

Федеральное государственное автономное образовательное учреждение
высшего образования «Национальный исследовательский университет
«Высшая школа экономики»

Факультет компьютерных наук
Основная образовательная программа
Прикладная математика и информатика

КУРСОВАЯ РАБОТА
ИССЛЕДОВАТЕЛЬСКИЙ ПРОЕКТ НА ТЕМУ
"ПРОГНОЗИРОВАНИЕ РОЗНИЧНОГО СПРОСА
НЕЙРОСЕТЕВЫМИ МЕТОДАМИ С ИСПОЛЬЗОВАНИЕМ
КЛАСТЕРИЗАЦИИ ТОРГОВЫХ ТОЧЕК"

Выполнил студент группы 192, 3 курса,
Тимонина Мария Александровна

Руководитель КР:
канд. техн. наук, доцент Департамента анализа данных и
искусственного интеллекта ФКН НИУ ВШЭ
Игнатов Дмитрий Игоревич

Москва 2022

Содержание

Аннотация	2
1 Введение	3
2 Обзор предметной области	5
3 Методология исследования	8
3.1 Устойчивость кластеризации магазинов	8
3.2 Прогнозирование товарного спроса	9
4 Эксперименты и результаты	11
4.1 Описание набора данных	11
4.2 Подготовка признаков	12
4.3 Кластеризация торговых точек	13
4.4 Прогнозирование рядов	15
5 Заключение	18

Аннотация

В этой работе рассматриваются способы комбинирования методов машинного обучения и нейросетевых подходов для прогнозирования спроса в сетевых магазинах по данным товарных чеков. Проверяется гипотеза о возможности улучшить результаты работы предсказательных моделей с помощью предварительного разбиения магазинов на кластеры. Проводится сравнительный анализ подходов к осуществлению кластеризации магазинов в рамках единой торговой сети с точки зрения устойчивости и объяснимости состава кластеров. В результате обучается модель, способная достаточно эффективно предсказывать спрос на популярные товары по данным сети-заказчика.

Abstract

This paper discusses how to combine machine learning and neural network approaches to predict demand on retail stores' products based on sales receipts. We test the hypothesis that it is possible to improve the results of predictive models by pre-partitioning stores into clusters. A comparative analysis of store clustering approaches in terms of cluster stability and explanatory power is conducted. As a result, a model able to predict the demand for popular products quite effectively is trained on a real 3-year data set of receipts.

Перечень ключевых слов

behavioral retail clustering

pairwise cluster stability

demand prediction, lstm

1 Введение

Розничные сети накапливают миллионы записей о работе своих магазинов каждый день, сохраняя информацию в формате чеков. Анализ покупательского спроса позволяет делать предположения об актуальных потребностях клиентов и трендах. Торговые предприятия широко используют классические модели машинного обучения на таких данных для снижения затрат на логистику, оптимизации структуры товарных каталогов, построения персональных рекомендаций.

В этой работе рассматриваются две связанные друг с другом задачи клиентской аналитики. Сначала различные методы кластеризации анализируются в применении к сложным объектам - описаниям торговых точек через историю покупок в них. Подходы сравниваются по внутренним и внешним метрикам, а также устойчивости полученных кластеров к сдвигу по времени. Далее рассматриваются способы прогнозирования спроса на отдельные товары в магазинах методами классического машинного обучения и нейросетей. Тестируется возможность улучшить качество предсказания с помощью предварительной кластеризации точек и обучения моделей для каждого из кластеров отдельно.

Объединение разных магазинов одного бренда в кластеры представляет самостоятельный интерес и актуальность для отделов продаж. На основании информации о схожем поведении клиентов в определенных районах или городах, маркетологи могут разрабатывать программы развития этих групп, формировать узконаправленные скидочные программы или объединять товары-комплементы, которые часто покупают вместе в наборы. В данном исследовании кластеризация точек происходит на основе реальной активности, особенностей покупательского спроса (снизу). Пространственные, демографические или административные особенности мест продаж (принаки сверху) используются для анализа качества получаемых кластеров, семантического обоснования.

Сетевые магазины заинтересованы в точном прогнозировании спроса следующих периодов для каждой своей торговой точки. Более точное предсказание с учетом динамики, сезонности и сложных рыночных факторов позволяет рассчитать требуемый объем закупок и оптимальнее распределять его по точкам, избегая дефицита или излишков на складах. При этом развитие нейросетевых подходов значительно частично избавляет исследователя от необходимости непосредственно генерировать признаки и позволяет учесть сложные зависимости в данных через слои скрытых переменных.

Таким образом, цель данного прикладного проекта - построение прогнозов по спросу на продукцию заказчика (сети розничных магазинов) на основании моделей кластеризации и предсказания временных рядов методами машинного и глубинного обучения. Среди задач, этапов работы следует выделить:

- изучение международных практик прогнозирования розничного спроса;
- определение ключевых метрик качества объединения магазинов в кластеры, отражающих специфику задачи;
- извлечение качественных признаков из необработанного набора чеков;
- тестирование известных решений и библиотек на новых данных;
- доработку основного решения за счёт комбинирования моделей и блоков, подбора гиперпараметров.

Работа структурирована следующим образом. Обзор предметной области приводится в Разделе 2. Далее следует описание набора данных и последовательность шагов по их предварительной подготовке, извлечению признаков. В Разделе 4 вводится метрика стабильности кластеризации, перечисляются рассматриваемые архитектуры предсказательных моделей. Раздел 5 посвящен описанию результатов эксперимента на реальном наборе данных сети строительных магазинов.

Код работы доступ в репозитории по ссылке:

2 Обзор предметной области

Большинство исследований на данных о покупках из товарных чеков концентрируются на решении задач сегментации клиентов и предсказания индивидуального спроса для улучшения рекомендаций. Среди покупателей каждого магазина есть представители разных клиентских групп. Соотношение, в котором эти группы делят все множество посетителей, уже дает нам простую модель признакового описания торговой точки. В то же время, подходы к выделению групп похожих покупателей можно попробовать напрямую применить к задаче кластеризации магазинов, определяя торговую точку как одного клиента в таких моделях.

В статье [Agarwal et al. \(2019\)](#) авторы проводят сравнительный анализ методов кластеризации для данных ритейла. Цель исследования - предложить устойчивый способ объединения магазинов в группы похожих как инструмент продвинутой аналитики для разработки регионо-центричной политики торговых сетей. В работе рассматриваются четыре техники мягкого выделения кластеров: самоорганизующаяся карта Кохонена (SOM), гауссово моделирование смеси (GMM), иерархическая кластеризация и метод нечетких С-средних (Fuzzy C-means). Устойчивость модели измеряется по параметру квартальной миграции, перемещения объектов выборки между кластерами на данных европейской сети продуктов питания, работающей в четырех регионах. В результате эксперимента ни один из методов не показывает лучший результат для всех рынков. Авторы дают рекомендацию подбирать метод кластеризации по набору данных, исходя из доступного набора признаков.

Иной подход к определению кластерной устойчивости описан в работе [Kuncheva and Vetrov \(2006\)](#). Рассматривается зависимость ансамблей из моделей К-средних от параметров случайной инициализации (выбора стартовых точек) и количества кластеров. В статье предлагается новый комбинированный индекс стабильности разбиения, который представляет собой сумму попарных стабильностей моделей и всего ансамбля. Парная стабильность

определяется как адаптированный индекс Рэнда (ориг. the adjusted Rand Index) между парами кластеров в ансамбле, усредненный по всем парам. Непарная стабильность определяется как энтропия матрицы консенсуса ансамбля. Экспериментальное сравнение проводится по десяти синтетическим и десяти реальным наборам данных. Результаты показали, что ансамбли в целом более стабильны для больших k . Взаимосвязь между устойчивостью и точностью для фиксированного k и случайной инициализации сильно зависит от набора данных. В текущей работе мы будем выбирать метод кластеризации, сравнивая их с помощью адаптированного индекса Рэнда.

Помимо классического машинного обучения в клиентской аналитике находят применение и методы обработки естественного языка. Использование моделей векторных представлений на данных чеков становится естественным, если набор товаров в одном чеке сопоставить последовательности слов в предложении. Этот подход был предложен в статье [Barkan and Koenigstein \(2015\)](#) и получил название Customer2Vec. Векторные представления для клиентов могут применяться как в качестве векторов признаков в моделях регрессии и классификации, так и для выделения семантики, поскольку Word2Vec способен выучивать смысловые отношения между объектами (man - woman, king - queen). Авторы статьи предполагают статичность среды, в которой предметы, входящие в один набор, считаются похожими, независимо от того, в каком порядке или времени они были куплены клиентом. Каждая пара предметов, которые находятся в одном и том же наборе, рассматривается как пара соседних слов, положительный пример. Таким образом, векторные представления данных в пространстве фиксированной размерности (как правило от 50 до 500) становятся удобным набором объектов для алгоритмов кластеризации. Однако, следует учитывать, что при переходе от последовательностей к множествам - неупорядоченному набору всех покупок клиента (в нашей постановке продаж магазина) - теряется пространственная и временная информация. Отдельный интерес представляют методы, комбинирующие алгоритмы кластеризации с моделями обучения с учителем. Так, в статье [Seleznev](#)

et al. (2018) (ссылается обзор [Katsov \(2020\)](#)) описывается схема совмещения Word2Vec и регрессионной модели, предсказывающей вектор бизнес-метрик. Левая и правая сети структурно независимы, но обучаются одновременно. Веса матрицы эмбедингов разделяются или копируются между сетями после каждой эпохи обучения. Таким образом, нейросетевые методы обработки текстов могут работать согласованно с задачей максимизации прибыли, увеличения конверсии ([Dogan et al. \(2018\)](#)).

Поведение потребителей в электронной коммерции можно описать с помощью сессий, последовательностей взаимодействий с интернет-магазином. В статье ([Żoła and Romański \(2017\)](#)) показано, что рекуррентные нейронные сети (РНС) являются удобным инструментом для моделирования и прогнозирования поведения потребителей. Эти модели работают с последовательностями различной не фиксированной длины и теоретически могут извлекать сколь угодно глубокие зависимости. В предлагаемом методе РНС применяются непосредственно к сериям сессий. Рассматриваются и тестируются на реальных данных крупной торговой онлайн-площадки два типа сетей. Первый тип принимает на вход последовательность фактов покупки, второй - описаний клиентских сессий от момента входа до завершения процесса выбора. Авторы показывают, что с помощью рекуррентных нейронных сетей можно не только предсказывать ожидаемую активность покупателей в следующих периодах, но также пользоваться их хорошей объясняющей способностью для интерпретации выводов модели.

В данной работе будет исследован способ улучшения качества прогноза потоварных продаж для разных товарных точек с помощью предварительной кластеризации. Для представления магазина как объекта выборки рассматривается два подхода - через агрегированные статистики товаров в чеках и через doc2vec эмбединги магазинов. Лучший метод кластеризации в каждом из подходов выбирается с помощью адаптированного индекса Рэнда. Далее предсказания с помощью рекуррентной модели LSTM сравниваются с базовой моделью для временных рядов prophet и подходов машинного обучения.

3 Методология исследования

3.1 Устойчивость кластеризации магазинов

Мы применяем попарный подход для измерения устойчивости разбиения сети магазинов на кластеры P_1, \dots, P_m в квартальном и годовом окне. Рассматриваются пары кластеров, поэтому этот индекс не следует путать с попарными метриками на элементах кластеризации.

Для каждой из $\frac{m(m-1)}{2}$ пар кластеров рассчитывается их мера схожести. Индекс устойчивости определяется как усредненное по всем таким парам значение схожести. Обозначим за $S(P_i, P_j)$ меру согласованности между двумя кластерами разных разбиений P_i и P_j . Индекс принимает значения от 0 до 1. Тогда индекс попарной устойчивости S_p это:

$$S_p = \frac{2}{m(m-1)} \sum_{1 \leq i, j \leq m, i < j} S(P_i, P_j) \quad (1)$$

Если разбиения совпадают (даже если нумерация кластеров изменена) степень их схожести будет равной 1. Если же кластеры независимы, то ожидаемое значение будет 0 для любого количества кластеров в разбиении.

Обозначим за A и B разбиения набора Z из N объектов на группы кластеров k_A и k_B соответственно, a_i - число объектов в i -ом кластере A , b_j - число объектов в j -ом кластере B . Будем считать A эталонным разбиением. Тогда введем также a_{ij} - количество объектов в пересечении i -ого кластера A , j -ого кластера B . Эта мера симметрична, она не зависит от перестановок меток. Адаптированный индекс Рэнда (the adjusted Rand Index, ARI) рассчитывается как

$$AR(A, B) = \frac{\sum_{i=1}^{k_A} \sum_{j=1}^{k_B} C_{a_{ij}}^2 - X_3}{\frac{1}{2}(X_1 + X_2) - X_3}, \quad (2)$$

$$X_1 = \sum_{i=1}^{k_A} C_{a_i}^2; X_2 = \sum_{j=1}^{k_B} C_{b_j}^2; X_3 = \frac{2}{m(m-1)} X_1 X_2$$

3.2 Прогнозирование товарного спроса

Исследуемые временные ряды формируются из данных чеков о продажах путем суммирования всех фактов продажи товара с учетом купленного количества в условных единицах. Эмпирически обнаружено, что временные ряды количества продаж большинства товаров обладают существенной недельной и годовой сезонной компонентой. Ее предварительное выделение позволяет улучшить качество рассматриваемых моделей.

Насколько ошибается модель в своих прогнозах, будем измерять через среднюю абсолютную ошибку в процентах (mean absolute percentage error).

$$MAPE = mean \left(\frac{y_i - \hat{y}_i}{y_i} \right), MAE = mean(y_i - \hat{y}_i)$$

В качестве базовой для предсказания значений временного ряда используется модель библиотеки **prophet**, предложенная в статье [Taylor and B. \(2017\)](#). Она признается достаточно точной, с ней часто сравниваются работы в области. Модель (3) задается следующим образом:

$$y_t = g_t + s_t + h_t + \varepsilon_t, \quad (3)$$

где слагаемые - это компоненты тренда (кусочно-линейная модель, модель насыщенного роста), сезонностей (через ряды Фурье), ошибки. Компонента h_t определяется через скалярное произведение векторов индикаторов выходного дня и изменений ряда в соответствующий выходной.

Другим перспективным подходом к предсказанию временных рядов является совмещение методов эконометрического анализа временных рядов и классического машинного обучения. В экспериментальной части работы модели градиентного бустинга (Extreme Gradient Boosting, XGBoost) и случайного леса (RandomForest) обучаются на наборе признаков соответствующих временных рядов. Они извлекаются с помощью Python библиотеки **tsfresh**.

Наконец, в этом исследовании применяется модель рекуррентной нейронной сети Долгая краткосрочная память (LSTM). Она была предложена в статье [Hochreiter and Schmidhuber \(1997\)](#). Обучение рекуррентных сетей в силу их последовательной структуры осложняется проблемами с обратным распространением ошибки. Из-за накопления множителей на пути назад первые слои сети могут получать недостаточный сигнал для обучения. Эта проблема решается в данной архитектуре через добавление более коротких связей между дальними нейронами.

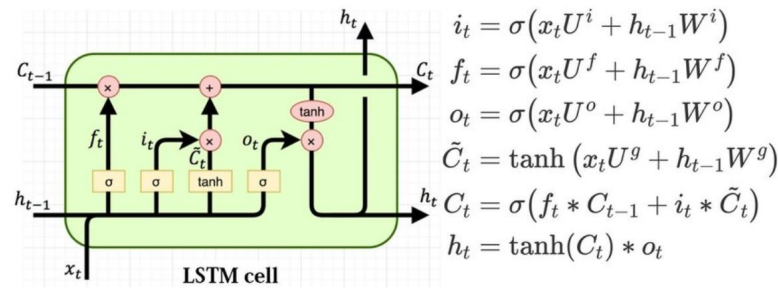


Рис. 3.1: Структура LSTM блока и уравнения коротких путей в нем

Если градиент взрывается, то применяют операцию обрезки градиента. Такую ситуацию можно отследить по скачку ошибки на обучающей выборке, превышении нормой градиента установленного порога.

4 Эксперименты и результаты

Программный код к проекту в виде Colab-Notebooks размещен в открытом репозитории на Github: <https://github.com/timmMariya/dl-sales-prediction>

4.1 Описание набора данных

Входной поток данных о клиентской активности представляет из себя полный набор чеков крупной сети строительных магазинов за три года (2018-2020). Каждая запись о покупке очередного товара содержит поля:

- Номер чека в единой системе
- Идентификатор товара в каталоге
- Идентификатор магазина
- Дата совершения покупки
- Единицы измерения (штуки, кг, литры)
- Количество товара в одной упаковке
- Количество купленных упаковок
- Выручка от сделки

База содержит 45605040 записей из 18,6 млн чеков по 500 ключевым для сети товарам с наивысшим спросом. Также был предоставлен доступ к базе с информацией о городе, районе, площади торгового зала каждого магазина сети и идентификаторах ближайших соседей.

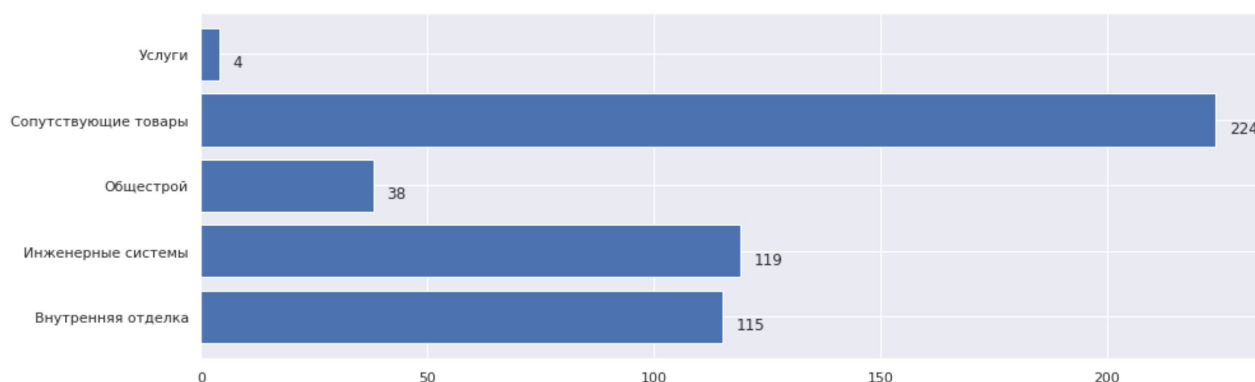


Рис. 4.1: Распределение товаров по категориям каталога верхнего уровня

Также доступен каталог сети магазинов, в котором по номеру товара можно получить информацию о полном названии товара в системе, единицах измерения (штуки, кг, литры), месте товара в иерархии каталога (4 уровня). Предварительный анализ данных показал, что товары из выборки распределяются между основными категориями в не пропорциональных частях (Рисунок 4.1).

4.2 Подготовка признаков

После удаления дубликатов записей в базе осталась информация из 43 миллионов чеков. Отследить и удалить из рассмотрения записи с ошибками человеческого фактора на кассе возможно только, если они содержат противоречивую информацию. К таким маркерам можно отнести отрицательное значение количества проданных единиц, отрицательную выручку или ситуации продажи по цене, на порядок отличающейся от нормальной для конкретного товара.

Для кластеризации магазинов были построены два вида признаковых описаний. Из данных чеков путем агрегирования были получены значения количества чеков, в которых фигурировал товар, количества проданного в условных единицах для товара, выручки от продаж каждого товара отдельно для каждого магазина. Весь анализируемый период был разбит на пять последовательных полугодовых блоков, между которыми измерялась степень попарной устойчивости кластеров.

Разведочный анализ данных о переменных показал, что данные о продажах распределены логнормально, поэтому количественные признаки были логарифмированы и затем отнормированы.

4.3 Кластеризация торговых точек

Эмпирическим путем было получено, что в пространстве потоварных признаков магазинов кластеры имеют существенно различающуюся плотность точек, поэтому плотностные алгоритмы типа DBSCAN дали кластеризацию низкого качества. Напротив, К-Means показал хорошие результаты, опишем зависимость целевых метрик от параметров подробнее.

Количество кластеров

В качестве метрик для подбора оптимального количества групп использовались стандартные индексы Силуэта и Данна. Как известно, Silhouette score измеряется в диапазоне $[-1, 1]$. Показатель, близкий к 1, означает, что кластеры очень плотные и хорошо разделены. Значение 0 означает, что кластеры перекрываются. Оценка менее 0 - данные, принадлежащие кластерам, могут быть ошибочными. Индекс Данна рассчитывается как отношение наименьшего межкластерного расстояния к наибольшему внутрикластерному расстоянию. Высокий показатель означает лучшую кластеризацию, так как наблюдения в каждом кластере расположены ближе друг к другу, в то время как сами кластеры находятся дальше друг от друга. Как видно на примере (Рисунок 4.2), значения индекса силуэта близки к единице, что говорит о качественной разделимости выборки. Наилучшие показатели достигаются при 4-6 кластерах. Будем только для этих значений проверять степень устойчивости разбиений.

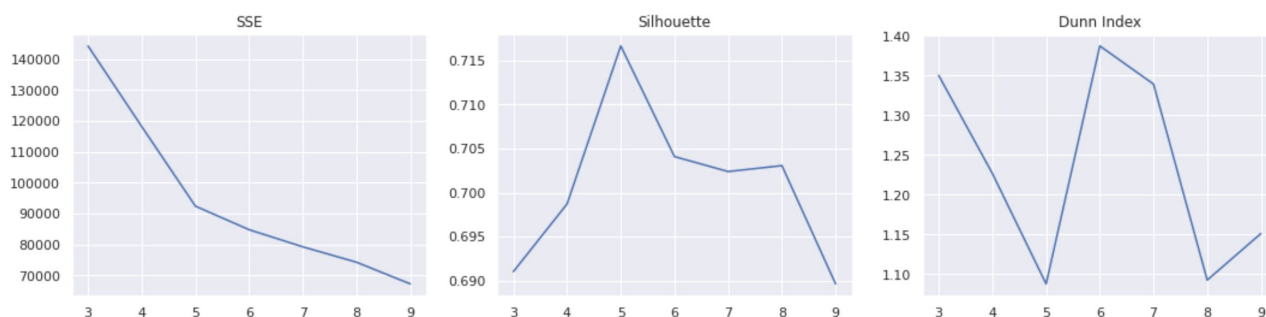


Рис. 4.2: Подбор числа кластеров по методу локтя. Пример

Случайная инициализация

Также удалось пронаблюдать изменчивость разбиения при запусках кластеризации K-Means с одинаковым количеством кластеров. Чтобы сгладить эффект зависимости, конечным решением стало использование ансамбля из 20 моделей с одинаковым количеством кластеров и голосованием.

Устойчивость разбиений

Адаптированный индекс Рэнда, как показатель устойчивости полученных разбиений, вычислялся для последовательных полугодий 2018-2020 годов и для пары первого и последнего блоков данных. Если в первые периоды устойчивость кластеров достаточно велика, то далее с началом пандемии можно наблюдать как возрастает расстояние между разбиениями. На примерах отдельных магазинов было установлено, что большая доля изменений вызвана существенными изменениями в режиме работы магазинов, спаде предложения, ограниченном режиме поставок.

	1-2	2-3	3-4	4-5	5-6	1-6
4 clusters	0.887851	0.777276	0.828571	0.561548	0.249676	0.091284
5 clusters	1.000000	0.843703	0.770932	0.617521	0.421842	0.123208
6 clusters	0.976040	0.825490	0.812901	0.622597	0.306128	0.163470

Рис. 4.3: Проверка полученных кластеров на устойчивость ко времени

Семантический смысл кластеров

В предлагаемом методе осмысленность получаемого разбиения может быть проанализирована через внешние признаки торговых точек, к примеру город, район, площадь торгового зала, метки ближайших соседей. Был проведен сравнительный анализ полученного разбиения с сегментацией "сверху" по перечисленным признакам. Наибольшая взаимосвязь разбиения наблюдается с параметром размера магазина - площади торгового зала и узкой направленностью (для лесопилок, только кровля, инструменты в центре города, все для дома в черте города). На основе полученной кластеризации, как и предпола-

галось в начале исследования, были составлены рекомендации о перераспределении ряда магазинов в зону ответственности других групп категорийного менеджмента и маркетинга в силу специфики спроса на местах.

4.4 Прогнозирование рядов

Те или иные подходы к предсказанию временных рядов показывают лучшие результаты в зависимости от свойств прогнозируемых переменных. Мы работаем со статистикой продаж различных товаров в их условных единицах. Некоторым категориям товаров свойственна выраженная годовая сезонность.

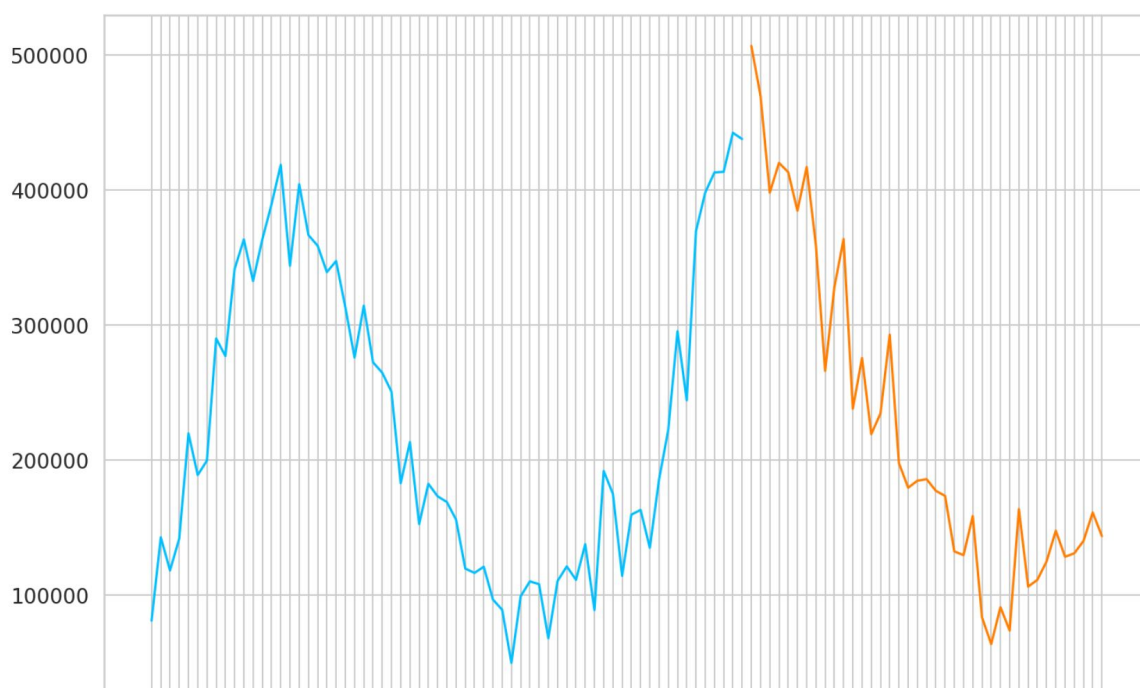


Рис. 4.4: Пример ряда из выборки. Позиция каталога: услуга доставки

В ходе экспериментов строится базовая модель с помощью библиотек `scikit-learn` и `prophet`. Далее градиентный бустинг и случайный лес обучаются на наборе признаков временного ряда, получаемого с помощью библиотеки `tsfresh`. В минимальном наборе признаков `tsfresh` `MinimalFCParameters` рассчитываются сумма значений, медиана, среднее, длина, стандартное отклонение, дисперсия, максимальное и минимальное значения каждого ряда. В расширенных наборах (`EfficientFCParameters`, `ComprehensiveFCParameters`) доступны оконные статистики, автокорреляции, линейный тренд, `p-value` па-

раметра наклона линии тренда в текущем скользящем окне и другие. Обширный набор является чрезмерным, однако благодаря интерпретируемости моделей есть возможность оценить вклад отдельных признаков и с помощью статистических тестов отобрать наиболее важные для конкретной задачи и особенностей набора данных.

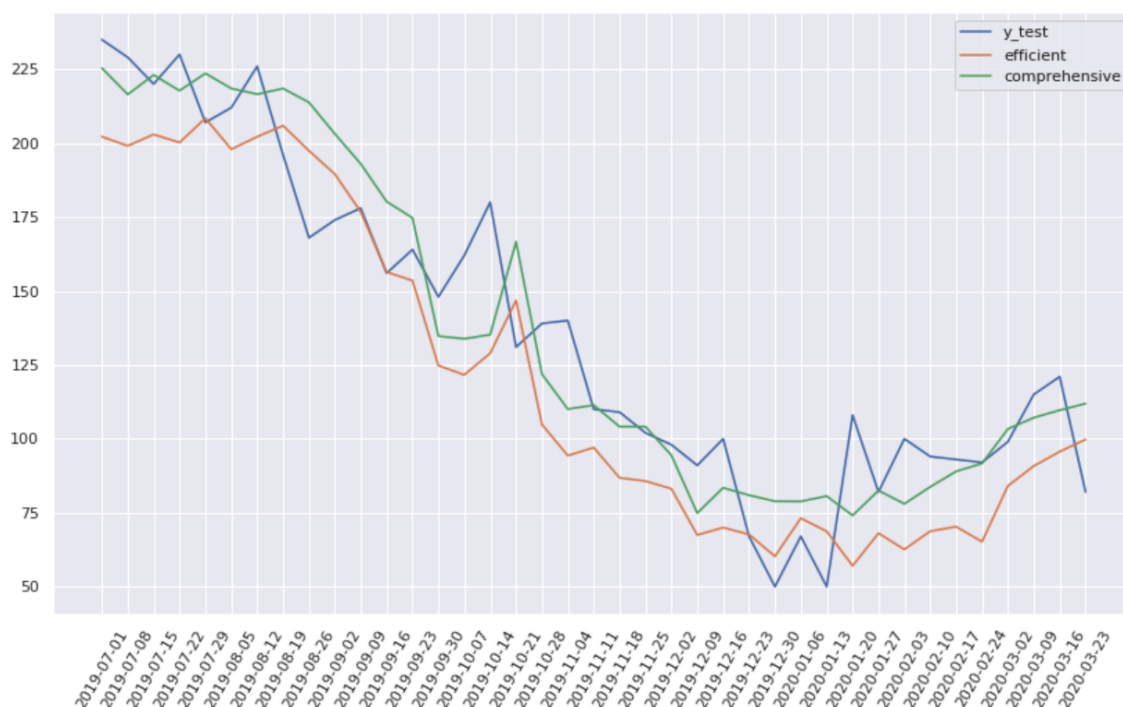


Рис. 4.5: Пример предсказания моделью машинного обучения на признаках, извлеченных tsfresh. Ряд для услуги доставки

Наилучший результат с точки зрения метрики MAPE показала модель градиентного бустинга на 300 деревьев, использующая оконные статистики различной ширины, ряды из годовой и недельной сезонности.

	minimal_features	efficient_features	comprehensive_optimized
MAE	29.623424	12.062095	10.562095
MAPE	0.536221	0.229053	0.129053

Рис. 4.6: Значения метрик качества моделей машинного обучения (лучший результат между случайным лесом и градиентным бустингом в каждом случае) по всей выборке временных рядов

Долгая краткосрочная память

С использованием фреймворка pytorch-lightning и библиотек для глубокого обучения строится модель, состоящая из LSTM блоков и нейронов активации. Рассматривается два подхода - обучение для предсказания спроса на некоторый товар по отрезкам временных рядов всех магазинов и только магазинов из одного кластера.

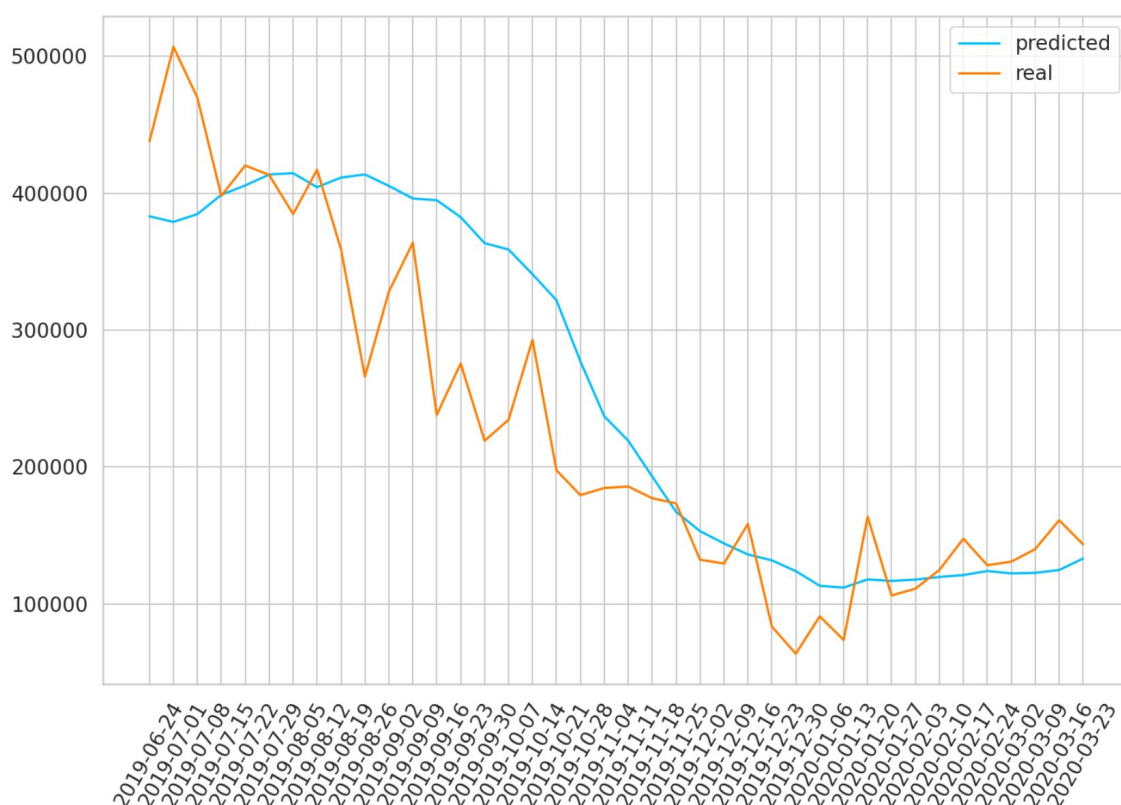


Рис. 4.7: Пример предсказания сетью на базе LSTM. Ряд для услуги доставки

Полученная сеть обучается предсказывать продажи следующего месяца по среднеквадратичной функции потерь, на задачу регрессии. Оказывается, что для кластеров, объединяющих магазины с большим количеством сезонных продаж - летние пики закупок строительных материалов для уличных работ, например - добавление в обучающую выборку фрагментов непохожих рядов ухудшает качество предсказания.

5 Заключение

В работе рассматриваются различные подходы к задаче предсказания продаж в разных магазинах единой торговой сети с применением кластеризации. Для использования рекуррентных нейронных сетей в качестве предсказательной модели требуется объемная обучающая выборка, увеличение которой с помощью исторических данных только по похожим магазинам позволяет достичь лучшего качества, чем при использовании всех имеющихся данных одновременно. Для моделей машинного обучения, основанных на решающих деревьях, напротив, большее разнообразие примеров помогает лучше разделять выборку.

Кластеризация магазинов одной сети представляет самостоятельный интерес для маркетологов, поскольку в торговых точках со схожей структурах спроса можно проводить общие скидочные программы, объединять характерные товары-комплементы в наборы. Поставленный эксперимент показал, что лучшее разбиение с точки зрения устойчивости кластеров дает метод K-Means, опережая плотностные алгоритмы и некоторые другие. Более того, он хорошо интерпретируем. Получившееся разбиение хорошо соотносится с данными о месторасположении магазинов, площади торговых залов.

Перспективным направлением для продолжения работы мы считаем усложнение структуры моделей нейронных сетей, использующих блоки LSTM, более тонкую настройку параметров, масштабирование обучающей выборки.

Список литературы

1. Kanika Agarwal, Mamta Rajnayak, and Prateek Jain. Comparative analysis of store clustering techniques in the retail industry. *International Journal of Applied Engineering Research*, 2019.
2. Oren Barkan and Noam Koenigstein. Item2vec: Neural item embedding for collaborative filtering. *Tel Aviv University*, 2015.
3. Dogan, Aycin, and Bulut. Customer segmentation by using rfm model and clustering methods: A case study in retail industry. *International Journal of Contemporary Economics and Administrative Sciences*, 2018.
4. Sepp Hochreiter and Jurgen Schmidhuber. Long short-term memory. *Neural computation*, 9:1735–80, 12 1997. doi: 10.1162/neco.1997.9.8.1735.
5. Ilya Katsov. Customer2vec: Representation learning for customer analytics and personalization. *TensorHouse AI for Enterprise Operations*, 2020.
6. L.I. Kuncheva and D.P. Vetrov. Evaluation of stability of k-means cluster ensembles with respect to random initialization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1798–1808, 2006. doi: 10.1109/TPAMI.2006.226.
7. Seleznev, Irkhin, and Kantor. Automated extraction of rider’s attributes based on taxi mobile application activity logs. *Yandex Taxi*, 2018.
8. SJ Taylor and Letham B. Forecasting at scale. *PeerJ Preprints* 5:e3190v2, 2017. doi: 10.7287/peerj.preprints.3190v2.
9. K. Żoła and B. Romański. User modeling using lstm networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1), 2017.