IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, June 2023

Advancements in Machine Learning Techniques for Traffic Flow Prediction in Autonomous Vehicles: A Comprehensive Review

Nagesh Lad[1] and Prof. Dr. Mitra V[2]
Department of E&TC[1,2]
Dr. D. Y. Patil Institute of Technology, Pimpri, Pune, India

Abstract: *Traffic flow prediction is a critical aspect of enabling the successful deployment of autonomous vehicles (AVs) in urban environments. Accurate and reliable traffic flow prediction plays a crucial role in allowing AVs to navigate efficiently and make informed decisions. Machine learning techniques have emerged as powerful tools for traffic flow prediction, leveraging large-scale datasets and complex models to capture the inherent dynamics of traffic patterns. This comprehensive review examines the state-of-the-art machine learning techniques used for traffic flow prediction in AVs, highlighting their strengths, limitations, and future research directions. The review also explores data sources and preprocessing techniques, performance evaluation metrics, case studies, and real-world applications. Furthermore, it discusses the challenges associated with traffic flow prediction for AVs, such as data scarcity and model interpretability, and identifies promising future research directions, including reinforcement learning and multimodal fusion. This review serves as a valuable resource for researchers, practitioners, and policymakers interested in advancing traffic flow prediction for autonomous vehicles.*

Keywords: Traffic flow prediction, Autonomous vehicles, Machine learning techniques, Advancements, Comprehensive review

I. INTRODUCTION

The deployment of autonomous vehicles (AVs) holds great promise for revolutionizing transportation by offering increased safety, efficiency, and convenience. However, for AVs to operate effectively in urban environments, accurate traffic flow prediction is essential. Traffic flow prediction involves forecasting the movement and behavior of vehicles on the road network, enabling AVs to make informed decisions and navigate efficiently through complex traffic scenarios.

Machine learning techniques have emerged as powerful tools in the field of traffic flow prediction. These techniques leverage large-scale datasets, complex models, and sophisticated algorithms to capture the underlying patterns and dynamics of traffic behavior. By analyzing historical traffic data, real-time sensor information, and external factors, machine learning models can provide reliable predictions of future traffic conditions.

This comprehensive review aims to provide an in-depth analysis of the advancements in machine learning techniques for traffic flow prediction in the context of autonomous vehicles. It will explore a wide range of methods employed in traffic flow prediction, including regression models, time series analysis, artificial neural networks, deep learning models, and ensemble techniques. Each technique will be evaluated in terms of its strengths, limitations, and suitability for AV traffic flow prediction.

Furthermore, the review will discuss the data sources commonly used in traffic flow prediction, such as traffic sensors, GPS data, and historical traffic records. It will also examine the pre-processing steps required to clean and prepare the data for machine learning models, ensuring the accuracy and reliability of the predictions.

To evaluate the performance of traffic flow prediction models, various evaluation metrics will be discussed, including mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R-squared). These metrics will enable the comparison and benchmarking of different techniques, providing insights into their effectiveness.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, June 2023

Moreover, the review will present notable case studies and real-world applications where machine learning techniques have been successfully applied to predict traffic flow in autonomous vehicles. These case studies will highlight the methodologies used, data sources employed, model architectures, and performance results, demonstrating the practicality and benefits of these techniques in real-world scenarios.

While machine learning techniques have shown promising results in traffic flow prediction, several challenges and future research directions remain. This review will address these challenges, including data scarcity, model interpretability, real-time prediction, and the integration of external factors, such as weather and events, into prediction models. It will also identify potential future research avenues, such as reinforcement learning and multimodal fusion, that can further enhance the accuracy and effectiveness of traffic flow prediction for autonomous vehicles.

In conclusion, this comprehensive review will provide a valuable resource for researchers, practitioners, and policymakers interested in understanding and advancing the field of traffic flow prediction for autonomous vehicles. By analyzing the advancements in machine learning techniques, data sources, evaluation metrics, and real-world applications, this review aims to facilitate the development of more accurate and reliable traffic flow prediction models, ultimately contributing to the successful deployment of autonomous vehicles in urban environments.

II. MACHINE LEARNING METHODOLOGIES

Shallow Machine Learning based methodologies were implemented for traffic congestion prediction, which showed improved prediction result comparing with probabilistic reasoning or other non-AI approaches.

2.1 Linear Regression

With the help of a multiple linear regression model and the mean absolute percentage error valuation approach, Lee et al. [1] were able to predict traffic congestion in relation to weather data with an accuracy of 84.8 percent.

In order to forecast the future traffic condition using linear regression and compare the anticipated traffic condition with the actual traffic condition, Putra et al. [2] planned to conduct an app poll among drivers on one of the routes close to Telkom University. The study's findings demonstrate that the suggested strategy accurately forecasts traffic conditions that are in the same class of degree of service as the actual conditions.

However, it's possible that linear regression doesn't always work to anticipate traffic congestion. Non-linear correlations between traffic factors cannot be modelled using linear regression. Other data-driven modelling methods, like ANN, kNN, and support vector regression (SVR), can be used to represent non-linear connections [3].

2.2 Support Vector Regression

In the pre-processing step, Satrinia and Saptawati [4] suggested utilising Map Matching in conjunction with a topological information approach. Map Matching generated a new trajectory that matched the road, and using those new trajectories, they estimated the speed for each road section. Support Vector Regression (SVR) was used by Satrinia and Saptawati [46] to forecast traffic speed, and the findings showed that Map Matching may be used to gain more accurate traffic speed information and SVR performs well in this regard.

Support vector regression (SVR) models, according to Philip et al. [3], forecast trip durations with a respectable level of accuracy, particularly when there is a dearth of data or a significant degree of data variability. In order to forecast the journey time for the present interval, Philip et al. [3] used travel times from the previous 40 min. They then developed SVR, ANN, and moving average models with 8 input variables and compared the results. The SVR model they utilised outperformed an Artificial Neural Network model and moving average technique under Indian traffic conditions, according to their research, and could be used to estimate the journey time at the next moment rather accurately using past travel time measurements.

2.3 Decision Trees

Finding rules or relationships that permit classification based on attributes is the goal of decision trees [5]. Decision trees, artificial neural networks, and nearest neighbours algorithms, according to Florido et al. [6], have been successfully used to anticipate traffic congestion at a specific site in Sevilla, Spain. The best traffic congestion forecast

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, June 2023

result among the three models was a decision tree [6]. According to Elleuch et al. [7], the findings were enhanced more by fusing the real-time GPS data, the anticipated congestion condition, and the anomalous events.

A comparison of decision tree, logistic regression, and neural network traffic congestion prediction systems was presented by Tamir et al. in their paper [8]. They trained and tested the model using 1,231,200 samples from five days of traffic data and the machine learning frameworks TensorFlow and Clementine [53]. In the Python programming environment, they claimed that decision trees outperform logistic regression and neural networks with an accuracy of 97.65 percent, while decision trees outperform neural networks and logistic regression with an accuracy of 95.9 percent, 96.9 percent, and 96.9 percent, respectively, in the Clementine environment. Decision Trees, according to Mystakidis and Tjortjis [9], are more accurate than Logistic Regression.

2.4 Gradient Boosting Decision Tree (GBDT)

Based on the Spark platform parallel Gradient Boosting Decision Tree algorithm, Bai et al. [10] proposed a method for forecasting urban road congestion. They demonstrated that the method can accurately forecast urban road congestion, shorten running time, increase prediction accuracy, and effectively assist with urban road management.

2.5 Random Forests

Because the random forest algorithm has the qualities of high resilience, high performance, and high practicability, Liu and Wu [11] advocated using it. As model input variables, they employed the weather conditions, time of year, season, unique road conditions, road quality, and holiday. The findings demonstrated that the traffic prediction model developed using the random forest classification algorithm had an accuracy of 87.5 percent, a low generalisation error, and could be effectively predicted. In addition, the calculation speed was quick, and it had a stronger applicability to the prediction of congested conditions.

Silva and Martins [12] asserted that Multiple Regression, KNearest Neighbors (KNN), Neural Network (Multilayer Perceptron), Random Forest, and Support Vector Machine were the models that would yield the best results for the situation at hand (SVM). They indicated that the Random Forest model would be the one that was most suited for the assigned task because it had the best overall scores and a really quick reaction time, giving it a significant advantage over the similarly scoring KNN that took over 10 minutes to do a task.

A short-term traffic congestion prediction system based on the random forest algorithm was put forth by Shenghua et al. [13]. For the simulation experiment, they employed high-speed road traffic data from the PeMS database in the United States. The experimental results showed that the accuracy of their approach was 94.36 percent, demonstrating the method's exceptional performance.

By enhancing the current random forest algorithm, Chen et al. [14] proposed a mixed forest prediction technique that takes into account the spatio-temporal correlation characteristics of the state of urban road traffic. For the classified forest, regression forest, post classified forest, and mixed forest algorithms, the best eigenvector and decision tree number combinations were chosen. The outcomes demonstrated the model's efficacy and potential for big data application.

2.6 K-Nearest Neighbors Algorithm (K-NN)

In order to forecast the effects of the second wave of traffic, Kwoczek et al. [15] modified the K-Nearest Neighbors (K-NN) method and searched among historical observations for the prior PSEs (cases) that shared the most similarities. According to Kwoczek et al. [15], one of the challenges in using the K-Nearest Neighbors algorithm was selecting the parameter k, which determined how many training examples needed to be in the feature space that were closest to each other.

Decision trees, artificial neural networks, and closest neighbours algorithms, according to Florido et al. [16], have been successfully used to predict traffic congestion at a specific site in Sevilla, Spain. Of the three models, NN had the worst prediction of traffic congestion.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, June 2023

III. DEEP LEARNING METHODOLOGIES

A neural network model using Suzhou's bike-sharing, holiday, and weather data was proposed by Wang et al. [17]. To confirm the use of the entity embedding in traffic issues, they forecasted the traffic flow of the bike sites using pertinent categorical data. The study's findings shown that entity embedding can significantly boost categorical variables' continuity, which boosts the predictive power of neural network models.

A method using ANN was proposed by Elleuch et al. [18] that considered real-time unpredictable occurrences like accidents in addition to historical GPS data. These events can have an impact on traffic bottlenecks. The outcome showed how well the suggested model predicted traffic congestion.

An artificial neural network (ANN) forecasting model was put forth by Olayode et al. [19] using the South African Road transportation system's traffic flow variables as a case study. These traffic data sets included a variety of different vehicle types as well as their speeds on the road as well as input and output variables for traffic density, time, and volume.

A method for forecasting traffic conditions was given by Hossain and Uddin [20] by establishing a relationship between two different datasets in accordance with time sequence, which identified groupings of traffic states with comparable patterns. They discovered that using the most data sets possible was advantageous and that the suggested ConvNet neural network performed admirably for classifying or segmenting images.

Using CNN, Bartlett et al. [21] suggested an online dynamical framework. The experiment's findings demonstrated that both short- and long-term temporal patterns increased prediction accuracy, and the suggested online dynamical framework outperformed a deep gated recurrent unit model in terms of prediction results by 10.8%. Their online learning framework's primary drawback was the ultimate loss of the long-term temporal patterns that were inherent in the training data.

IV. CONCLUSION

One of the biggest issues in today's society is the amount of traffic and the congestion it causes. The number of vehicles on our route is increasing alarmingly daily, resulting in delays and frustration for drivers caught in traffic jams. For the prediction of traffic congestion, some research initiatives in the past used probabilistic reasoning techniques like fuzzy logic and hidden Markov models. But more recently, a growing number of research groups have used methods based on machine learning or deep learning to predict traffic congestion. Deep learning-based approaches for predicting traffic congestion have particularly gained popularity during the past five years. There are still a lot of opportunities to utilise other machine learning or deep learning based methodologies for traffic congestion prediction. A high interest is currently given to spatial-temporal modelling by considering network characteristics from the model. On top of that, more types of datasets can be utilised for traffic congestion prediction. Semi-supervised deep learning models can be a good way to utilize more types of datasets for traffic congestion prediction.

REFERENCES

[1]. Lee, J., Hong, B., Lee, K., Jang, Y.J., 2015. A prediction model of traffic congestion using weather data, in: 2015 IEEE International Conference on Data Science and Data Intensive Systems, pp. 81–88.

[2]. Putra, N.A.P., Lhaksmana, K.M., Wahyudi, B.A., 2018. An android application for predicting traffic congestion using polling method. 2018 6th International Conference on Information and Communication Technology, ICoICT2018, 178–183

[3]. Philip, A.M., Ramadurai, G., Vanajakshi, L., 2018. Urban arterial travel time prediction using support vector regression. Transportationin Developing Economies 2018 4:1 4, 1–8.

[4]. Satrinia, D., Saptawati, G.A., 2018. Traffic speed prediction from gps data of taxi trip using support vector regression. Proceedings of 2017 International Conference on Data and Software Engineering, ICoDSE 2017 2018-January, 1–6. URL: https://ieeexplore.ieee.org/ document/8285869

[5]. Asencio-Cortés, G., Florido, E., Troncoso, A., Martínez-Álvarez, F., 2016. A novel methodology to predict urban traffic congestionwith ensemble learning. Soft Computing 2016 20:11 20, 4205–4216.

[6]. Florido, E., Castaño, O., Troncoso, A., Martínez-Álvarez, F., 2015. Data mining for predicting traffic congestion and its application to spanish data. Advances in Intelligent Systems and Computing 368, 341–351.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 5, June 2023

[7]. Elleuch, W., Wali, A., Alimi, A.M., 2016. Intelligent traffic congestion prediction system based on ann and decision tree using big gps traces. Advances in Intelligent Systems and Computing 557, 478–487.
[8]. Tamir, T.S., Xiong, G., Li, Z., Tao, H., Shen, Z., Hu, B., Menkir, H.M., 2020. Traffic congestion prediction using decision tree, logistic regression and neural networks. IFAC-PapersOnLine 53, 512–517.
[9]. Mystakidis, A., Tjortjis, C., 2020. Big data mining for smart cities: Predicting traffic congestion using classification. 11th International Conference on Information, Intelligence, Systems and Applications, IISA 2020
[10]. Bai, X., Feng, Y., Li, L., Zhang, L., 2020. Research on prediction of urban road congestion based on spark-gbdt. 2020 IEEE 5th Interna- tional Conference on Intelligent Transportation Engineering, ICITE 2020, 102–106
[11]. Liu, Y., Wu, H., 2018. Prediction of road traffic congestion based on random forest. Proceedings - 2017 10th International Symposium on Computational Intelligence and Design, ISCID 2017 2, 361–364.
[12]. Silva, C., Martins, F., 2020. Traffic flow prediction using public transport and weather data: A medium sized city case study. Advances in Intel- ligent Systems and Computing 1160 AISC, 381–390.
[13]. Shenghua, H., Zhihua, N., Jiaxin, H., 2020. Road traffic congestion prediction based on random forest and dbscan combined model. Proceed- ings - 2020 5th International Conference on Smart Grid and Electrical Automation, ICSGEA 2020, 323–326
[14]. Chen, Z., Jiang, Y., Sun, D., 2020. Discrimination and prediction of traffic congestion states of urban road network based on spatio-temporal correlation. IEEE Access 8, 3330–3342.
[15]. Kwoczek, S., Martino, S.D., Nejdl, W., 2014. Predicting and visualizing traffic congestion in the presence of planned special events. Journal of Visual Languages and Computing 25, 973–980. S1045926X14001219.
[16]. Florido, E., Castaño, O., Troncoso, A., Martínez-Álvarez, F., 2015. Data mining for predicting traffic congestion and its application to spanish data. Advances in Intelligent Systems and Computing 368, 341–351.
[17]. Wang, B., Shaaban, K., Kim, I., 2019. Reveal the hidden layer via entity embedding in traffic prediction. Procedia Computer Science 151, 163–170.
[18]. Elleuch, W., Wali, A., Alimi, A.M., 2016. Intelligent traffic congestion prediction system based on ann and decision tree using big gps traces. Advances in Intelligent Systems and Computing 557, 478–487.
[19]. Olayode, I.O., Tartibu, L.K., Okwu, M.O., 2021. Prediction and modeling of traffic flow of human-driven vehicles at a signalized road intersection using artificial neural network model: A south african road transportation system scenario. Transportation Engineering 6, 100095.
[20]. Hossain, M.A., Uddin, M.N., 2019. Forecast upcoming traffic states by exploiting nearest junctions and big data. 1st International Conference on Advances in Science, Engineering and Robotics Technology 2019, ICASERT 2019
[21]. Bartlett, Z., Han, L., Nguyen, T.T., Johnson, P., 2019. A novel online dynamic temporal context neural network framework for the prediction of road traffic flow. IEEE Access 7, 153533–153541.