

Прогнозирование спроса розничных сетей продаж

Добролюбова Ольга, 622 группа

18 октября 2018 г.

1 Основные сведения

Актуальность исследования

Проблематика прогнозирования спроса розничных сетей состоит в следующем:

- Стандартные алгоритмы не справляются с моделированием спроса из-за высокой волатильности спроса, взаимозависимости между товарами и другими проблемами, которые связаны со специфичностью продаж в этой области.
- Компания имеет данные лишь за небольшой промежуток времени по причине того, что, к примеру, компания недавно вышла на рынок или же попросту не проводила сохранение данных за большой промежуток времени;
- Характеристики товара предоставлены в ограниченном виде (часто компании не желают раскрывать внутреннюю информацию аналитику или, по причине непонимания работы алгоритма прогноза, компания не сохраняла необходимую для прогнозирования информацию).

Цель

Целью работы является конструирование наилучшей модели прогнозирования спроса на товары розничных сетей магазинов.

Гипотеза

Гипотеза состоит в том, что для всех розничных сетей можно подобрать единую модель прогнозирования спроса, которая будет давать наилучший прогноз вне зависимости от особенностей розничных сетей, а также от экономической ситуации в стране.

Альтернативная гипотеза состоит в том, что особенности каждой розничной сети могут требовать индивидуального подхода к анализу данных.

Задачи

Задачи исследования состоят в следующем:

- Выявление наиболее подходящих моделей прогнозирования, а также метрик оценки прогноза
- Построение прогноза на основе эмпирических данных
- Подсчет ошибок прогноза и сравнение качества моделей на основе полученных ошибок прогнозов, с учетом макроэкономических показателей и особенностей розничной сети

- Корректировка размера периода прогнозирования и количества данных для обучения для нахождения наилучшего соотношения для каждой модели
- Выбор наилучшей модели на основе ошибок прогноза и максимального наилучшего периода прогнозирования
- Конструирование модели, которая даст лучший результат
- Выявление факторов оказывающих наибольшее влияние на качество прогноза путем исключения их из лучшей модели и сравнения качества прогноза с изначальной моделью

2 Данные

Эмпирические данные взяты с сайта Kaggle¹. Для исследования используются данные следующих розничных сетей продаж:

- Wal-Mart Stores, Inc - крупнейшая в мире американская розничная сеть продаж, занимающаяся продажами продовольственных и промышленных товаров;
- Rossman - европейский магазин парфюмерно-галантерейных и аптекарских товаров.

Таблица 1: Имеющиеся данные для Walmart и Rossman.

Магазин	Walmart	Rossman
Период	05.02.2010 - 01.11.2012	01.01.2013 - 31.07.2015
Индикаторы	Store Date Dept Weekly_Sales IsHoliday Temperature Fuel_Price CPI Unemployment	Store Date Sales Customers Open StateHoliday SchoolHoliday StoreType Assortment CompetitionDistance CompetitionOpenSince Promo Promo2 Promo2Since PromoInterval

¹<https://www.kaggle.com>

Таблица 2: Описание индикаторов.

Assortment	тип ассортимента: базовый/расширенный/дополнительный
CompetitionDistance	количество километров до ближайшего магазина конкурентов
CompetitionOpenSince	дата, когда был открыт ближайший магазин конкурентов
CPI	индекс ИПЦ
Customers	количество клиентов за каждый день
Date	дата
Dept	идентификатор отдела
Fuel_Price	стоимость бензина в регионе
IsHoliday	индикатор является ли неделя праздничной
Open	индиктор открытого/закрытого магазина
Promo	индикатор акций в магазине
PromoInterval	месяцы начала глобальных акций Promo2
Promo2	глобальная акция для всех магазинов определенного типа
Promo2Since	дата начала участия магазина в глобальной акции Promo2
Sales	продажи за каждый день продаж
SchoolHoliday	индиктор открытых/закрытых государственных школ
StateHoliday	государственный праздник: праздничный день/Пасха/Рождество
Store	уникальный идентификатор магазина
StoreType	индикатор одного из 4-х типов магазина
Temperature	средняя температура в регионе
Unemployment	уровень безработицы
Weekly_Sales	недельные продажи

3 Модели

В работах, в которых изучается проблематика прогнозирования спроса, сравнивается либо прогнозы деревьев, либо ARIMA с прогнозами рекуррентных нейронных сетей.

Среди деревьев отмечается преимущества прогноза Random Forestt[9], в случае сравнения моделей ARIMA и рекуррентных нейронных сетей, в ряде работ отмечается преимущество нейронных сетей[15][4]. Однако данное преимущество может быть связано с размерностью тренировочной и тестовой выборки, а также с отсутствием правильных преобразований для модели ARIMA[15][3][13].

Таблица 3: Преимущества и недостатки моделей для выполнения задач.

Модель	Плюсы	Минусы
Decision Tree	Нет тонкой настройки	Проблема «переобучения»
Random Forest		Большой вес параметрам с подкатегориями
Gradient Boosted Trees		
LSTM	Устойчивость к временным лагам ²	Чувствительность к преобразованию данных «Угадывание» параметров
MLP	«Гибкость»	
SARIMA	Процедуры подбора наилучших параметров	Условие стационарности для ряда Чувствительность к преобразованию данных

Таблица 4: Метрики для проверки качества моделей.

Метрика	Расшифровка	Формула
MSE	Среднее квадратичное отклонение	$\frac{\sum_{i=1}^n y_i - x_i}{n}$
RMSE	Корень из MSE	$\sqrt{\frac{\sum_{i=1}^n y_i - x_i}{n}}$
MAD	Среднее абсолютное отклонение	$\sup_{1 \leq i \leq n} y_i - x_i $

К недостаткам нейронных сетей и деревьев следует отнести и проблему «черного ящика» - полученные результаты прогноза не поддаются интерпретации.

4 Текущие результаты для деревьев и нейронной сети

Приведены результаты оценки для магазина Walmart модели со всеми индикаторами для 1 магазина и одного типа товара (Store = 1, Dept = 5):

$$\begin{aligned} Weekly_Sales_i = & \alpha_1 Temperature_i + \alpha_2 Fuel_Price_i + \alpha_3 CPI_i \\ & + \alpha_4 Unemployment_i + \alpha_5 IsHoliday_i + \varepsilon_i, \end{aligned} \quad (1)$$

и для модели с лагом:

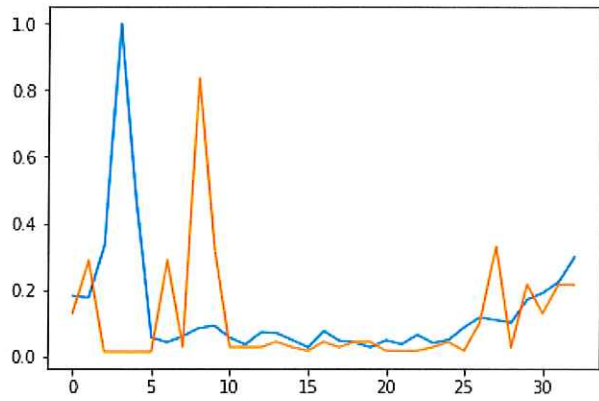
$$\begin{aligned} Weekly_Sales_i = & \alpha_0 Weekly_Sales_{i-1} + \alpha_1 Temperature_i + \alpha_2 Fuel_Price_i \\ & + \alpha_3 CPI_i + \alpha_4 Unemployment_i + \alpha_5 IsHoliday_i + \varepsilon_i. \end{aligned} \quad (2)$$

Таблица 5: Метрики моделей.

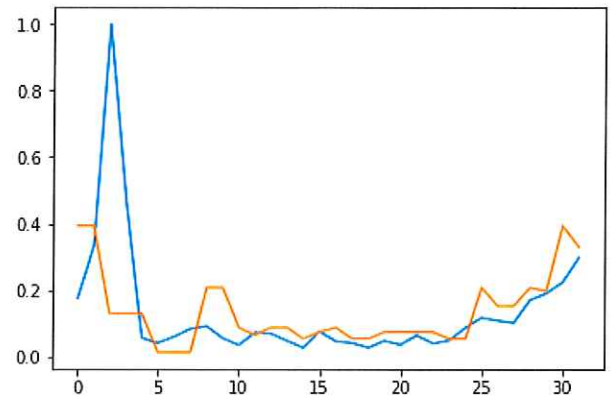
Модель	MSE	RMSE	MAD	Модель с лагом	MSE	RMSE	MAD
Decision Tree	0.063	0.251	2.199	Decision Tree	0.032	0.179	2.161
Random Forest	0.03	0.174	0.821	Random Forest	0.018	0.137	0.669
Gradient Boosted	0.031	0.176	0.741	Gradient Boosted	0.02	0.144	0.697
LSTM	0.03	0.174	0.523	LSTM	0.02	0.144	0.322
MLP	0.028	0.168	0.376	MLP	0.034	0.184	0.263

На графиках представлены нормализованные данные действительных значений (синяя линия) и спрогнозированных (оранжевая линия). Прогноз осуществлялся на основании 110 недель на 33 недели вперед.

²по сравнению с другими нейронными сетями.

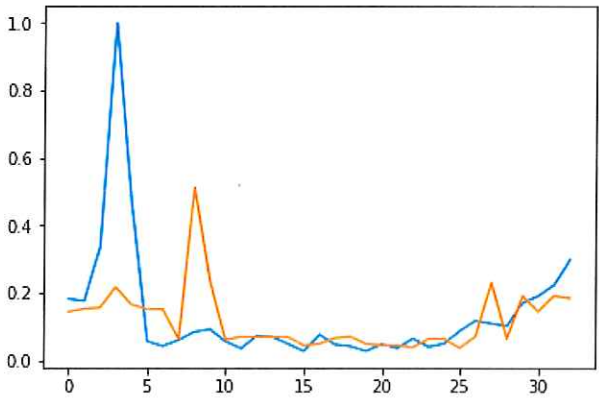


а) Модель без лага

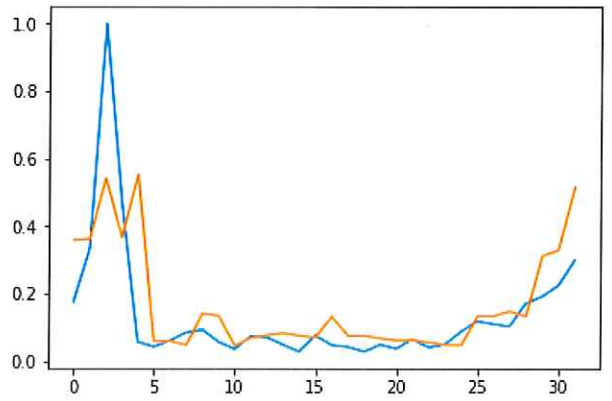


б) Модель с лагом

Рис. 1: Модель Decision Tree.

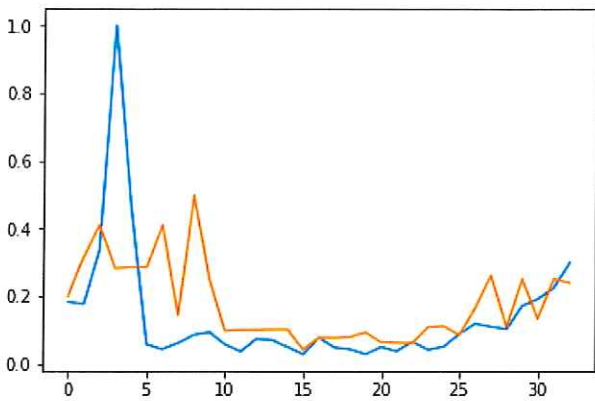


а) Модель без лага

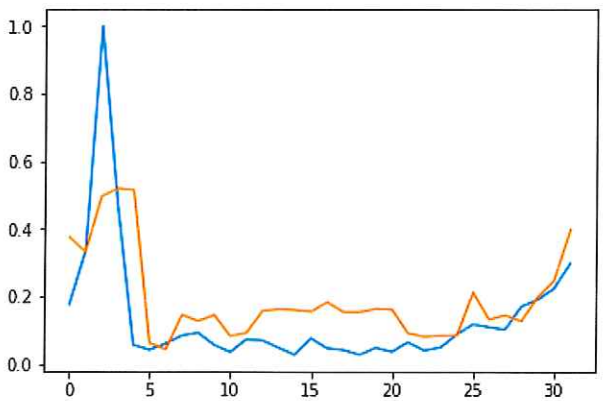


б) Модель с лагом

Рис. 2: Модель Random Forest.



а) Модель без лага



б) Модель с лагом

Рис. 3: Модель Gradient Boosted.

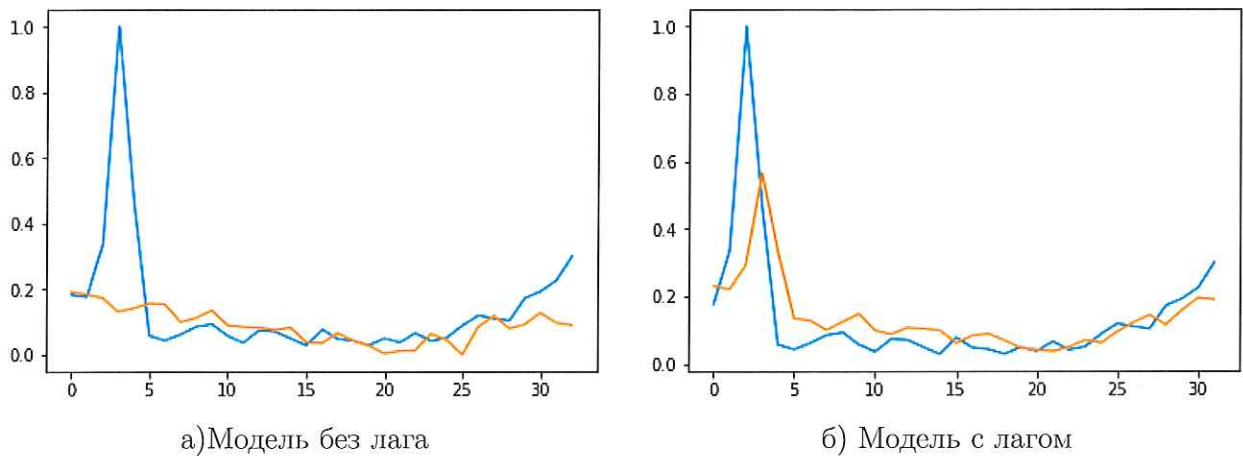


Рис. 4: Модель LSTM.

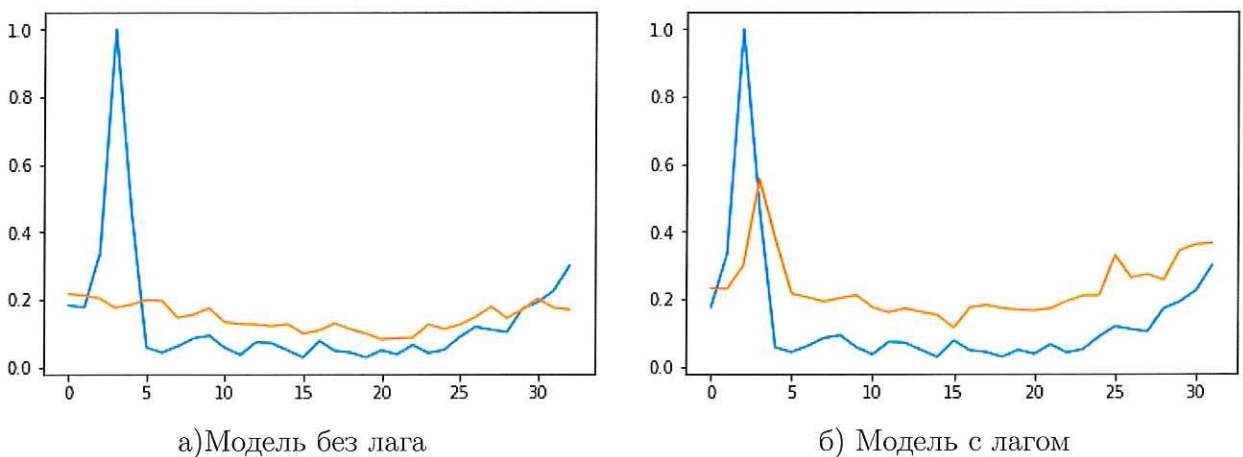


Рис. 5: Модель MLP.

Список литературы

- [1] Классификация многомерных наблюдений [Книга] / авт. Айвазян С.А., Бежаева З.И. и Староверов О.В. - М.: Статистика, 1974.
- [2] Временные ряды [Книга] / авт. Кендалл М. - М.: Финансы и статистика, 1981. Выбор модели и парадоксы прогнозирования [Журнал] / авт. Иццоки О. - Квантиль, 2006 г.
- [3] A hybrid SARIMA wavelet transform method for sales forecasting [В Интернете] / авт. Choi Tsan-Ming, Yu Yong, Au Kin-Fan. // Science Direct. - 2011 г. - <https://www.sciencedirect.com/science/article/pii/S0167923610002356>.
- [4] A Solution to Forecast Demand Using Long Short-Term Memory Recurrent Neural Networks for Time Series Forecasting [Журнал] / авт. Goyal Adarsh [и др.].
- [5] An ARIMA Supply Chain Model [Журнал] / авт. Gilber Kenneth.. - Management Science, 2005 г.
- [6] An LSTM-Based Dynamic Customer Model for Fashion Recommendation [В Интернете] / авт. Heinz Sebastian, Bracher Christian, Vollgraf Roland. // Cornell University Library. - 2017 г. - <https://arxiv.org/abs/1708.07347>.
- [7] Business Cycles, Indicators and Forecasting [Журнал] / авт. Stock J. H., Watson M. W. - 1993 : University of Chicago Press.

- [8] Deep Recurrent Neural Networks for Product Attribute Extraction in eCommerce [Журнал] / авт. Majumder Bodhisattwa Prasad [и др.]. - ArXiv, 2018 г.
- [9] Demand forecasting with high dimensional data: the case of SKU retail sales forecasting with intra-and inter-category promotional information [Журнал] / авт. Ma Shaohui, Fildes Robert, Huang Tao. - European Journal of Operational Research, 2015 г.
- [10] Demand Prediction Using Machine Learning Methods and Stacked Generalization [Журнал] / авт. Tugay Resul, Oguducu Sule Gunduz. - ResearchGate, 2017 г.
- [11] Forecasting Retail Food Prices under Current Conditions [Журнал] / авт. Clauson Annette L. - American Journal of Agricultural Economics, 1997 г.
- [12] Linear Adjustment Costs and Seasonal Labor Demand: Evidence from Retail Trade Firms [Журнал] / авт. Anderson Patricia M. - The Quarterly Journal of Economics, 1993 г.
- [13] Oral-Care Goods Sales Forecasting Using Artificial Neural Network Model [В Интернете] / авт. Vhatkar Sangeeta, Dias Jessica. // Science Direct. - 2016 г. - <https://www.sciencedirect.com/science/article/pii/S1877050916001629>.
- [14] Python и машинное обучение [Книга] / авт. Рашка Себастьян.. - ДМК Пресс, 2017.
- [15] Sales Forecasting with Partial Recurrent Neural Networks: Empirical Insights and Benchmarking Results [В Интернете] / авт. Müller-Navarra Moritz, Lessmann Stefan, Voss Stefan. // IEEE Computer Society. - 2015 г. - <https://www.computer.org/csdl/proceedings/hicss/2015/7367/00/7367b108.pdf>.
- [16] Short-Run Predictive Models for Retail Meat Sales [Журнал] / авт. Marion B. W., Walker F. E.. - American Journal of Agricultural Economics, 1978 г.
- [17] The Leader in Global Forecasting and Quantitative Analysis. Overview [Журнал] / авт. Oxford Economics. - Oxford: Oxford Economics, 2014 г.
- [18] The Locational Impact of Wal-Mart Entrance: A Panel Study of the Retail Trade Sector in West Virginia [Журнал] / авт. Hicks Michael J., Wilburn Kristy. - Center for Business and Economic Research Lewis College of Business, 1999 г.
- [19] Анализ временных рядов [Журнал] / авт. Канторович Г.Г. - Экономический журнал Высшей школы экономики, 2002 г.
- [20] Нейронные сети. Полный курс [Книга] / авт. Хайкин С. - Вильямс, 2018.
- [21] Методы эконометрики: учебник/С.А. Айвазян [Книга] / авт. Айвазян С.А. - М.: Магистр: ИНФРА-М, 2010.