

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

УДК 004.855.5

Тузина
Екатерина Олеговна

**Система прогнозирования очередей на границе
на базе нейронной сети**

АВТОРЕФЕРАТ

на соискание степени

магистра информатики и вычислительной техники

по специальности 1-40 81 04 – Обработка больших объемов информации

Научный руководитель

Пилецкий И.И.

к.ф.-м.н., доцент

Минск 2020

ВВЕДЕНИЕ

Прогнозирование применяется во многих областях человеческой деятельности, таких как наука, управление, экономика, производство и множество других сфер.

Понятие прогнозирования не имеет четко очерченных границ и активно взаимодействует со смежными задачами анализа данных. Одно из определений прогнозирования было дано Г. Тейлом: «некоторое суждение относительно неизвестных, особенно будущих, событий».

Прогнозирование является одним из ключевых моментов при принятии решений и направлено на определение тенденций динамики конкретного объекта или события на основе ретроспективных данных, т.е. анализа его состояния в прошлом и настоящем.

Необходимость прогноза может быть обусловлена стремлением предугадать значение некоторых показателей в будущем или оценить показатели некоторого объекта, основываясь на известных данных о нем. Точность любого прогноза обусловлена объемом истинных (верифицированных) исходных данных и периодом их сбора, объемом не верифицированных исходных данных и периодом их сбора, а также свойствами системы, объекта, подвергающихся прогнозированию, и методиками, подходами к прогнозированию.

Вариантов практического применения прогноза на сегодняшний день достаточно много: это предсказания на фондовых рынках, предсказание результатов спортивных игр, оценка недвижимости, предсказание экономических показателей и так далее.

Актуальность выбранной темы обусловлена важностью прохождения границы вовремя и сокращении затраченного времени. Данный аспект важен как для сферы грузоперевозок, так и для физических лиц, водителей легковых автомобилей.

Значение грузоперевозок с помощью автомобильного транспорта возрастает ежедневно. Именно этот транспорт востребован и используется каждый день. Перевозки играют одну из важнейших ролей в развитии экономики и международных отношений страны. Развитие сферы грузоперевозок в свою очередь эффективно влияет на развитие экономики. Крупный бизнес активно использует услуги грузоперевозок, как на внутреннем рынке, так и на международной арене. При любых грузоперевозках очень важна доставка грузов вовремя.

Зачастую водители, совершающие международные поездки, вынуждены много времени проводить в очередях на границе, от нескольких часов до недель

(для водителей грузовых автомобилей). Ситуация с очередями в пунктах пропуска не стабильная и очереди автомобилей зависят от неизвестных факторов. Также решение задачи прогнозирования количества автомобилей в очереди пункта пропуска сможет помочь сотрудникам Государственного пограничного комитета Республики Беларусь в планировании распределения ресурсов.

Способы реализации программного обеспечения, используемого для осуществления прогнозов, можно условно разделить на два класса: реализующее статистические методы прогнозирования и использующее технологии нейронных сетей.

Рядом научных исследований было доказано, что использование нейронных сетей для прогнозирования дает некоторые преимущества относительно статистических прогнозов, так как нейронные сети являются более чувствительными к параметрам.

Способности нейронной сети к прогнозированию напрямую следуют из ее способности к обобщению и выделению скрытых зависимостей между входными и выходными данными. После обучения сеть способна предсказать будущее значение некой последовательности на основе нескольких предыдущих значений и/или существующих в настоящий момент факторов [5].

В данной работе рассматривается решение задачи прогнозирования количества грузовых и легковых автомобилей в очереди пункта пропуска с построением модели нейронной сети.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Целью работы является разработка решения для задачи прогнозирования количества автомобилей в пункте пропуска в момент времени. Результатом должен быть набор моделей, способных спрогнозировать одно и несколько значений, являющихся количеством автомобилей, в будущем.

Решение должно быть реализовано на языке программирования Python. В работе используются самостоятельно полученные данные с сайта Государственного пограничного комитета Республики Беларусь о количестве грузовых и легковых автомобилей в пункте пропуска. Точность решения является более приоритетной по сравнению со скоростью его работы.

Для достижения поставленной цели необходимо решить следующие задачи: проанализировать современное состояние области машинного обучения и анализа данных, выделить этапы решения задачи и подходы к решению каждого этапа, определить подход к сбору и обработке данных и осуществить,

провести эксперименты, сравнить результаты и сделать вывод о том, какими методами и моделями машинного обучения можно наилучшим образом (в заданных ограничениях) решить поставленную задачу.

Объектом исследования является выбранная прикладная задача.

Предметом исследования – применение алгоритмов машинного обучения для решения поставленной задачи.

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

Исследуемые в работе подходы для прогнозирования позволяют предположить ситуацию в пункте пропуска, что дает возможность физическим лицам и компаниям, занимающимся международными грузоперевозками, использовать полученную информацию с целью сокращения времени, затраченного на прохождение границы, и доставки товаров вовремя. Также решение задачи прогнозирования количества автомобилей в очереди пункта пропуска сможет помочь сотрудникам Государственного пограничного комитета Республики Беларусь в планировании распределения ресурсов.

В качестве прикладной задачи была выбрана задача прогнозирования количества автомобилей в пункте пропуска. В рамках исследовательской работы над выбранной задачей была изучена предметная область машинного обучения и временных рядов. Было разработано решение для прогнозирования количества автомобилей в будущем в момент времени и прогнозирования интервала значений в будущем.

Личный вклад соискателя

Результаты, приведенные в диссертации, получены соискателем лично.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, трех глав, заключения, списка использованных источников и приложений. В первой главе находится обзор предметной области: нейронных сетей, временных рядов и задачи прогнозирования временных рядов. Во второй главе приведена информация об используемых технологиях и их возможностях. В третьей главе дано последовательное описание этапов работы над решением поставленной задачи: разбиение на подзадачи (получение и подготовка данных, построение модели нейронной сети) и подходы к решению каждой подзадачи. Приведено также сравнение результатов работы моделей для трех задач прогнозирования,

а также сравнение моделей для задач прогнозирования для грузовых и легковых автомобилей.

Общий объем диссертации – 61 страницу. Работа содержит 15 формул и 26 рисунков. Библиографический список включает 32 наименования.

ОСНОВНОЕ СОДЕРЖАНИЕ

Во введении приведено обоснование актуальности задач прогнозирования в целом и рассматриваемой задачи конкретно. Приведено описание проблемы очередей на границе, как для людей, стоящих в очереди, так и для сотрудников пункта пропуска. В первой главе рассмотрены искусственные нейронные сети, в частности большое внимание уделяется рекуррентным нейронным сетям и их особенностям. Приведено описание временных рядов и их характеристик, понятие задачи прогнозирования и классификация прогнозов. Во второй главе рассмотрены язык программирования и библиотеки, использованные в работе для построения моделей, и облачный сервис Google Colaboratory. В третьей главе задача разбивается на два основных этапа: получение и подготовка данных, построение и обучение нейронной сети. На первом этапе были получены данные о количестве автомобилей на пропускном пункте для грузовых и легковых автомобилей с официального сайта Государственного пограничного комитета Республики Беларусь, данные проанализированы и обработаны. На втором этапе сначала были построены модели для грузовых автомобилей для прогнозирования следующего значения (использован один признак), прогнозирования значения через двое суток в будущем и прогнозирования интервала значений на следующие сутки (использовано несколько признаков). При исследовании моделей было выявлено, что лучше всего модели справляются с прогнозированием значений в наиболее близком будущем. Структуры моделей, полученные для грузовых автомобилей, были использованы для легковых автомобилей. Для легковых модели показали себя хуже, что связано с большим количеством нулевых данных. В заключении приведены краткие результаты работы и намечены возможные дальнейшие шаги для решения задачи.

ЗАКЛЮЧЕНИЕ

В ходе проведенной работы решалась задача прогнозирования количества автомобилей в пункте пропуска на примере пункта пропуска Брузги на границе с Республикой Польша. Набор данных получен с сайта Государственного пограничного комитета Республики Беларусь и подготовлен в рамках данной работы.

После получения необходимых данных построены модели для прогнозирования на основе одного и нескольких признаков, для решения следующих задач:

- прогнозирование следующего значения;
- прогнозирование значения через двое суток в будущем;
- прогнозирование интервала значений на следующие сутки.

Лучше всего показала себя модель для прогнозирования последующего значения, ошибка этой модели на тестовых данных меньше, чем для других моделей. Также иллюстрация работы модели для прогнозирования интервала значений в будущем показала, что более точно прогнозируются значения ближе к началу интервала. Проанализировав результаты прогнозов модели значения через двое суток в будущем и части прогнозов модели для прогнозирования интервала значений, относящейся к более позднему будущему, можно сделать вывод, что с отдалением во времени прогнозируемого выходного значения от входных значений точность прогноза снижается.

Для легковых автомобилей модели показали себя хуже и прогнозирование даже на следующий шаг не оказалось успешным. Модель плохо справляется с прогнозом в связи с большим количеством (около 60%) нулевых данных.

На данный момент для рассматриваемой задачи возможно дать краткосрочный прогноз на несколько шагов вперед. Стоит также отметить вопрос качества данных на сайте Государственного пограничного комитета Республики Беларусь. Подсчеты ведутся человеком, система не автоматизирована, что может повлечь ошибки. После изменений в системе подсчета и точности данных станет возможным следующий шаг в работе и улучшение построенных моделей.