УДК 656.073+004.048

Проблемы подготовки данных для обучения нейросети с целью анализа и прогнозирования тарифов на морском транспорте

А. О. Баранникова,
канд. ист. наук, научный сотрудник Научной лаборатории пространственной логистики Морского государственного университета им. адм. Г. И. Невельского,

А. К. Вороненко,
эксперт Научной лаборатории пространственной логистики МГУ им. адм. Г. И. Невельского

В сложных и непредсказуемо изменяющихся внешних условиях, влияющих на российский рынок транспортно-логистических услуг, с учетом внутренней неустойчивости его участников и масштабными непрекращающимися изменениями структуры, в особенности на Дальнем Востоке РФ, крайне важно не только увеличить скорость обработки документов и обмен информацией между участниками процессов, но и прогнозировать изменение цен на перевозку.

Одна из задач логистического оператора — формирование конкурентной цены на перевозку и сопутствующие логистические услуги. Она формирует ряд подзадач, включающих мониторинг рынка перевозок, анализ возможностей использования подвижного состава и др.

Поскольку контейнерный рынок является частично закрытым, несмотря на превалирующее использование линейных морских перевозок и регулярных контейнерных поездов, сбор данных о ценах и анализ коммерческих особенностей представляет здесь определенные сложности.

После февраля 2022 г. с дальневосточного рынка логистических услуг ушли крупные контейнерные перевозчики, которые задавали тренд на бóльшую открытость, единые стандарты, цифровизацию и т. п. Одновременно наблюдается активный вход на рынок небольших китайских компаний, а также активизация российских экспедиторов по организации своих собственных линий и маршрутов. Эти тренды вместе с наблюдаемым глобальным ростом цен на контейнерные перевозки приводят к снижению качества услуг, что еще более усложнило анализ стоимости и выбор подходящих исполнителей для участников транспортного процесса, не имеющих собственного подвижного состава.

Отсутствие единообразия в вопросах публикования цен и оформления коммерческих предложений, сложность анализа стоимостей и наличия подвижного состава и доступного оборудования в необходимых локациях не только усложняют обработку входящей и исходящей документации и выбор подходящего тарифного плана и перевозчика, но также не позволяют спрогнозировать поведение цен как на краткосрочную, так и на долгосрочную перспективу.

На фоне сложностей с аналитикой прогнозирование стоимости транспортных перевозок выглядит еще более комплексной задачей, хотя потребность в прогнозе стоимости перевозок актуальна для всех участников рынка — как для стратегирования и планирования стоимости транспортных расходов в цене товара грузовладельцев, так и для анализа позиций конкурентов в транспортной среде и других задач. Частично решить проблему прогнозирования стоимости перевозок может внедрение нейросетей, которые зарекомендовали себя мощным и эффективным инструментом для анализа, прогнозирования и моделирования и применяются во многих отраслях — экономике, медицине, инженерии и т. д. **[1–5]**. Нейросети считаются популярным инструментом благодаря тому, что они не программируются, а обучаются, что делает их применение относительно простым.

Сеть самостоятельно определяет степень влияния тех или иных факторов на результат операций, и чем больше данных в нее поступает, тем более высока вероятность получения от нее требуемых

результатов **[6]**. В то же время за счет этого свойства эффективность работы нейросети зависит от качества данных, на которых она обучена.

Помимо количества и качества обучающих данных на процесс обучения и последующей работы нейросети влияет ряд факторов, связанных с особенностями ее архитектуры, параметрами обучения, качеством кода и т. д. Но, прежде чем экспериментировать с выбранной базовой архитектурой нейросети, представляется логичным обеспечить высокое качество и достаточный объем данных для обучения.

При малом объеме обучающих данных или при их зашумленности и неструктурированности самая лучшая нейросеть с идеальной архитектурой не будет демонстрировать удовлетворительные результаты. Таким образом, подготовка качественных входных данных в достаточном объеме становится приоритетным, но при этом наиболее длительным и сложным процессом в рамках обучения нейросети.

Рассмотрим проблемы, возникающие на начальном этапе подготовки данных для обучения нейросети для анализа и прогнозирования тарифов на морском транспорте. Некоторые из них будут проиллюстрированы практическими примерами.

Нейросеть и сбор массива данных для ее обучения

Для выполнения задач исследования разработан многослойный перцептрон, подходящий для задач регрессии — прогнозирования непрерывного числового значения на основе входных данных. Модель нейросети с тремя скрытыми слоями построена с использованием рабочего пространства Keras в среде Python. Задача, определенная для нейросети, — прогнозировать тарифы на конкретные перевозки в следующем месяце текущего года. Набор данных для обучения нейросети представляет собой ретроспективные данные о тарифах на морские контейнерные перевозки между портами Дальнего Востока России и Китая за последние пять лет.

Первым шагом в подготовке данных является их сбор. В настоящее время существует большое количество открытых наборов обучающих данных (датасетов), содержащих статистику, изображения, видео, тексты, биржевые цены и т. д. Такие наборы уже подготовлены и размечены, т. е. могут быть использованы для обучения нейросетей. Однако с учетом уникальности задач, возлагаемых на конкретную сеть для решения прикладной задачи, для ее обучения, как правило, требуются специфические, возможно даже уникальные данные. Готовые наборы подходят скорее для тренировки, хотя и могут быть использованы для обучения конкретных нейросетей.

Например, датасеты, содержащие размеченные изображения, подходят для обучения сверточных нейросетей, которые распознают и анализируют изображения. Статистические данные по экономикам разных стран могут быть задействованы для обучения нейросети, прогнозирующей макроэкономические показатели и т. д.

Если речь идет о разработке нейросети, которая будет выполнять относительно узкие задачи типа прогнозирования тарифов на морском транспорте на определенных маршрутах, для ее обучения, скорее всего, потребуется обширный массив данных частных компаний. Несмотря на то, что линейное судоходство осуществляется по заранее объявленному расписанию и тарифам, эта информация не является полностью открытой. Судоходные компании, работающие в контейнерном сегменте, используют следующие практики:

- информация предоставляется по специальной регистрации на официальном сайте, включая, например, предварительную заявку или подписанный договор;
- пересылка детальной информации производится по заявке на перевозку или по факту совместной работы с грузовладельцем;
- тарифы продаются через агентов (очень распространенная практика в КНР), которые распространяют их через рассылки от своего имени.

Периодичность индексации тарифов или публичных рассылок варьируется каждой компанией самостоятельно в зависимости от внутренней политики и изменения внешних факторов. Так, в период с февраля 2022 г. по ноябрь 2023 г. часть компаний индексировала тарифы на контейнерные перевозки практически ежедневно.

Еще одну проблему составляет тот факт, что многие публичные датасеты нельзя использовать в коммерческих целях, что становится проблемой при создании коммерческого программного обеспечения на основе нейросети. В связи с этим процесс сбора входных данных для обучения значительно усложняется.

Помимо относительной закрытости данных, возникает проблема их объема. Чтобы нейросеть смогла эффективно обучиться, находить нелинейные закономерности в данных и выдавать максимально точные прогнозы, необходимо обеспечить достаточный объем этих данных. При этом показатель зависит от задач, возлагаемых на конкретную нейросеть, и может составлять от 10 тыс. до миллиона и более строк данных. Но чем крупнее набор данных, тем лучших результатов можно достичь при обучении нейросети, особенно если данные содержат много уникальных категориальных значений. С точки зрения прогнозирования важна также периодичность данных. Например, для прогнозирования тарифов в будущем месяце необходимо иметь ретроспективные данные за каждый месяц хотя бы последних 3–5 лет, чтобы модель могла выявить и запомнить сезонные и циклические паттерны.

Получение необходимого объема данных о тарифах становится сложной задачей по ряду причин:

- недостаточная цифровизация бизнес-процессов на транспорте в целом и в частности в области транспортной логистики и экспедирования: многие участники могут вести учет перевозок и расчет стоимости в ручном режиме без должной автоматизации и накопления электронных данных;
- коммерческая тайна: компании могут несвоевременно предоставлять подробную информацию или вообще отказывать в доступе к ней из-за опасений конкурентной борьбы, а также в силу особенностей национального законодательства;
- отсутствие систематического хранения данных в некоторых звеньях логистической цепочки: в то время, как судоходные компании сохраняют архив стоимостей перевозок, посредники, через которых действуют некоторые судоходные линии, не ведут баз в принципе, не говоря уже о хранении больших данных.

Нормализация данных

После сбора определенного объема данных следующим необходимым этапом подготовки становится их нормализация. Информация, связанная с морскими перевозками, часто характеризуется неоднородностью форматов, стилей и структуры документов, отсутствием привязки к каким-либо стандартам.

Порты КНР - Порты РФ														
При перевозке в СОС ктк, город сдачи порожнего указывать при размещении букинга														
СТАВКИ НА ФИДЕРНЫЕ СУДА 内贸驳船报价	ПОРТ ВЫХОДА 始发港	ПОРТ ПРИБЫТИЯ 目的港	СУДНО 航次	20GP SOC	40GP/HC SOC	20GP COC (выдача в Xiamen/Shantou штраф за отмену букинга 500USD) 厦门/汕头放箱 取消费500$	20GP COC (выдача в Tianjin/Jinzhou/Weifang/ Qingdao/Taicang/Shanghai/ Ningbo/Guangzhou/Shenzhen/ Dalian 天津/锦州/潍坊/青岛/太仓/上海/宁波/广州/深圳/大连放箱	40GP/HC COC (выдача в Taicang//Tianjin/ Dalian/Jinzhou/ Weifang) 太仓/天津/大连/锦州/潍坊放箱	40GP/HC COC (выдача в Ningbo) 宁波放箱	40GP/HC COC (выдача в Qingdao) 青岛放箱	40GP/HC COC (выдача в Shanghai) 上海放箱	40GP/HC COC (выдача в Guangzhou) 广州放箱	40GP/HC COC (выдача в Shenzhen) 深圳放箱	40GP/HC COC (выдача в Xiamen/Shantou штраф за отмену букинга 500USD) 厦门/汕头放箱 取消费500$
40HC/20GP GUANGZHOU-SHANGHAI 6100RMB/4710RMB 40HC/20GP GUANGZHOU-TAICANG 5800RMB/4450RMB 40HC/20GP GUANGZHOU-QINGDAO 5400RMB/4170RMB	LIANGYUNGANG/ SHANGHAI/NINGBO/ 连云港/上海/宁波	ВСК 东方港	HUI DA9-049N,051N HUI FA-049N,051N,053N	$3 150	$5 050	$4 650	$4 050	$6 600	$6 500	$5 800	$6 400	$6 400	$6 350	$7 050
40HC/20GP NINGBO-SHANGHAI 4700RMB/3100RMB 40HC/20GP NINGBO-TIANJIN 4150RMB/2850RMB	GUANGZHOU(XINSHA)/ XIAMEN/YANTIAN/ RIZHAO 广州(新沙)//厦门/盐田/日照	ВСК 东方港	XIN HE HE 2211N,2213N	$3 150	$4 600	$4 650	$4 050	$6 500	$6 300	$5 600	$6 200	$6 200	$6 200	$6 900
	TIANJIN/RIZHAO 天津/日照	ВСК 东方港	LOA HARMONY 2209N,2211N /XXH1-045N,047N	$3 150	$4 500	$4 650	$4 050	$5 900	$5 800	$5 500	$5 900	$5 900	$5 850	$6 550
40HC/20GP XIAMEN-QINGDAO 4450RMB/3050RMB 40HC/20GP XIAMEN-SHANGHAI 4810RMB/3810RMB 40HC/20GP XIAMEN-TAICANG 4530RMB/3330RMB	GUANGZHOU(XINSHA)/ XIAMEN/YANTIAN/ RIZHAO/TIANJIN/ TAICANG 广州(新沙)//厦门/盐田/日照/天津/太仓	ВАНИНО 瓦尼诺	XXH1-045N,047N LOA HARMONY 2209N,2211N XIN HE HE 2211N,2213N XXH2-035N	$2 900	$4 300	$4 200	$3 900	$5 500	$5 400	$5 300	$5 500	$5 300	$5 500	$5 800
40HC/20GP DALIAN-QINGDAO 4850RMB/3600RMB	BUSAN 釜山	ВАНИНО 瓦尼诺	XXH2-033N,035N	$2 600	$3 700	/	/	/	/	/	/	/	/	/
40HC/20GP DALIAN-SHANGHAI 7210RMB/5035RMB 40HC/20GP DALIAN-TAICANG 6920RMB/4730RMB	BUSAN 釜山	ВСК 东方港	XXH1-045N,047N LOA HARMONY 2209N,2211N; XIN HE HE 2211N,2213N	$2 950	$4 400	/	/	/	/	/	/	/	/	/

备注（Remarks）：

1、Включен фрахт на условиях FILO

2、Города сдачи порожних контейнеров:

20GP COC Москва, Санкт-Петербург,Красноярск, Новосибирск, п.Восточный, Благовещенск, Хабаровск, Иркутск, Омск, Екатеринбург - DROP OFF = 0

40HC COC Москва, Санкт-Петербург, Красноярск, Новосибирск, п.Восточный, Ванино - DROP OFF = 0

40HC COC Екатеринбург DROP OFF 200USD/ктк

3、Включены THC в порту прибытия

4、Пользование COC КТК (20GP, 40HC)

С даты выдачи порожнего ктк до даты сдачи груженого ктк в порт выхода в КНР - 14 суток бесплатного пользования, с 30-х суток выставляется утеря 7500USD/40HC, 5500USD/20GP

С даты выгрузки в порту назначения до даты сдачи порожнего ктк - 45 суток (условия CY-CY), с 75-х суток выставляется утеря 7500USD/40HC, 5500USD/20GP

С даты прибытия на конечную станцию назначения и до даты сдачи порожнего ктк (условия CY-FOR) - 14 суток бесплатного пользования, с 30-х суток выставляется утеря 7500USD/40HC, 5500USD/20GP

По истечению вышеуказанных сроков бесплатного пользования выставляется штраф за СНП в размере - 12USD/40HC/сутки, 8USD/20GP/сутки

5、Не включено: локальные расходы в порту выхода, досмотр, оформление доп специфических док-в на груз, плата за простои, возникшие по независящим от нашей компании причинам, хранение на станции назначения, страховка, раскредитация на станции назначения.

1、此运费为FILO条款

2、空箱可还箱城市：

20GP COC 莫斯科，圣彼得堡。克拉斯诺亚尔斯克、新西伯利亚、东方港、布拉戈维申斯克、哈巴罗夫斯克、伊尔库茨克、鄂木斯克，叶卡 - 免平衡费

40HC COC 莫斯科、圣彼得堡、克拉斯诺亚尔斯克、新西伯利亚、东方港、瓦尼诺 免平衡费

40HC COC 叶卡捷琳堡 DROP OFF 200USD/集装箱

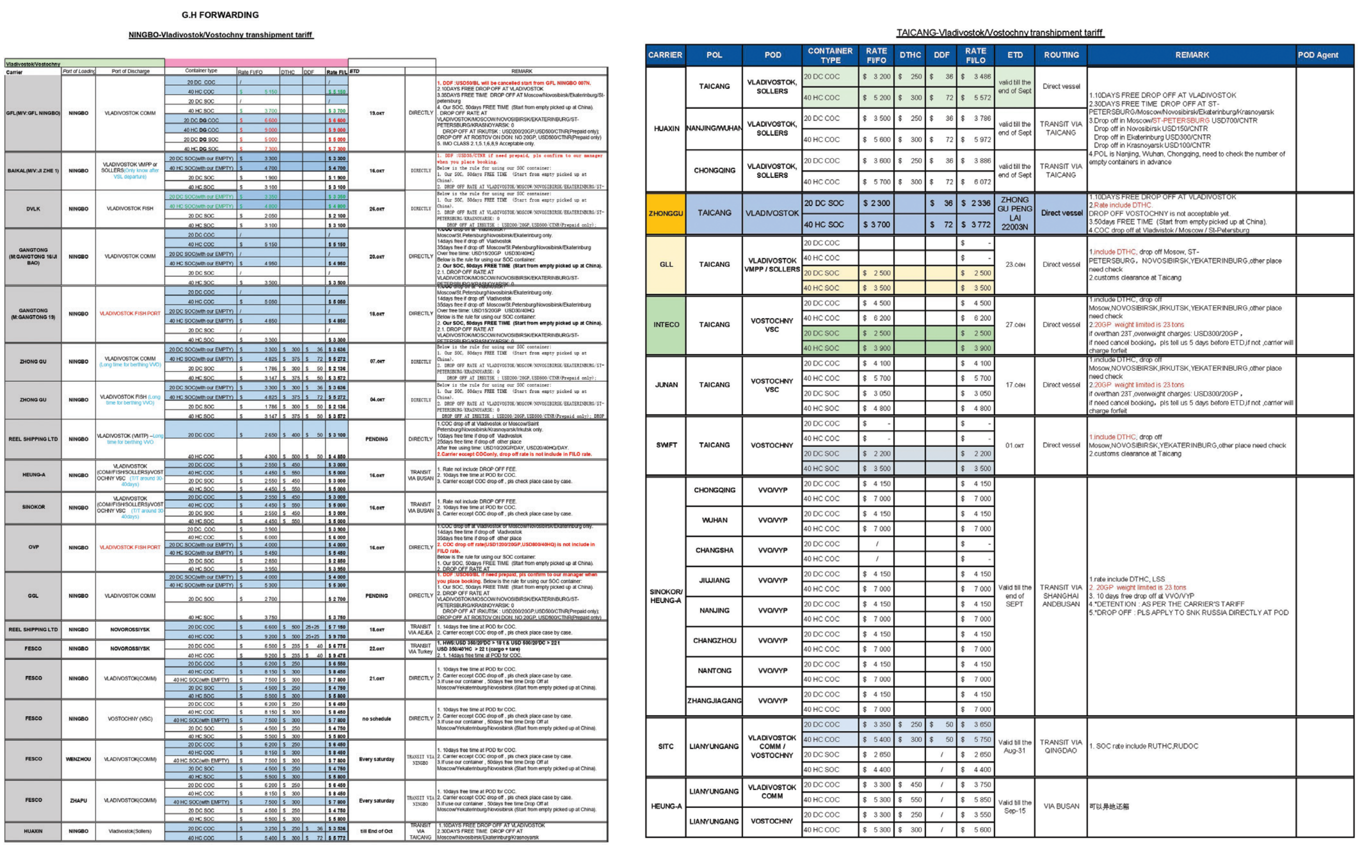

Рис. 1. Примеры тарифов китайских агентов

Одни операторы публикуют в открытом доступе файлы с тарифами в формате pdf, другие распространяют по электронной почте файлы в xlsx и т. д. Данные небольших логистических операторов зачастую не структурированы, зашумлены, содержат грамматические ошибки и пустые, незаполненные ячейки.

При этом экспедиторские компании из Китая предлагают больший перечень опций доставки контейнеров судоходными компаниями (по сравнению с российскими экспедиторами), поэтому рассмотрим обработку обучающих данных на примере их рассылки тарифов в формате MS-Excel. В частности, приведем xlsx файлы с тарифами, предоставляемыми разными операторами (*рис. 1*).

Форматы таблиц в разных листах каждого файла сильно различаются по количеству столбцов и их названиям,

данные в ячейках неоднородны, сами таблицы содержат много лишней информации и пустых ячеек. Это объясняется отсутствием у логистических операторов единого согласованного формата хранения и предоставления данных в сфере транспорта и логистики даже в пределах одной страны.

Если же речь идет о международных перевозках, то различия в форматах данных только усиливаются. Таким образом, после сбора необходимых данных возникает задача их очистки и приведения к единому формату. Чтобы нейросеть могла обучиться на данных, они должны быть структурированы, согласованы, очищены от лишнего шума и не содержать большого количества пропущенных значений.

Следует отметить, что в настоящее время существуют готовые платформы и модули для подготовки и реструктуризации данных, однако они отличаются высокой стоимостью подписки (достигающей нескольких тысяч долларов в год) и потому подходят, как правило, для крупных компаний, которые имеют дело с большими массивами информации и могут позволить себе такие расходы. Для обработки меньшего объема данных существуют универсальные скрипты, способные автоматизировать процесс подготовки данных из разных источников. Однако создание универсального скрипта для логистов становится сложной задачей, если документы имеют разную структуру (как в приведенном примере).

Даже относительно простые варианты — вроде объединения нескольких файлов xlsx в один для обучения нейросети, написание кода, который автоматически бы отбирал значения из указанных колонок или удалял определенные ячейки, — в данном случае не могут быть реализованы без предварительной ручной обработки неструктурированных таблиц.

Для документов с иными форматами и структурой существует ряд техник, способных облегчить процесс подготовки данных: применение инструментов оптического распознавания текста (OCR) для извлечения данных из сканированных документов или изображений (что полезно при работе с нередактируемыми файлами в pdf-формате); алгоритмы для распознавания и приведения различных форматов дат, денежных единиц к единому стандарту, нормализация имен и названий и т. д. Однако в большинстве случаев при работе с распознанными, но неструктурированными документами требуется ручная обработка таблиц, что усложняет и увеличивает время на процесс очистки и подготовки данных **[7]**.

Какой бы тщательной ни была подготовка данных, в них могут содержаться ошибки, допущенные людьми, например, опечатки, различные варианты названий и имен, в том числе и на разных языках. Данная проблема может быть решена путем создания словарей и файлов нормализации имен в используемой программной среде. В нашем случае создан файл Python name_normalization.py, в котором приводятся к одному знаменателю названия и имена, и импортирован в код нейросети: from name_normalization import port_names.

Преобразование данных для машинного обучения

После очистки и нормализации данных необходимо преобразовать их в формат, подходящий для подачи на вход нейронной сети. Обучающие данные могут быть загружены в код нейросети в форматах csv, xlxs, или в виде датафрейма Pandas (в среде Python). В этом случае используем заполненную вручную CSV-таблицу с данными, которую можно загрузить с помощью следующей строки кода:

data = np.genfromtxt('costs_data.csv', delimiter=';', dtype=str).

Наша обучающая выборка представлена *n*-ным количеством строк и семью столбцами переменных (*рис. 2*).

Таким образом, про каждую из n перевозок известно семь характеристик: дата, пункт отправления, пункт назначения, оператор, тип контейнера, размер контейнера, стоимость. Целевая переменная — стоимость — прогнозируется на остальных, входных переменных. В коде нейросети определяем столбцы входных признаков и целевой переменной:

input_cols = [0, 1, 2, 3, 4, 5]
target_col = 6.

В связи с тем, что подготовленные для обучения данные содержат категориальные (строковые) признаки, мы закодировали их в числовые значения, чтобы они могли быть использованы алгоритмами машинного обучения **[8, 9]**. Для данной модели наиболее оптимален метод One-Hot Encoding, который позволяет сохранить всю информацию о категориальных признаках и представить их в виде, подходящем для обучения нейронной сети. Он также позволяет модели выявлять нелинейные зависимости между категориальными признаками и целевой переменной. Его можно импортировать следующим образом:

from sklearn.preprocessing import OneHotEncoder.

Альтернативой One-Hot Encoding является подготовка данных вручную — получение уникальных значений для каждого входного признака, создание списка бинарных признаков. Данный подход позволяет контролировать процесс преобразования данных и вносить коррективы на начальном этапе обучения нейросети. В связи с тем, что к данной модели применяется максимально контролируемое обучение, а набор данных пока не большой, выбрано ручное кодирование признаков.

Сначала определены уникальные значения для каждого входного признака (столбца) в наборе данных. Затем для каждой строки данных создан список бинарных признаков (0 или 1), где единица соответствует значению, которое присутствует в этой строке.

На завершающем этапе данные разделены на обучающую, валидационную и тестовую выборки в пропорциях 60/20/20. Обучающая выборка используется непосредственно для обучения модели, а тестовая — для ее финальной

Дата	Оператор	Пункт отправления	Пункт назначения	Тип контейнера	Размер контейнера	Стоимость
01.09.2022	HUAXIN	TAICANG	SOLLERS	COC	20	3486
01.09.2022	HUAXIN	TAICANG	SOLLERS	COC	40	5572
01.09.2022	HUAXIN	NANJING	SOLLERS	COC	20	3786
01.09.2022	HUAXIN	NANJING	SOLLERS	COC	40	5972
01.09.2022	HUAXIN	WUHAN	SOLLERS	COC	20	3786
01.09.2022	HUAXIN	WUHAN	SOLLERS	COC	40	5972
01.09.2022	HUAXIN	CHONGQING	SOLLERS	COC	20	3886
01.09.2022	HUAXIN	CHONGQING	SOLLERS	COC	40	6072
01.09.2022	ZHONGGU	TAICANG	VLADIVOSTOK	SOC	20	2336

Рис. 2. CSV-таблица с данными для обучения нейросети

оценки после обучения. Валидационная выборка используется для настройки гиперпараметров модели и промежуточной оценки во время обучения. Ее добавление также предотвращает переобучение модели:

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42).
X_train, X_val, y_train, y_val = train_test_split (X_train, y_train, test_size=0.25, random_state=42).

Следует отметить, что создание качественных и объемных данных для обучения нейросети не гарантирует автоматически отсутствия проблем в будущем. Обучение нейросети — непрерывный процесс и после ее базового обучения будут необходимы новые данные, которые она могла бы обрабатывать, оценивать и учиться на них.

Соответственно пользователь нейросети вновь столкнется с указанными проблемами сбора и обработки информации. Решить проблему могла бы выработка единого стандарта документов хотя бы между участниками отдельной логистической цепочки, интеграция внутренних систем компаний через API или своевременный обмен актуальной информацией о тарифах, поступающей в корпоративные системы. Наряду с проблемой качества, это частично решило бы и проблему количества данных, которая в данный момент существует в сфере транспорта и логистики **[9]**.

Заключение

На пути к цифровизации на транспорте возникают разнообразные проблемы: высокая стоимость внедрения новых решений; торможение отработанных технологических процессов; необходимость длительной обкатки существующих продуктов, не пригодных в готовом виде для использования конкретной компанией ввиду дефицита IT-решений для транспортной отрасли в целом и для экспедиторской и логистической деятельности в частности.

Уход крупных западных компаний, успешно использовавших ERP и BMP системы, в том числе в логистическом сегменте, еще более усложнил ситуацию. Рынок же отечественных решений, потребность в которых крайне высока и в других отраслях, не поспевает за спросом. Выходом из ситуации могут быть поэтапные разработки программных продуктов для каждого элемента внутри сегментов транспортной логистики или экспедирования, включая возможности искусственного интеллекта.

Подготовка данных является критически важным этапом при обучении нейронных сетей для анализа и прогнозирования тарифов на морском транспорте. Решение проблем, связанных с получением, нормализацией и преобразованием данных, в том числе с использованием категориального преобразования, позволяет значительно повысить производительность и точность прогнозов нейронной сети.

Несмотря на очевидные сложности, связанные с ограниченным объемом данных для обучения, существуют различные подходы, которые позволяют эффективно преодолевать данные ограничения. Так, диверсификация источников данных, применение алгоритмических методов увеличения выборки, а также совершенствование архитектуры нейронных сетей способны компенсировать нехватку информации и обеспечить построение точных прогнозных моделей. Т

Источники

1. Иващенко М. Г., Нестеренков С. Н., Ситников А. В. Эффективное использование нейронных сетей в решении задач автоматизации в логистике // BIG DATA и анализ высокого уровня: сборник научных статей X Международной научно-практической конференции в двух частях, 13 марта 2024 г., Минск. Минск: Белорусский государственный университет информатики и радиоэлектроники, 2024. С. 368–372. EDN: MQURLR.
2. Ли Ш., Сюе В. Особенности применения современных цифровых технологий для оптимизации логистических // Russian Economic Bulletin. 2024. Т. 7. № 4. С. 119–126. DOI: 10.58224/2658-5286-2024-7-4-119-126. EDN: QWWTFJ.
3. Дмитриев А. В. Цифровизация транспортно-логистических услуг на основе применения технологии дополненной реальности // Вестник ЮУрГУ. Серия: Экономика и менеджмент. 2018. № 2. URL: https://cyberleninka.ru/article/n/tsifrovizatsiya-transportno-logisticheskih-uslug-na-osnove-primeneniya-tehnologii-dopolnennoy-realnosti (дата обращения: 30.11.2024).
4. Баранникова А. О., Вороненко А. К., Смирнов С. М. О некоторых аспектах применения технологий Индустрии 4.0 в морских портах // Транспортное дело России. 2024. № 2. С. 122–126. EDN: DFLKEM.
5. Астраханцева И. А., Кутузова А. С., Астраханцев Р. Г. Рекуррентные нейронные сети для прогнозирования региональной инфляции // Научные труды Вольного экономического общества России. 2020. № 3. URL: https://cyberleninka.ru/article/n/rekurrentnye-neyronnye-seti-dlya-prognozirovaniya-regionalnoy-inflyatsii (дата обращения: 01.12.2024).
6. Афанасьев С., Смирнова А., Ахметсафин И., Молоканов И. LGD-модели для розничного кредитования. Часть 2: разработка «ядра» модели // Риск-менеджмент в кредитной организации. 2021. № 4.
7. Качановский Ю. П., Коротков Е. А. Предобработка данных для обучения нейронной сети // Фундаментальные исследования. 2011. № 12–1. С. 117–120. EDN: OFYRIV.
8. Жанаева С. Б. К вопросу о подготовке данных при разработке модели нейронной сети // Вестник СибГУТ И. 2022. Т. 16. № 4. С. 69–79. DOI: 10.55648/1998-6920-2022-16-4-69-79. EDN: DJHMSW.
9. Маслов И. А. Оптическое распознавание символов в информационных системах и проблемы внедрения // E-Scio. 2023. № 3 (78). URL: https://cyberleninka.ru/article/n/opticheskoe-raspoznavanie-simvolov-v-informatsionnyh-sistemah-i-problemy-vnedreniya (дата обращения: 29.11.2024).