции. – 2020. – С. 325-328.

7. Преображенский Ю. П. Некоторые проблемы автоматизации процессов / Ю. П. Преображенский В сборнике: Техника

и технологии: пути инновационного развития. Сборник научных трудов 8-й Международной научно-практической конференции. Юго-Западный государственный университет. – 2019. – С. 62-64.

8. Попова Е. В. Исследование алгоритма dispose-паттерна при принятии решений управления памятью в клиент-компонентной модели .NET / Е. В. Попова // Моделирование, оптимизация и информационные технологии. –2020. – Т. 8. – № 3 (30). – С. 30-31.

THE USE OF DATA MINING TECHNOLOGY FOR DATA FORECASTING ALGORITHM ON THE PORTAL OF PUBLIC SERVICES

©2020 A. P. Preobrazhensky, V. O. Logvina

Voronezh Institute of High Technologies (Voronezh, Russia)

Information received on the portal of public services requires the use of special methods for its processing and subsequent analysis. Data Mining technologies are used to implement complex tasks, through the solutions of which the forecasting of indicators is carried out on the portal of public services.

Keywords. Data Mining, forecasting problem, neural network, forecasting period, forecasting horizon, forecasting interval, neural network forecasting algorithm