

УДК 004

Сравнительный анализ моделей машинного обучения в задаче прогнозирования загруженности парковок

Д.Я. Матросов[✉], А. Нурулла

МИРЭА – Российский технологический университет, Москва, Россия

В статье рассматривается прикладной подход к прогнозированию краткосрочной загруженности городских парковочных зон на примере Москвы. Для анализа временных рядов процентной заполняемости парковок сравниваются статическая модель и две рекуррентные нейронные сети. Данные о загрузке парковок проходят этапы ресемплирования, интерполяции пропусков и инженерии признаков. Модели обучаются на исторических данных с разделением на тренировочную, валидационную и тестовую выборки по времени; валидация проводится с использованием ранней остановки. Эксперименты показали, что рекуррентные модели превосходят статическую по качеству прогноза. Практическая значимость работы заключается в возможности интеграции предложенных моделей в систему интеллектуального управления парковками, что позволит снижать время поиска машиномест и оптимизировать нагрузку на городские дороги в реальном времени.

Ключевые слова: городские парковки, прогнозирование временных рядов, ARIMA, LSTM, GRU, интеллектуальные транспортные системы, прикладной анализ данных.

Comparative Analysis of Machine Learning Models in the Task of Car Park Congestion Forecasting

D.Ya. Matrosov[✉], A. Nurulla

MIREA – Russian Technological University, Moscow, Russia

The article considers an applied approach to forecasting short-term occupancy of urban parking zones on the example of Moscow. A static model and two recurrent neural networks are compared to analyse time series of parking occupancy percentage. Parking occupancy data undergoes the stages of resampling, skip interpolation and feature engineering. The models are trained on historical data with separation into training, validation and test samples by time; validation is performed using early stopping. Experiments show that the recurrent models outperform the static model in terms of prediction quality. The practical significance of the work lies in the possibility of integrating the proposed models into an intelligent parking management system, which will reduce the time of searching for parking spaces and optimise the load on urban roads in real time.

Keywords: urban car parks, time series forecasting, ARIMA, LSTM, GRU, intelligent transport systems, applied data analysis.

Введение

Проблема эффективного использования парковочного пространства становится все более актуальной в современных мегаполисах. Исследования показывают, что значительная доля городских пробок возникает из-за поиска парковочных мест – водители могут тратить до нескольких десятков минут на поиски свободного места, что повышает расход топлива и уровень загрязнения окружающей среды [1]. Особенно остро эта проблема проявляется в крупных городах: по данным индекса загруженности дорог TomTom Москва стабильно входит в число самых перегруженных трафиком

городов мира [2–7]. В связи с этим возрастают требования к интеллектуальным системам управления парковочным пространством, которые будут способны оптимизировать использование уже имеющейся инфраструктуры.

Несмотря на наличие ряда исследований, посвященных прогнозированию загруженности парковок, прикладные реализации таких моделей пока немногочисленны и имеют ограничения. Современные методы глубокого обучения – рекуррентные нейронные сети, ансамблевые модели – до сих пор недостаточно изучены в контексте прогнозирования загруженности городских парковок, особенно с учетом влияния внешних факторов.

Цель исследования – сравнить эффективность моделей глубокого обучения и классического статистического анализа в задаче прогноза временных рядов загруженности городских парковок (на примере г. Москвы) и оценить практическую значимость полученных результатов. Для достижения этой цели в работе реализован комплексный подход, включающий сбор и обработку реальных данных, получаемых из API Московского транспорта и экспериментальное сравнение ряда алгоритмов прогнозирования. Предполагается, что результаты позволят заложить основу для интеллектуальной системы мониторинга и прогноза парковочной загрузки – эффективного инструмента для водителей и городских служб, позволяющего оптимизировать транспортную инфраструктуру.

Обзор литературы

Анализ текущей литературы показывает растущий интерес к методам анализа данных в управлении транспортной инфраструктурой. Существующие подходы к прогнозированию загруженности парковок можно разделить на три основные категории: классические статистические модели, алгоритмы машинного обучения и методы глубокого обучения. Традиционные статистические модели временных рядов – например, ARIMA, SARIMA, модели экспоненциального сглаживания – долгое время являлись основными инструментами прогноза. Так, A. Dunning применила классическую модель ARIMA для построения системы предсказания загруженности городских парковок [8]. Преимуществом подобных подходов считается высокая интерпретируемость результатов и относительно низкая вычислительная сложность. Однако у этих методов имеются существенные ограничения: они недостаточно чувствительны к выбросам в данных (резким всплескам или падениям показателей загруженности) и плохо справляются с нелинейными зависимостями и сложными периодическими паттернами, характерными для временных рядов парковочной активности [9]. В последние годы для прогнозирования в транспортной сфере активно применяются алгоритмы машинного обучения: Random Forest, метод опорных векторов, градиентный бустинг и др. Результаты ряда исследований демонстрируют высокую точность таких моделей. В частности, Gutmann и соавторы с помощью XGBoost добились очень низкой ошибки первого рода ($\sim 0,1045$) при прогнозе загруженности грузовых парковок на 2 часа вперед. По коэффициенту детерминации модель также показала достойное качество: $R^2 \approx 0,886$ при горизонте прогнозирования 30 минут. Эти результаты подтверждают эффективность алгоритмов машинного обучения в задачах краткосрочного прогноза. Рекуррентные нейронные сети (RNN), особенно их архитектуры LSTM (Long Short-Term Memory) и GRU (Gated Recurrent Unit), представляют новое поколение методов анализа временных рядов. Благодаря механизму «внутренней памяти» RNN автоматически улавливают как краткосрочные, так и долгосрочные зависимости во временном ряде, что позволяет точнее прогнозировать сложные динамические процессы. В то время как классические модели

требуют предположений о стационарности или независимости наблюдений, рекуррентные сети обучаются напрямую на последовательности данных, учитывая полный контекст ряда [12]. Применительно к прогнозированию парковочной загрузки методы глубокого обучения демонстрируют превосходство над более простыми моделями. Например, согласно недавнему исследованию, LSTM-модель обеспечила более высокий R^2 и меньшую среднюю абсолютную ошибку по сравнению с XGBoost даже при прогнозе на час вперед [11]. Модель GRU также подтвердила преимущества рекуррентных сетей в анализе парковочной активности [13]. В других работах успешно применялись LSTM для предсказания занятости парковочных мест, показывая высокую точность при соответствующей настройке сети [14]. Таким образом, современные RNN обладают потенциалом более точно моделировать сложные временные ряды загрузки парковок по сравнению с традиционными подходами.

Несмотря на то, что существует множество публикаций, посвященных прогнозированию загруженности парковок, до сих пор недостаточно исследован комплексный подход к анализу парковочной системы в масштабе целого города с использованием современных методов глубокого обучения.

Методология исследования

Сбор данных и предварительная обработка

В качестве источника данных в работе используются реальные данные о загруженности городских парковок, полученные через открытый API Московского транспорта. Этот источник предоставляет актуальную информацию о проценте занятых парковочных мест в режиме реального времени с высоким разрешением по времени, что обеспечивает детальное представление динамики загрузки парковок.

Исходные временные ряды загруженности были очищены и нормализованы для последующего анализа. Предварительная обработка включала следующие этапы:

1. *Ресемплирование*: приведение всех временных рядов к единому дискретному шагу в 30 минут. Это устраняет пропуски и выравнивает частоту измерений по всем парковкам.

2. *Интерполяция коротких разрывов*: пропущенные значения длительностью до 3 часов подряд восстанавливались линейной интерполяцией. То есть подряд идущие не более 6 пустых точек заполнялись по линейному тренду.

3. *Заполнение концов ряда*: если в начале или конце ряда оставались пропуски, которые невозможно интерполировать (отсутствие предыдущих или последующих данных), они заменялись ближайшим известным значением (метод forward/backward fill).

В результате применения этих процедур удалось получить максимально незашумленные непрерывные временные ряды, описывающие загруженность парковок при минимальном искажении их трендов.

Инженерия признаков

Для более точного прогнозирования были добавлены дополнительные признаки, характеризующие временные и календарные зависимости, помимо целевого показателя *occupancy_rate* (процент занятости парковки):

- час суток (*hour*): от 0 до 23 – отражает время дня, чтобы учитывался суточный цикл спроса на парковку;
- день недели (*day_of_week*): от 0 (понедельник) до 6 (воскресенье) – учитывает различия между буднями и выходными;

- день месяца (*day_of_month*): от 1 до 31 – может улавливать эффекты конкретных дат (например, начало месяца, праздники);
- месяц (*month*): от 1 до 12 – для учета сезонных факторов (например, летние или зимние месяцы);
- признак выходного дня (*is_weekend*): бинарный индикатор (1 для субботы/воскресенья, 0 иначе) – явно различает выходные и будни;
- признак рабочего времени (*is_working_hour*): бинарный индикатор рабочего времени (например, 8:00–20:00) – различает дневную и ночную загрузку.

Кроме того, циклические временные признаки (час и день недели) были преобразованы с помощью синусоидальных функций, чтобы модель корректно учитывала их цикличность. Например, для часа суток вычислялись $hour_sin = \sin(2\pi * hour / 24)$ и $hour_cos = \cos(2\pi * hour / 24)$, аналогично для дня недели с периодом 7. Эти синусоиды позволяют непрерывно представить циклическое время (23:00 плавно переходит в 0:00, воскресенье – в понедельник). Таким образом, был сформирован новый набор признаков, включающий целевое значение *occupancy_rate* и расширенный контекст временных факторов для каждого наблюдения.

Архитектуры моделей

В качестве представителя классических подходов выбрана модель ARIMA. ARIMA сочетает три компонента: авторегрессию (порядок p), интегрирование (разность порядка d) и скользящее среднее (порядок q). В данном исследовании модель ARIMA строится исключительно на основе значения *occupancy_rate*, поэтому для нее используется только целевой временной ряд. Оптимальные параметры (p, d, q) определялись с помощью функции *auto_arima* из библиотеки *pmdarima*, перебирающей сочетания параметров по обучающей выборке. После определения оптимальных параметров модель ARIMA обучается на тренировочном сегменте данных, а затем используется для предсказания значений на горизонте тестовой выборки.

Одной из выбранных рекуррентных моделей стала LSTM, доказавшая свою эффективность при анализе последовательных данных [14]. Сеть LSTM способна запоминать долгосрочные зависимости, что критически важно для парковочных данных, где суточные и недельные паттерны повторяются регулярно. Разработанная LSTM-сеть состоит из следующих слоев. Во входной слой подается последовательность длины *seq_length* с набором признаков (описанных выше) за каждый временной шаг. Затем следуют два слоя LSTM: первый LSTM-слой содержит 64 ячейки, второй – 32 (значения выбраны эмпирически). Для экспериментов предусматривалась возможность сделать слои двунаправленными (Bidirectional) для учета контекста как в прямом, так и обратном направлении по времени; в случае одностороннего LSTM учитывалась только прямая временная зависимость. Первый LSTM-слой настроен на возврат всей последовательности выходов (*return_sequences=True*), чтобы следующий рекуррентный слой получил полный временной контекст. Второй LSTM-слой возвращает только последний выход (свертывая временную ось до выходного вектора фиксированной длины). Dropout-слои с коэффициентом 0,2 вставлены после каждого LSTM-слоя для предотвращения переобучения. Наконец, выходной полносвязный слой (*Dense*) преобразует выход второго LSTM-слоя в одно предсказываемое значение (парковочная загрузка на следующий предсказываемый промежуток времени, либо на несколько промежутков,

если задается многошаговый прогноз). Такая архитектура позволяет модели улавливать как краткосрочные колебания, так и более длительные тренды в данных.

Архитектура GRU во многом аналогична описанной LSTM-сети, но основана на более простых ячейках GRU. Модель GRU выбрана как более вычислительно эффективная альтернатива LSTM при сопоставимой моделирующей способности [13]. GRU-ячейка объединяет механизмы забывания и обновления информации, что упрощает сеть и снижает число параметров. В реализованной сети GRU также использованы два скрытых слоя с 64 и 32 единицами соответственно. Как и в случае LSTM, исследовалась возможность двунаправленной конфигурации слоев, первый GRU-слой возвращает полную последовательность выходов для передачи контекста дальше, второй – выдаёт финальный скрытый вектор. Аналогично, после каждого GRU-слоя применен Dropout с долей 0,2 отключаемых нейронов. Выходной слой – полносвязный, возвращает прогнозное значение загрузки. В совокупности, такая двухслойная GRU-сеть содержит меньше параметров, чем LSTM, и обучается быстрее, при этом сохраняя большую часть способности LSTM-модели выявлять сложные паттерны во временном ряде.

Обучение моделей и схема валидации

Перед подачей в модели данные были трансформированы в набор обучающих примеров вида (X, y) . Каждому прогнозируемому значению сопоставляется вектор входных признаков за несколько предыдущих временных шагов. В эксперименте длина входной последовательности (*seq_length*) и горизонт прогнозирования (*horizon*) выбирались как гиперпараметры. Например, для прогноза на 1 час вперед (*horizon* = 2 при 30-минутном шаге) и *seq_length* = 48, каждому целевому значению соответствует последовательность предшествующих данных за последние 24 часа. Из исходного временного ряда формируется трехмерный массив X (образцы $\times$ время $\times$ признаки) и соответствующий вектор целевых значений y .

При разделении выборок строго соблюдался временной порядок наблюдений. Весь массив подготовленных последовательностей сортировался по времени; последние 20% образцов отводились под тестовую выборку, полностью изолированную от процесса тренировки моделей. Из оставшихся данных выделялась валидационная выборка размером 10% (последний сегмент), остальные ~70% использовались как обучающая выборка. Таким образом, обеспечено, что ни одна модель не «видит» будущее при обучении или подборе гиперпараметров – это имитация реальных условий эксплуатации, когда прогноз строится лишь на основе данных прошлого и настоящего времени.

Валидационная выборка применялась для настройки гиперпараметров и мониторинга переобучения. Реализована техника ранней остановки (*early stopping*): после каждой эпохи обучения на тренировочных данных вычислялась ошибка модели на валидационной выборке, и, если в течение 8 последовательных эпох качество не улучшалось, обучение прекращалось досрочно, восстанавливались веса лучшей эпохи. Такой подход предотвращает деградацию модели на тренировочных данных и обеспечивает выбор наиболее обобщающей версии. Кроме того, по метрикам на валидации осуществлялся выбор окончательной конфигурации гиперпараметров моделей.

При обучении нейросетевых моделей были использованы современные методы регуляризации, направленные на повышение обобщающей способности сети. В частности, введение упомянутых слоёв Dropout существенно снижает риск переобучения за счет случайного зануления части нейронов. Механизм ранней

остановки также служит для предотвращения переобучения и был рекомендован в работах по обучению RNN на временных данных [15]. Выбор этих техник основан на их доказанной эффективности для рекуррентных нейронных сетей, что подтверждено, например, в работе W. Zaremba и соавторов.

Метрики оценки

Для количественной оценки точности прогнозов используются три распространенные метрики регрессии: средняя абсолютная ошибка (MAE), корень среднеквадратичной ошибки (RMSE) и коэффициент детерминации (R^2). Каждая метрика отражает разные аспекты качества модели.

Метрика MAE (Mean Absolute Error) – среднее значение абсолютной разности между предсказанными $\hat{y}_t$ и фактическими значениями y_t за выборку из n наблюдений.

В контексте парковок MAE легко интерпретируется в процентных пунктах загруженности. Например, MAE = 4 означает, что в среднем прогноз отличается от факта на 4%. Метрика оценивает все отклонения линейно, без квадратичного возрастания штрафа за большие ошибки, поэтому она устойчива к единичным выбросам и удобна для прикладной интерпретации.

Метрика RMSE (Root Mean Squared Error) представляет собой квадратный корень из средней квадратичной ошибки.

RMSE, в отличие от MAE, сильнее реагирует на крупные отклонения из-за возведения ошибки в квадрат перед усреднением. В задаче парковочного прогноза высокая RMSE указывает на наличие значительных просчетов в отдельные моменты (например, во время резких пиков утром или вечером), что может быть критично при практическом использовании модели.

Метрика R^2 (коэффициент детерминации) – доля дисперсии зависимой переменной, объясненная моделью. Эта метрика находится в диапазоне от 0 до 1 (или может быть отрицательной, если модель хуже тривиального предсказания).

В проведенных экспериментах коэффициент детерминации служит обобщающей метрикой: например, $R^2 = 0,8$ означает, что модель объясняет ~80% наблюдаемой вариации загруженности парковки, а ~20% остаются необъясненными.

Результаты

Для оценки моделей была сформирована тестовая выборка из реальных данных о загруженности парковок за 4 дня (24–27 апреля 2025 г.). Горизонт прогнозирования – 1 час (2 шага по 30 мин). Все модели обучались на предшествующих исторических данных вплоть до начала тестового периода. Ниже представлены сравнительные результаты предсказаний моделей ARIMA, LSTM и GRU на тестовом отрезке, а также метрики их качества.

На тестовых данных рекуррентные нейросети продемонстрировали существенное преимущество над классической ARIMA в точности прогнозирования. Однако между двумя архитектурами RNN выявлены некоторые различия. В частности, отмечено, что модель LSTM менее чувствительна к резким изменениям загруженности по сравнению с GRU. Иными словами, при внезапных пиках или провалах в загрузке парковки LSTM сглаживает эти колебания, тогда как GRU стремится сильнее повторить их амплитуду. Это видно при рассмотрении прогнозов на примере отдельной парковки со сложным, сильно меняющимся паттерном загрузки: GRU достаточно точно пытается следовать за скачками, в то время как LSTM дает более сглаженный прогноз. На рисунках 1 и 2 показаны реальные и предсказанные значения загрузки для

такой парковки (зеленая линия – фактическая загруженность, синяя пунктирная – прогноз на 1 час вперед). Видно, что в моменты резких пиков (например, утренние и вечерние часы) прогноз LSTM чуть отстает от фактического роста, в то время как GRU реагирует почти мгновенно, пусть и с некоторой погрешностью. В целом же оба алгоритма достаточно хорошо моделируют общую динамику, различия проявляются именно в реакции на экстремумы.

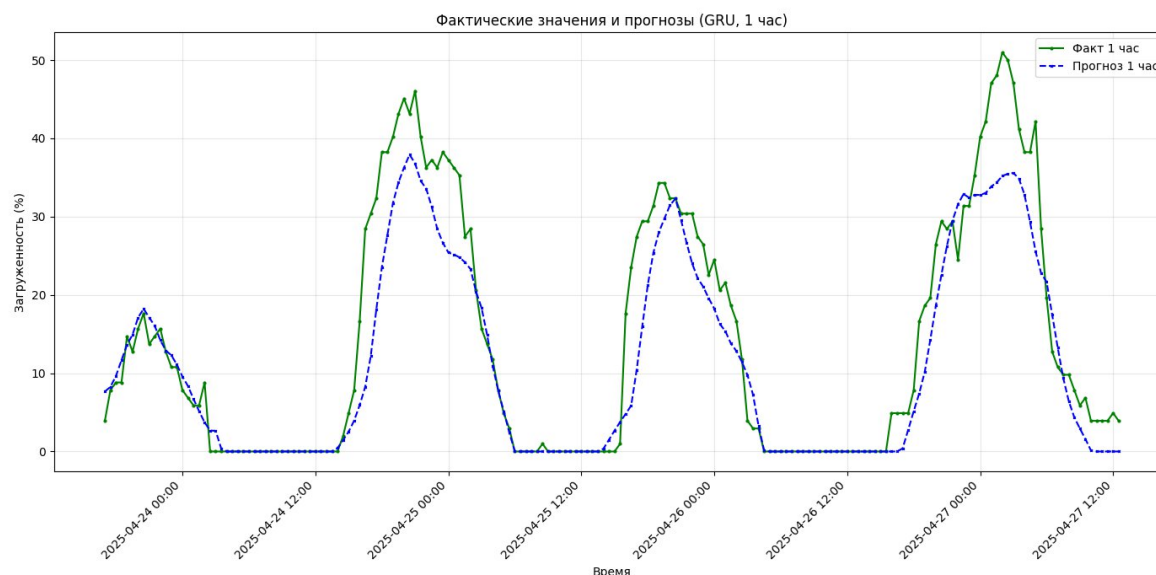


Рисунок 1. Сравнительный график для предсказаний модели GRU

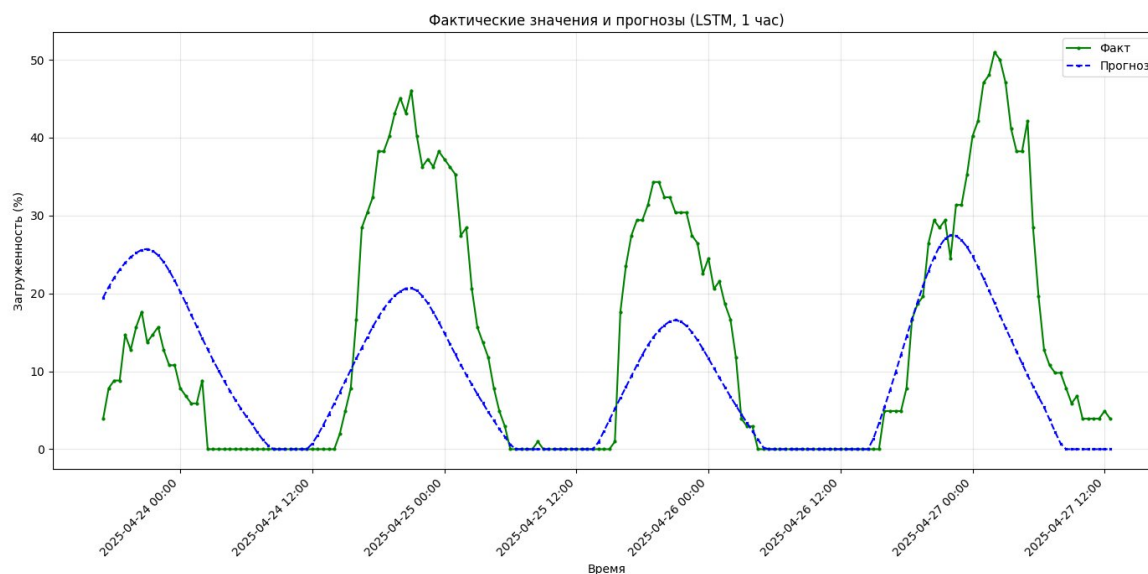


Рисунок 2. Сравнительный график для предсказаний модели LSTM

В среднем по многим парковкам качество предсказаний LSTM и GRU оказалось сопоставимым. На парковках с более регулярным, предсказуемым трендом обе модели дают почти идентичные результаты, успешно улавливая суточные пики и спады. Это подтверждается на другом примере (рис. 3 и 4), где показаны прогнозы GRU и LSTM для парковки с устойчивым суточным циклом загрузки. Здесь обе модели практически

совпадают между собой и с фактическими данными, демонстрируя высокую точность на каждом пике утренней и вечерней загрузки.

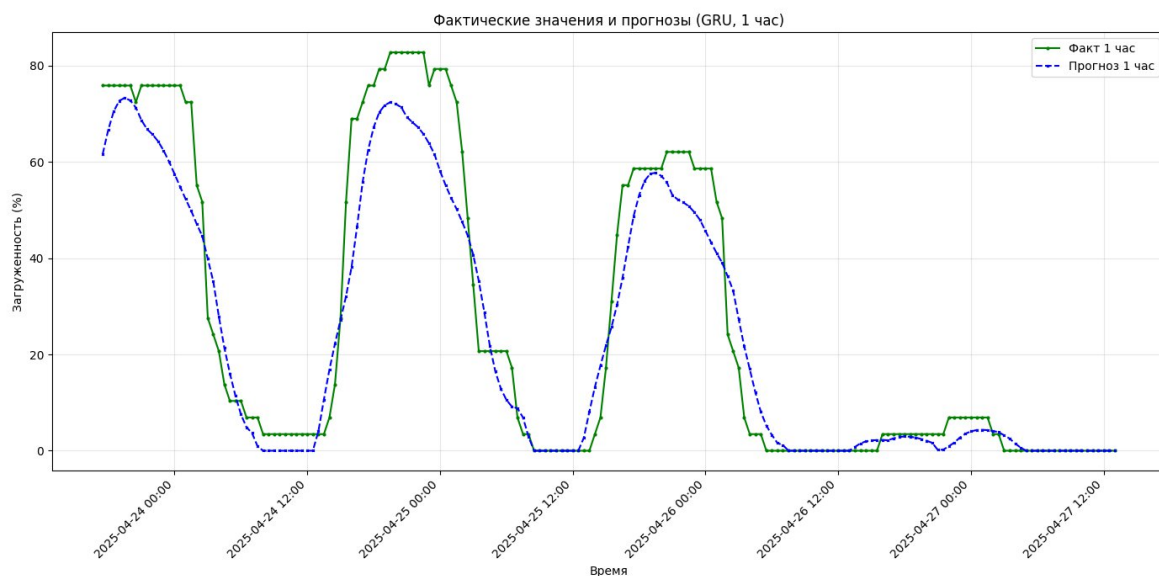


Рисунок 3. Сравнительный график для предсказаний модели GRU

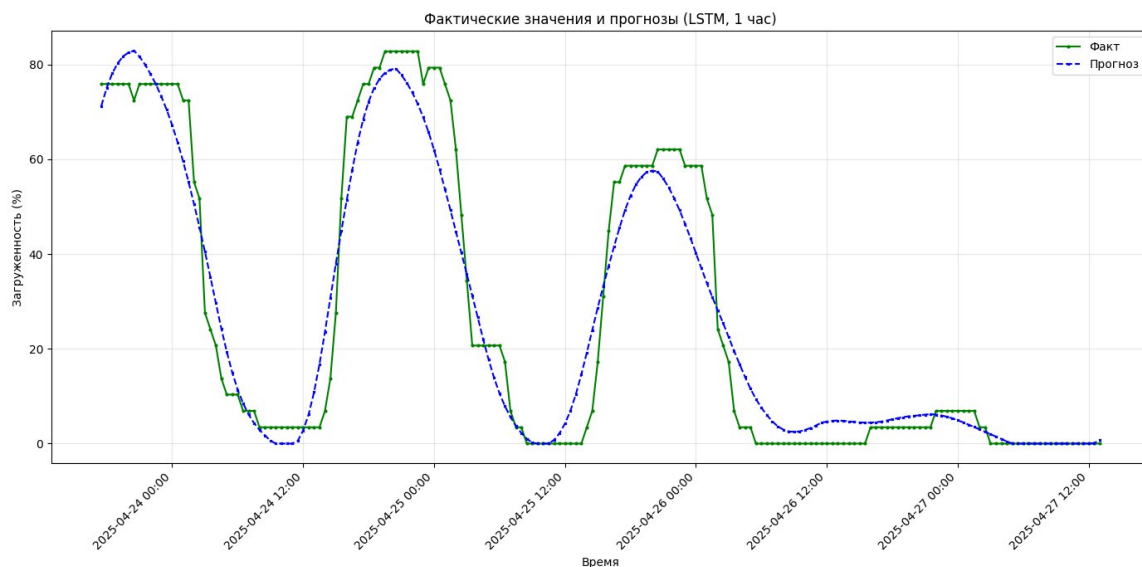


Рисунок 4. Сравнительный график для предсказаний модели LSTM

Модель ARIMA справилась с прогнозом хуже, чем современные нейросети. Ее предсказания в ряде случаев значительно отклонялись от фактической загруженности, особенно на парковках с нестандартными паттернами (где зависимость загрузки от времени и других факторов сложна). К основным причинам относительно слабого результата ARIMA можно отнести ограниченный объем обучающей выборки и отсутствие в модели механизмов учета сложных нелинейных влияний. Тем не менее, ARIMA задала важный ориентир для сравнения и подтвердила необходимость применения более мощных методов. Для наглядности на рисунках 5–7 приведено сравнение всех трех моделей по значениям метрик MAE, RMSE и R^2 на тестовой

выборке. Эти графики отчетливо показывают превосходство рекуррентных сетей над классическим подходом: ошибки LSTM и GRU существенно ниже, а коэффициент детерминации – значительно выше, чем у ARIMA.

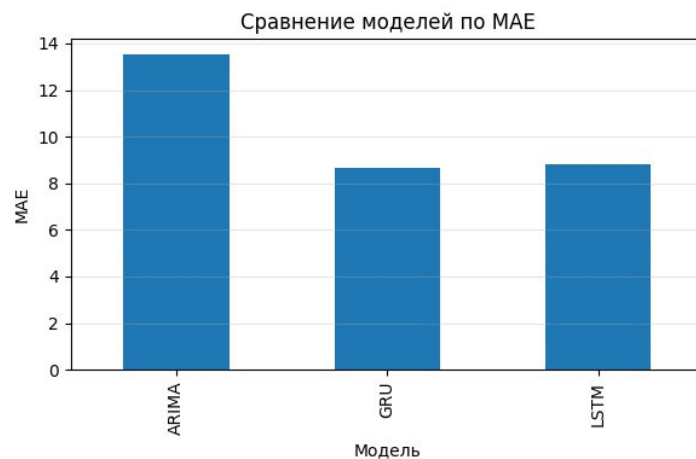


Рисунок 5. Показатели средней абсолютной ошибки

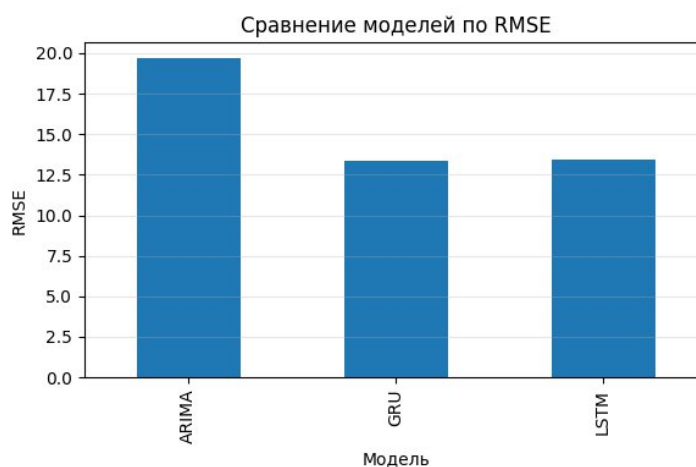


Рисунок 6. Показатели корня среднеквадратической ошибки

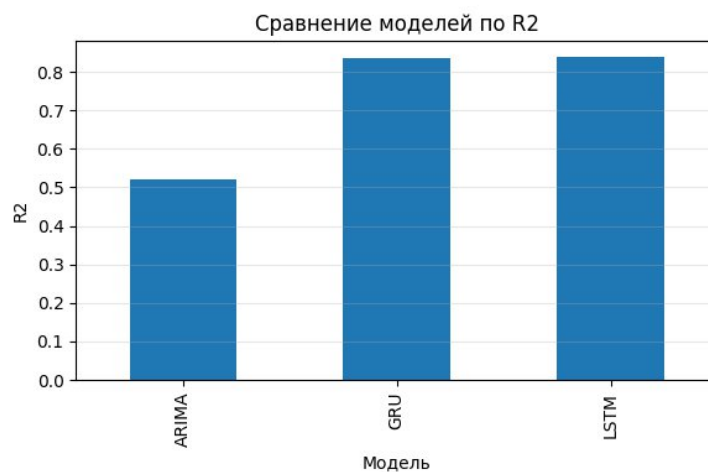


Рисунок 7. Показатели коэффициента детерминации

В количественном выражении, средняя абсолютная ошибка для нейросетевых моделей составила порядка 8–9 процентных пунктов, тогда как для ARIMA превышала 13 (рис. 5). Аналогично, RMSE нейросетей находился на уровне ~12–13 (в тех же единицах), что значительно ниже ~19 у ARIMA (рис. 6). Коэффициент детерминации R^2 достигал 0,8 у моделей GRU и LSTM против лишь 0,5 у ARIMA (рис. 7). Иными словами, RNN объясняют около 80% вариации парковочной загрузки, тогда как ARIMA – лишь около половины, что согласуется с выводами других исследователей о превосходстве глубокого обучения [11].

Несмотря на заметно более высокую точность моделей LSTM и GRU по сравнению с ARIMA, даже нейросети имеют определенную погрешность. В наших экспериментах средняя ошибка прогноза составила около 10% от реальных значений (MAE ~8–10). Это означает, что остаются случаи, когда предсказания отклоняются от факта примерно на десятую часть – в контексте парковок такой разброс может быть существенен, например, при пиковых нагрузках (разница между 90% и 100% загруженности парковки определяет наличие свободных мест). Для детального анализа ошибок была выбрана одна из парковок, качество прогноза на которой оказалось близким к медианному по выборке (средняя ошибка около 10%). На рисунке 8 приведено разбиение предсказанных значений загрузки этой парковки (для LSTM и GRU) в зависимости от фактических значений. Видно, что облако точек и крестов группируется вдоль диагонали, что подтверждает отсутствие существенного смещения у моделей – обе они в среднем не переоценивают и не недооценивают загрузку. Однако разброс вокруг идеальной линии свидетельствует о наличии ошибок: при низкой фактической загрузке (20–30%) модели иногда давали завышенные прогнозы ~30%, и наоборот, при фактической 60% отдельные прогнозы около 50%. В целом разброс не превышает $\pm 10\%$ от диагонали, что соответствует рассчитанным значениям MAE.

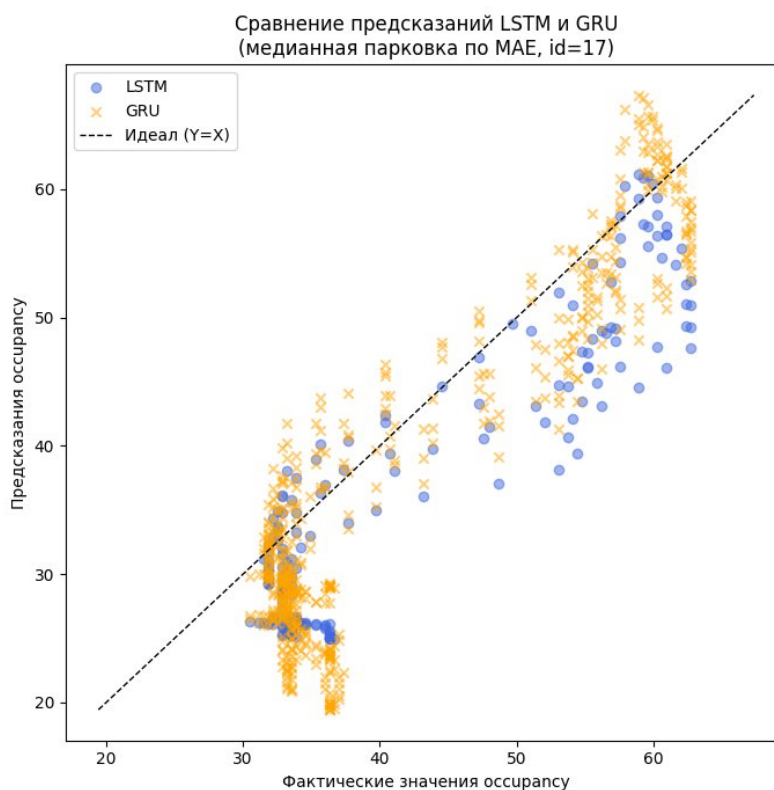


Рисунок 8. Сравнение предсказаний загрузки парковки № 17

Наконец, помимо точности предсказаний, в практической задаче важно учитывать компромисс между качеством и вычислительной сложностью моделей. Разработанная система должна быть потенциально применима в реальном времени и масштабируема на большое число парковок, поэтому сравнивались также затраты ресурсов для каждого алгоритма.

Рекуррентные нейросети достигают значительно более высокого значения $R^2 \approx 0,8$, чем ARIMA ($\sim 0,5$), то есть существенно лучше объясняют поведение системы. Однако плата за это – возросшие вычислительные требования. Обучение LSTM-сети заняло ~ 74 минуты, GRU – ~ 57 минут, тогда как ARIMA обучалась за 42 минуты (на том же аппаратном обеспечении). Объем памяти, занимаемый сохраненной моделью LSTM (442 КБ), более чем в 2 раза превышает размер ARIMA (213 КБ), GRU – примерно в 1,5 раза больше ARIMA. Важным практическим наблюдением является то, что скорость получения прогноза у обученных нейросетей оказалась выше, чем у ARIMA: единичный прогностический шаг (на всех парковках датасета) выполнялся за ~ 23 – 27 секунд для GRU/LSTM против ~ 30 секунд для ARIMA. Таким образом, при внедрении в систему реального времени RNN-модели, хоть и требуют больше ресурсов на этапе обучения, позволяют быстрее обновлять прогноз состояния парковок, что критично для оперативного реагирования.

Совмещая результаты по точности и сложности, можно сделать вывод, что модель GRU демонстрирует наилучший баланс показателей в нашем случае. Она чуть менее точна, чем LSTM ($R^2 = 0,81$ vs $0,83$), но обучается быстрее и потребляет меньше памяти, при практически одинаковой скорости работы на прогноз. LSTM занимает второе место по интегральной эффективности, а ARIMA – лишь на третьем, показывая худшую точность при ненамного меньших ресурсных затратах.

Обсуждение

Полученные результаты подтвердили, что для задачи краткосрочного прогнозирования загруженности парковок глубокие нейронные сети существенно превосходят классические статистические модели. Рекуррентные архитектуры (LSTM, GRU) достигли коэффициента детерминации порядка 0,8, тогда как ARIMA – около 0,5. Это означает, что, учитывая только прошлые значения временного ряда, нейросети способны захватывать большую часть закономерностей парковочного процесса, в то время как возможности ARIMA ограничены предположением о линейности и стационарности процесса. Данный вывод согласуется с тенденцией применения RNN в других областях временного прогнозирования, где они демонстрировали более высокую точность [11, 12].

Несмотря на общее преимущество нейросетей, важно отметить их ограничения. Во-первых, для обучения глубоких моделей требуется значительный объем данных. В наших экспериментах использовались длительные временные ряды с подробной историей; однако, если речь идет о новом парковочном объекте или районе, где исторических данных мало, эффективность LSTM/GRU может снизиться. Во-вторых, даже обученные модели показывают среднюю ошибку порядка 10%, что может быть критично в условиях, когда требуется очень точный прогноз (например, при близкой к полной загрузке парковки каждый процент имеет значение). В частности, выявлено, что наибольшие затруднения у моделей вызывают периоды с аномальными ситуациями – например, дни массовых мероприятий, экстремальные погодные условия, праздники. В такие моменты фактическая загруженность может резко отклоняться от типичных

паттернов, и модели, обученные преимущественно на регулярных данных, ожидаемо теряют точность.

Для преодоления этих ограничений необходимы дальнейшие улучшения. Интеграция дополнительных данных могла бы повысить устойчивость моделей к нестандартным ситуациям. Например, можно включить в систему информацию о городских событиях (концерты, матчи, фестивали), данные о погоде, сведения о дорожном трафике – все эти факторы прямо влияют на спрос на парковку. Первые шаги в этом направлении предпринимаются в современных исследованиях, комбинирующих нейросетевые модели с внешними признаками (так, Zhao и Zhang добавляли категориальные признаки домохозяйств и объекты POI для улучшения прогноза парковки с помощью GCN-GRU модели [13]). Кроме того, следует рассмотреть механизмы онлайн-обучения или адаптации модели к меняющимся условиям. В реальной эксплуатации система должна регулярно обновлять свои прогнозные модели по мере поступления новых данных, чтобы не терять актуальность при изменении транспортной обстановки или появлении новых закономерностей.

Данное исследование имеет прикладную ценность, так как на его основе может быть разработана интеллектуальная система управления парковочным пространством. Продемонстрированный подход способен быть реализован как на уровне городской инфраструктуры (например, в центре управления парковочным хозяйством мегаполиса), так и в виде мобильного приложения для конечных пользователей. Такая система позволит в режиме реального времени информировать водителей о прогнозируемой загруженности парковок в точке назначения, рекомендовать оптимальные парковки с учетом времени прибытия, а городским службам – перераспределять нагрузку (например, перенаправлять поток машин на менее загруженные зоны или динамически регулировать тарифы). Внедрение подобной системы поможет решить ряд острых проблем современных мегаполисов: сократить время поиска парковочного места, снизить лишнюю нагрузку на дороги (меньше кружения в поисках парковки – меньше пробок), уменьшить расход топлива и выбросы от автомобилей, повысить общий комфорт городской среды за счет снижения стресса у водителей.

Заключение

В рамках данного исследования разработаны и экспериментально проверены модели прогнозирования загруженности городских парковок на основе данных API Московского транспорта. Был проведен сравнительный анализ различных подходов к моделированию временных рядов – от классической статистической модели ARIMA до современных рекуррентных нейронных сетей LSTM и GRU – и выявили их преимущества и недостатки в контексте задачи краткосрочного прогноза парковочной загрузки.

Полученные результаты убедительно показывают, что рекуррентные нейронные сети существенно превосходят классические модели по точности прогнозирования. Модели GRU и LSTM достигли $R^2 \approx 0,8$, что указывает на их способность объяснить ~80% вариации данных, тогда как ARIMA ($R^2 \approx 0,5$) значительно уступает. Таким образом, использование глубокого обучения в задаче предсказания занятости парковок оправдано и приводит к заметному улучшению качества прогноза.

В то же время нейросетевые модели имеют ряд ограничений. Для достижения оптимальных результатов им требуется большой объем исторических данных, что может быть проблематично при запуске системы на новых локациях. Кроме того, даже

лучшие модели демонстрируют среднее отклонение порядка 10% от фактической загрузки, что может оказаться критичным в периоды пиковой нагрузки. Анализ ошибок показал, что наибольшие сложности возникают при прогнозировании в нестандартных условиях (массовые мероприятия, праздники, экстремальная погода), когда поведение автовладельцев отличается от обычного. Эти наблюдения указывают на направления улучшения моделей.

Перспективы развития включают несколько важных шагов. Во-первых, планируется расширить набор входных данных за счет интеграции внешних источников информации. Во-вторых, необходимо внедрить механизмы динамической адаптации моделей к изменяющимся условиям – например, периодическое дообучение на новых данных. В-третьих, интересным направлением является разработка модулей персонализации рекомендаций для конечных пользователей: учитывая привычки водителя, его типичные маршруты и предпочтения. Наконец, важна проработка вопросов интеграции прогнозной системы с существующими городскими информационными системами и популярными навигационными сервисами, что упростит внедрение технологии и повысит охват аудитории.

Таким образом, проведенное исследование вносит вклад в теоретическое понимание методов прогнозирования временных рядов в области городской мобильности, а также создает практическую основу для реализации интеллектуальных систем управления парковочным пространством, способных существенно улучшить транспортную ситуацию в современных городах.

СПИСОК ИСТОЧНИКОВ

1. Shoup D.C. Cruising for Parking / D.C. Shoup // Transport Policy. – 2006. – Vol. 13, Iss. 6. – P. 479–486.
2. TomTom Traffic Index 2016 – The Results Are In! Mexico City Takes Crown of 'Most Traffic Congested City' in World from Istanbul // TomTom [Электронный ресурс]. – URL: <https://corporate.tomtom.com/static-files/5289c5aa-310c-4965-a4c0-516760a8a6fd> (дата обращения: 04.03.2025).
3. TomTom Traffic Index 2017: Mexico City Retains Crown of 'Most Traffic Congested City' in World // TomTom [Электронный ресурс]. – URL: <https://corporate.tomtom.com/static-files/7b98ada1-ec2c-4da0-8579-f83ac0f31d88> (дата обращения: 04.03.2025).
4. TomTom Traffic Index 2018 [Электронный ресурс]. – URL: <https://traffic-index-docs.s3-eu-west-1.amazonaws.com/TomTomTrafficIndex-Ranking-2018-full.pdf> (дата обращения: 04.03.2025).
5. TomTom Traffic Index 2019 [Электронный ресурс]. – URL: <https://nonews.co/wp-content/uploads/2020/02/TomTom2019.pdf> (дата обращения: 04.03.2025).
6. TomTom Traffic Index 2020 [Электронный ресурс]. – URL: <https://traffic-index-docs.s3-eu-west-1.amazonaws.com/TomTomTrafficIndex-Ranking-2020-full.pdf> (дата обращения: 04.03.2025).
7. TomTom Traffic Index 2021 [Электронный ресурс]. – URL: <https://nonews.co/wp-content/uploads/2022/02/TomTom2021.pdf> (дата обращения: 04.03.2025).
8. United States Patent Application Publication No. US 2004/0032342 A1. Method and System for Projecting Dynamic Parking Availability Based on an Ongoing Survey for Remote Lots With High Demand: Pub. Date: 19.02.2004 / Dunning A.E.
9. Parking Spaces Demand Forecasting Based on ARIMA Model / H. Chen, M. Yang, Yu. Zhang, H. Wen // ICCMS '23: Proceedings of the 2023 15th International Conference on Computer Modeling and Simulation. – New York: Association for Computing Machinery, 2023. – P. 133–137.

10. Truck Parking Occupancy Prediction: XGBoost-LSTM Model Fusion / S. Gutmann, Ch. Maget, M. Spangler, K. Bogenberger // *Frontiers in Future Transportation*. – 2021. – Vol. 2. – URL: <https://doi.org/10.3389/ffutr.2021.693708> (дата обращения: 27.03.2025).
11. Short-Term Prediction of On-Street Parking Occupancy Using Multivariate Variable Based on Deep Learning / M. Lyu, Ya. Ji, Ch. Kuai, Sh. Zhang // *Journal of Traffic and Transportation Engineering*. – 2024. – Vol. 11, Iss. 1. – P. 28–40.
12. Multi-Horizon Air Pollution Forecasting with Deep Neural Networks / M. Arsov, E. Zdravevski, P. Lameski // *Sensors*. – 2021. – Vol. 21, No. 4. – URL: <https://doi.org/10.3390/s21041235> (дата обращения: 27.03.2025).
13. Zhao X. Enhancing Predictive Models for On-Street Parking Occupancy: Integrating Adaptive GCN and GRU with Household Categories and POI Factors / X. Zhao, M. Zhang // *Mathematics*. – 2024. – Vol. 12, No. 18. – URL: <https://doi.org/10.3390/math12182823> (дата обращения: 27.03.2025).
14. Wang Zh. Research on Parking Space Prediction Based on Long Short-Term Memory / Zh. Wang, W. Xue, X. Hei // *2019 International Conference on Intelligent Computing, Automation and Systems (ICICAS)*. – IEEE, 2019. – P. 392–396.
15. Zaremba W., Sutskever I., Vinyals O. Recurrent Neural Network Regularization // *arXiv [Электронный ресурс]*. – URL: <https://arxiv.org/abs/1409.2329> (дата обращения: 27.03.2025).

ИНФОРМАЦИЯ ОБ АВТОРАХ

Матросов Данил Ярославович, студент, МИРЭА – Российский технологический университет, Москва, Россия.

e-mail: matrosov.d.y@edu.mirea.ru

Нурулла Амин, студент, МИРЭА – Российский технологический университет, Москва, Россия.

e-mail: nurulla.a@edu.mirea.ru