Vol.15 No.3(2024),58-69
DOI: https://doi.org/10.61841/turcomat.v15i3.14778

An approach for two-dimensional convolutional neural networks for hourly passenger boarding demand prediction based on uneven smart-card data

Dr. K. Jayarajan [1], Banda Laxmiprasanna [2], Chinthireddy Shravya [3], Akkamgari Laxmi Prasanna [4]

[1] Assistant Professor, Department of Information and Technology, Malla Reddy Engineering College for Women, Autonomous, Hyderabad

[2],[3],[4] Student, Department of Information and Technology, Malla Reddy Engineering College for Women, Autonomous, Hyderabad, bandalaxmiprasanna.2004@gmail.com, chinthireddyshravya@gmail.com, sweetyreddy815@gmail.com

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 numbers at 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 traveling instances to add into a synthetic training dataset containing more evenly distributed traveling and non-traveling examples, we suggest using deep generative adversarial networks (Deep-GAN). Next, a deep neural network, or 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 training set with a better similarity so diversity and, therefore, a stronger prediction capability, according to a comparison of the Deep-GAN's performance with other conventional resampling techniques. The study emphasizes the importance

Vol.15 No.3(2024),58-69
DOI: https://doi.org/10.61841/turcomat.v15i3.14778

of the issue and offers helpful recommendations for enhancing the quality of the data and model performance for individual travel behavior analysis and travel behavior prediction.

INTRODUCTION

The National Natural Science Foundation from Russia (Project No. 71890972/71890970) is acknowledged for its partial financial assistance. Support for Tianli Tang comes from the Jiangsu Funding Programme for Outstanding Fellows Talent and the National Natural Science Fund of China's Key Project (No. 52131203). The UKRI Future Leader the Fellowship, UK, provided Charisma Choudhury with financial assistance. [MR/T020423/1-NEXUS]. The Ministry of Education's Social and Humanities Sciences Foundation (18YJC630190) and Zhejiang Province's Natural Science Foundation (LQ18G030012) both provide assistance for Yuanyuan Wang. For supplying our smart-card data for the present study, Changsha Longxiang Bus Ltd., Ltd. is also appreciated by the authors. Contact T. Tang at T-Tang@seu.edu.cn, the School of Transport, Southeastern University, Nanjing, 211189, China. C. Choudhury and R. Liu are employed at the University of Leeds' Centre for Transport Studies, located in Leeds, LS2 9JT, UK. (E-mails: C.F. Choudhury@leeds.ac.uk, R. Liu@its.leeds.ac.uk) At Napier University in Edinburgh, Edinburgh, E H10 5DT, United Kingdom, A. Fonzone is employed at the Transport Research Center (email: a.fonzone@napier.ac.uk). The author Y. Wang may be reached at wangyuan@zufe.edu.cn, via email at the School for Business Administration, Zhejiang College of Financial and Economic Sciences, Hangzhou, 310018, China. correspondence: R.Liu@its.leeds.ac.uk Urbanization is causing population growth quickly, which in turn is driving up the need for travel and causing negative consequences including air pollution and traffic jams. A common solution to these transportation issues is public transit, which is recognized as an environmentally friendly and sustainable form of mobility. Buses have long been a mainstay in the passenger transportation industry as a traditional public transit option. Bus services, however, are not up to par due to inconsistent journey times, bus bunching, and congestion. Since ride-hailing services have become more popular in recent years, this has led to a decline in bus usage in several cities.The operators of buses need to figure out how to make their vehicles operate better and project a more appealing

Vol.15 No.3(2024),58-69
DOI: https://doi.org/10.61841/turcomat.v15i3.14778

image in order to maintain and grow passenger numbers. Modern bus systems may have much better level-of-service and dependability thanks to advanced operation and administration, which also contributes to higher bus ridership. This calls for figuring out how passenger demand varies both spatially and temporally and adjusting the supply side as needed. Automated fare collection is the original purpose of the smart-card system. As the system also logs boarding details, such as who boards, where, and while smart-card data has grown into an easily available and valuable source of data for spatiotemporal demand analysis, public transportation planning, and additional analysis of lowering emissions for sustainable transportation. Easily observing the passenger movement at bus stops as well as on bus lines allows us to determine the geographical and temporal features of bus journeys based on the data collected from smart cards. It is still quite difficult to mechanically extract meaningful information from huge data, however. Large-scale smart-card dataset analysis has found a useful new tool in machine learning approaches in recent years. Liu et al., for example, used a model with decision trees to identify important characteristics in the prediction of passenger flow in public transportations. Using a neural network model, Zuo et al. developed a three-stage framework to predict individual access in bus systems. We have shown in our very own recent study that using machine learning methods in conjunction with smartcard data may be a potent strategy for forecasting the temporal and geographical trends in bus boarding. Overall, when averaged over all travelers, the projections proved to be quite accurate. Our study has also highlighted the problems with data imbalance that arise when attempting to forecast traveler behavior down to the precise spatial-temporal details and individual traveler level. One example of a rare occurrence would be the boarding of a single smart-card holder in a particular stop within a specific time window (such as an hour). The majority of the data would indicate negative instances (such as non-traveling instead of boarding at that bus stop throughout this time window), with very few positive instances (such as traveling and boarding at that location at this time). When it comes to forecasting travel behavior at the individual traveler and high spatial-temporal detail levels, such imbalances in the data may drastically lower the effectiveness and precision of machine learning models. In order to address the issue of data imbalance when trying to disaggregate boarding demand (i.e., individual travelers boarding actions during each hour of

Vol.15 No.3(2024),58-69
DOI: https://doi.org/10.61841/turcomat.v15i3.14778

the day), we propose an over-sampling method in this study called Deep-generative adversarial nets (Deep-GAN) model. This model was initially created in the context of image generation. Our findings indicate a significant improvement in prediction accuracy when using a more balanced and synthesized database. More benchmarking is done to compare the suggested strategy, which is based on the Deep GAN method, to other resampling techniques, such as the Synthetic Minority Oversampling Technique and Random Under-Sampling, and it exhibits better performance. This is the arrangement for the remainder of the paper. In transport studies, the main resampling techniques are reviewed in Section II along with their uses. The particular problem with data imbalance in hourly boarding demand prediction is explained in Section III. Using a deep neural network, or DNN, to forecast each smart-card holder's boarding behavior (boarding or not boarding) at any time of day, Section IV employs a Deep-GAN to offer a produced, more balanced training sample. Part VI discusses the outcomes after Section V implements the suggested methodology to a case study from the real world. The paper's major conclusions and contributions are finally outlined in Section VII, which also offers suggestions for future research.

RELATED WORK

"Timetable coordination of first trains in urban railway network: A case study of beijing,"

In this study, a model of first train schedule coordination for urban railway networks based on line and transfer station significance is provided. To solve the proposed model, a mathematical programming solution is employed together with the development of a sub-network connection approach. To ensure that our proposed model works as intended, we simulate both a genuine Beijing urban train network and a basic test network. The results show that by drastically reducing the connection time, the suggested approach is beneficial in enhancing the transfer performance.

"Predicting peak load of bus routes with supply optimization and scaled shepard interpolation: A newsvendor model,"

The regularity and vehicle capacity of public transportation routes have a significant impact on the quality of service provided. The aggregate values of these factors go toward the expenses related

Vol.15 No.3(2024),58-69
DOI: https://doi.org/10.61841/turcomat.v15i3.14778

to route operations in addition to expenses related to passenger comfort, such congestion and waiting. Our innovative method for solving the issue integrates operator and passenger expenses into a broader 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 transportation service or to comply with additional 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 categorized 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 utilize a case study with sensitivity analysis to show the model's applicability.

"Artificial intelligence in railway transport: Taxonomy, regulations and applications,"

In most technical fields, artificial intelligence (AI) has become more and more prevalent, and railway transportation is no exception. However, there's a chance that, like many other categories, railway practitioners will become lost in the myriad of new terms and their associated ambiguities, missing out on the true potential and opportunities presented by, to name just a few of the most promising AI-related fields, machine learning, artificial seeing, and big data analytics. This study aims to provide railway scholars and practitioners with an introduction to the fundamental ideas and potential uses of artificial intelligence. In order to achieve this, this paper offers a structured taxonomy that will assist practitioners and researchers in understanding AI methods, domains, fields of study, and applications—both generally and in relation to specific railway applications like traffic management, maintenance, and autonomous driving. Additionally highlighted are the crucial facets of ethics and the capacity of AI to explain railway operations. Relevant research covering both planned and actual applications has supported the link between AI principles and railway subdomains in order to hint in some interesting possibilities.

"A review on co-benefits of mass public transportation in climate change mitigation,"

Vol.15 No.3(2024),58-69
DOI: https://doi.org/10.61841/turcomat.v15i3.14778

Due to the desired win-win outcomes of such initiatives 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 transportation 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. The environmental advantages of public transportation, particularly in terms of lower air pollution in cities, have been the exclusive subject of several studies.

"Bus od matrix reconstruction based on clustering wi-fi probe data,"

The design, running, and administration of the urban transportation system all depend heavily on the calculation of the demand for passengers across the whole city. 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 transport patronage 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 upon the incomplete OD matrix & 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 measured reality 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

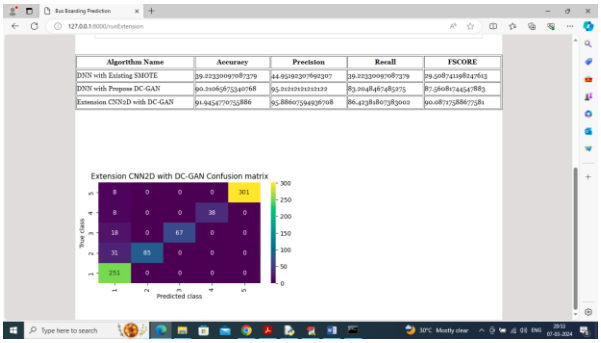
1) **User Login:** Username and password admin and admin may be used to log in to the system.
2) **Process Dataset:** This module loads the dataset, cleans it, normalizes it, and then applies DCGAN to address any imbalance issues.

Vol.15 No.3(2024),58-69
DOI: https://doi.org/10.61841/turcomat.v15i3.14778

3) **Existing Smote:** The correctness of the process data will be tested using test data once the SMOTE algorithm has run.
4) **Propose Deep-GAN DNN:** Using test data, the DNN algorithm calculates accuracy after being trained with DEEP GAN produced 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 create a comparison graph between each algorithm and show the amount of test data and anticipated boarding after that.

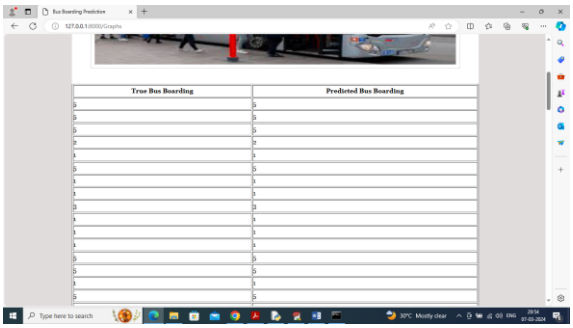
RESULT AND DISCUSSION

In above screen propose DNN with DC-GAN generated data got 90% accuracy and in confusion matrix graph can see all diagnol values are correct prediction count and remaining blue boxes has very few incorrect prediction count. Now click on 'Extension DC-GAN CNN2D' link to train extension algorithm and get below page

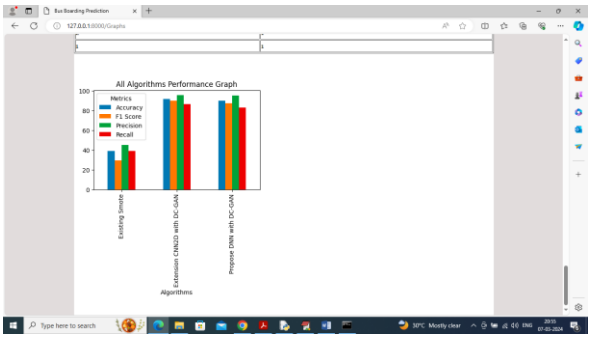


In above screen extension got 91% accuracy and can see other metrics also. Now click on 'Comparison Graph' link to get below graph

Vol.15 No.3(2024),58-69
DOI: https://doi.org/10.61841/turcomat.v15i3.14778



In above screen first column contains Test data boarding person count and second column contains Extension prediction boarding persons and scroll down above page to view below graph



In above graph x-axis represents algorithm names and y-axis represents accuracy and other metrics in different bars and in all algorithms extension got high accuracy.

CONCLUSION

We encountered data imbalance while attempting to anticipate passenger boarding behavior within a time frame using actual bus smart-card data, which served as the impetus for our investigation. In this study, we suggested a Deep-GAN to enhance a DNN based forecasting algorithm for individual boarding behavior by oversampling the traveling instances and rebalancing the rate of traveling 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 ratios within the instruction dataset, we discovered that, on average, the model performs better with more unbalanced data, with the biggest increase occurring at a 1:5 ratio of both positive and negative examples. The high incidence of imbalance will result in deceptive load profiles and the perfectly balanced data may

Vol.15 No.3(2024),58-69
DOI: https://doi.org/10.61841/turcomat.v15i3.14778

overestimate ridership at peak hours from the standpoint of the hour 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 undersampling 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 public transportation system and suggests a deep learning method called Deep-GAN to address it. • The actual and synthetic traveling instances produced by Deep-GAN alongside additional over-sampling techniques were evaluated for differences in similarity or diversity in this research. By assessing the effectiveness of the subsequent travel behavior prediction model, it also examined various resampling techniques for the purpose of improving data quality. This is the initial validation and assessment of the various data resampling techniques based on actual data from the public transportation system. • Unlike prior travel demand prediction challenges, this work creatively modeled individual boarding behavior. This individual-based model may provide more information on passenger behavior than the widely used aggregated forecast, and the findings will help with the study regarding similarities and heterogeneities. Predictive models will advance in sophistication as computer power and technology advance. The bus network and bus routes will eventually give way to individual travel behavior as the focus 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 transportation. It is expected that unbalanced data will be a difficulty for future prediction use in public transportation systems. Our study suggests a Deep-GAN algorithm to deal with the problem of data imbalance in traveler behavior 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

Vol.15 No.3(2024),58-69
DOI: https://doi.org/10.61841/turcomat.v15i3.14778

expertise in handling comparable data imbalance problems, particularly in public transportation. It should be mentioned that even with Both GAN and DNN models' excellent performance, there remain 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 total amount of the dataset can benefit by classifying passengers (using clustering techniques, for example). Third, boarding behavior's spatiotemporal properties are not taken into account by the Deep GAN as it is now. Enhancing the quality of produced dummy traveling instances and the effectiveness of the subsequent predictive models may be achieved by tailoring the networks of producer and a discriminator in GAN according to the boarding behavior 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 characteristics and 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 modeling boarding behavior, even in its present state. The results may assist public transportation authorities in raising the system's efficiency and quality of service by providing a more accurate forecast of boarding behavior. Better alighting or transfer behavior prediction, for example, is only one more way it may be applied to other aspects of public transportation use behavior.

.

REFERENCES

Vol.15 No.3(2024),58-69
DOI: https://doi.org/10.61841/turcomat.v15i3.14778

[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. Besinovi ˇ c, L. De Donato, F. Flammini, R. M. Goverde, 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," Transportmetrica 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.

[7] A. Fonzone, J.-D. Schmocker, 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. Schmocker, W. Sun, A. Fonzone, 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, "Ridesourcing behavior 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.

Vol.15 No.3(2024),58-69
DOI: https://doi.org/10.61841/turcomat.v15i3.14778

[11] Z. Chen, K. Liu, J. Wang, and T. Yamamoto, "H-convlstm-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. Sorratini, R. Liu, and S. Sinha, "Assessing bus transport teliability 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," Transportmetrica 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.