10.48047/jocaaa.2024.32.01.28

CNN2D Based Model for Prediction of Hourly Boarding Demand of Bus Passengers using Imbalanced Records from Smart-Cards

Mr K.Jummelal[1], Bhavana Vemparala [2], K.Naga Sahithi[3], B.Prathyusha[4]
[1] Assistant Professor, Department of Computer Science and Engineering, Sridevi Women's Engineering College, Hyderabad, India
Email: jummelalphd@gmail.com
[2,3,4,]B.Tech Student, Department of Computer Science and Engineering, Sridevi Women's Engineering College, Hyderabad, India

ABSTRACT:

An invaluable resource for understanding passenger boarding patterns and forecasting future travel demand is the tap-on smart-card data. Positive instances, on the other hand—boarding at a given bus stop at a certain time—are less common than negative instances when looking at the smart-card data (or instances) by boarding stops and by time of day. Machine learning algorithms that are used to estimate hourly boarding counts from a certain location have been shown to be much less accurate when the data is imbalanced. Before using the smart-card data to forecast bus boarding demand, this research tackles the problem of data imbalance in the data. To create fake travelling cases to add to a synthesised training set with more evenly distributed travelling and non-traveling examples, we suggest using deep generative adversarial networks (Deep-GAN). Next, a deep neural network (DNN) is trained on the synthetic dataset to predict which instances from a given stop in a certain time frame will travel and which ones won't. According to the findings, resolving the data imbalance problem may greatly enhance the predictive model's functionality and make it more accurate in predicting ridership profiles. The suggested strategy may create a synthetic dataset for training with a better similarity and variation and, therefore, a stronger prediction capability, according to a comparison of the Deep-GAN's performance with other conventional resampling techniques. In order to improve data quality and model performance for individual travel

10.48047/jocaaa.2024.32.01.28

behaviour analysis and travel behaviour prediction, the article emphasises the importance of the subject and offers helpful advice.

INTRODUCTION

We are grateful for the funding provided by the Chinese National Natural Sciences Foundation (Project No. 71890972/71890970). The Jiangsu Funding Programme of Excellent Postdoctoral Talent and the National Natural Sciences Foundation of China's Key Project (No. 52131203) both provide funding for Tianli Tang. Charisma Choudhury is grateful for the financial assistance she received from the UKRI Future Leader Fellowship. As [MR/T020423/1-NEXUS]. The Social and Humanities Research Fund of the Ministry during Education (18YJC630190) and the Foundation for Natural Science of the Province of Zhejiang (LQ18G030012) both provide assistance for Yuanyuan Wang. Additionally, the authors are grateful to Hunan Longxiang Buses Co., Ltd. for supplying the smart-card data used in this investigation. Southeast University, Nanjing, which 211189, China is home to T. Tang, who may be reached by email at T-Tang@seu.edu.cn. The educational institution of Leeds, Leeds in LS2 9JT, United Kingdom's Institute of Transport Studies is home to R. Liu and C. Choudhury. Email addresses: C.F. Choudhury@leeds.ac.uk and R. Liu@its.leeds.ac.uk Email address: a.fonzone@napier.ac.uk. A. Fon zone is employed at the Railway Research the Institute, Napier University of Edinburgh, Edinburgh, EH10 5the DT, United Kingdom. * The corresponding author is Y. Wang, who works at the Zhejiang University and Finance and Economics, School about Business Administration, Hangzhou, 310018, China (email: wangyuan@zufe.edu.cn). Email address: R.Liu@its.leeds.ac.uk

Urbanization's quick advancement brings with it an increase in the people living in urban areas, a rise in travel demand, and the negative consequences of air pollution and traffic jams. Public transit is well known for being an environmentally friendly and sustainable way to go around town and solve transportation issues like these. Buses, being a traditional form of public transit, have always been essential for carrying passengers. Low levels of bus services are the

10.48047/jocaaa.2024.32.01.28

result of crowded buses, inconsistent journey times, and bus bunching. Due in large part to the rise of hailing a ride service in recent years, bus ridership has declined in several cities. Bus operators need to figure out how to boost the bus's appeal and image while maintaining and growing its ridership. Bus ridership may be increased by using advanced operating and management strategies that can greatly enhance service dependability and quality of service. To do this, it is required to comprehend the temporal and geographical fluctuations in demand for travellers and to modify the supply side as needed. The original purpose of the smart-card system was automated fare collecting. Since the system keeps track of who boards buses, where they board, and where, smart-card data has evolved into a ready-made useful data source for spatiotemporal demand analysis, public transportation planning, and additional analysis of emission reductions for sustainable transportation. The passenger movement at bus stops along with bus lines may be readily seen using the smart-card data, which allows us to infer the temporal and geographical aspects of bus journeys. Still, there is a great deal of difficulty in automatically extracting relevant information from huge data. Machine learning methods have been a successful and efficient way to analyse big smart-card datasets in recent years. In the case of transit passenger flow prediction, Liu et al. used a decision tree model to capture important characteristics. To predict each person's accessibility in bus systems, Zuo et al. developed a three-phase framework using a neural network model. Our own recent study shows that using smartcard data in conjunction with machine learning methods may be a potent strategy for forecasting the temporal and geographical patterns of van boarding. When the forecasts were averaged across all passengers, they were shown to be quite accurate overall. Nevertheless, our study has also highlighted the problems with data imbalance that arise when attempting to forecast passenger behaviour down to a number of individual travellers and minute spatiotemporal characteristics. If a single smart-card holder boards a bus at a particular stop within a given time window (say, an hour), for example, that is an uncommon occurrence; the majority of records would indicate negative instances (i.e., non-boarding, or not getting on at this bus's stop throughout this

10.48047/jocaaa.2024.32.01.28

time window), with very few positive instances (boarding at this location at this time). These problems with data imbalance have the potential to greatly impair the effectiveness and precision of machine learning algorithms that are used to forecast traveller conduct at the scale of individual passengers and minute spatial-temporal details. This serves as the driving force behind the current study, in which we propose an over-sampling technique, the deep creative clashing nets (Deep-GAN) model (which was originally developed at the field of image generation) to solve the problem of data imbalance in the prediction of disaggregate boarding demand (i.e., the boarding behaviour of individual passengers throughout the day). We demonstrate a considerable improvement in prediction accuracy with the synthesised and more balanced database. It is shown that the suggested strategy, which is based on the Deep GAN method, performs better when compared to alternative resampling techniques, such as the Random Under-Sampling and Synthetic Minority Oversampling Techniques. The remainder of the document is arranged as follows. The principal techniques for resampling and their uses in transport studies are reviewed in Section II. The precise data imbalance problem with hourly boarding demand prediction is explained in Section III. In Section IV, the individual smart-card holders' boarding behaviours (boarding or not boarding) are predicted by a deep neural network (DNN) using a Deep-GAN to produce a synthesised, more equitable training data sample. A case study from the real world is applied to the suggested strategy in Section V, and the outcomes are described in Section VI. Section VII concludes by summarising the major conclusions and findings from this work and making suggestions for further research.

RELATED WORK

A newsvendor model for forecasting bus route peak loads using supply optimisation and scaling Shepard interpolation

The schedule and vehicle capacity of public transport lines have a significant impact on the quality of service provided. The aggregate values of these factors go towards the expenses related to route management as well as expenses related to passenger

10.48047/jocaa.2024.32.01.28

comfort, such congestion and waiting. Our proposed method for solving the issue integrates operator and passenger expenses into a generalised newsvendor model. The expenses associated with waiting and congestion are borne by the passengers, while the operator bears the costs associated with vehicle size, unsold tickets, and lost revenue. In order to provide a minimum level of public transport service or fulfil other regulatory issues, the regulator may set limits, such as the maximum vehicle capacity and the maximum average waiting time for passengers. The newsvendor model offers several benefits: (a) expenses are categorised as surpluses (empty seats) and shortages (overcrowding); (b) the model displays optimal results simultaneously for frequency as well as vehicle size; (c) a quick and effective algorithm is created; and (d) the model presumes stochastic demand and is not limited to a particular distribution. We utilise an instance investigation and a sensitivity evaluation to show the model's applicability.

Taxonomy, rules, and uses of machine learning in railway transportation

The majority of engineering fields are seeing a rise in the use of artificial intelligence (AI), and train travel is no exception. But because there are so many new terms and definitions related to them, there's a chance that practitioners in the railway industry, along with a number of other fields, will become bogged down in the ambiguities and blurred lines, missing out on the true potential and opportunities presented by, among other highly promising AI-related fields, machine learning, artificial seeing, and big data analytics. The purpose of this study is to familiarise railway scholars and practitioners with the fundamental ideas and potential uses of artificial intelligence. In order to achieve that goal, this paper provides a structured taxonomy that will aid scholars and industry professionals in comprehending artificial intelligence (AI) methods, domains, specialties, and applications—both broadly and in direct relation to railway applications like traffic control, maintenance, and autonomous driving. Additionally, the significance of ethics and the capacity of AI to explain railway operations are presented. In order to offer some indications of prospective possibilities, pertinent research addressing both planned and current

10.48047/jocaaa.2024.32.01.28

applications has reinforced the relationship between AI principles and railway subdomains.

An assessment of the additional advantages of mass transit for reducing climate change

Due to the desired win-win outcomes of such measures towards local as well as global aims, the scale of co-benefits from policies addressing warming mitigations has been actively pushed. This study examines research on the quantifiable health and environmental co-benefits of several public transit scenarios. To assess the papers from 2004 to August 2015, a systematic review was carried out. Nine of the 153 articles that were found met every requirement in this evaluation. A lot of research has only looked at the positive effects on the environment, particularly the decrease in air pollution caused by public transit in urban areas.

The calculation of passenger demand for the whole city is essential for the planning administration, and operation of the urban transportation system. The bus od grid reconstruction is based on cluster wi-fi probe data. In this research, traces of smartphone users are gathered using one of the latest crowd sourcing datasets, the Wi-Fi probe data. We create an OD matrix reconstruction framework that includes features extraction for transit consumption and K-means clustering to separate transit users from non-transit users. The partial OD matrix is more dependable with such a structure. Next, based on the incomplete OD matrices and the number of people boarding and alighting, a probabilistic estimate approach of bus OD matrix rebuilding is given. In Suzhou, China, a field study on bus line 5 was conducted. The OD-level discrepancy between the suggested approach and the observed ground truth is 0.5-1.5 passengers per stop, indicating the reliability of the proposed OD matrix reconstruction method.

METHODOLOGY

To implement this project, we have designed following modules as web application

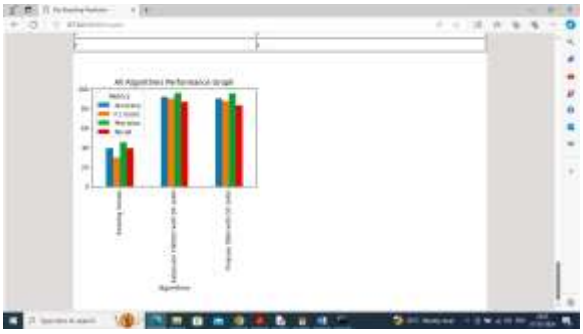
10.48047/jocaaa.2024.32.01.28

1) User Login: user can login to system using username and password as admin and admin
2) Process Dataset: this module will load dataset and then normalize, clean and apply DCGAN to handle imbalance issue
3) Existing Smote: process dataset will be run with SMOTE algorithm and then test accuracy with test data
4) Propose Deep-GAN DNN: DNN algorithm get trained on DEEP GAN generated data and then calculate accuracy on test data
5) Extension Deep-GAN CNN2D: CNN2D algorithm get trained on DEEP GAN gene rated data and then calculate prediction accuracy
6) Comparison Graph: will plot comparison graph between all algorithms and then display number of test data boarding and predicted boarding

RESULT AND DISCUSSION

In above screen click on 'User Login' link to get below page

In above screen extension got 91% accuracy and can see other metrics also. Now click on 'Comparison Graph' link to get below graph



In above graph x-axis represents algorithm names and y-axis represents accuracy and

10.48047/jocaaa.2024.32.01.28

other metrics in different bars and in all algorithms, extension got high accuracy

CONCLUSION

We encountered data imbalance while attempting to anticipate passenger boarding behaviour within a time frame using actual bus smart-card data, which served as the impetus for our work. In this study, we suggested a Deep-GAN to enhance a DNN based model of prediction of individual board behaviour by oversampling travel instances and rebalancing the rate for travelling and non-traveling instances using the smart-card dataset. Applying the models to actual smart-card data gathered from seven bus routes in Changsha, China, allowed for an evaluation of Deep-GAN's performance. When we compared the various imbalance ratio in the model's training dataset, we discovered that, on average, the model performs better with more unbalanced data, with the biggest increase occurring at a 1:5 ratio between both positive and negative examples. The high incidence of imbalance will result in deceptive load profiles and the perfectly balanced data may overestimate ridership at peak hours from the standpoint of the hourly breakdown of bus ridership forecast accuracy. The model performs better when both over- and under-sampling are done, according to a comparison of several resampling techniques. Among the over-sampling techniques, Deep GAN gets the highest recall and accuracy ratings. While a forecasting model trained using Deep GAN data does not perform appreciably better than other resampling techniques, Deep GAN has the potential to greatly enhance both the predictive model's performance and the quality of the training dataset, particularly in situations where under sampling is inappropriate for the data. This work has the following contributions: This research is the first to address the problem of data imbalance in the system of public transportation and suggests a deep learning method called Deep-GAN to address it. • The actual and synthetic travelling instances produced by Deep-GAN and other over-sampling techniques were evaluated for differences in resemblance and diversity in this research. By assessing the effectiveness of the subsequent travel behaviour prediction model, it also examined various resampling techniques for the purpose of improving data quality. This is the first validation and assessment of the various data resampling techniques based on actual

10.48047/jocaaa.2024.32.01.28

data from the public transport system. • Unlike prior travel demand prediction problems, this research creatively modelled individual boarding behaviour. This individual-based model may provide more information on passenger behaviour than the widely used aggregated forecast, and the findings will help with the study of the parallels and heterogeneities. Predictive models will advance in sophistication as computer power and technology advance. The bus network or bus lines will eventually give way to personal travel behaviour as the aim in the area of demand prediction for public transportation systems. The digital twin on the public transportation system is one example of how this innovation may significantly improve planning and administration for public transport. It is expected that unbalanced data will be a difficulty for future prediction in public transit systems. Our study suggests a Deep-GAN model to deal with the problem of data imbalance in traveller behaviour prediction. The validation using real-world data demonstrated that, in comparison to previous resampling techniques, Deep-GAN demonstrated a superior capacity to handle the problem of data imbalance and advantages the prediction models. This study offers managers and academics invaluable expertise in handling comparable data imbalance problems, particularly in public transit. It should be mentioned that even with Deep GAN or DNN models' excellent performance, there are still some restrictions. First, the oversampling is the only purpose for which Deep-GAN is used in this study. Nonetheless, a hybrid version of Deep GAN exists in which negative cases are under-sampled and positive examples are over-sampled. Future studies will be motivated to evaluate the hybrid Deep-GAN's performance due to the encouraging outcomes of the Deep-GAN oversampling. Second, the prediction in this work is made at the person level, leading to an explosion of data and higher computational complexity. Reducing the size the dataset can benefit by classifying passengers (using clustering techniques, for example). Third, boarding behaviour’s spatiotemporal properties are not taken into account by the Deep GAN as it is now. Enhancing the quality of produced mock travelling instances or the effectiveness of the subsequent predictive models may be achieved by tailoring the networks comprising generator and a discriminator in

10.48047/jocaaa.2024.32.01.28

GAN according on the boarding behaviour features. Ultimately, the suggested Deep GAN autonomously chose the data augmentation features and variations. Thus, the enhancements are probably not at their best. Further gains are probably possible if the features plus the ideal imbalance ratio are chosen together, although this would increase computing complexity. This can be put to the test later. Likewise, it has been postulated that the ideal imbalance rate for Deep GAN corresponds to the ideal rate for various other resampling techniques. Further study is required to test this notion. This study highlights the significant advancement provided by the Deep-GAN approach in resolving the data imbalance problem while simulating boarding behaviour, even in its present state. The results may assist public transport authorities in raising the system's efficiency and quality of service by providing a more accurate forecast of boarding behaviour. It is also possible to expand its application to other aspects of the conduct of using public transport, such as improved alighting and transfer behaviour prediction.

REFERENCES

[1] X. Guo, J. Wu, H. Sun, R. Liu, and Z. Gao, “Timetable coordination of first trains in urban railway network: A case study of Beijing,” Applied Mathematical Modelling, vol. 40, no. 17, pp. 8048–8066, 2016.

[2] W. Wu, P. Li, R. Liu, W. Jin, B. Yao, Y. Xie, and C. Ma, “Predicting peak load of bus routes with supply optimization and scaled Shepard interpolation: A newsvendor model,” Transportation Research Part E: Logistics and Transportation Review, vol. 142, p. 102041, 2020.

[3] N. Barinov ˇ c, L. De Donato, F. Flammini, R. M. Loverde, Z. Lin, R. Liu, ´ S. Marrone, R. Nardone, T. Tang, and V. Vittorini, “Artificial intelligence in railway transport: Taxonomy, regulations and applications,” IEEE Transactions on Intelligent Transportation Systems, 2021.

[4] S. C. Kwan and J. H. Hashim, “A review on co-benefits of mass public transportation in climate change mitigation,” Sustainable Cities and Society, vol. 22, pp. 11–18, 2016.

[5] Y. Wang, W. Zhang, T. Tang, D. Wang, and Z. Liu, “Bus od matrix reconstruction based on clustering wi-fi probe data,” Transporter B: Transport Dynamics, pp. 1–16, 2021, Doi: 10.1080/21680566.2021.1956388.

[6] S. J. Berrebi, K. E. Watkins, and J. A. Laval, “A real-time bus dispatching policy to minimize passenger wait on a high frequency route,” Transportation Research Part B: Methodological, vol. 81, pp. 377–389, 2015.

10.48047/jocaaa.2024.32.01.28

[7] A. Fon zone, J.-D. Schmoke, and R. Liu, "A model of bus bunching ¨ under reliability-based passenger arrival patterns," Transportation Research Part C: Emerging Technologies, vol. 59, pp. 164–182, 2015.

[8] J. D. Schmoke, W. Sun, A. Fon zone, and R. Liu, "Bus bunching ¨ along a corridor served by two lines," Transportation Research Part B: Methodological, vol. 93, pp. 300–317, 2016.

[9] D. Chen, Q. Shao, Z. Liu, W. Yu, and C. L. P. Chen, "Ride sourcing behaviour analysis and prediction: A network perspective," IEEE Transactions on Intelligent Transportation Systems, pp. 1–10, 2020.

[10] E. Nelson and N. Sadowsky, "Estimating the impact of ride-hailing app company entry on public transportation use in major us urban areas," The B.E. Journal of Economic Analysis & Policy, vol. 19, no. 1, p. 20180151, 2019.

[11] Z. Chen, K. Liu, J. Wang, and T. Yamamoto, "H-conflate-based bagging learning approach for ride-hailing demand prediction considering imbalance problems and sparse uncertainty," Transportation Research Part C: Emerging Technologies, vol. 140, p. 103709, 2022.

[12] R. Liu and S. Sinha, "Modelling urban bus service and passenger reliability," 2007.

[13] J. A. Suratin, R. Liu, and S. Sinha, "Assessing bus transport reliability using micro-simulation," Transportation Planning and Technology, vol. 31, no. 3, pp. 303–324, 2008.

[14] Y. Wang, W. Zhang, T. Tang, D. Wang, and Z. Liu, "Bus od matrix reconstruction based on clustering wi-fi probe data," Transporter B: Transport Dynamics, pp. 1–16, 2021.

[15] Y. Hollander and R. Liu, "Estimation of the distribution of travel times by repeated simulation," Transportation Research Part C: Emerging Technologies, vol. 16, no. 2, pp. 212–231, 2008.

[16] W. Wu, R. Liu, and W. Jin, "Modelling bus bunching and holding control with vehicle overtaking and distributed passenger boarding behaviour," Transportation Research Part B: Methodological, vol. 104, pp. 175–197, 2017.