

УДК 656.025, 656.072, 656.027
DOI 10.26118/2782-4586.2025.90.11.016

Иванова Любовь Николаевна
Санкт-Петербургский государственный морской технический университет
Ананченко Игорь Викторович
Санкт-Петербургский государственный технологический институт
Зудилова Татьяна Викторовна
Университет ИТМО

**Комбинированная модель прогнозирования для повышения
экономической эффективности региональных
пассажирских перевозок**

Аннотация. В работе предложен новый метод для прогнозирования пассажиропотока с применением комбинированной модели. Эффективное развитие регионального транспортного сектора экономики зависит от прогнозирования пассажиропотока. Для повышения качества пассажирских перевозок необходимо построение логистических моделей и разработка точных методов для прогнозирования. Решаемая задача прогнозирования пассажиропотока особенно актуальна в условиях ускоренного роста населения и туристического потока, расширения городских границ, развития инфраструктуры. Качественное прогнозирование пассажирского потока позволит эффективно управлять транспортом, снижать затраты. Точный прогноз пассажирского потока обеспечивает комфортные условия для пассажиров. Для точного прогнозирования в работе предложен новый комбинированный метод, объединяющий преимущества различных моделей прогнозирования: градиентного бустирования, ансамбля деревьев решений и метода Холта-Уинтерса. Для комбинированной модели выполняется оптимальный выбор весов на основе характеристик моделей. Для вычислительного эксперимента авторами разработана программа на Python. Результаты эксперимента показали повышение точности прогнозирования комбинированной модели на 15% по сравнению с градиентным бустированием, и на 19% по сравнению с экспоненциальным сглаживанием.

Ключевые слова: пассажирские перевозки, прогнозирование пассажиропотока, комбинированная модель, машинное обучение, повышение эффективности.

Ivanova Lubov Nikolaevna
Saint Petersburg State Marine Technical University
Ananchenko Игорь Викторович
Saint Petersburg State Technological Institute
Zudilova Tatyana Viktorovna
ITMO University

A combined forecasting model to improve the economic efficiency of regional passenger transportation

Abstract. This paper proposes a new method for forecasting passenger traffic using a combined model. Effective development of the regional transport sector depends on passenger traffic forecasting. Improving the quality of passenger transportation requires building logistics models and developing accurate forecasting methods. The problem of forecasting passenger traffic is particularly relevant in the context of accelerated population and tourist growth, expanding city boundaries, and infrastructure development. High-quality passenger traffic forecasting enables efficient transport management and cost reduction. Accurate passenger traffic forecasting ensures comfortable conditions for passengers. To achieve accurate forecasting, this paper proposes a new combined method that combines the advantages of various forecasting models: gradient boosting, decision tree ensemble, and the Holt-Winters method. For the combined model, optimal weight selection is performed based on the characteristics of the models. The authors developed a Python program for the computational experiment. The results of the experiment showed an increase in forecasting accuracy for the combined model by 15% compared to gradient boosting, and by 19% compared to exponential smoothing.

Keywords: passenger transportation, passenger flow forecasting, combined model, machine learning, efficiency improvement.

Введение. Региональные транспортные перевозки существенно влияют на социально-экономическое развитие региона. Для эффективного планирования логистических маршрутов и управления логистикой в регионе необходимы точные прогнозы. Качественное прогнозирование пассажиропотока важно для управления перевозками, снижения расходов и повышения уровня сервиса для пассажиров. Применяемым для прогнозирования статистическим методам не хватает гибкости и точности. Для неравномерного, периодического пассажирского потока с высокой динамичностью необходимо применение комбинированных методов статистики и машинного обучения. Для высокого качества сервиса пассажирских перевозок и оптимизации региональной транспортной инфраструктуры существенно точное прогнозирование пассажирского потока и создание комбинированных моделей, ориентированных на обрабатываемые динамические данные. Для анализа и прогнозирования динамических данных с наличием периодических колебаний применяют статистические методы: линейная регрессия, авторегрессия, модель скользящих средних, экспоненциального сглаживания, модели автокоррелированных интегрированных скользящих средних (ARIMA), метод Холта-Уинтерса, двойное экспоненциальное сглаживание, экстраполяционные методы. Также, широко применяются в прогнозировании методы машинного обучения и нейронные сети: градиентное бустирование (Gradient Boosting), ансамбль деревьев решений (Random Forest), глубокие нейронные сети (Deep Neural

Networks), рекуррентные нейронные сети (Recurrent Neural Networks), сверточные нейронные сети (Convolutional Neural Networks), LSTM (Long Short-Term Memory networks), Prophet, k-ближайших соседей (k-Nearest Neighbors).

Однако и те, и другие методы имеют свои ограничения, недостатки и сложную настройку параметров моделей. Среди множества методов машинного обучения можно отметить хорошую точность для методов: случайного леса, градиентного бустирования. Однако и такие модели имеют проблемы переобучения и чувствительности к колебаниям и выбросам в данных. Именно поэтому необходимо разрабатывать новые комбинированные методы, объединяющие преимущества методов статистических, машинного обучения, и компенсирующих недостатки друг друга. В исследовании авторы предлагают новую комбинированную модель, которая объединяет модели машинного обучения случайного леса, градиентного бустирования и статистическую модель тройного экспоненциального сглаживания. В комбинированной модели результат прогноза строится на основании показателей качества прогноза моделей с применением весовых коэффициентов. Такой подход повышает качество прогнозирования динамических данных.

1. Обзор литературы

В настоящее время большое число научных работ посвящено современным методам и моделям прогнозирования пассажиропотока. Приведем обзор современных работ авторов.

В статье [1] разработан гибридный метод прогнозирования пассажиропотока метро. Метод оценивает миграционные процессы и обеспечивает высокое качество прогнозов, учитывая различные источники данных.

В статье [2] представлена методика для прогнозирования пассажиропотока с учетом повторяющихся поездок. Предложенная методика повышает точность прогнозирования.

В статье [3] предложен подход для объединения методов машинного обучения. Целью исследования является увеличение точности прогнозирования пассажирского потока на железнодорожных вокзалах.

В статье [4] авторы осуществили внедрение комбинированного метода прогнозирования спроса на авиаперевозки. Суть подхода основана на сером прогнозировании и нечетком интеграле Шоке, существенно улучшившего точность предиктивной аналитики.

В проведенных исследованиях [5] подтверждено положительное влияние использования поискового индекса Baidu на повышение точности прогнозирования пассажиропотока метро. Успех данных экспериментов открывает новые возможности для повышения качества предиктивной аналитики.

В статье [6] авторами разработан современный метод прогнозирования пассажиропотока метро. Предложенный способ основан на индексах

поисковых систем и технологиях глубокого обучения, что обеспечивает существенное повышение точности прогнозов.

В статье [7] разработана и представлена модель Interaction Graph Network (IG-Net) для высокоточного прогнозирования пассажиропотока на станциях метро. В результате исследований модель показала значительную эффективность на практических примерах из Сучжоу.

В результате исследований [8] создана мощная модель SARIMA-SVM для прогнозирования пассажиропотока, которая продемонстрировала исключительную эффективность на конкретных примерах (город Пекин).

В результате исследований [9] создана надежная среднесрочная модель прогнозирования спроса на авиаперевозки, которая оценивает экономические и исторические тенденции рынка.

В статье [10] авторами описана современная пространственная хэш-графовая сверточная сеть STHMGCN для предиктивной аналитики пассажиропотока автобусов. Модель подтвердила свою эффективность и широкий потенциал применения в сфере интеллектуального транспорта в результате выполненных исследований.

В статье [11] авторами применена упрощенная, но высокоэффективная модель прогнозирования пассажиропотока с использованием глубокой нейронной сети и архитектуры ST-LSTM. Разработанная авторами модель показала точные и надежные результаты на данных из индийского города Кочи.

В статье [12] предложена уникальная модель НАТС для оптимального согласования расписания высокоскоростных поездов и воздушных рейсов. Следует упомянуть, что данная модель обеспечивает максимальное привлечение пассажиропотока и сокращение изменений исходного расписания.

В статье [13] авторами разработана усовершенствованная ансамблевая модель с использованием алгоритма SEEMDAN и сети CNN-LSTM. Такая модель позволяет качественно извлекать важные временные и пространственные свойства данных и, кроме того, улучшить точность прогнозов на 39%.

В статье [14] авторами разработана гибридная архитектура прогнозирования исходящего пассажиропотока метро, которая включает ансамблевое и трансферное обучение. Помимо этого, модель обладает высочайшими показателями точности и надежности.

В статье [15] проведен подробный обзор методов прогнозирования пассажиропотока в железнодорожном транспорте. Помимо этого, выделены ключевые недостатки выбранной тематики и сформулированы перспективные направления исследований.

В статье [16] представлена инновационная динамическая пространственно-временная сеть TADSTN для прогнозирования пассажиропотока в транспорте. По итогам экспериментов, данная сеть показала точные и стабильные результаты на реальных массивах данных.

В статье [17] разработана прогрессивная трансформаторная модель прогнозирования пассажиропотока между транспортными узлами. Характерную для длинных временных серий проблему всплеска и исчезновения градиентов модель успешно преодолела.

В статье [18] предложена принципиально новая трехуровневая модель TMFO-AGGRU для прогнозирования пассажиропотока метро. Данная модель характеризуется высокой точностью и скоростью расчетов.

В статье [19] разработана модель STCNN для краткосрочного прогнозирования пассажиропотока. Предложенная модель показала наибольшую точность прогнозов на особо важных маршрутах.

В статье [20] выполнено подробное сравнение девяти популярных методов глубокого обучения для прогнозирования пассажиропотока, которое доказывает преимущества двунаправленных моделей.

На основе рассмотренных работ можно отметить, что прогнозирование пассажиропотока требует разных подходов в зависимости от множества факторов и неоднородностью пассажиропотока. В работах применяются различные методы машинного обучения, глубокого обучения, такие как CNN, LSTM, графовые нейронные сети. Качество прогноза существенно зависит от точности настройки выбранной модели для конкретной задачи прогнозирования.

2. Комбинированный метод прогнозирования

Комбинированная модель прогнозирования строится на основе метода статистического и машинного обучения.

Метод включает следующие основные этапы.

На первом этапе происходит предобработка и подготовка данных. Производится проверка пропущенных временных данных, очистка данных от шумов и выбросов.

На втором этапе применяются три основные модели для прогнозирования. Выполняется настройка параметров моделей для подготовленного набора данных.

Для комбинированного метода выбраны три модели:

1. Модель случайного леса использует ансамбль обучения, построенный на деревьях решений. Для уменьшения ошибки каждое дерево решения участвует в коллективном голосовании.

2. Модель градиентного бустирования для уменьшения ошибки использует последовательное улучшение моделей, создавая последовательность деревьев. Данная модель имеет высокий уровень точности при большом наборе данных.

3. Модель Холта-Уинтерса применяет экспоненциальное сглаживания для временных рядов с наличием тренда и сезонностью. Модель эффективна для повторяющихся процессов.

На этом этапе каждая модель создает свой прогноз и выполняется оценка качества прогноза на части данных.

На следующем этапе на основе метрик качества прогноза для трех моделей рассчитываются веса каждой модели для композитного прогноза.

Формула композитного прогноза записывается в виде (1):

$$F(t)=w1 \cdot R(t)+w2 \cdot G(t)+w3 \cdot H(t), \quad (1)$$

где: $F(t)$ - итоговый прогноз на момент времени t ,

$R(t)$, $G(t)$, $H(t)$ - прогнозы для трех моделей (случайный лес, градиентное бустирование, Холта-Уинтерса),

$w1$, $w2$, $w3$ — веса на основании метрик оценки точности отдельных моделей.

Важной особенностью комбинированного метода является автоматический подбор весов на основе оценки среднеквадратической ошибки на контрольной выборке данных. Комбинированная модель определяет вклад каждой модели в итоговый прогноз. Для автоматического подбора весов применен метод оптимизации Нелдера — Мида, который определяет баланс между вкладами трех моделей: случайного леса, градиентного бустирования и Холта-Уинтерса. Таким образом обеспечивается наилучшее качество прогноза.

3. Вычислительный эксперимент для комбинированного метода

Для выполнения эксперимента авторами подготовлен набор данных за первую половину 2025 года. Данные содержат объемы пассажирских перевозок и учитывают дополнительные временные признаки, такие как время дня, дни недели, месяца, выходные, наличие праздничных дней. Данные факторы влияют на интенсивность пассажирского потока.

Для проведения эксперимента авторами составлены программы для Python с применением библиотек: sklearn, statsmodels, matplotlib, pandas, scipy, numpy.

Настройка параметров для каждой из моделей: случайного леса, градиентного бустирования и модели Холта-Уинтерса выполнена для выбора оптимальной конфигурации. Для оценки результатов использовались метрики точности и качества прогноза: MSE, RMSE, MAE. На основе метрик для комбинированной модели автоматически рассчитывались коэффициенты для весов моделей.

Опишем основные шаги алгоритма программы для Python.

1. Вначале выполняется импорт библиотек. Для проведения анализа необходимы библиотеки pandas, numpy, для визуализации matplotlib, для моделей машинного обучения sklearn, для статистических моделей statsmodels.

2. Далее проводим загрузку подготовленных текстовых данных из файла формата csv. Выполняется чтение из файла данных пассажиропотока и создание датафрейма. При этом учитываются различные временные признаки.

3. На следующем этапе проводим верификацию данных. Чтобы сделать данные пригодными для моделирования, выполняется проверка на пропущенные значения и фильтрация отдельных выбросов в данных.

4. Выполняется разбиение данных на тренировочные и тестовые наборы.

Общий набор данных делится на части: тренировочную (80%) и тестовую (20%). Первая часть используется для обучения моделей, а вторая часть для оценки качества прогнозов.

5. Проводится определение оптимальных гиперпараметров для каждой из моделей градиентного бустирования и случайного леса.

Модель градиентного бустирования настраивается по скорости обучения, глубины деревьев, минимальной численности узлов. Аналогично, для модели случайного леса подбираются количество деревьев, глубина, численность разделения.

6. На следующем этапе модели градиентного бустирования и случайного леса обучаются с оптимальными гиперпараметрами.

7. Выполняется прогнозирование методом Хольта-Уинтерса. Применяются для статистической модели параметры с аддитивным трендом и сезонностью. Выполняется подбор оптимальных гиперпараметров модели: тренд, сезонность, сглаженность.

8. Выполняется создание комбинированной модели. Для этого результаты прогнозирования трех моделей комбинируются. Определяются веса трех моделей методом оптимизации Нелдера — Мида на основе метрик качества.

Решается оптимизационная задача минимизации среднеквадратической ошибки при ограничениях. В результате комбинированная модель учитывает преимущества каждой из трех моделей.

9. Выполняется расчет качества прогнозов с помощью стандартных метрик: среднеквадратической ошибки (MSE), корня среднеквадратической ошибки (RMSE) и средней абсолютной ошибки (MAE). В результате метрики показывают точность каждого прогноза и комбинированного прогноза.

10. Для наглядности выполняется визуализация результатов и построения графиков на прогнозируемом интервале для всех трех моделей и прогноз комбинированной модели.

Разработанная программа демонстрирует комплексный подход к решению задачи прогнозирования пассажиропотока, используя статистические методы и машинное обучение. Комбинированная модель включает преимущества каждого метода.

В результате экспериментов на основании рассчитанных метрик качества определено, что комбинированный метод превосходит по точности на 15 % метод градиентного бустирования, на 13% метод случайного леса и на 19% метод Холта-Уинтерса.

4. Обсуждение результатов эксперимента

Результат эксперимента можно объяснить компенсацией недостатков отдельных моделей в новом комбинированном методе. Также эффективен в комбинированном методе подбор весов на основе метрик. Кроме того, для учета периодических и сезонных факторов существенна правильная настройка параметров трех выбранных моделей.

Совместное применение статистических методов и машинного обучения увеличивает точность прогнозирования пассажиропотоков. Для определения весов в комбинированной модели также возможно применять более сложные алгоритмы для заданных условий. Такой подход позволяет развивать комбинированный метод. Результаты эксперимента подтверждают

эффективность комбинации трех моделей для прогнозирования пассажиропотока. Применение нового метода позволит транспортным компаниям точно оценивать динамику пассажиропотока. Также, метод позволяет применять более сложные алгоритмы определения весов моделей. Учет дополнительных факторов позволяет повысить чувствительность комбинированной модели. Разработанная авторами программа предоставляет практический инструмент для прогнозирования пассажиропотока. Применение комбинированного метода в задачах прогнозирования пассажиропотока позволит повысить эффективность регионального транспорта.

Заключение

В работе предложен новый метод, комбинирующий модели градиентного бустирования, случайного леса и Холта-Уинтерса. Для комбинированного метода определяются веса трех моделей методом оптимизации Нелдера — Мида на основе метрик точности. Результаты эксперимента показали повышение точности для комбинированной модели на 15% по сравнению с методом градиентного бустирования и на 19% по сравнению с экспоненциальным сглаживанием. Для реализации эксперимента разработаны программы на Python. Исследование показывает перспективность комбинированных методов для прогнозирования пассажиропотоков. Объединение статистических методов и машинного обучения повышает точность прогноза. Определено направление развития метода для учета в комбинированной модели различных факторов. Разработанная программа предоставляет инструмент для качественного прогнозирования в транспортном секторе экономики.

Список источников

1. Cheng Y., Li H., Sun S., Liu W., Jia X., Yu, Y. Short-term subway passenger flow forecasting approach based on multi-source data fusion // *Information Sciences*. – 2024. – Т. 679. – С. 121109.
2. Cheng Z., Trépanier M., Sun L. Incorporating travel behavior regularity into passenger flow forecasting // *Transportation Research Part C: Emerging Technologies*. – 2021. – Т. 128. – С. 103200.
3. Chuwang D. D., Chen W., Zhong M. Short-term urban rail transit passenger flow forecasting based on fusion model methods using univariate time series // *Applied Soft Computing*. – 2023. – Т. 147. – С. 110740.
4. Hu Y. C. Air passenger flow forecasting using nonadditive forecast combination with grey prediction // *Journal of Air Transport Management*. – 2023. – Т. 112. – С. 102439.
5. Jin K., Sun S., Li H., Zhang F. A novel multi-modal analysis model with Baidu Search Index for subway passenger flow forecasting // *Engineering Applications of Artificial Intelligence*. – 2022. – Т. 107. – С. 104518.
6. Li H., Jin K., Sun S., Jia X., Li Y. Metro passenger flow forecasting through multi-source time-series fusion: An ensemble deep learning approach // *Applied Soft Computing*. – 2022. – Т. 120. – С. 108644.

7. Li P., Wang S., Zhao H., Yu J., Hu L., Yin H., Liu Z. IG-Net: An interaction graph network model for metro passenger flow forecasting // *IEEE Transactions on Intelligent Transportation Systems*. – 2023. – T. 24. – №. 4. – C. 4147-4157.
8. Li W., Sui L., Zhou M., Dong H. Short-term passenger flow forecast for urban rail transit based on multi-source data // *EURASIP Journal on Wireless Communications and Networking*. – 2021. – T. 2021. – №. 1. – C. 9.
9. Lundaeva K. A., Saranin Z. A., Pospelov K. N., Gintciak, A. M. Demand Forecasting Model for Airline Flights Based on Historical Passenger Flow Data // *Applied Sciences*. – 2024. – T. 14. – №. 23. – C. 11413.
10. Luo D., Zhao D., Ke Q., You X., Liu L., Ma H. Spatiotemporal hashing multigraph convolutional network for service-level passenger flow forecasting in bus transit systems // *IEEE Internet of Things Journal*. – 2021. – T. 9. – №. 9. – C. 6803-6815.
11. Mulerikkal J., Thandassery S., Rejathalal V., Kunnankody D. M. D. Performance improvement for metro passenger flow forecast using spatio-temporal deep neural network // *Neural Computing and Applications*. – 2022. – T. 34. – №. 2. – C. 983-994.
12. Tan Y., Li Y., Wang R., Mi X., Li Y., Zheng H., Wang Y. Improving synchronization in high-speed railway and air intermodality: Integrated train timetable rescheduling and passenger flow forecasting // *IEEE Transactions on Intelligent Transportation Systems*. – 2022. – T. 23. – №. 3. – C. 2651-2667.
13. Wang J., Wang R., Zeng, X. Short-term passenger flow forecasting using CEEMDAN meshed CNN-LSTM-attention model under wireless sensor network // *IET Communications*. – 2022. – T. 16. – №. 10. – C. 1253-1263.
14. Wang X., Zhu C., Jiang J. A deep learning and ensemble learning based architecture for metro passenger flow forecast // *IET Intelligent Transport Systems*. – 2023. – T. 17. – №. 3. – C. 487-502.
15. Xue Q., Zhang W., Ding M., Yang X., Wu J., Gao Z. Passenger flow forecasting approaches for urban rail transit: A survey // *International Journal of General Systems*. – 2023. – T. 52. – №. 8. – C. 919-947.
16. Yi P., Huang F., Wang J., Peng J. Topology augmented dynamic spatial-temporal network for passenger flow forecasting in urban rail transit // *Applied Intelligence*. – 2023. – T. 53. – №. 21. – C. 24655-24670.
17. Yue M., Ma S. LSTM-based transformer for transfer passenger flow forecasting between transportation integrated hubs in urban agglomeration // *Applied Sciences*. – 2023. – T. 13. – №. 1. – C. 637.
18. Zhang Y., Chen Y., Wang Z., Xin, D. TMFO-AGGRU: A graph convolutional gated recurrent network for metro passenger flow forecasting // *IEEE Transactions on Intelligent Transportation Systems*. – 2023. – T. 25. – №. 3. – C. 2893-2907.
19. Zhang Y., Sun K., Wen D., Chen D., Lv H., Zhang Q. Deep learning for metro short-term origin-destination passenger flow forecasting considering section capacity utilization ratio // *IEEE Transactions on Intelligent Transportation Systems*. – 2023. – T. 24. – №. 8. – C. 7943-7960.

20. Zhang Y., Wang X., Xie J., Bai, Y. Comparative analysis of deep-learning-based models for hourly bus passenger flow forecasting // Transportation. – 2024. – Т. 51. – №. 5. – С. 1759-1784.

Сведения об авторах

Иванова Любовь Николаевна. старший преподаватель кафедры управления бизнес-технологиями инженерно-экономического факультета, Санкт-Петербургский государственный морской технический университет, Санкт-Петербург, Россия

Ананченко Игорь Викторович, кандидат технических наук, доцент, доцент кафедры системного анализа и информационных технологий, Санкт-Петербургский государственный технологический институт (технический университет), Санкт-Петербург, Россия

Зудилова Татьяна Викторовна, кандидат технических наук, доцент, доцент факультета прикладной информатики, Университет ИТМО, Санкт-Петербург, Россия

Information about the authors

Ivanova Lubov Nikolaevna, Senior Lecturer, Department of Business Technology Management, Faculty of Engineering and Economics, Saint Petersburg State Marine Technical University, Saint Petersburg, Russia

Ananchenko Igor Viktorovich, PhD, assistant professor, Saint Petersburg State Technological Institute (Technical University), Saint Petersburg, Russia

Zudilova Tatyana Viktorovna, PhD, assistant professor of the Faculty of AI, National Research University ITMO, Saint Petersburg, Russia