

УДК 646.21

Построение краткосрочного прогноза количества вагонов на станции и путях необщего пользования

В. А. Ламехов, Е. К. Коровяковский

Петербургский государственный университет путей сообщения Императора Александра I, Россия, 190031, Санкт-Петербург, Московский пр., 9

Для цитирования: Ламехов В. А., Коровяковский Е. К. Построение краткосрочного прогноза количества вагонов на станции и путях необщего пользования // Известия Петербургского государственного университета путей сообщения. СПб.: ПГУПС, 2024. Т. 21, вып. 1. С. 199–216. DOI: 10.20295/1815-588X-2024-01-199-216

Аннотация

Цель: провести анализ исходных данных, определить пропуски и выбросы в данных, разделить данные на временные интервалы, вычислить коэффициенты корреляции, частичной автокорреляции, кросс-корреляции, проанализировать тренд и сезонность полученных временных рядов. Используя авторегрессионные модели, модели машинного обучения, нейронечеткие модели, построить прогнозы временного ряда и определить качество полученных прогнозов. **Методы:** плотность точек, автокорреляция, частичная автокорреляция, кросс-корреляция, тест Фостера — Стюарта, тест Дикки — Фуллера, ARMA, SARMA, MLP, Encoder-Decoder LSTM, TSK, Fuzzy-Partitions, SCRG, Transformers. **Результаты:** получены оценки точности прогнозов выбранных моделей, сопоставлены результаты работы прогнозных моделей обученных на разных выборках исходных данных. Сделаны выводы об эффективности использования исследуемых прогнозных моделей. **Практическая значимость:** исследована способность выбранных моделей к построению краткосрочных прогнозов количества вагонов на станции, проанализированы факторы, влияющие на точность получаемых прогнозов.

Ключевые слова: нечеткие нейронные сети, нейронные сети, авторегрессионные модели, анализ работы станции, прогнозирование.

Введение

Наличие отклонений технологического времени совершения операций с вагонами на объектах железнодорожного транспорта, примыкающих к железнодорожным станциям, от нормативного, предусмотренного в Инструкции о порядке обслуживания и организации движения на железнодорожном пути необщего пользования (Инструкция) или Единым технологическим процессом (ЕТП), делает невозможным использование данного нормативного времени для построения точных прогнозов работы исследуемых объектов. Решением данной проблемы могут

выступать различные прогнозные модели, использующие временные ряды. Потребность в построении прогнозных моделей на железнодорожном транспорте рассмотрена в работе [1].

В данной работе рассматривается использование статистических методов, методов машинного обучения, с использованием нейронных сетей различных архитектур, в том числе нечеткие нейронные сети. Программное обеспечение, использованное для построения нечетких нейронных сетей и их адаптации для прогнозирования временных рядов на

железнодорожном транспорте, приведено в [2]. Использование алгоритмов нечеткой математики приведено в работах [3, 4].

Периодичность сбора данных, точность внесенной информации в информационные системы ОАО «РЖД», специфика работы станций и путей необщего пользования (ПНОП) требует специальной предобработки данных для использования в прогнозных моделях.

Характеристика полученных исходных данных

В данном разделе рассматриваются исходные данные, полученные из системы «Автоматизированное рабочее место анализа отчетных и пономерных данных» (АРМ ОНД) ОАО «РЖД».

Для использования получаемых исходных данных, собираемых с использованием системы АРМ ОНД ОАО «РЖД», необходимо провести их предобработку. Предобработка исходных данных разделена на несколько этапов:

1. Приведение исходных данных к дискретным временным рядам методом накопления. На данном этапе решается задача определения интервала наблюдения величин. Так как для оценки взаимосвязи значений временных рядов используется коэффициент корреляции или автокорреляции, основанный на коэффициенте корреляции Пирсона, то для получения несмещенной и эффективной оценки коэффициента корреляции выборка должна быть распределена нормально.
2. Анализ полученных временных рядов на наличие выбросов. На данном этапе анализируются полученные временные ряды. Используя группу предложенных

методов: метод поиска выбросов, основанный на алгоритме плотности точек, и метод скользящего среднего, определяются и заполняются обнаруженные в полученных временных рядах выбросы.

3. Анализ коэффициентов автокорреляции выбранных интервалов временных рядов. Для определения возможности прогнозирования одномерных временных рядов определяются коэффициенты автокорреляции и частичной автокорреляции, а также выделяются их значимые компоненты для выбранных интервалов разбиения исходных временных рядов.
4. Анализ наличия тренда и сезонности в рассматриваемых временных рядах. Используя статистические тесты Фостера – Стюарта и Дикки – Фуллера, анализируется наличие трендовых составляющих в среднем и дисперсии рассматриваемых временных рядов, временные ряды проверяются на стационарность.

Для построения прогнозных моделей была собрана повагонная информация о работе грузовой станции Октябрьской железной дороги с примыкающим ПНОП с рабочим парком превышающим 400 вагонов в сутки. Данный ПНОП характеризуется преобладанием погрузки над выгрузкой вагонов и коэффициентом сдвоенных операций близким к единице.

В табл. 1 приведена часть посуточных данных о работе станции, полученных из АРМ ОНД. Данные собраны за период с 16.07.2022 по 16.10.2023.

Исследуемых исходных данных достаточно для того, чтобы составить равноотстоящие интервальные временные ряды прибытия,

ТАБЛИЦА 1. Посуточные показатели работы станции

Дата	Грузовые транзитные без переработки	Грузовые транзитные с переработкой	Местные	Погрузка, тонны	Погрузка, вагоны	Выгрузка, вагоны	Всего занято при грузовых операциях	Всего освобождено вагонов при грузовых операциях
16.07.2022	0	44	435	31422	444	22	444	22
17.07.2022	0	14	494	31275	444	55	444	55
18.07.2022

отправления, подачи, уборки и погрузки вагонов на станции. Необходимость приведения исходных данных к формату временных рядов с точки зрения прогнозирования будущих значений прибытия, отправления, подачи, уборки и погрузки вагонов обусловлена малой значимостью большинства параметров исходных данных, схожестью значений временных параметров исследуемых исходных данных, а также наличием аномальных значений в исходных данных.

В табл. 2, 3 и 4 приведены обезличенные данные об операциях и времени операций, производимых с каждым вагоном на станции и ПНОП, на основании которых строятся временные ряды для их последующей обработки и использования в прогнозных моделях.

На рис. 1 приведены графики наличия местных вагонов на станции, прибытия, отправления, погрузки вагонов. Значения, представленные на графиках временных рядов, получены для первого временного ряда из посуточных показателей работы станции, представленных в табл. 1; для последующих временных рядов — путем группировки значений исходных данных по 24-часовым интервалам.

В графиках, полученных путем группирования по интервалам, наблюдаются выбросы. В ноябре и декабре 2022 года, мае 2023 года отсутствуют данные для нескольких последовательных значений временного ряда. Отсутствующие данные о прибытии вагонов влияют на снижение значений погрузки вагонов на ноябрь 2022 года для временного ряда погрузки вагонов.

На рис. 2 приведено сопоставление временных рядов местных вагонов на станции и погрузки вагонов. На данном рисунке отчетливо видны отклонения значений временных рядов.

Помимо указанных выше выбросов, в исходных данных наблюдаются значения параметров, которые не могут быть объяснены технологией работы ПНОП, указанной в Инструкции. Так, например, в табл. 5 приведены данные о времени погрузки вагонов на ПНОП, которые совпадают для множества последовательно погруженных вагонов.

В исследуемых повагонных данных также наблюдаются значения-дубликаты.

Для полученных исходных данных невозможно восстановить точное время погрузки и выгрузки вагонов на ПНОП.

ТАБЛИЦА 2. Данные о нахождении вагонов на станции

Номер вагона	Дата и время начала простоя	Простой вагонов, всего	Дата и время окончания простоя	На ПНОП, на ответственности ОАО «РЖД»	На ПНОП, на ответственности собственника	На ПНОП, в отстое	Дата и время начала простоя на ПНОП	Простой вагона на ПНОП	Дата и время окончания простоя на ПНОП	На ПНОП, на ответственности ОАО «РЖД»	На ПНОП, на ответственности грузоотправителя
5#####	16.07 15:20	34,85	18.07 2:11	23,26	0,00	0,00	16.07 18:05	11,58	17.07 5:40	0,01	11,56
5#####

ТАБЛИЦА 3. Данные о погрузке вагонов

Номер вагона	Груз	Администрация собственника	Род вагона	Тип парка	Местный/ транзитный	Дата грузовой операции	Масса груза, кг	Станция назначения вагона	Код грузоотправителя	Код грузополучателя
5#####	РУДА ЖЕЛЕЗНАЯ	РЖД	ПВ	собственный	транзитный	18.07 4:52	70600	КОШТА	###	###
5#####

ТАБЛИЦА 4. Данные о выгрузке вагонов

Номер вагона	Груз	Администрация собственника	Род вагона	Тип парка	Местный/ транзитный	Дата грузовой операции	Масса груза, кг	Станция назначения вагона	Код грузоотправителя	Код грузополучателя
5#####	НЕФТЬ	РЖД	ЦС	собственный	местный	18.07 2:04	58300	####	####	####
5#####

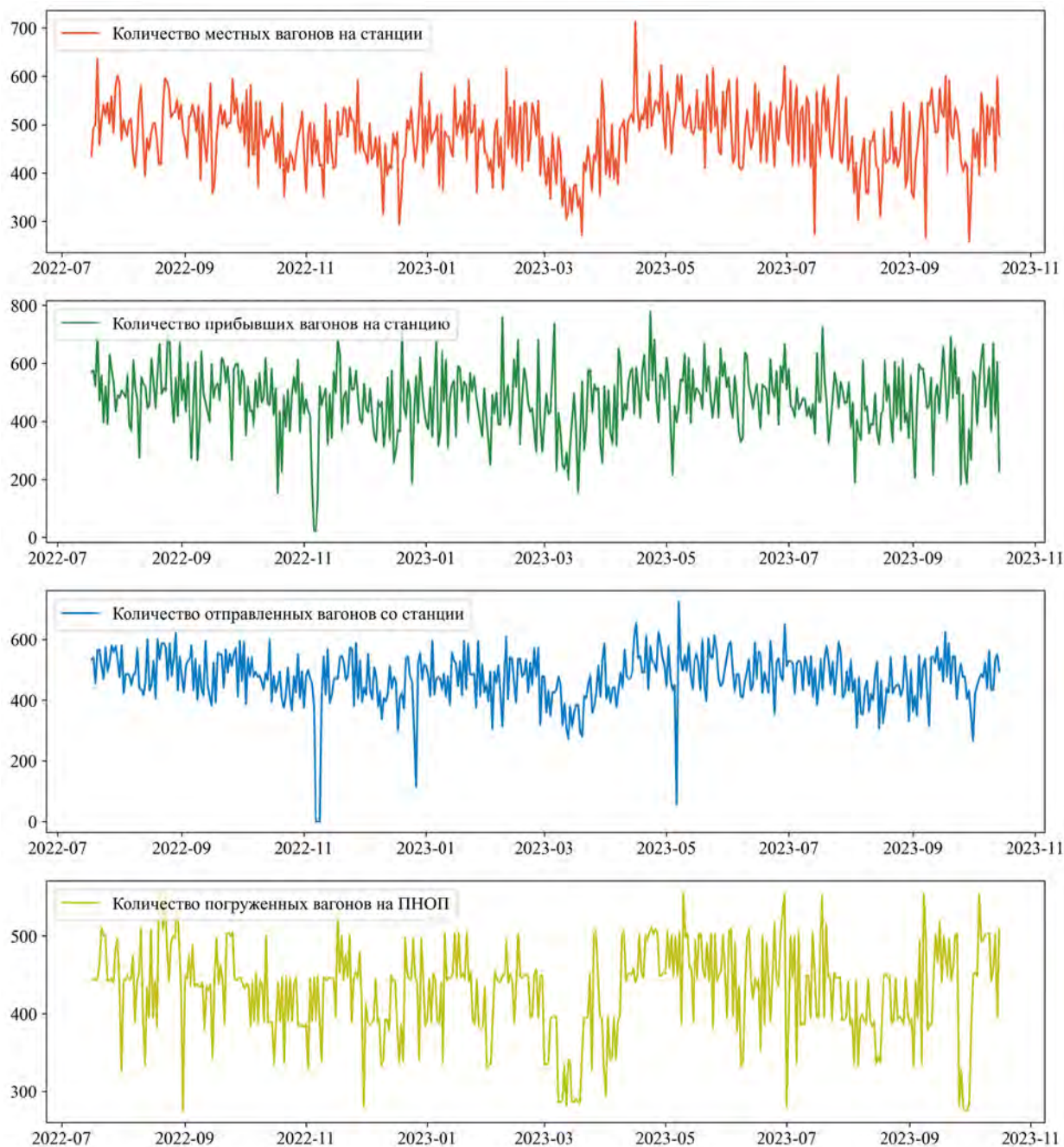


Рис. 1. Временные ряды наличия местных вагонов, прибытия, отправления, погрузки вагонов

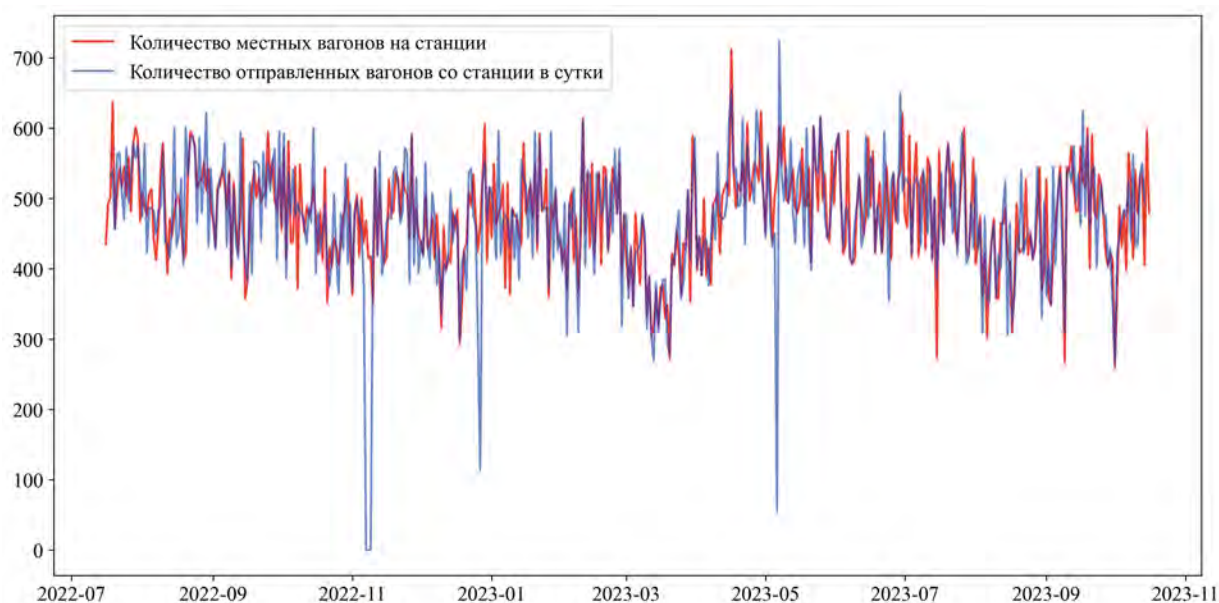


Рис. 2. Временные ряды наличия местных вагонов и отправления вагонов

ТАБЛИЦА 5. Данные о времени погрузки вагонов на ПНОП

Номер вагона	Дата грузовой операции	Масса груза, кг	Номер вагона	Дата грузовой операции	Масса груза, кг	Номер вагона	Дата грузовой операции	Масса груза, кг
5#####	18.07.22 15:10	74400	5#####	18.07.22 15:10	69100	5#####	18.07.22 15:11	76500
6#####	18.07.22 15:10	74400	5#####	18.07.22 15:10	69100	6#####	18.07.22 15:11	76600
5#####	18.07.22 15:10	74700	5#####	18.07.22 15:10	69600	5#####	18.07.22 15:11	76700

Предобработка полученных исходных данных

В данном разделе описаны анализ и предобработка полученных исходных данных, наличие особенностей в исходных данных, препятствующих использованию всего набора данных для построения прогнозов.

Для получения как можно большего количества исходных данных временного ряда подбирается размер интервала, на котором будут группироваться повагонные исходные данные. Интервал подбирается таким образом, чтобы

распределение случайной величины соответствовало нормальному закону распределения. Для оценки наиболее значимых компонент временных рядов, количества ретроспективных данных, подаваемых в модель для построения прогноза, используется коэффициент автокорреляции и частичной автокорреляции. Так как данные коэффициенты основаны на коэффициенте корреляции Пирсона, то для получения несмещенной и эффективной оценки коэффициента корреляции выборка должна быть распределена нормально.

На рис. 3 приведены гистограммы распределения значений получаемой случайной величины в зависимости от интервала разбиения исходных данных.

Как видно из гистограмм, наиболее приближенные к нормальному закону распределения значения случайной величины получены при значении интервала, равного 12, однако для увеличения числа значений временного ряда и обучения выбранных моделей будет использоваться интервал, равный 8 часам.

Для автоматизации определения выбросов в полученных временных рядах используется метод на основе плотности точек. Суть метода заключается в том, что для каждой точки, для всех возможных окон заданного размера,

содержащих эту точку, определяется наличие количества соседей точки, удовлетворяющих функции расстояния. Если такое количество соседей превышает установленную величину, то точка не является выбросом.

На рис. 4 приведены значения временного ряда, определенные как выбросы, для временного ряда прибытия вагонов на станцию по 8-часовым интервалам. Полученные значения временных рядов, определенные как выбросы, заменяются методом скользящего среднего.

Для определения наличия тренда и сезонности в данных проводятся тесты Фостера – Стюарта и Дикки – Фуллера. Для оценки количества ретроспективных данных, подаваемых в модели для получения прогноза,

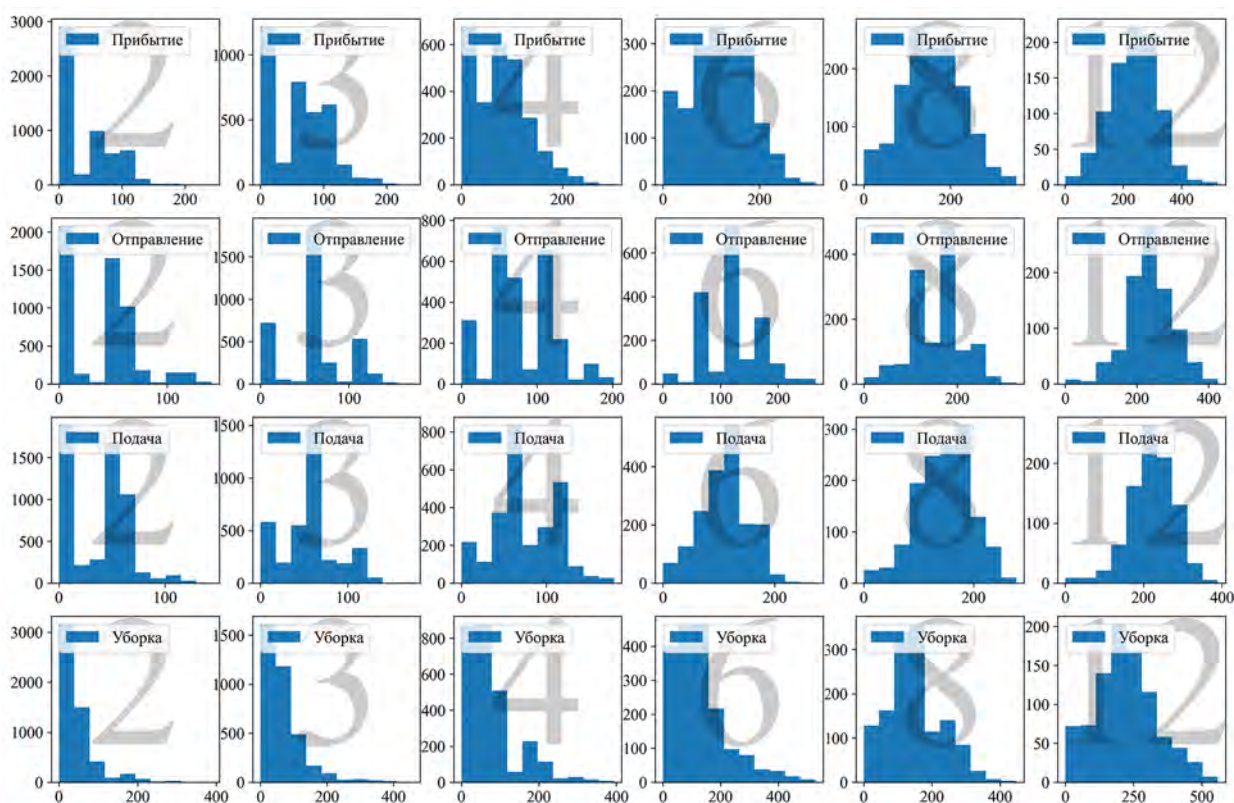


Рис. 3. Гистограммы распределения случайной величины при различных временных интервалах

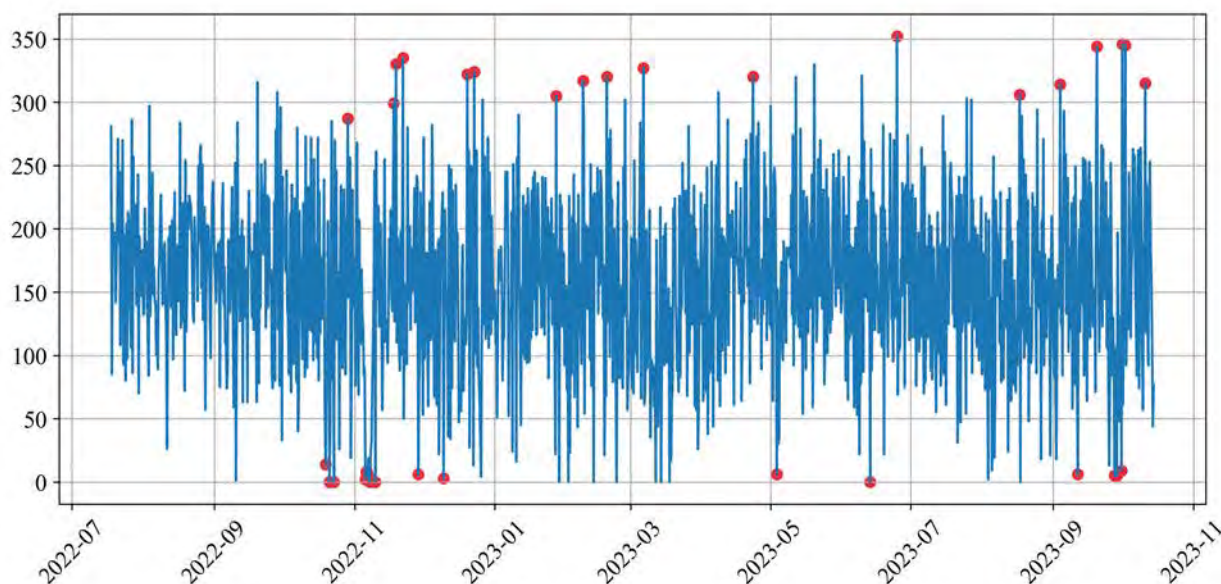


Рис. 4. Выбросы для временного ряда прибытия вагонов

строятся функции автокорреляции и частичной автокорреляции значений временного ряда.

Так как специфика работы ПНОП может подразумевать изменения технологии, связанной с техническими изменениями на ПНОП, проведением плановых ремонтных работ на ПНОП, станции или железнодорожных линиях, изменением конъюнктуры груза, рода подвижного состава, то использование всего временного ряда за отчетный период может вызывать возникновение погрешностей при прогнозировании. Поэтому для последующих исследований прогнозных модели будут обучаться на данных, полученных делением исходных на 3-, 6-, 9-месячные отрезки, помимо полного набора данных.

Очевидно, что при постоянной неизменной технологии работы как станции, так и ПНОП должна наблюдаться одинаковая зависимость между количеством прибывших, отправленных, поданных и убранных вагонов. Изменения данной зависимости во

времени могут свидетельствовать об изменении технологии работы. Для наглядности сделанных выводов на рис. 5 приведены значения коэффициента кросс-корреляции пар временных рядов прибытия и подачи вагонов на ПНОП, подачи и уборки вагонов с ПНОП и уборки, и отправления вагонов со станции, полученных для исследуемых временных интервалов.

На рис. 6 приведены графики частичной автокорреляции для временных рядов прибытия, отправления, подачи и уборки для каждого рассматриваемого временного периода.

Как видно из рис. 5 и 6, с увеличением числа исходных данных, используемых для построения коэффициентов частичной автокорреляции и кросс-корреляции для некоторых временных рядов, наблюдается изменение степени значимости компонент, что свидетельствует о неоднородности технологии работы исследуемых станции и ПНОП за рассматриваемый период. Таким образом,

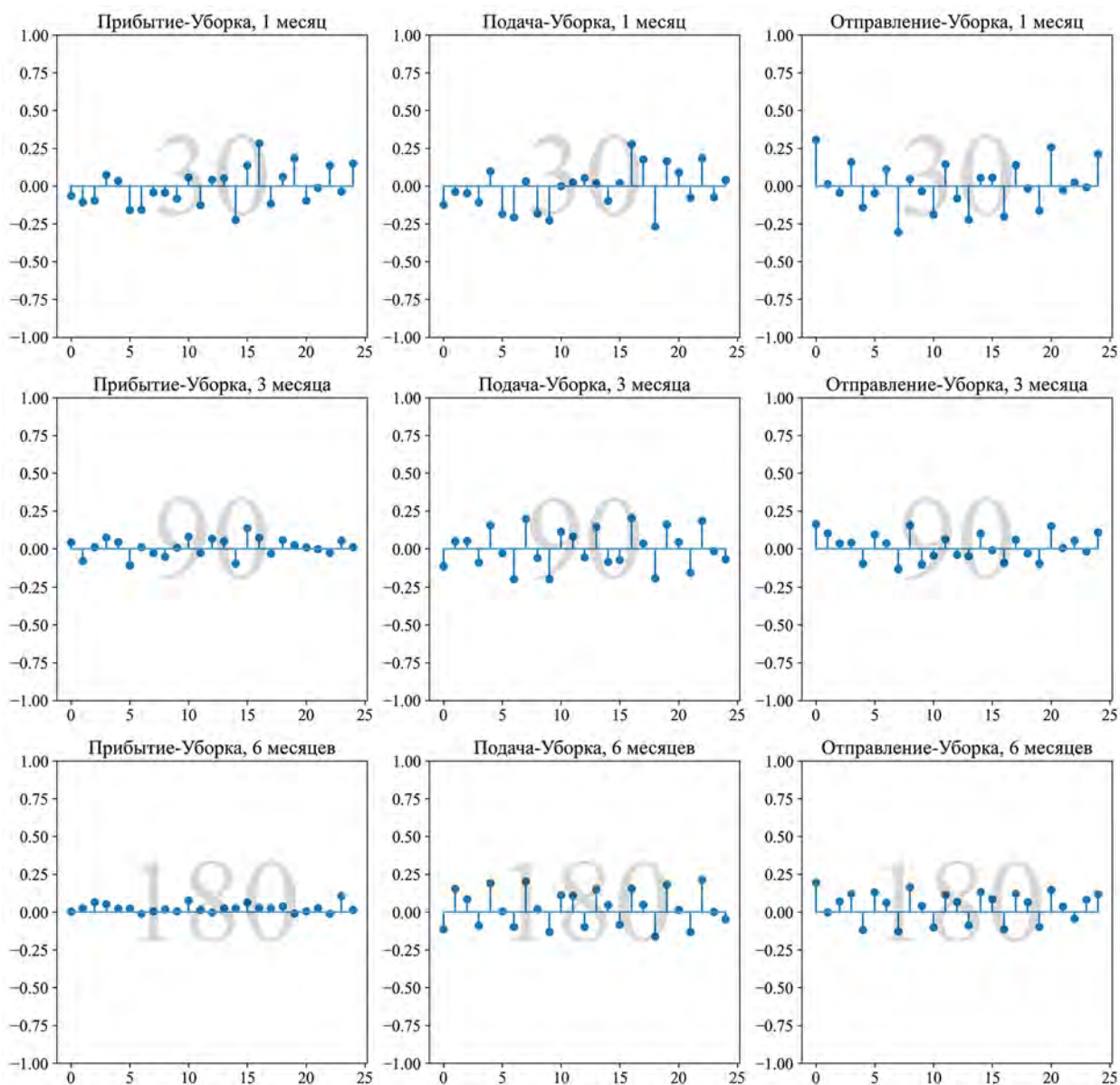


Рис. 5. Графики коэффициента кросс-корреляции для исследуемых пар временных рядов для каждого рассматриваемого временного периода

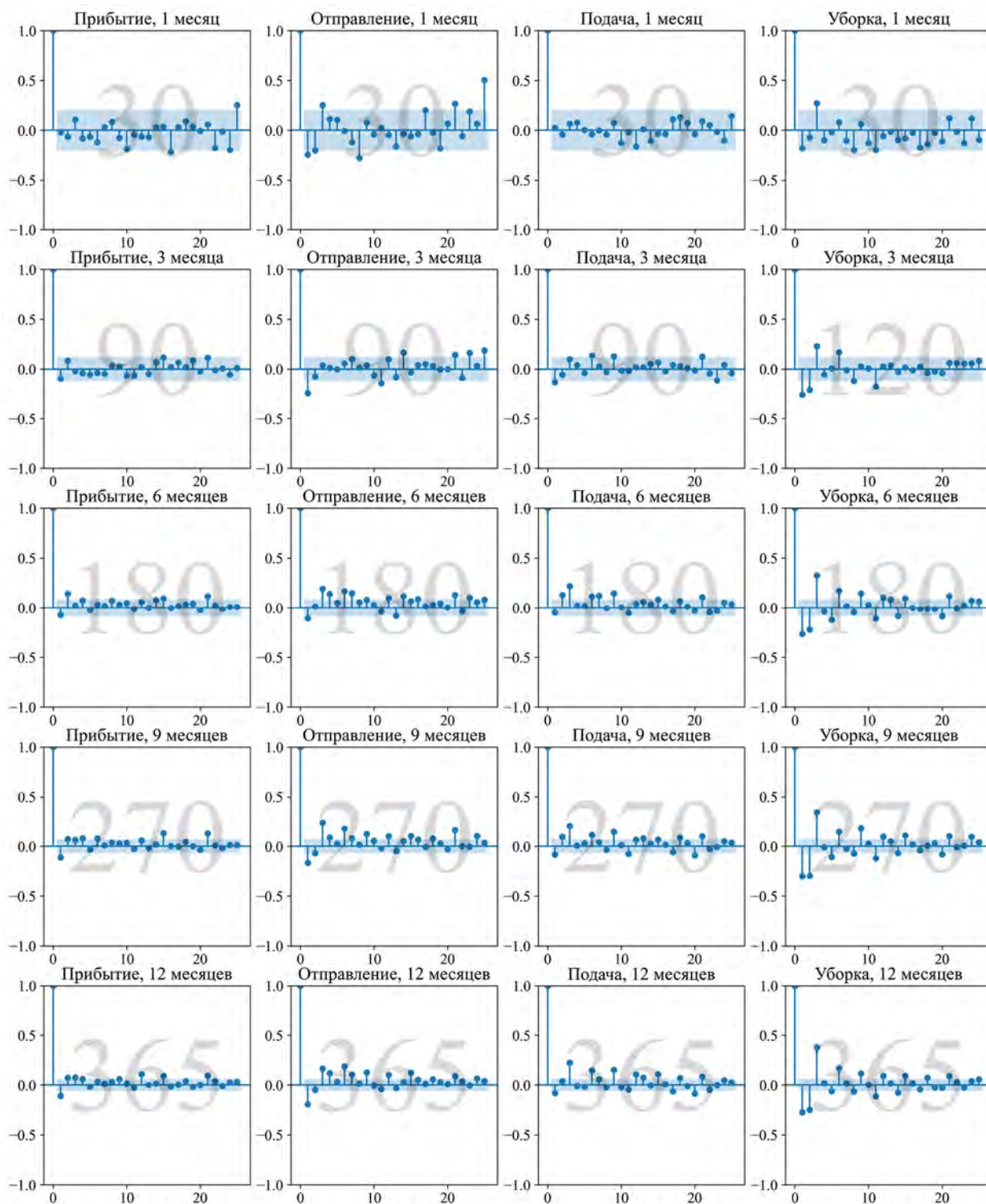


Рис. 6. Графики коэффициента частичной автокорреляции для временных рядов прибытия, отправления, подачи и уборки вагонов для каждого рассматриваемого временного периода

использование всего набора данных для построения краткосрочного прогноза может приводить к ухудшению качества прогноза.

Проведенные тесты Фостера – Стюарта и Дикки – Фуллера показали отсутствие сезонных и трендовых компонент в исследуемых временных рядах за весь отчетный период.

Основываясь на полученных выше исходных данных, необходимо построить краткосрочный прогноз. Для построения прогнозов будут использоваться временные ряды прибытия, отправления и погрузки вагонов, так как для них наблюдаются различные закономерности изменения величины коэффициентов автокорреляции и частичной автокорреляции для исследуемых временных периодов.

Необходимо определить, использование временных рядов какой величины рассматриваемого периода дает наилучший результат прогноза. Следует определить дальность прогноза для различных прогнозных моделей.

Методы, используемые в данной работе, разделены на три группы. В качестве базовых методов, результаты применения которых будут использоваться в качестве точки отсчета качества получаемых прогнозов, используются авторегрессионные методы (ARMA). Вторая группа методов использует нечеткие нейронные модели типа TSK [8, 9, 10] с различными функциями определения параметров нечетких правил. Третья группа методов использует нейронные сети: многослойный перцептрон, сети долгой краткосрочной памяти (LSTM) различных архитектур.

Таким образом, для построения прогнозов будут использоваться ARMA-, TSK-, MLP-, Encoder-Decoder LSTM-, Transformer-модели,

обученные на временных рядах, полученных разбиением исходных данных на 8-часовые интервалы, за разные интервалы времени — от 3 до 12 месяцев.

Построение прогнозов

В данном разделе рассматриваются прогнозы, полученные выбранными моделями. Для наиболее показательных случаев приводятся графики прогнозных и действительных значений временных рядов. В качестве критериев оценки качества получаемых прогнозов выбраны MSE-, RMSE-, MAPE- и MSPE-метрики. Приведены значения метрик для наиболее показательных значений полученных прогнозов. Обзор методов, используемых для прогнозирования временных рядов, приведен в работе [5].

В табл. 6 приведены результаты прогнозов ARMA-моделей для временных рядов прибытия, отправления, погрузки вагонов за полгода.

На рис. 7 представлен типичный результат прогноза ARMA-моделей.

В табл. 7 приведены результаты прогнозов MLP-моделей для временных рядов прибытия, отправления, погрузки вагонов за полгода.

На рис. 8 представлен результат прогноза MLP-моделей для временных рядов прибытия, отправления вагонов за полгода.

В табл. 8 приведены результаты прогнозов Encoder-Decoder LSTM-моделей для временных рядов прибытия, отправления, погрузки вагонов за 8 месяцев.

На рис. 9 представлен результат прогноза Encoder-Decoder LSTM-моделей для временных рядов прибытия, отправления вагонов за 8 месяцев.

ТАБЛИЦА 6. Результаты прогнозов ARMA за полгода

Временной ряд	MSE	RMSE	MAPE	MSPE
Дальность прогноза — 20 значений				
Прибытие вагонов	2650,82	51,49	39,79	63,08
Отправление вагонов	1107,13	33,27	22,91	47,86
Погрузка вагонов	1701,55	41,25	34,81	59,0
Дальность прогноза — 10 значений				
Прибытие вагонов	2705,04	52,01	34,54	58,77
Отправление вагонов	964,22	31,05	21,32	46,17
Погрузка вагонов	833,16	28,86	23,62	48,61

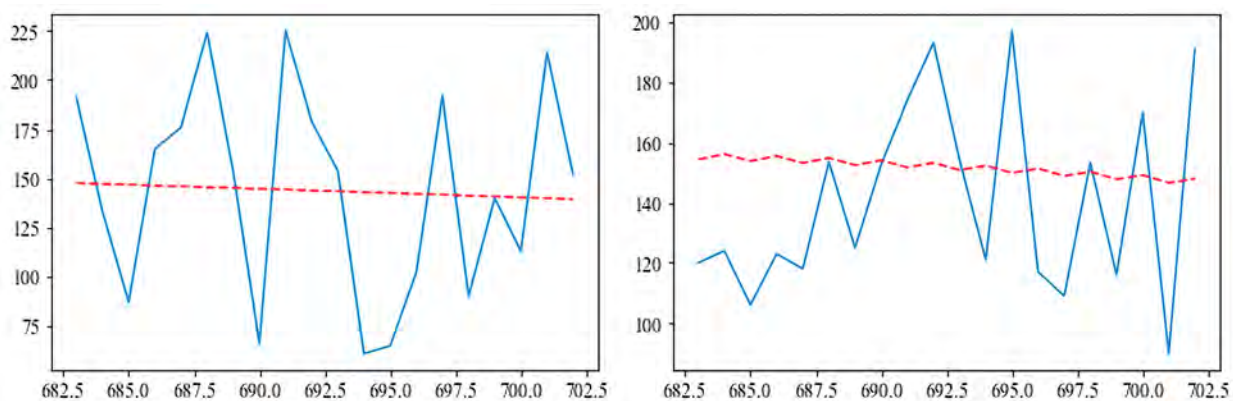


Рис. 7. Прогноз ARMA-моделей

ТАБЛИЦА 7. Результаты прогнозов MLP-моделей за полгода

Временной ряд	MSE	RMSE	MAPE	MSPE
Дальность прогноза — 20 значений				
Прибытие вагонов	2789,46	52,82	38,1	61,72
Отправление вагонов	870,67	29,51	18,46	42,97
Погрузка вагонов	1675,66	40,93	21,14	45,98
Дальность прогноза — 10 значений				
Прибытие вагонов	3101,38	55,69	35,01	59,17
Отправление вагонов	563,65	23,74	14,55	38,14
Погрузка вагонов	1002,1	31,66	15,17	38,95

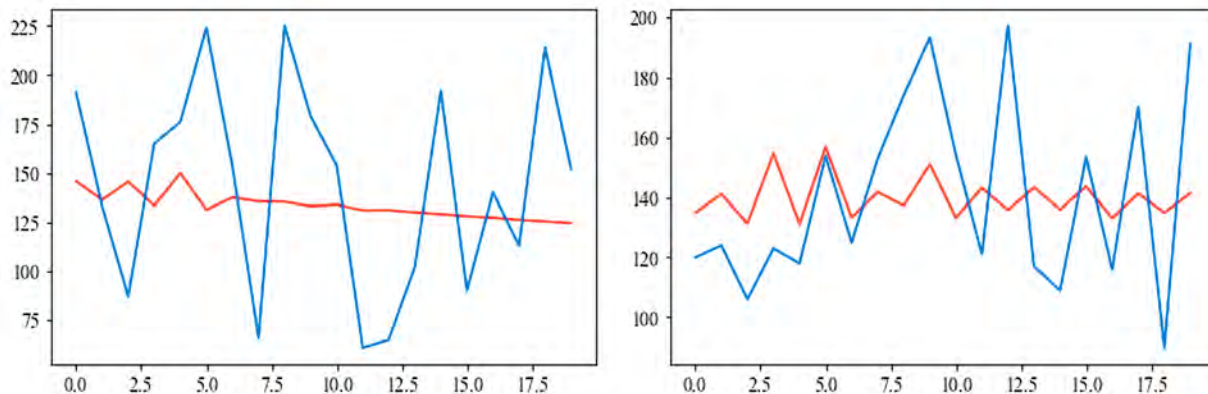


Рис. 8. Прогноз MLP-моделей

ТАБЛИЦА 8. Результаты прогнозов Encoder-Decoder LSTM-моделей за 8 месяцев

Временной ряд	MSE	RMSE	MAPE	MSPE
Дальность прогноза — 20 значений				
Прибытие вагонов	2318,65	48,15	34,99	59,15
Отправление вагонов	519,58	22,79	13,0	36,06
Погрузка вагонов	1881,79	43,38	27,47	52,41
Дальность прогноза — 10 значений				
Прибытие вагонов	1763,31	41,99	41,19	64,18
Отправление вагонов	210,68	14,51	7,58	27,54
Погрузка вагонов	2394,03	48,93	31,63	56,24

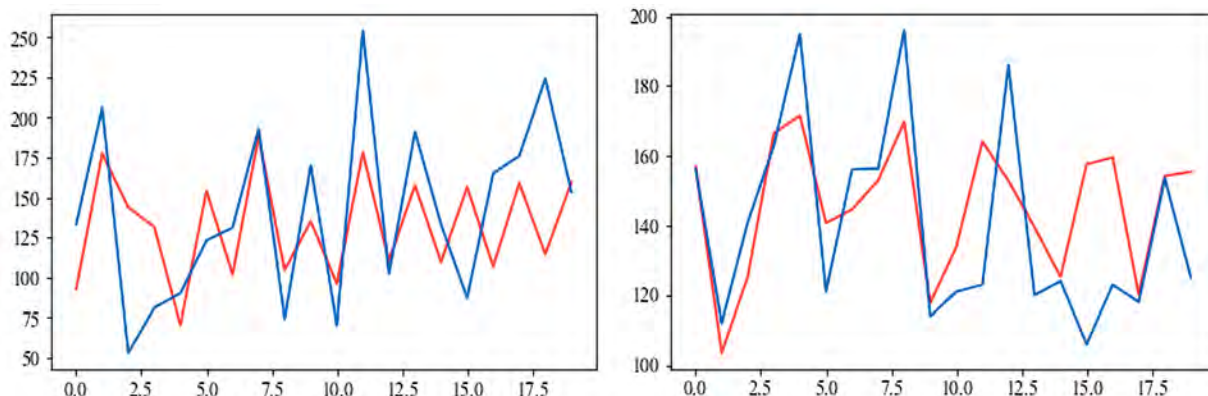


Рис. 9. Прогноз Encoder-Decoder LSTM-моделей за 8 месяцев

На рис. 10 приведены результаты прогноза Encoder-Decoder LSTM-моделей для временных рядов за весь исследуемый период.

Результаты прогнозов, полученные с использованием Encoder-Decoder LSTM-моделей, являются лучшими среди всех рассматриваемых моделей. Это связано как с отсутствием необходимости в обучении на больших наборах данных, так и с лучшей возможностью запоминания краткосрочных зависимостей.

Результаты прогнозов, полученные с использованием TSK-моделей и моделей транс-

формеров, приведены в табл. 9 и 10. Использование нейронечетких моделей и различных алгоритмов их обучения рассмотрено в работах [6–9]. Использование трансформеров для прогнозирования временных рядов рассмотрено в работе [10].

Несмотря на перспективу использования и повсеместное распространение моделей трансформеров, в последнее время результаты прогноза существенно хуже результатов Encoder-Decoder LSTM-моделей.

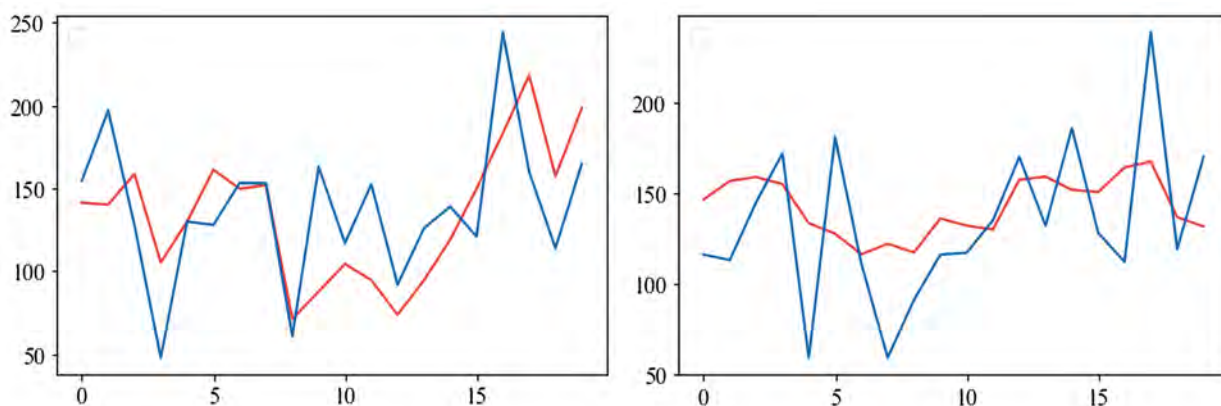


Рис. 10. Прогноз Encoder-Decoder LSTM-моделей за весь исследуемый период

ТАБЛИЦА 9. Результаты прогнозов TSK-моделей за весь исследуемый период

Временной ряд	MSE	RMSE	MAPE	MSPE
Дальность прогноза — 20 значений				
Прибытие вагонов	2386,19	48,85	39,41	62,78
Отправление вагонов	1193,17	34,54	21,6	46,48
Погрузка вагонов	2777,88	52,71	35,54	59,62
Дальность прогноза — 10 значений				
Прибытие вагонов	3101,38	55,69	35,01	59,17
Отправление вагонов	563,65	23,74	14,55	38,14
Погрузка вагонов	1002,1	31,66	15,17	38,95

ТАБЛИЦА 10. Результаты прогнозов Transformer-моделей за весь исследуемый период

Временной ряд	MSE	RMSE	MAPE	MSPE
Дальность прогноза — 20 значений				
Прибытие вагонов	2789,46	52,82	41,87	66,38
Отправление вагонов	870,67	29,51	18,02	37,64
Погрузка вагонов	1515,24	38,93	31,00	52,54
Дальность прогноза — 10 значений				
Прибытие вагонов	2989,28	54,67	38,17	64,95
Отправление вагонов	616,75	24,83	13,64	29,53
Погрузка вагонов	1202,12	34,67	34,08	70,14

Заключение

Использование прогнозных моделей на железнодорожном транспорте при прогнозировании количественных показателей работы станций и ПНОП связано с уникальными для данной отрасли трудностями. Непостоянство технологии работы станций и ПНОП, отсутствие возможности автоматизированного сбора данных о грузовых операциях с вагонами, небольшое количество релевантных данных для обучения моделей, множество факторов, влияющих на идентичность технологии работы станции и ПНОП во времени, являются ключевыми ограничивающими факторами для построения точных прогнозов.

Разбиение исходных данных на почасовые интервалы позволяет избавиться от зависимости от неточно внесенных времен выполнения операций с вагонами на станции и ПНОП, использования промежутка исходных данных определенного размера позволяет избежать учета устаревшей технологии работы станции и ПНОП, которая будет влиять на точность получаемых прогнозов.

Улучшение качества прогнозов может быть связано с использованием многомерных временных рядов. Так, например, наличие кросс-корреляции между временными рядами прибытия и отправления вагонов, подачи и уборки свидетельствует о влиянии значительной одной случайной величины на другую.

Для временных рядов, описывающих разные технологические процессы, происходящие на станции, наблюдаются различные результаты при использовании прогнозных моделей. Так, для временного ряда отправления вагонов со станции при использовании Encoder-Decoder LTSM-моделей ошибка прогнозирования может достигать 7 %, а для временного ряда прибытия вагонов на станцию не опускается ниже 34 %. Это связано с различным количеством экзогенных факторов, влияющих на исследуемые показатели.

В последующих работах будет решена задача применения многомерных временных рядов при построении прогнозов работы объектов железнодорожного транспорта и использование полученных прогнозов в имитационных моделях.

Библиографический список

1. Ламехов В. А. Алгоритм построения прогнозной модели транспортно-логистической деятельности на основе применения нечетких нейронных сетей / В. А. Ламехов, Е. К. Коровяковский // Бюллетень результатов научных исследований. 2022. № 3. С. 137–150. DOI: 10.20295/2223-9987-2022-3-137-150.
2. Свидетельство о государственной регистрации программы для ЭВМ № 2022684797 Российская Федерация. Программа автоматизированного определения количества нечетких правил и параметров antecedентов и консеквентов нечетких нейронных сетей типа TSK: № 2022684298: заявл. 07.12.2022: опубл. 19.12.2022 / В. А. Ламехов; заявитель Федеральное государственное бюджетное образовательное учреждение высшего образования «Петербургский государственный университет путей сообщения Императора Александра I».
3. Dolgoplov P., Konstantinov D., Rybalchenko L., et al. Optimization of train routes based on neuro-fuzzy modeling and genetic algorithms. *Procedia Comput Sci*, 2019. 149, 11–18. DOI: 10.1016/j.procs.2019.01.101.
4. Ermakova A. V. Application of fuzzy mathematics for choosing maintenance intervals for non-public railway tracks / A. V. Ermakova // *Nexo Revista Científica*. 2021. Vol. 34, no. 6. P. 1885–1891. DOI: 10.5377/nexo.v34i06.13194. EDN DBYXRY.
5. Makridakis S., Spiliotis E., Assimakopoulos V. Statistical and Machine Learning forecasting methods : Concerns and ways forward. *PLoS ONE*, 2018. 13 (3) : e0194889. DOI: 10.1371/journal.pone.0194889.
6. Cárdenas J. J., García A., Romeral J. L., et al. Evolutive ANFIS training for energy load profile forecast for an IEMS in an automated factory. *ETFA*. 2011. P. 1–8. DOI: 10.1109/ETFA.2011.6059079.
7. Zhou Y., Guo S., Chang F. Explore an evolutionary recurrent ANFIS for modelling multi-step-ahead flood forecasts. *Journal of Hydrology*, 2019.
8. Wei C., Chen T. and Lee S. k-NN Based Neuro-fuzzy System for Time Series Prediction, 2013 14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing. 2013. P. 569–574. DOI: 10.1109/SNPD.2013.68.
9. Pousinho H. M. I., Mendes V. M. F., Catalão J. P. S. A hybrid PSO-ANFIS approach for short-term wind power prediction in Portugal, *Energy Conversion and Management*. 2011. Vol. 52, iss. 1. P. 397–402. ISSN 0196–8904.
10. Zeng, A., Chen, M., Zhang, L., et al. Are Transformers Effective for Time Series Forecasting? *AAAI Conference on Artificial Intelligence*, 2022. DOI: 10.48550/arXiv.2205.13504.

Дата поступления: 24.12.2023

Решение о публикации: 12.02.2024

Контактная информация:

ЛАМЕХОВ Владимир Андреевич — аспирант;
lamekhovva@gmail.com

КОРОВЯКОВСКИЙ Евгений Константинович —
канд. техн. наук, доцент; ekorsky@mail.com

Construction of short-term forecast of the number of railcars at the stations and non-public routes. Results and analysis

V. A. Lamekhov, E. K. Korovyakovsky

Emperor Alexander I Petersburg State Transport University, 9, Moskovsky pr., Saint Petersburg, 190031, Russia

For citation: Lamekhov V. A., Korovyakovsky E. K. Construction of short-term forecast of the number of railcars at the stations and non-public routes. Results and analysis // Proceedings of Petersburg Transport University. 2024. Vol. 21, iss. 1. P. 199–216. (In Russian) DOI: 10.20295/1815-588X-2024-01-199-216

Abstract

Objective: collect raw data for building predictive models. Analyze the initial data, identify data outliers and outliers, divide the data into time intervals, calculate correlation coefficients, partial autocorrelation, cross-correlation, analyze the trend and seasonality of the obtained time series. Using autoregressive models, machine learning models, neuro-fuzzy models to build forecasts of time series and determine the quality of the obtained forecasts. **Methods:** point density, autocorrelation, partial autocorrelation, cross-correlation, Foster-Stewart test, Dickey-Fuller test, ARMA, MLP, Encoder-Decoder LSTM, TSK, Fuzzy-Partitions, SCRG, Transformers. **Results:** we obtained estimates of the prediction accuracy of the selected models, compared the results of the predictive models trained on different samples of initial data. Conclusions are made about the efficiency and methods of building predictive models. **Practical significance:** the significance of building accurate predictive models for the key quantitative indicators of stations and non-public routes operation is shown. The factors influencing the accuracy of the obtained forecasts are analyzed.

Keywords: fuzzy neural networks, neural networks, autoregressive models, railway station performance analysis, forecasting.

References

1. Lamekhov V. A. Algoritm postroenija prognoznoj modeli transportno-logisticheskoj dejatel'nosti na osnove primenenija nechetkih nejronnyh setej / V. A. Lamekhov, E. K. Korovyakovskij // B'ulleten' rezul'tatov nauchnyh issledovanij. 2022. № 3. S. 137–150. DOI: 10.20295/2223-9987-2022-3-137-150. (In Russian)
2. Svidetel'stvo o gosudarstvennoj registracii programmy dlja JeVM № 2022684797 Rossijskaja Federacija. Programma avtomatizirovannogo opredelenija kolichestva nechetkih pravil i parametrov antecedentov i konsekvantov nechetkih nejronnyh setej tipa TSK: № 2022684298: zajavl. 07.12.2022: opubl. 19.12.2022 / V. A. Lamekhov; zajavitel' Federal'noe gosudarstvennoe bjudzhetnoe obrazovatel'noe uchrezhdenie vysshego obrazovanija "Peterburgskij gosudarstvennyj universitet putej soobshhenija Imperatora Aleksandra I". (In Russian)
3. Dolgopolov P., Konstantinov D., Rybalchenko L., et al. Optimization of train routes based on neuro-fuzzy modeling and genetic algorithms. *Procedia Comput Sci*, 2019. 149, 11–18. DOI: 10.1016/j.procs.2019.01.101.
4. Ermakova A. V. Application of fuzzy mathematics for choosing maintenance intervals for non-public railway tracks / A. V. Ermakova // *Nexo Revista Científica*. 2021. Vol. 34 (6). P. 1885–1891. DOI: 10.5377/nexo.v34i06.13194. EDN DBYXRY.
5. Makridakis S., Spiliotis E., Assimakopoulos V. Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS ONE*, 2018. 13 (3): e0194889. DOI: 10.1371/journal.pone.0194889.
6. Cárdenas J. J., García A., Romeral J. L., et al. Evolutionary ANFIS training for energy load profile forecast for an IEMS in an automated factory. *ETFA*. 2011. P. 1–8. DOI: 10.1109/ETFA.2011.6059079.

7. Zhou Y., Guo S., Chang F. Explore an evolutionary recurrent ANFIS for modelling multi-step-ahead flood forecasts. *Journal of Hydrology*, 2019.

8. Wei C., Chen T. and Lee S. k-NN Based Neuro-fuzzy System for Time Series Prediction, 2013 14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing. 2013. P. 569–574. DOI: 10.1109/SNPD.2013.68.

9. Pousinho H. M. I., Mendes V. M. F., Catalão J. P. S. A hybrid PSO-ANFIS approach for short-term wind power prediction in Portugal, *Energy Conversion and Management*. 2011. Vol. 52. Iss. 1. P. 397–402. ISSN 0196–8904.

10. Zeng, A., Chen, M., Zhang, L., et al. Are Transformers Effective for Time Series Forecasting? *AAAI Conference on Artificial Intelligence*, 2022. DOI: 10.48550/arXiv.2205.13504.

Received: 24.12.2023

Accepted: 12.02.2024

Author's information:

Vladimir A. LAMEKHOV — Postgraduate Student;
lamekhovva@gmail.com

Evgeny K. KOROVYAKOVSKY — PhD in Engineering, Associate Professor; ekorsky@mail.com