

Санкт-Петербургский государственный университет

Направление «Математическое обеспечение и администрирование
информационных систем»

Профиль «Системное программирование»

Сергей Иванович Андреев

Агрегация сведений и оценка параметров грузовых маршрутов на основе методов машинного обучения в условиях информационного дефицита

Курсовая работа

Научный руководитель:
доц. каф. информатики,
к.т.н. Абрамов М. В.

Санкт-Петербург
2020

Оглавление

Введение	3
1. Обзор предметной области	6
1.1. Актуальность анализа параметров грузовых маршрутов	6
1.2. Сравнение аналогов	8
1.3. Выводы по главе	9
2. Методы и подходы	10
2.1. Описание используемых программных средств	10
2.2. Требования к модели машинного обучения	10
3. Детали реализации	12
3.1. Архитектура программного комплекса	12
3.2. Сбор датасета	14
Заключение	16
Список литературы	17
Приложение А: список терминов	18

Введение

Актуальность темы. В последнее время грузовые перевозки занимают большую долю транспортно-логистических услуг в России. Например, в 2017 году общемировой показатель доли грузовых услуг в области логистики составлял 55%, в то время как в России он составлял 88% [4]. Это означает, что отрасль грузовых перевозок в России занимает серьезное место на рынке логистических услуг. На момент 2020 года насчитывается более 220 крупных грузовых транспортных компаний [6]. Многие из этих компаний имеют так называемые калькуляторы доставки — веб-сервисы, позволяющие рассчитать стоимость предоставления услуг данной компании для заданных исходных параметров грузового маршрута, то есть стоимость и сроки доставки заданного габаритами, весом и прочими характеристиками груза из заданных пунктов. Такие калькуляторы очень важны для компаний, предоставляющих грузовые транспортные услуги [5]. Например, они позволяют снизить нагрузку на операторов колл-центров, автоматизировать процессы в компании, привлечь пользователей использовать услуги данной компании. В то же время, клиентам таких компаний зачастую бывает быстрее и удобнее воспользоваться такими калькуляторами для получения данных об интересующем грузовом маршруте, чем выяснять эту информацию у оператора горячей линии. В последнее время наблюдается рост популярности калькуляторов доставки. На рисунке 1 показана динамика количества запросов "калькулятор доставки" в поисковом сервисе yandex.ru за период с января 2019 по декабрь 2020. Данные с сайта wordstat.yandex.ru.

При таком обилии грузовых транспортных компаний у клиента возникает проблема выбора оптимального по цене или срокам доставки перевозчика. Для этого клиенту необходимо внести исходные данные о грузовом маршруте во все калькуляторы доставки выбранных им для сравнения компаний, подождать окончания обработки запроса на каждом сервисе, сравнить полученные результаты. Данный алгоритм действий достаточно тяжело выполнять вручную, если хочется иметь для сравнения большое количество перевозчиков. Для решения этой пробле-

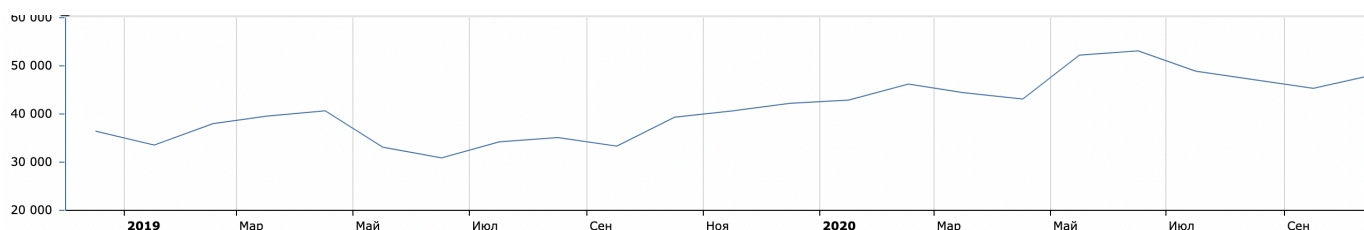


Рис. 1: Динамика количества запросов "Калькулятор доставки" в поисковом сервисе yandex.ru за период с января 2019 по декабрь 2020. Данные с сайта wordstat.yandex.ru.

мы существуют агрегаторы калькуляторов доставки. Данные сервисы предоставляют клиенту интерфейс для ввода исходной информации о грузовом маршруте и функциональность по автоматическому поиску информации о данном грузовом маршруте на наборе калькуляторов доставки.

Однако, на данный момент существует не так много различных агрегаторов, в существующие агрегаторы обладают рядом недостатков. Во-первых, обычно такие агрегаторы представляют из себя отдельный сайт, в то время, как пользователю было бы гораздо удобнее иметь доступ к сервису, содержащему в себе ряд полезных логистических решений, а не искать для каждой логистической задачи свой сервис. Во-вторых, существующие агрегаторы никак не анализируют полученную информацию о стоимости и сроках доставки, а просто выводят ее пользователю. Анализ этой информации позволил бы понимать, завышена ли цена или сроки доставки на конкретный маршрут, или же наоборот, полученное предложение является выгодным, а также спрогнозировать ожидаемую цену и сроки доставки для маршрутов, информация о которых не доступна в соответствующем калькуляторе доставки. Случается, что грузовая компания может осуществить доставку по некоторому маршруту, но данная информация по тем или иным причинам не отображается калькулятором доставки. Это может случаться из-за неудачной программной реализации калькулятора, или особенностей маршрута, требующих дополнительного уточнения данных у оператора для определения цены и сроков доставки. В таком случае пользователю было

бы удобно изначально иметь представление о том, сколько будет стоить доставка по специфическому маршруту.

Объектом исследования являются параметры грузовых транспортных маршрутов, а **предметом исследования** являются методы предсказания цены и сроков доставки для данных маршрутов.

Целью данной работы автоматизация оценки параметров грузовых маршрутов через агрегацию доступных сведений на основе методов машинного обучения в условиях информационного дефицита.

Для достижения данной цели были сформулированы и решены следующие задачи:

- провести обзор предметной области и сравнение аналогов;
- спроектировать и реализовать прототип модуля программ для агрегации данных о грузовых транспортных маршрутах;
- собрать данные о параметрах грузовых транспортных маршрутов, провести обработку данных;
- на основе собранных данных построить несколько моделей машинного обучения, решающих задачу предсказания цены и сроков доставки для заданного грузового маршрута;
- выбрать критерии для сравнения моделей, выявить самую подходящую модель;
- реализовать модель, провести апробацию и эксперименты
- Интегрировать систему в сервис cargotime.ru ¹

¹Единая грузовая служба – это проект, цель которого объединить профессиональных логистов, крупных перевозчиков и частных клиентов в общем намерении взаимовыгодно сотрудничать друг с другом.

1. Обзор предметной области

В данной главе подчеркивается актуальность анализа параметров грузовых маршрутов, а также проводится сравнение основных агрегаторов калькуляторов доставки.

1.1. Актуальность анализа параметров грузовых маршрутов

Калькуляторы доставки позволяют клиентам грузовых транспортных компаний рассчитывать стоимость и сроки доставки для интересующих грузовых маршрутов, совершая при этом минимум действий. Такие данные обычно требуются профессиональным логистам, индивидуальным предпринимателям и частным клиентам. Однако, калькуляторы нацелены на расчет параметров только в рамках услуг, предоставляемых конкретной грузовой компанией. Существуют сервисы, агрегирующие информацию о грузовом маршруте с множества различных калькуляторов доставки.

При поиске расчете параметров грузового маршрута в агрегаторе калькуляторов доставки пользователю было бы удобно понимать, предложенная цена является завышенной, или заниженной, то есть выгодной. На основании этой информации клиент может попробовать отложить грузоперевозку, либо попробовать поторговаться с грузоперевозчиком ради получения скидки. Выводы о справедливости цены возможно делать на основании данных о прочих грузовых транспортных маршрутах. Анализ данных грузовых маршрутов позволил бы также прогнозировать параметры маршрутов, данные о которых по тем или иным причинам не доступны в калькуляторе доставки грузовой компании. Зачастую, это не означает, что по данному маршруту доставка невозможна; это означает, что детали такого маршрута необходимо уточнять у оператора. Но уточнение данных у оператора даже для небольшого количества компаний может занять продолжительное время, которое будет потрачено впустую, если цена или сроки доставки

по такому маршруту будут сильно завышены. Предсказание параметров поможет пользователю отфильтровать список компаний, сразу отбросить те из них, которые с большой вероятностью его не устроят, и выяснять подробности доставки у оператора для компаний, которые удовлетворили его по предварительной оценке.

Подобная идея уже реализована в агрегаторе авиаперевозчиков Aviasales ². Анализируя методами машинного обучения данные об авиабилетах различных авиакомпаний, сервис предсказывает цену на авиабилеты для дат с отсутствующими еще данными по авиабилетам. Также сервис анализирует цену на авиабилет, определяя, вероятно ли падение цены в дальнейшем [3].

²<https://www.aviasales.ru/>

1.2. Сравнение аналогов

На данный момент существует ряд различных агрегаторов калькуляторов доставки. Такие сервисы полезны профессиональным логистам или частным клиентам, желающим быстро получить сводку рассчитываемых параметров грузового маршрута для множества грузовых компаний. Набор некоторых популярных в настоящее время агрегаторов калькуляторов доставки представлен в таблице 1.

Таблица 1: Популярные агрегаторы калькуляторов доставки

Сервис	Деклар. число перевозчиков	Инф. о перевозчиках	Доп. логистические сервисы	Анализ результатов	Стоимость использования
Cargo guru ³	30	+	-	нет	бесплатно
eDost ⁴	56	+	-	нет	бесплатно
MainMine ⁵	61	+	+	нет	бесплатно
Catapulto ⁶	20	-	+	дем.-я лучшего по цене или по срокам варианта	бесплатно
Агрегатор сборных грузов ⁷	4	+	-	нет	бесплатно
C6V.RU ⁸	13	-	-	нет	частично бесплатно

Сравнение проводилось с целью проанализировать наличие и применимость текущих решений с точки зрения имеющейся функциональности, затрат, возможности расчета параметров грузового маршрута

³<https://cargo.guru>

⁴<http://www.edost.ru>

⁵<https://mainmine.ru/>

⁶<https://catapulto.ru/calc/>

⁷<https://asgruz.ru/partners/>

⁸<https://c6v.ru/price>

для популярных перевозчиков.

Выбранные для сравнения параметры позволили проанализировать сервисы на объем поддерживаемых калькуляторов доставки, наличие подробных данных о перевозчиках, наличие дополнительных логических сервисов и функциональности по анализу полученных результатов, например, справедливости цены или сроков доставки. Большая часть сервисов не имеет подробной информации о перевозчиках, которые используются при агрегации, либо площадка, на которой они расположены, не имеет больше никаких логистических сервисов. Некоторые из сервисов имеют практически не стилизованный и как следствие не очень привлекательный и удобный интерфейс, например, eDost. При этом некоторые сервисы имеют ошибки реализации, например, MainMine не показывает сроки доставки для больше чем половины калькуляторов доставки при достаточно простых исходных параметрах грузового маршрута: пункт отправления – Москва, пункт назначения – Санкт-Петербург, объем – 1м^3 , вес – 100кг.

Таким образом, существующие сервисы имеют сильно ограниченную функциональность, и имеют проблемы, ухудшающие удобство их использования пользователями.

1.3. Выводы по главе

В данной главе описана актуальность анализа параметров грузовых маршрутов агрегаторами калькуляторов доставки. Проведено сравнение известных агрегаторов калькуляторов доставки, выявлены их достоинства и недостатки.

2. Методы и подходы

В данной главе представлена теоретическая часть, послужившая фундаментом для решения поставленных задач.

2.1. Описание используемых программных средств

При выборе технологий для реализации программного продукта нужно учитывать, что продукт должен быть легко встраиваем в стороннюю площадку, содержащую и другие логистические решения. Выбранные технологии должны позволять писать надежный и качественный код. Кроме того, выбор технологий должен позволять использовать готовые библиотеки с основными методами машинного обучения.

Для реализации программного продукта был выбран язык TypeScript и среда исполнения NodeJS. Данные технологии позволяют писать надежный и лаконичный объектно-ориентированный код, а полученное программное ядро, содержащее бизнес-логику, можно легко и быстро встроить на веб-сервер.

2.2. Требования к модели машинного обучения

Поставленная в работе задача предсказания цены и сроков доставки может быть решена различными моделями машинного обучения [2]. Более формально задачу можно сформулировать следующим образом: необходимо реализовать модель, принимающую в качестве входных данных пункт отправления, пункт назначения, габариты груза: длина, высота, ширина; вес груза, количество грузовых мест, либо данные, которые возможно получить из нижеперечисленных. Например, расстояние между пунктом отправления и назначения, численность населения в пункте отправления и т.д. Кроме того, входными данными является название компании, для которой необходимо сделать предсказание, а также тип грузоперевозки. Предусмотрены следующие типы грузоперевозки:

- "Авто" – обыкновенная автомобильная грузоперевозка.

- "Авто экспресс" – автомобильная грузоперевозка с увеличенной ценой, но как наименьшими возможными сроками доставки.
- "Авиа" – грузоперевозка с использованием воздушного транспорта.
- "ЖД" – грузоперевозка с использованием железнодорожного транспорта

Выходными данными модели являются цена грузоперевозки и цена грузоперевозки с учетом доставки от и до терминала.

3. Детали реализации

В данной главе описаны архитектура, подробности реализации программного продукта, его модулей. Кроме того, в этой главе приводятся подробности проведения тестирования, апробации и интеграции программного продукта на площадку `cargotime.ru`⁹.

3.1. Архитектура программного комплекса

Программный комплекс можно разделить на несколько частей:

- Клиентская часть
- Серверная часть
 - Сетевой клиент
 - Агрегатор калькуляторов доставки
 - Модель предсказания цены и сроков доставки

Рассмотрим подробнее серверную часть программного продукта. На рис. 2 изображена диаграмма компонентов серверной части.

Модуль Core

Данный модуль отвечает за агрегацию данных о грузовом маршруте из различных калькуляторов доставки.

Класс `ParsersManager` предоставляет интерфейс данного модуля и управляет процессом получения данных от калькуляторов доставки. Класс `ParsersFactory` представляет из себя элемент паттерна "Фабрика", создавая конкретные реализации абстрактного класса `Parser`. Таким образом, `ParsersManager` абстрагируется от конкретных реализаций абстрактного класса `Parser`, используя объявленный в нем методы.

Перечисленные сущности оперуют объектами вида `IRequest` и `IResponse` – стандартизированные форматы исходных и рассчитываемых данных

⁹Единая грузовая служба – это проект, цель которого объединить профессиональных логистов, крупных перевозчиков и частных клиентов в общем намерении взаимовыгодно сотрудничать друг с другом.

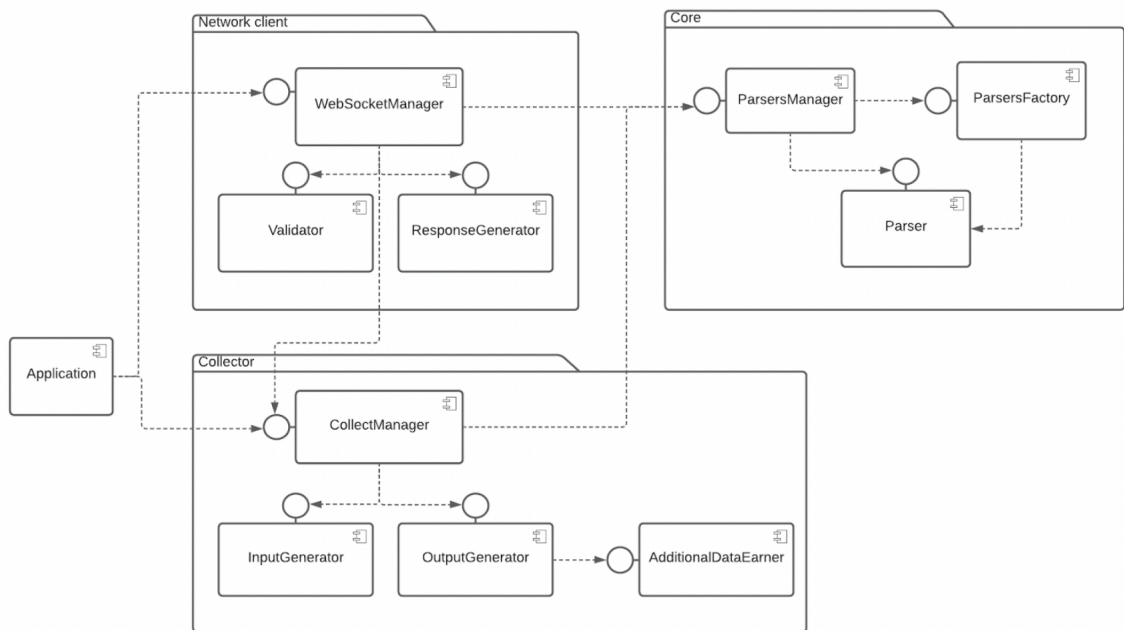


Рис. 2: Диаграмма компонентов программного продукта

грузового маршрута соответственно. Конкретные реализации Parser реализуют функциональность по преобразованию объекта IRequest в специфичный для данного калькулятора доставки вид, отправки запроса на веб-сервер калькулятора соответствующей грузовой компании и преобразования полученного ответа в стандартный для системы вид – IResponse.

Модуль Network client

Данный модуль отвечает за обработку входящих подключений, проверяя корректность исходных параметров грузового маршрута, и используя интерфейс модуля Core для получения результата.

Процесс агрегации может занять некоторое время (от 3 до 20 секунд), а ответ от веб-серверов одних компаний может прийти значительно раньше, чем от других. В связи с этим, каждые рассчитанные данные грузового маршрута для конкретной компании необходимо отправить клиенту сразу, не дожидаясь окончания поиска. Поэтому для сетевого взаимодействия было решено использовать протокол вебсокетов [1]. Данный протокол позволяет серверу после установки соединения с клиентом быть инициатором отправки данных, чего не позволяет

протокол HTTP.

Класс `WebSocketManager` иницирует веб-сокет сервер и ожидает подключений клиентов. При подключении клиента корректность исходных данных грузового маршрута проверяется классом `Validator`, и в случае корректных входных данных `WebSocketManager` создает экземпляр `ParsersManager` и начинает асинхронный поиск. После получения очередного результата он преобразуется в подходящий для отправки клиенту формат классом `ResponseGenerator`, после чего `WebSocketManager` отправляет клиенту порцию данных.

Модуль `Collector`

Данный модуль отвечает за генерацию датасета. `CollectManager` управляет процессом генерации: создает экземпляр `InputGenerator`, который в зависимости от входных данных считывает массив исходных параметров грузовых маршрутов из файла, либо генерирует его в соответствии с конфигурационным файлом. Далее для всех элементов массива `ParserManager` получает ответы калькуляторов, которые `OutputGenerator` записывает в выходной файл.

3.2. Сбор датасета

Реализованное ядро программного комплекса (модуль агрегации `Core`) спроектировано независимым от сетевого контроллера, поэтому может использоваться отдельно от него, благодаря чему его можно удобно использовать для сбора датасета. Скорость сбора данных напрямую зависит от скорости сети, скорости работы серверов калькуляторов доставки различных грузоперевозчиков, поэтому сбор датасета может занимать некоторого времени.

Для сбора пробного датасета были выбраны входные данные, состоящие из всевозможных комбинаций пяти крупных городов России, трех размеров для параметров груза (малая/средняя/крупная длина/ширина/высота грузовых мест). В дальнейшем планируется оптимизировать набор входных данных для лучшей точности модели. Полученный выходной дата-

сет состоит из 20 000 записей. Каждая запись имеет следующие признаки: название пункта отправления, название пункта назначения, тип грузоперевозки, название компании-перевозчика, длина груза, ширина груза, высота груза, вес груза, количество грузовых мест, стоимость грузоперевозки, стоимость грузоперевозки с учетом доставки от и до терминала, минимальное количество суток доставки, максимальное количество суток доставки.

Дальнейшая реализация сервиса AdditionalDataEarner позволит с помощью google maps api добавить в выходные данные информацию о расстоянии между пунктом назначения и отправления, а также численность населения этих пунктов.

Заключение

В рамках данной работы были достигнуты следующие результаты:

- Проведен обзор предметной области и сравнение аналогов
- Спроектирован и реализован программный модуль – агрегатор данных о грузовых транспортных маршрутах.
- Собраны данные о параметрах грузовых транспортных маршрутов

Научная новизна. Все результаты, выносимые на защиту, являются новыми. Собран специфичный для данной области датасет, состоящий из записей, содержащих входные и выходные параметры грузовых маршрутов для российских грузоперевозчиков, а также некоторые дополнительные полезные данные (расстояние между пунктами грузоперевозки, численность населения). Проведено сравнение методов предсказания цены грузоперевозки.

Теоретическая и практическая значимость. Разработанные методы, модели и программные решения позволяют проводить автоматизированную агрегацию информации о грузовых маршрутах и оценку их параметров. Внедрение программного комплекса в сервис cargotime.ru позволяет пользователям решать еще больший спектр различных задач из области логистики, что определенно способствует развитию исследований по данной тематике и развитию данной области в целом.

Список литературы

- [1] I. Fette A. Melnikov. The WebSocket Protocol. RFC 6455. — 2011. — URL: <https://tools.ietf.org/html/rfc6455> (дата обращения: 28.04.2020).
- [2] Schutt Cathy O'Neil & Rachel. Doing Data Science», publisher =.
- [3] Как технологии искусственного интеллекта помогают Aviasales расти: семь примеров. — 2018. — URL: <https://vc.ru/growth/51391-kak-tehnologii-iskusstvennogo-intellekta-pomogayut-aviasale> (дата обращения: 28.11.2020).
- [4] Л.М. Капустина. Россия на мировом рынке транспортно-логистических услуг. — Сборник Уральского государственного экономического университета, Екатеринбург, 2017.
- [5] ПОГУДИНА ЕЛЕНА АЛЕКСАНДРОВНА ПУПЕНЦОВА СВЕТЛАНА ВАЛЕНТИНОВНА. РАЗРАБОТКА ТРАНСПОРТНОГО КАЛЬКУЛЯТОРА ПО ЛОГИСТИКЕ ДЛЯ ТИПОГРАФИИ ООО «ПРИНТ СЕВЕР», publisher =.
- [6] Популярные российские грузовые транспортные компании. — 2020. — URL: https://shipadvisor.ru/transport/?PAGEN_1=2 (дата обращения: 28.11.2020).

Приложение А: список терминов

Грузовой маршрут — совокупность данных о пункте отправления, пункте назначения, габаритах груза, его весе, количестве грузовых мест, компании, осуществляющей перевозку, сроках доставки и цене доставки между соответствующими терминалами, а также между заявленными географическими пунктами.

Исходные параметры грузового маршрута — совокупность данных о пункте отправления, пункте назначения, габаритах груза, его весе, количестве грузовых мест.

Рассчитываемые параметры грузового маршрута — совокупность данных о компании, осуществляющей перевозку, сроках доставки и цене доставки между соответствующими терминалами, а также между заявленными географическими пунктами.

Калькулятор доставки — программный сервис, зачастую являющийся веб-приложением, предоставляющий пользователю интерфейс для получения рассчитываемых параметров грузового маршрута по исходным параметрам грузового маршрута при использовании услуг грузоперевозки фиксированной транспортной компании.

Агрегатор калькуляторов доставки — программный сервис, зачастую являющийся веб-приложением, предоставляющий пользователю интерфейс для получения множества рассчитываемых параметров грузового маршрута при использовании услуг различных транспортных компаний по исходным параметрам грузового маршрута.

Перевозчик — грузовая транспортная компания, имеющая собственный калькулятор доставки.