

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.383

Клебанов
Дмитрий Александрович

Алгоритмы прогнозирования времени доставки на основе
моделей машинного обучения

АВТОРЕФЕРАТ
на соискание академической степени
магистра

по специальности 1-40 80 04 – Информатика и технологии программирования

Научный руководитель
Сиротко Сергей Иванович
к.т.н., доцент

Минск 2024

КРАТКОЕ ВВЕДЕНИЕ

В постоянно меняющемся мире современной логистики доставка на "последней миле" стала важнейшей областью деятельности, что отражает растущий потребительский спрос на быстрые, надежные и эффективные услуги доставки. Доставка на "последней миле" – заключительный этап на пути товара со склада к порогу потребителя, часто является самым сложным и дорогостоящим этапом цепочки поставок.

Процесс доставки на "последней миле" имеет огромное значение, поскольку напрямую влияет на удовлетворенность и лояльность клиентов. Эффективная доставка "последней миле" может выделить бизнес на конкурентном рынке, обеспечивая своевременные и бесперебойные поставки и тем самым повышая качество обслуживания клиентов в целом.

Доставка "последней миле" сопряжена с многочисленными трудностями. Городские пробки, узкие сроки доставки, непредсказуемые схемы движения транспорта и высокие затраты, связанные с доставкой в разные места, являются существенными препятствиями. Необходимость гибкого подхода к выбору времени и мест доставки для потребителей еще больше усложняет процесс. Также одной из основных финансовых проблем при доставке по "последней миле" является высокая стоимость доставки. Для решения этих проблем предприятия все чаще обращаются к технологическим инновациям.

Для реализации платформы доставки необходимо множество различных компонент. В их числе есть алгоритмы построения и оптимизации маршрутов, которые позволяют эффективно использовать доступное системе курьерское время. Алгоритмы ценообразования позволяют не только контролировать прибыль компании, но и могут использоваться в качестве рычага управления спросом и предложением. Например, повышение цен или предоставление скидок на доставку в непииковые часы могут помочь лучше управлять ресурсами.

Прогнозирование расчетного времени прибытия (ETA) необходимо для алгоритмов построения маршрутов и ценообразования, а также предоставления клиентам точных сроков доставки. Точное прогнозирование ETA не только повышает прозрачность и доверие к компании благодаря коммуникации точного времени доставки клиенту, но и непосредственно влияет на эффективность платформы доставки.

Диссертационная работа посвящена разработке алгоритмов прогнозирования времени доставки на основе моделей машинного обучения, которые могут использоваться в основе ключевых алгоритмов платформы доставки.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Целью диссертационной работы является разработка алгоритмов и программного обеспечения для решения задач прогнозирования времени доставки.

Для достижения поставленной цели необходимо решить следующие задачи:

- 1) Провести анализ предметной области, определить особенности доставки и сложность решения задачи прогнозирования времени доставки.
- 2) Провести анализ существующих методов и подходов для прогнозирования временных рядов.
- 3) Разработать методы и алгоритмы прогнозирования времени доставки.
- 4) Провести сравнительный анализ реализованных алгоритмов времени доставки.

Объектом исследования являются алгоритмы прогнозирования временных рядов на основе моделей машинного обучения.

Предметом исследования является использование алгоритмов прогнозирования временных рядов на основе машинного обучения для решения задачи прогнозирования времени доставки.

Основной *гипотезой*, положенной в основу диссертационной работы, является возможность прогнозирования времени доставки с помощью моделей машинного обучения. Время доставки может зависеть от множества различных факторов. Основное влияние на время доставки могут оказывать факторы по региону, заявкам, пользователям и исполнителям. Совокупность данных, содержащих эти факторы, допускает возможность эффективного прогнозирования времени доставки, что, в конечном счете, может повысить эффективность платформ, предоставляющих услуги доставки.

Связь работы с приоритетными направлениями научных исследований и запросами реального сектора экономики

Работа выполнялась в соответствии с научно-техническим заданием и планом работ кафедры «Информатика и технологии программирования» по теме «Алгоритмы прогнозирования времени доставки на основе моделей машинного обучения» (ГБ №210-С 31.01.2024, № ГР 256241, научный руководитель НИР – С. И. Сиротко).

Личный вклад соискателя

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя С. И. Сиротко, заключается в формулировке целей и задач исследования.

Опубликованность результатов диссертации

По теме диссертации опубликовано 2 работы в сборниках трудов и материалов международных конференций.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, трех глав, заключения, списка использованных источников, списка публикаций автора и приложения. В первой главе представлен анализ предметной области, выявлены основные существующие проблемы и сложности в рамках тематики исследования. Вторая глава посвящена анализу подходов к прогнозированию временных рядов, которые могут быть применены при разработке алгоритмов прогнозирования времени доставки. В третьей главе предложены алгоритмы прогнозирования времени доставки, реализован сбор данных и алгоритмы, представлены результаты экспериментальных исследований, сделан сравнительный анализ алгоритмов.

Общий объем работы составляет 60 страниц, из которых основного текста – 40 страниц, 17 рисунков на 10 страницах, 6 таблиц на 5 страницах, список использованных источников из 21 наименований на 2 страницах и 1 приложения на 8 страницах.

ОСНОВНОЕ СОДЕРЖАНИЕ

Во **введении** определена область и указаны основные направления исследования, показана актуальность темы диссертационной работы, дана краткая характеристика исследуемых вопросов, обозначена практическая ценность работы.

В **первой главе** был проведен анализ предметной области, описана доставка на “последней миле”, сформулированы операционные и технические сложности ее реализации, рассмотрены основные технологии, необходимые для реализации платформы доставки и сделана постановка задачи.

Доставка на “последней миле” – это заключительный этап цепочки поставок или процесса доставки, когда товар перемещается из транспортного

узла или склада в конечный пункт назначения, как правило, в частное жилье или розничный магазин. В основе технологий ценообразования доставки, оптимизации маршрутов и коммуникации времени прибытия курьера лежит технология прогнозирования времени доставки.

Понятие маршрута можно определить как упорядоченное множество точек, для которых известны координаты, номер заявки от клиента и дополнительная полезная информация от клиента. В одном маршруте могут находиться точки разных заявок разных клиентов. Определим понятие курьера как исполнителя маршрута.

Курьер может обладать различными характеристиками:

- иметь термосумку для перевозки товаров, требующих поддержание постоянного температурного режима
- иметь возможность доставлять товар до здания или же до двери квартиры;
- иметь личный или корпоративный транспорт различного вида: автомобиль, велосипед, мотоцикл;
- иметь грузовой транспорт с персоналом для разгрузки товара и установки.

Тогда задачей прогнозирования времени доставки является предсказание множества моментов времени T_i прибытия курьера на точку i при условии заданного маршрута, курьера и стартового времени T_0 .

Во **второй главе** был проведен анализ существующих методов машинного обучения и подходов прогнозирования временных рядов, которые могут быть использованы для решения поставленной задачи.

Метод градиентного бустинга позволяет построить сильную модель путем объединения нескольких слабых моделей, обычно моделей деревьев решений. Благодаря этому градиентный бустинг имеет высокую точность прогнозирования в сравнении с другими классическими методами машинного обучения. При этом, данный метод предоставляет возможность интерпретации полученного результата относительно исходных признаков.

Нейронные сети – это вычислительные структуры, которые для обработки сигналов используют явления, аналогичные происходящим в нейронах живых существ. Обученная на ограниченном множестве данных, нейронная сеть способна обобщать полученную информацию и показывать хорошие результаты на новых данных, не использовавшихся в процессе обучения, что является ее преимуществом в сравнении с классическими методами машинного обучения.

В **третьей главе** была проведена подготовка данных, включающая в себя фильтрацию, определение целевой переменной и обогащение данных дополнительными признаками. Также были реализованы алгоритмы для предсказания времени доставки, проведены эксперименты с подбором

гиперпараметров моделей машинного обучения, определены метрики качества и проведен анализ полученных результатов.

В качестве исходных данных был взят набор данных из открытых источников. LaDe – первый общедоступный набор данных о доставках "последней мили", содержащий миллионы посылок из реальной отрасли. На основе проведенного анализа данных было принято решение реализовать признаки по трем различным ключам агрегации:

- признаки области вокруг начальной точки;
- признаки области вокруг конечной точки;
- признаки по курьеру.

Область вокруг точек определялась с помощью библиотеки NZ от компании Uber, которая предоставляет обширный набор инструментов для работы с геоданными. В качестве области был выбран гексагон разрешения 8, к которой относится точка. В свою очередь, по каждому из ключей агрегации были подсчитаны статистики по времени, скорости и пройденному расстоянию за прошедшие 12 доставок. Также был добавлен признак потенциально затрачиваемого времени, который вычислялся как скорость за прошедшие 12 доставок умноженная на расстояние между начальной и конечной точкой. Дополнительные признаки в совокупности с уже существующими должны повысить эффективность и интерпретируемость моделей.

При реализации нейронной сети на основе LSTM был выбран подход, при котором за каждым LSTM слоем следует Dropout слой с целью предотвращения переобучения. Вероятность зануления в слое Dropout будет подобрана в пределах от 0.1 до 0.3. Было решено взять 4 пары LSTM-Dropout слоев, размер каждого слоя (N, 50), где N – количество подряд идущих дней в признаках исходного набора данных. Последние слои производят переход сначала на другую размерность (N, 50) \rightarrow 50, а затем и в выходной результат.

При реализации нейронной сети на основе CNN был выбран подход, при котором за каждым Conv слоем следует Pooling слой с целью постепенного уменьшения исходной размерности с целью уменьшения переобучения и ускорения сходимости модели. Размерность Pooling слой будет понижать в 2 раза (окно 2). Было решено взять 3 пары Conv -Pooling слоев, размер каждого слоя по оси X зависит от N, где N – количество подряд идущих дней в признаках исходного набора данных, и с каждой следующей парой слоев уменьшается в 2 раза. По оси Y размерности были выбраны 64, 128, 64, чего должно быть достаточно для описания сложных нелинейных зависимостей в данных. После пар слоев Conv-Pooling используется слой Flatten, преобразующий исходные данные в вектор, к которому в дальнейшем применяется слой Dropout для снижения переобучения модели. Вероятность зануления в слое Dropout будет

подобрана в пределах от 0.1 до 0.3. Последний слой применяет результирующую функцию активации.

При реализации нейронной сети на основе CNN-LSTM был выбран подход, при котором сначала используется 3 пары Conv-Pooling слоев для описания сложных нелинейных зависимостей в данных, затем слой Flatten, преобразующий исходные данные в вектор, после чего 2 пары LSTM-Dropout слоев, а также последний слой, применяющий результирующую функцию активации. Размерности по оси Y для CNN-Pooling и LSTM-Dropout пар слоев были сохранены – (64, 128, 64) и (50, 50) соответственно.

В качестве классических методов машинного обучения были рассмотрены градиентный бустинг и линейная регрессия. Использовалась реализация градиентного бустинга их библиотеки CatBoost от компании Яндекс. В качестве тривиального алгоритма был выбран метод скользящего среднего.

В качестве метрик качества использовалась MAE – средняя абсолютная ошибка (формула 3.1):

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - y| \quad (3.1)$$

Также использовалась метрика RMSE – корня из среднеквадратичной ошибки (формула 3.2):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y)^2} \quad (3.2)$$

Для каждого метода был произведен подбор гиперпараметров и были выбраны лучшие модели по метрике качества RMSE. Рассмотрим итоговые результаты различных методов предсказания (таблица 3.1). Стоит отметить, что добавление признаков повысило качество классических моделей машинного обучения: линейной регрессии и градиентного бустинга. При этом на качестве нейронных сетей добавление признаков не повлияло. Это связано с тем, что реализованные нейронные сети получают на вход данные из прошлого и имеют высокую способность к обобщению. Также стоит отметить, что модель на основе LSTM по качеству не сильно отличается от модели на основе CatBoost-a. Лучшей моделью оказалась CNN-LSTM, которая сочетает в себе способность лучше адаптируется к актуальным изменениям от LSTM и способность к поиску глубоких зависимостей от CNN. Стоит отметить, что модель градиентного бустинга CatBoost справилась с задачей лишь на 1% хуже в сравнении с CNN-LSTM, при этом требует меньшее количество вычислительных ресурсов, что может оказаться решающим фактором в выборе в пользу использования данной модели в платформе доставки.

Таблица 3.1 – Итоговые результаты моделей на отложенной выборке

Набор данных	Модель	Метрика	Значение
Без доп. признаков	MA	RMSE	18.715267
Без доп. признаков	MA	MAE	350.261216
Без доп. признаков	MA	ACC10	0.657328
Без доп. признаков	MA	ACC20	0.816464
Без доп. признаков	Linear	RMSE	3.616133
Без доп. признаков	Linear	MAE	13.07642
Без доп. признаков	Linear	ACC10	0.641243
Без доп. признаков	Linear	ACC20	0.878183
С доп. признаками	Linear	RMSE	3.566232
С доп. признаками	Linear	MAE	12.718012
С доп. признаками	Linear	ACC10	0.675364
С доп. признаками	Linear	ACC20	0.883364
Без доп. признаков	Catboost	RMSE	3.158024
Без доп. признаков	Catboost	MAE	9.973117
Без доп. признаков	Catboost	ACC10	0.787759
Без доп. признаков	Catboost	ACC20	0.899121
С доп. признаками	Catboost	RMSE	3.133210
С доп. признаками	Catboost	MAE	9.817007
С доп. признаками	Catboost	ACC10	0.791573
С доп. признаками	Catboost	ACC20	0.901696
Без доп. признаков	LSTM	RMSE	3.134499
Без доп. признаков	LSTM	MAE	9.825082
Без доп. признаков	LSTM	ACC10	0.900837
Без доп. признаков	LSTM	ACC20	0.899121
С доп. признаками	LSTM	RMSE	3.1338
С доп. признаками	LSTM	MAE	9.820704
С доп. признаками	LSTM	ACC10	0.791156
С доп. признаками	LSTM	ACC20	0.900894
Без доп. признаков	CNN-LSTM	RMSE	3.125103
Без доп. признаков	CNN-LSTM	MAE	9.766272
Без доп. признаков	CNN-LSTM	ACC10	0.792371
Без доп. признаков	CNN-LSTM	ACC20	0.901534
С доп. признаками	CNN-LSTM	RMSE	3.125173
С доп. признаками	CNN-LSTM	MAE	9.766703
С доп. признаками	CNN-LSTM	ACC10	0.792358
С доп. признаками	CNN-LSTM	ACC20	0.901528

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

1) Предложена архитектура программной системы для решения задачи прогнозирования времени доставки.

2) Предложен метод прогноза времени доставки на основе комбинированной модели машинного обучения, сочетающий в себе достоинства нескольких подходов к решению задачи прогноза времени доставки.

3) С использованием обучения с учителем удалось достичь высокого качества предсказания времени доставки. По метрике качества MAE модель на основе LSTM лучше адаптируется к актуальным изменениям и решает задачу прогнозирования по сравнению с моделью на основе CNN, которая более склонна к поиску глубоких зависимостей. При этом модель CNN-LSTM, являясь комбинацией предыдущих, имеет преимущества обеих моделей, что позволяет ей делать прогноз с более высоким качеством.

4) Благодаря разработке дополнительных признаков удалось достичь высокого качества модели градиентного бустинга, которая может использоваться в качестве альтернативы комбинированной модели CNN-LSTM, так как потребляет значительно меньше вычислительных ресурсов ценой небольшой просадки качества.

Рекомендации по практическому использованию результатов

1) Полученные результаты формируют теоретическую и практическую базу для разработки и интеграции ПО компьютерных систем для решения задач прогнозирования времени доставки. Они могут быть использованы для модернизации и дальнейшего развития существующих систем.

2) Разработанные методы и алгоритмы прогнозирования времени доставки могут применяться в основе алгоритмов построения маршрутов доставки для оптимизации потенциальной экономии курьерского времени, могут применяться в основе алгоритмов ценообразования для оценки затрачиваемого курьером времени, которое необходимо для поддержания баланса спроса и предложения вне зависимости от внешних факторов.

3) Результаты работы могут использоваться для коммуникации пользователям сервисов, предоставляющих доставку, о предполагаемом времени доставки, потенциальных временных интервалов в момент создания заявки, а также коммуникации курьерам о затрачиваемом времени в пути.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1-А. Клебанов Д.А. Использование гибридных моделей машинного обучения для анализа временных рядов / Д.А. Клебанов // Международная научно-практическая конференция «Advances in Science and Technology». – Москва, Российская Федерация, 2022. – с. 29.

2-А. Клебанов Д.А. Прогнозирование цен акций с помощью нейронных сетей / Д.А. Клебанов // Компьютерные системы и сети: материалы 59-ой научной конференции аспирантов, магистрантов и студентов. – Минск: БГУИР, 2023. – с. 161-162.