© November 2024 | IJIRT | Volume 11 Issue 6 | ISSN: 2349-6002

A Review on Traffic Forecast Based on Machine Learning Regression Models

Rajeshwari Pawar[1] Prof. Snehlata Mishra[2]
Department of Computer Science and Engineering[1,2]
SAGE University, Indore[1,2]

***Abstract:* In the recent times, there has been a significant surge in the utilization of statistical and evolutionary algorithms in the development of intelligent traffic systems, particularly within the realm of Intelligent Transportation Systems (ITS). This involves the exploration of various methodologies to predict or estimate traffic volume in specific geographic areas under varying conditions. The objective is to effectively monitor and manage substantial traffic flows, a formidable challenge in urban and semi-urban areas globally. The forecasting task is intricate due to the inherently random and uncorrelated nature of the data. Establishing a clear functional relationship, such as through correlation or regression analysis, is rarely predefined. Consequently, conventional statistical algorithms are being investigated in a preliminary phase to adapt parameters according to the dynamic statistical properties of the input data.**

This paper provides a thorough examination of the necessity for evolutionary statistical algorithms in addressing traffic forecasting challenges, highlighting key points from existing literature. Additionally, it presents an extensive review of conventional statistical algorithms employed thus far. The paper concludes by elucidating performance metrics used to assess the efficacy of these algorithms. This comprehensive review aims to establish a foundational understanding for future research endeavors in this domain.

***Keywords:* Intelligent Traffic System (ITS), Machine Learning, Deep Learning, Regression Models.**

I. INTRODUCTION

The global population boom and mass migration of huge segments of the population into urban and semi-urban regions has raised the demand for automated and intelligent traffic monitoring systems. An important metric in the examination of transportation operations and highway performance is traffic volume, or throughput. Highly detailed traffic volume data is essential for locating congested areas, helping with traffic re-distribution, and putting accident avoidance plans into action [1]. Moreover, it serves as the disaggregated source for determining the average daily traffic throughout the year (AADT) [2].

At the network level, AADT provides a gauge for the total amount of highway capacity used, suggests the quality of service provided by the roads, and may be applied to trend analysis, project prioritisation, and highway planning [3]. Currently, sensors like inductive loop detectors, radar detectors, and/or continuous counting stations (CCS) are the primary sources of traffic count (volume) [4]. However, given budgetary constraints, deploying sensors with a wide network coverage might be costly and impracticable, particularly in rural locations [5]. Because of this, during the past ten years, the question of how to spatially estimate or anticipate traffic flow has become more interesting in place of huge sensor deployment. The following are some examples of common uses [6]:

- Navigation & Route Optimization
- Intelligent Parking
- Illumination.
- Mishap Identification
- Mishaps on roads

Typically, the traffic speed can be modelled as [7]:

$$Speed = f(time, lane\ congestion, other\ variables) \quad (1)$$

The most common diagrammatic representation of traffic speed analysis happens to be the horseshoe diagram which essentially represents the relation among:

- Vehicles per hour (count)
- Average Speed

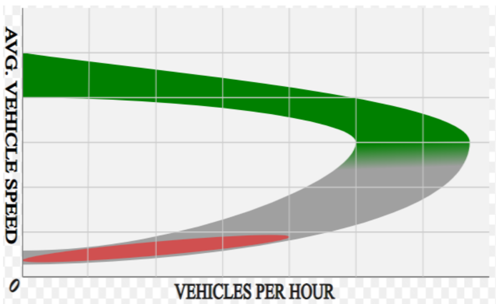

Fig.2 Horse shoe diagram

(Source:https://en.m.wikipedia.org/wiki/File:Speed-flow_horseshoe_diagram_traffic_congestion)

The horseshoe diagram is an essential instrument in analyzing traffic speed, especially in comprehending the dispersion of vehicle speeds throughout various segments of a road [8]. Its purpose is to visually depict the movement and behavior of traffic in different situations, emphasizing locations where there are changes in speed [9]. Through the utilization of this graphic, traffic engineers and analysts can acquire valuable understanding regarding the fluctuations in speed, a critical factor in the enhancement of road safety and the optimization of traffic management [10].

The main significance of the horseshoe diagram comes in its capacity to detect and examine fluctuations in speed that may result in possible safety risks. The uniformity of traffic speed along a roadway is influenced by various factors such as road design, traffic density, and environmental conditions [11]. The horseshoe diagram aids in identifying specific spots where speeds surpass or fall short of the average, potentially indicating a necessity for revisions to speed limits, enhancements in road design, or heightened enforcement [12]. This graphic is especially valuable for identifying crucial regions such as bends, junctions, or congested zones where events connected to speed are more prone to happen.

There are several models to forecast traffic speeds based on regression analysis. The analysis of such models is presented next [13]:

III. BASELINE APPROACHES FOR TRAFFIC FORECASTING

This section presents the common models used for traffic forecasting:

Traffic speed prediction plays a crucial role in intelligent transportation systems, facilitating improved traffic control, decreased congestion, and heightened road safety [14]. Machine learning algorithms are often used for this task because they can effectively analyze intricate patterns in traffic data. These models have the ability to forecast real-time or future traffic speeds by assessing multiple parameters, including historical speed data, meteorological conditions, and traffic volume [15].

Models of linear regression

Linear regression is a basic and widely utilized machine learning technique for predicting traffic speed [16]. It operates by establishing a direct correlation between the dependent variable (traffic speed) and one or more independent factors (such as time of day, weather, and traffic flow) [17]. Although linear regression models are straightforward to apply and understand, they may not accurately represent the non-linear connections that are frequently observed in traffic data. However, they can be valuable as a fundamental model or in situations when the data demonstrates a prominent linear pattern [17].

Decision Trees and Random Forests

Decision trees are a widely used technique for predicting traffic speed. They operate by partitioning the data into subsets according to the input feature values, resulting in a hierarchical arrangement of judgments that ultimately yield a prediction [18]. Random forests, composed of numerous decision trees collaborating, are highly successful in mitigating overfitting and enhancing prediction accuracy [19]. These models provide the ability to manage intricate and non-linear connections in traffic data, rendering them appropriate for highly dynamic traffic settings [20].

Support Vector Machines (SVM)

Support Vector Machines (SVM) are robust models utilized for classification and regression tasks, such as predicting traffic speed. SVM operates by identifying the most advantageous hyperplane that effectively divides the data into distinct classes or predicts a continuous value [21]. SVM can be employed in traffic speed prediction to represent intricate connections between traffic speed and diverse influencing factors. While Support Vector Machines (SVMs) are efficient for datasets of small to medium size, they can be computationally burdensome when dealing with huge datasets.

Neural Networks and Deep Learning

Neural networks, especially deep learning models, have become popular in traffic speed prediction because they can effectively represent intricate and non-linear connections. These models are composed of numerous layers of interconnected neurons that analyze and acquire knowledge from data [22]. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are particularly advantageous for predicting traffic speed because they can effectively capture the sequential patterns in the data, making them well-suited for forecasting time-series data. Nevertheless, neural networks necessitate substantial quantities of data and processing resources, which might pose a constraint [23].

Gradient Boosting Machines

Gradient Boosting Machines (GBM) are a type of ensemble learning algorithm that enhances accuracy by combining the predictions of numerous weak learners, usually decision trees [24]. GBM models, such as XGBoost and LightGBM, have demonstrated significant potential in predicting traffic speed. This is attributed to their capacity to effectively handle extensive datasets and intricate feature relationships. Additionally, they exhibit resilience against overfitting and frequently surpass other models in terms of predictive accuracy [25].

III. PREVIOUS WORK

The previous work in the domain is cited in this section

Wang et al. [26] proposed a novel graph-based spatio-temporal autoencoder that utilizes an encoder-decoder architecture to predict traffic speed in a spatio-temporal context, specifically addressing the issue of missing values. More precisely, we consider imputation and prediction as separate tasks that we train one after the other. This approach helps us minimize the negative effects of imputation on the raw data used for prediction and also speeds up the model training process. In addition, we employ graph convolutional layers that incorporate a self-adaptive adjacency matrix to simulate spatial dependencies. We also leverage gated recurrent units to facilitate temporal learning. In order to assess the suggested model, we do thorough case studies on two actual traffic datasets. These datasets exhibit two distinct missing patterns and encompass a broad and realistic range of missing rates, spanning from 20% to 80%. The experimental findings consistently show that the model performs better than the current state-of-the-art methods for predicting traffic with missing values. Additionally, the model maintains a steady level of performance across different missing circumstances and prediction horizons.

Ma at al. [27] proposed involves applying the ARIMA statistical time series model to analyze the residuals obtained from the MLP machine learning algorithm. This allows for accurate prediction of traffic conditions across a network. The method is referred to as NN-ARIMA. The residual time series of the neural network is used to extract location-specific traffic features through the use of ARIMA. Subsequently, the neural network MLP is employed to capture the network-scale co-movement pattern of all traffic flows. The experiment data indicate that applying ARIMA analysis to the neural network residuals as a post-processing step leads to a significant improvement in the accuracy of traffic state forecasts, specifically in terms of reducing mean squared error.

Haghighat et al. [28] proposed that (ITS) has been greatly improved by utilizing deep learning techniques, which have transformed problem areas that were previously tackled using analytical or statistical methods, and have even extended into previously unexplored domains. These technological developments have resulted in decreased maintenance expenses, improved safety and security along transportation routes, optimized operations in public transportation and ride-sharing industries, and driven research on autonomous vehicles to unparalleled levels. In addition, deep learning has made traffic control and planning simpler. The objective of this study is to demonstrate the advancements in research on intelligent transportation systems (ITS) that have been made possible by the use of deep learning. It seeks to provide a thorough analysis and detailed insight into the application of deep learning models in ITS.

Hydari et al. [29] proposed that data-driven approaches have provided a fresh opportunity for studying control-based systems in several fields such as power, robotics, transportation, and the Internet of Things. The integration of data-driven applications with transport systems is essential in modern transportation applications. This study investigates the latest applications of deep reinforcement learning (RL) in traffic control.

The text focuses on a detailed examination of traffic signal control (TSC) applications that utilize reinforcement learning (RL) techniques. These applications have been extensively studied in existing research. A comprehensive examination is conducted on different formulations of issues, settings for reinforcement learning (RL), and simulation environments for traffic signal control (TSC). Deep reinforcement learning (RL) models have been extensively studied in the literature for various autonomous driving applications.

Dazango et al. [30] proposed that a specific set of pathways in space-time, which have the lowest cost and are the shortest, may solve any well-defined traffic problem involving kinematic waves. These pathways have a unique metric and are associated with a flow-density relationship that is concave. The equi-cost contours represent the paths of vehicles. If the flow-density relation is strictly concave, the set of waves will have a unique collection of shortest routes. Shocks, if they happen, are the curves where the shortest routes end in the solution area. The revised formulation significantly streamlines kinematic wave theory while broadening its scope of application. One such approach is to employ space-time shortcuts as a model to depict the movement of slow buses, which can be challenging to deal with using existing methods.

The performance evaluation metrics are:

$$MAPE = \frac{100}{M} \sum_{i=1}^{N} \frac{E-E_i|}{i} \quad (1)$$

The accuracy of prediction is computed as:

$$Ac = 100 - \frac{100}{M} \sum_{i=1}^{N} \frac{E-E_i|}{i} \% \quad (2)$$

Here,
n is the number of errors
i is the iteration number
E is the actual value
E_i is the predicted value

V. CONCLUSION

From the previous discussion, it is evident that smart and intelligent traffic systems are essential for effectively managing the high traffic volumes in urban and semi-urban areas worldwide, which are typically caused by major population shifts. Although there are several different uses for smart traffic monitoring, including optimizing routes and navigation, managing parking, controlling lighting, detecting accidents, identifying road anomalies, and managing infrastructure, it is crucial to consider the calculation of traffic volume across the entire network. This study provides a comprehensive analysis of contemporary statistical methods used to forecast the volume of traffic throughout a whole network. It focuses on important features of previous research in this area. The objective is to lay the groundwork for creating efficient and precise prediction models for traffic volumes in the realm of intelligent traffic and transportation systems.

REFERENCES

[1] Z. Diao et al., "A Hybrid Model for Short-Term Traffic Volume Prediction in Massive Transportation Systems," in IEEE Transactions on Intelligent Transportation Systems, 2018, vol. 20, no. 3, pp. 935-946.

[2] A Pompigna, F Rupi, “Comparing practice-ready forecast models for weekly and monthly fluctuations of average daily traffic and enhancing accuracy by weighting methods”, Journal of Traffic and Transportation Engineering , Elsevier 2017, vol. 5, Issue 4, pp. 239-253.

[3] I. M. Wagner-Muns, I. G. Guardiola, V. A. Samaranayke and W. I. Kayani, "A Functional Data Analysis Approach to Traffic Volume Forecasting," in IEEE Transactions on Intelligent Transportation Systems, 2017, vol. 19, no. 3, pp. 878-888.

[4] AS Soliño, ALL Galera,, “Measuring uncertainty of traffic volume on motorway concessions: a time-series analysis”, Transportation Research Procedia, Elsevier 2017, Volume 27, pp.3-10.

[5] D. Wang, Q. Zhang, S. Wu, X. Li and R. Wang, "Traffic flow forecast with urban transport network," 2016 IEEE International Conference on Intelligent Transportation Engineering (ICITE), 2016, pp. 139-143.

[6] K. Xie and R. Li, "A combined forecasting method for traffic volume," 2016 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), 2016, pp. 95-100

[7] M Meng, J Zhang, YD Wong, PH Au, “Effect of weather conditions and weather forecast on cycling travel behavior in Singapore”, International Journal of Sustainable Transportation, Taylor and Francis, 2016, vol.10, issue.0, pp. 773-780.

[8] M Moniruzzaman, H Maoh, W Anderson, “Short-term prediction of border crossing time and traffic volume for commercial trucks: A case study for the Ambassador Bridge”. Transportation Research Part C: Emerging Technologies, Elsevier 2016, Vol. 63, pp. 182-194.

[9] J Tang, G Zhang, Y Wang, H Wang, F Liu, “A hybrid approach to integrate fuzzy C-means based imputation method with genetic algorithm for missing traffic volume data estimation”, Transportation Research Part C: Emerging

Technologies, Elsevier 2015, Volume 51, pp.29-40.

[10] K. C. Dey, A. Mishra and M. Chowdhury, "Potential of Intelligent Transportation Systems in Mitigating Adverse Weather Impacts on Road Mobility: A Review," in IEEE Transactions on Intelligent Transportation Systems, 2015, vol. 16, no. 3, pp. 1107-111.

[11] M Xu, Y Di, H Ding, Z Zhu, X Chen, H Yang, “AGNP: Network-wide short-term probabilistic traffic speed prediction and imputation”, Communications in Transportation Research, Elsevier 2023, vol.3, 100099.

[12] A Sharma, A Sharma, P Nikashina, V Gavrilenko, “A graph neural network (GNN)-based approach for real-time estimation of traffic speed in sustainable smart cities”, Sustainability, MDPI 2023, vol.15, no.5, pp.1-25.

[13] R Qaddoura, MB Younes, “Temporal prediction of traffic characteristics on real road scenarios in Amman”, Journal of Ambient Intelligence and Humanized Computing, Springer, 2023. Vol.14, pp. 9751–9766.

[14] Y Zhang, T Zhao, S Gao, M Raubal, “Incorporating multimodal context information into traffic speed forecasting through graph deep learning”, International Journal of Geographical Information Science, Taylor and Francis, 2023, vol.37, no.9, pp. 1909-1935.

[15] Y. Gao, C. Zhou, J. Rong, Y. Wang and S. Liu, "Short-Term Traffic Speed Forecasting Using a Deep Learning Method Based on Multitemporal Traffic Flow Volume," in IEEE Access, vol. 10, pp. 82384-82395, 2022.

[16] C. Ma, Y. Zhao, G. Dai, X. Xu and S. -C. Wong, "A Novel STFSA-CNN-GRU Hybrid Model for Short-Term Traffic Speed Prediction," in IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 4, pp. 3728-3737, April 2022.

[17] X. Xu, T. Zhang, C. Xu, Z. Cui and J. Yang, "Spatial–Temporal Tensor Graph Convolutional Network for Traffic Speed Prediction," in IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 1, pp. 92-103,2023.

[18] C Xu, A Zhang, C Xu, Y Chen, “Traffic speed prediction: spatiotemporal convolution network based on long-term, short-term and spatial features”, Applied Intelligence, Springer, 2022, vol.52, pp. 2224–2242.

[19] Z Zhang, Y Li, H Song, H Dong, “Multiple dynamic graph based traffic speed prediction method”, Neurocomputing, Elsevier, 2021, vol.461, pp. 109-117.

[20] A. Abdelraouf, M. Abdel-Aty and J. Yuan, "Utilizing Attention-Based Multi-Encoder-Decoder Neural Networks for Freeway Traffic Speed Prediction," in IEEE Transactions on Intelligent Transportation Systems, 2021, vol. 23, no. 8, pp. 11960-11969.

[21] J. J. Q. Yu, C. Markos and S. Zhang, "Long-Term Urban Traffic Speed Prediction With Deep Learning on Graphs," in IEEE Transactions on Intelligent Transportation Systems, 2021, vol. 23, no. 7, pp. 7359-7370.

[22] YL Hsueh, YR Yang, “A short-term traffic speed prediction model based on LSTM networks”, International Journal of Intelligent Transportation Systems Research, Springer 2021, vol.19, pp. 510–524.

[23] M. Cao, V. O. K. Li and V. W. S. Chan, "A CNN-LSTM Model for Traffic Speed Prediction," 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), Antwerp, Belgium, 2020, pp. 1-5.

[24] X. Yang, Y. Yuan and Z. Liu, "Short-Term Traffic Speed Prediction of Urban Road With Multi-Source Data," in IEEE Access, vol. 8, pp. 87541-87551, 2020.

[25] X. Meng et al., "D-LSTM: Short-Term Road Traffic Speed Prediction Model Based on GPS Positioning Data," in IEEE Transactions on Intelligent Transportation Systems, 2020 vol. 23, no. 3, pp. 2021-2030.

[26] A. Wang, Y. Ye, X. Song, S. Zhang and J. J. Q. Yu, "Traffic Prediction With Missing Data: A Multi-Task Learning Approach," in IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 4, pp. 4189-4202

[27] T Ma, C Antoniou, T Toledo, “Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast”, Journal of Transportation Research Part, Emerging Technologies, Elsevier 2022, Volume 111, pp. 352-372.

[28] Haghighat, A.K., Ravichandra-Mouli, V., Chakraborty, P. et al. Applications of Deep Learning in Intelligent Transportation Systems. J. Big Data Anal. Transp, Springer 2021, vol. 2, 115–145.

[29] A. Haydari and Y. Yılmaz, "Deep Reinforcement Learning for Intelligent Transportation Systems: A Survey," in IEEE Transactions on Intelligent Transportation Systems, 2020, vol. 23, no. 1, pp. 11-32.

[30] CF Daganzo, A variational formulation of kinematic waves: basic theory and complex boundary conditions, Transportation Research Part B: Methodological, Elsevier 2019, vol.39, Issue 2, pp.187-196