IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⋇ Impact Factor 7.39 ⋇ Vol. 11, Issue 7, July 2022

DOI: 10.17148/IJARCCE.2022.11749

PREDICTING BUS PASSENGER FLOW AND PRIORITIZING INFLUENTIAL FACTORS USING MULTI-SOURCE DATA: SCALED STACKING GRADIENT BOOSTING DECISION TREES

Lavanya D[1], Prof.M.N.Chandan[2]

PG Scholar, Dept. of MCA, PES College of Engineering, Mandya, Karnataka, India[1]

Assistant Professor, Dept. of MCA, PES College of Engineering, Mandya, Karnataka, India[2]

Abstract: Making informed judgments and maximising the use of the available transportation resources are made easier with accurate bus passenger flow forecasting. A wide range of elements connected to the travel environment that have an impact on passenger flow can be identified using data from various sources. A decent prediction model should resolve the accompanying multicollinearity problem in addition to fully using the latent information hidden in multisource data. Based on this concept, we offer a special scaled stacking gradient boosting decision tree (SS-GBDT) model to anticipate bus passenger flow. The SS-GBDT consists of the prior feature generation module and the following GBDT prediction module. The stacking method was used in the prior module to provide a number of improved multi-source data characteristics using a few basic models with comparable performance. By using a quasi-attention-based mechanism, we explicitly develop a scaled stacking approach (precision-based scaling and time-based scaling). The prediction module improves prediction performance by using the newly developed characteristics as input to calculate the passenger flow using a GBDT model with layered data. On two actual bus routes in Guangzhou, China, the plan is tested. Considering the results, it can be concluded that SS-GBDT is superior in terms of prediction stability and accuracy. Additionally, it is better suited to handle the multicollinearity issue with multisource data. The variables that impact predicting passenger flow can also be sorted. When there are sizable amounts of data, the prediction model is flexible and scalable, enabling the integration of a number of influencing factors.

INTRODUCTION

Popular modes of transportation like public transportation are necessary for complex transportation tasks. For the transit system's operational management and planning, the forecast of passenger flow forms the basis. Passenger flow forecasts could be classified into long-term and short-term levels depending on the anticipated time frame. The first primarily focuses on the development of the transit system, such as line alignment and bus stop location, while the second tries to forecast the expected passenger flow in the near future for bus operational management. Passenger flow can be predicted at both the route and stop levels, depending on the level of detail. The article presents a short-term route-level passenger flow forecast that seeks to forecast the number of passengers who will utilise public transportation over the next transportation along a bus route during the day.

LITERATURE SURVEY

Optimal circular bus routes planning for transit network design problem in urban areas

This paper presents a deterministic solution methodology for Transit route Network Design Problem (TrNDP). The main objective is to design a set of bus routes minimizing both users and operator costs while satisfying some constraints; such as minimum demand trips coverage, maximum bus route length and route directness. The proposed methodology provides an efficient set of circular closed bus routes. The formulation of the methodology consists of three parts; 1- representation of transit route network and input data, 2- representation of transit route network objectives and constraints as mathematical programming, 3- the structure of the solution methodology for bus route design. The methodology structure has been tested through Mandl's benchmark network problem. The test results showed that the methodology developed in this study is able to improve a given network solution in terms of number of constructed routes, transit service coverage, transfer directness, and solution reliability. Based on the presented methodology, a more robust network optimization tool would be produced for public transportation planning purposes.

Public Transportation Network Design and Frequency Setting: Pareto Optimality through Alternating-Objective Genetic Algorithms

The transportation network design and frequency setting problem concerns the optimization of transportation systems comprising fleets of vehicles serving a set amount of passengers on a predetermined network (e.g., public transport systems). It has been a persistent focus of the transportation planning community while, its NP-hard nature continues to present obstacles in designing efficient, all-encompassing solutions. In this paper, we present a new approach based on an alternating-objective genetic algorithm that aims to find Pareto optimality between user and operator costs. Extensive computational experiments are performed on Mandl's benchmark test and prove that the results generated by our algorithm are 5–6% improved in comparison to previously published results for Pareto optimality objectives both in regard to user and operator costs. At the same time, the methods presented are computationally inexpensive and easily run on office equipment, thus minimizing the need for expensive server infrastructure and costs. Additionally, we identify a wide variance in the way that similar computational results are reported and, propose a novel way of reporting benchmark results that facilitates comparisons between methods and enables a taxonomy of heuristic approaches to be created. Thus, this paper aims to provide an efficient, easily applicable method for finding Pareto optimality in transportation networks while highlighting specific limitations of existing research both in regards to the methods used and the way they are communicated.

Route optimization of urban public transportation

In this paper we show the optimization process of urban public transportation routes based on operations research techniques. This is shown in the outline of the development and importance of public transportation planning, its stages, its design and models. We present the design of networks of bus routes showing the overview and background of suitable optimization models for the public transportation system. We developed an optimization model minimizing transfers and we discuss the results according to the proposed theory. The article ends with the main conclusions and recommendations found in the study to improve the route optimization of urban public transportation.

Smart Urban Transit Systems: From Integrated Framework to Interdisciplinary Perspective

Urban transit system is an important part of city transportation, which is an interdisciplinary industry, including traffic engineering, operation research, and computer science. To provide smart services for passengers while applying the new technologies, it is necessary to build an optimal transit network and transit service. A smart transit system is processed from strategic planning, tactical planning, operational planning, transit evaluation to marketing and policy. For each stage, large quantities of related literature have been introduced from different perspectives. The aim of this research is to document the main smart urban transit models, topics and implementations for future references and research in each stage. For the planning part, this paper first summarized the objectives, constraints, algorithms, and implications of the models currently in use and classified the objectives and constraints with classic category and new category. The prominent topics and potential research were captured clearly when comparing the two categories. The methodologies for solving those models were proposed and the genetic algorithm and simulated annealing have been mostly used, which will be helpful for filling the gaps for further research. Despite of the model updates, this study also summarized the application trends such as integrated network design in strategic planning, synchronization and timetable recovery from disruption in tactical and operational planning. To improve the transit system and service, evaluation models on service reliability, service accessibility, timetable robustness, and energy consuming are proposed, which highlight the gap between the idealized service and the real service. Some flexible fare scheme, investments, and commercial strategies are discussed in the financial part. The conclusion highlighted the future scope of the smart urban transit in passenger demand management, travel information service, facility and service optimization and shared mobility, in order to make it more convenient for the passengers and more friendly to the environment.

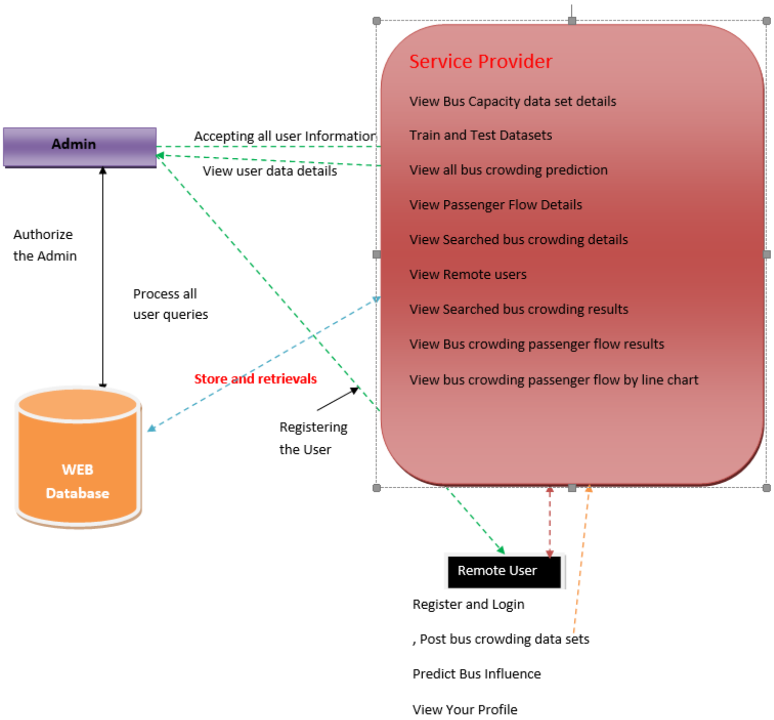
BACKGROUND STUDY

The goal of non-parametric techniques, in contrast to parametric procedures, is to construct a nonlinear relationship between the input and output variables without any prior knowledge. In the transportation industry, artificial neural network (ANN) models are becoming extensively used because they can manage complex interactions in datasets. With ANN, there is a possibility of either over- or under-fitting, though. Support vector machine (SVM) and support vector regression (SVR) models, two further non-parametric models, may be used to circumvent the limitations of neural networks. These models can also handle problems with nonlinearity, limited samples, high dimension, local minima, and

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022
DOI: 10.17148/IJARCCE.2022.11749

over-fitting. To examine train arrival delays, Markovi'c et al. used SVR. Recently, Barbour et al. employed SVR to forecast when freight goods would arrive.

METHODOLOGY



The system suggests utilising a unique scaled stacking gradient boosting decision tree (SS-GBDT) model to predict the flow of passengers on buses using a range of source datasets. The preceding feature generation module and the subsequent GBDT prediction module are components of the SS-GBDT. A few basic models with comparable performance were employed in the preceding module's stacking procedure to provide a number of enhanced multi-source data features.

In specifically, we introduce a quasi-attention-based technique to develop a scalable stacking strategy (precision-based scaling and time-based scaling). By estimating the passenger flow using a GBDT model with layered data and the newly produced characteristics as input, the prediction module can increase prediction performance. The suggested concept is tested on two real bus routes in Guangzhou, China. The findings indicate that SS-GBDT has a higher level of prediction accuracy and stability. Additionally, given data from multiple sources, it is better able to deal with the multicollinearity problem. The factors influencing forecasting passenger flow can also be sorted by this method. A wide range of influencing factors can be included in the prediction model because it is adaptable and expandable, even in the presence of significant volumes of data.

- **Service Provider**
- **View and Authorize Users**
- **Remote User**

CONCLUSION

Planning and managing transportation operations requires the ability to predict short-term passenger flow correctly and consistently. In the age of the internet of things, big data from multiple sources can be used to learn more about a variety of influencing elements. In this study, we suggest a scalable approach that can manage the difficulties of multi-source data to anticipate short-term bus passenger flow. Scaled stacked gradient boosting decision trees (SS-GBDT), a new machine learning model, can combine data from several sources and create new features while separating the impacts of their interactions. A quasi-attention based strategy, particularly precision-based average weighting and time weighting, improves the stacking process. The prediction module uses the newly created features as input to anticipate the passenger

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 7, July 2022
DOI: 10.17148/IJARCCE.2022.11749

flow using the GBDT model and stacking data. This method utilises the potential effectiveness of the multi-source approach and combines the benefits of several machine learning models.

BIBLIOGRAPHY

[1] B. Yu, Z.-Z. Yang, P.-H. Jin, S.-H. Wu, and B.-Z. Yao, "Transit route network design-maximizing direct and transfer demand density," *Transp. Res. C, Emerg. Technol.*, vol. 22, pp. 58–75, Jun. 2012.
[2] W. Wu, R. Liu, W. Jin, and C. Ma, "Stochastic bus schedule coordination considering demand assignment and rerouting of passengers," *Transp. Res. B, Methodol.*, vol. 121, pp. 275–303, Mar. 2019.
[3] W.Wu, R. Liu, and W. Jin, "Modelling bus bunching and holding control with vehicle overtaking and distributed passenger boarding behaviour," *Transp. Res. B, Methodol.*, vol. 104, pp. 175–197, Oct. 2017.
[4] Z. Dai, X. C. Liu, Z. Chen, R. Guo, and X. Ma, "A predictive headwaybased bus-holding strategy with dynamic control point selection: A cooperative game theory approach," *Transp. Res. B, Methodol.*, vol. 125, pp. 29–51, Jul. 2019.
[5] M. Rahman, S. Yasmin, and N. Eluru, "A joint panel binary logit and fractional split model for converting route-level transit ridership data to stop-level boarding and alighting data," *Transp. Res. A, Policy Pract.*, vol. 139, pp. 1–16, Sep. 2020.
[6] X. Ma, X. Zhang, X. Li, X.Wang, and X. Zhao, "Impacts of free-floating bikesharing system on public transit ridership," *Transp. Res. D, Transp. Environ.*, vol. 76, pp. 100–110, Nov. 2019.
[7] N. S. Ngo, "Urban bus ridership, income, and extreme weather events," *Transp. Res. D, Transp. Environ.*, vol. 77, pp. 464–475, Dec. 2019. 8] P. N. E. Chee, Y. O. Susilo, and Y. D. Wong, "Determinants of intentionto- use first-/last-mile automated bus service," *Transp. Res. A, Policy Pract.*, vol. 139, pp. 350–375, Sep. 2020.
[9] C. Ding, X. J. Cao, and P. Næss, "Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in oslo," *Transp. Res. A, Policy Pract.*, vol. 110, pp. 107–117, Apr. 2018.
[10] N. G. Polson and V. O. Sokolov, "Deep learning for short-term traffic flow prediction," *Transp. Res. C, Emerg. Technol.*, vol. 79, pp. 1–17, Jun. 2017.

