A Survey on Annual Average Daily Traffic (AADT) Driven Models for Traffic Forecasts

Manju Chouhan[1], Prof. Vinay W. Deulkar [2]
Department of Civil Engineering[1,2]
JIT, Borawan, India.[1,2]

Abstract:- ***Off late, statistical and evolutionary algorithms are being extensively used for smart traffic systems whose sub-section is Intelligent Transportation Systems (ITS). For that purpose, different methods are being explored to estimate or forecast the traffic volume in a geographic are under different prevailing conditions so as to monitor and control massive traffic volumes, which has become a serious challenges in urban and even semi-urban areas worldwide. The prediction or forecasting problem is challenging since the nature of the data is extremely random and uncorrelated. A clear functional relationship in terms of correlation or regression analysis is seldom preset. Hence conventional statistical algorithms are being explored in the pretest which an adapt parameters as per the changing statistical properties of the fed data. This paper presents a comprehensive review on the need for evolutionary statistical algorithms for traffic forecasting problems and also cites the salient points of the existing literature. Moreover a comprehensive review of existing statistical algorithms used hitherto, are also cited. Finally the performance metrics are explained to evaluate the performance of such algorithms. This comprehensive review is expected to serve as a baseline for further research in the domain.***

Keywords: Smart Traffic Monitoring, Intelligent Traffic systems (ITS), Evolutionary Statistical Methods, Performance Metrics.

I. Introduction

As population explosion and mass exodus of large chunks of population towards urban and semi-urban areas have become prevalent throughout the world, the necessity for automated and smart traffic monitoring systems have also increased [1]. Traffic volume or throughput serves as a crucial indicator in highway performance and transportation operation analysis. Highly- granular traffic volume provides key information in identifying congested roadways, assisting traffic re-distribution, and implementing accident prevention strategies. Furthermore, it is the dis-aggregated source for calculating annual average daily traffic (AADT) [2]. AADT at the network level offers a measure of overall utilization of highway facility, implies the level of service of roads, and can be used for highway planning, trend studies, and project prioritization currently, traffic count (volume) is mainly obtained from sensors such as inductive loop detectors, radar detectors and/or continuous counting stations (CCS [3]). Yet installing sensors with a large network coverage can be impractical and expensive given budget constraint, especially in rural areas. As a result, how to spatially estimate/predict traffic volume to substitute massive sensor deployment has been an intriguing topic over the past decade. Some typical applications can be [4]:

1) Route Optimization & Navigation.
2) Smart Parking.
3) Lighting.
4) Accident Detection.
5) Road Anomalies.
6) Infrastructure Management.

II. Existing Evolutionary Statistical Models

Due to the need of large data sets to be analyzed, it is necessary to use computational tools which are fast, accurate and can handle copious amounts of data [5]. Evolutionary statistical algorithms are a set of such algorithms which show the aforesaid characteristics [6]. Evolutionary Statistical algorithms try to mimic the human attributes of thinking which are [7]:

1) Parallel data processing
2) Self-Organization
3) Learning from experiences

Some of the commonly used techniques are discussed below [8]-[9]:

1) Statistical Regression: These techniques are based on the time series approach based on the fitting problem that accurately fits the data set at hand. The approach

generally uses the auto-regressive models and means statistical measures. They can be further classified as:
a) Linear
b) Non-Linear
Mathematically:
Let the time series data set be expressed as:

$$\boldsymbol{Y = \{Y1, Y2 \ldots \ldots \ldots \ldots \ldots \ldots .. Yt\}} \quad \textbf{(1)}$$

Here,
Y represents the data set
t represents the number of samples
Let the lags in the data be expressed as the consecutive differences.
The first lag is given by:

$$\boldsymbol{\Delta Y_1 = Y_{t-1}} \quad \textbf{(2)}$$

Similarly, the j^{th} lag is given by:

$$\boldsymbol{\Delta Y_j = Y_{t-j}} \quad \textbf{(3)}$$

2) Correlation based fitting of time series data: The correlation based approaches try to fit the data based on the correlation among the individual lags. Mathematically it can be given by:

$$\boldsymbol{A_t = corr(Y_t, Y_{t-1})} \quad \textbf{(4)}$$

Here,
Corr represents the auto-correlation (which is also called the serial correlation)
Y_t is the t^{th} lagged value
Y_{t-1} is the $(t-1)^{st}$ lagged value
The mathematical expression for the correlation is given by

$$\boldsymbol{corr(Y_t, Y_{t-1}) = \frac{conv(Y_t, Y_{t-1})}{\sqrt{varY_t, varY_{t-1}}}} \quad \textbf{(5)}$$

Here,
Conv represents convolution given by:

$$\boldsymbol{conv\{x(t), h(t)\} = \int_{t=1}^{\infty} x(\vartheta)h(t-\vartheta)d\vartheta} \quad \textbf{(6)}$$

Here,
ϑ is a dummy shifting variable for the entire span of the time series data
t represents time
Y_t is the t^{th} lagged value
Y_{t-1} is the $(t-1)^{st}$ lagged value
X is function 1
H is function 2
Var represents the variance given by:

$$\boldsymbol{var(X) = X_i - E(X)} \quad \textbf{(7)}$$

Here,
Xi is the random variable sample
E represents the expectation or mean of the random variable X

3) Finite Distribution Lag Model (FDL): This model tries to design a finite distribution model comprising of lags fitted to some distribution such as the normal or lognormal distributions. Mathematically:

$$\boldsymbol{Y_t = \alpha_t + \delta_1 z_1 + \cdots \ldots \ldots \ldots \ldots . \delta_t z_t + \mu_t} \quad \textbf{(8)}$$

Here,
Y_t is the time series data set
α_t is a time dependent variable
δ_1 is a time-varying co-efficient
z is the variable (time variable)
t is the time index
μ_t is the time dependent combination-coefficient

4) Artificial Neural Networks (ANN): In this approach, the time series data is fed to a neural network resembling the working of the human based brain architecture with a self-organizing memory technique.

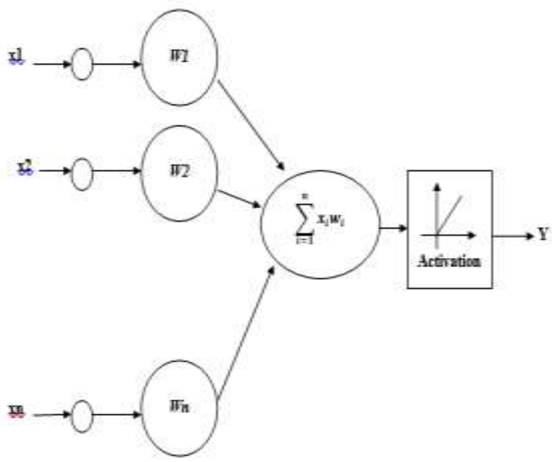

Fig.3 Mathematical Model of Neural Network

The approach uses the ANN and works by training and testing the datasets required for the same. The general rule of the thumb is that 70% of the data is used for training and 30% is used for testing. The neural network can work on the fundamental properties or attributes of the human brain i.e. parallel structure and adaptive self-organizing learning ability. Mathematically, the neural network is governed by the following expression:

$$Y = f\left(\sum_{i=1}^{n} X_i \cdot W_i + \theta_i\right) \qquad (9)$$

Here,
X_i represents the parallel data streams
W_i represents the weights
θ represents the bias
f represents the activation function

The second point is critically important owing to the fact that the data in time series problems such as sales forecasting may follow a highly non-correlative pattern and pattern recognition in such a data set can be difficult. Mathematically:

$$x = f(t) \qquad (10)$$

Here,
x is the function
t is the time variable.
The relation f is often difficult to find being highly random in nature.
The neural network tries to find the relation f given the data set (D) for a functional dependence of x(t).
The data is fed to the neural network as training data and then the neural network is tested on the grounds of future data prediction. The actual outputs (targets) are then compared with the predicted data (output) to find the errors in prediction. Such a training-testing rule is associated for neural network. The conceptual mathematical architecture for neural networks is shown in the figure below where the input data is x and fed to the neural network [10].

III. Previous Work

This section highlights the existing work in the domain, along with its salient features:

Han et al. [11] proposed that traffic volume data are crucial for effective traffic management, infrastructure development, and demand forecasting. This study addresses the challenges associated with traffic volume data collection, including, notably, equipment malfunctions that often result in missing data and inadequate anomaly detection. This model incorporates both raw and adjusted traffic volume data from 2017 to 2019, employing long short-term memory (LSTM) techniques to manage data discontinuities. A power function was integrated to simulate various error correction scenarios, thus enhancing the model's resilience to prediction inaccuracies. The performance of the model was evaluated using certain metrics, such as the mean absolute error, the root mean squared error, and the coefficient of determination, thus validating the effectiveness of the deep learning approach in refining traffic volume estimations.

Part et al. [12] proposed that hourly traffic volume prediction is now emerging to mitigate and respond to hourly-level traffic congestion augmented by deep learning techniques. Incorporating meteorological data into the forecasting of hourly traffic volumes substantively improves the precision of long-term traffic forecasts. Nonetheless, integrating weather data into traffic prediction models is challenging due to the complex interplay between traffic flow, time-based patterns, and meteorological conditions. This paper proposes a graph convolutional network to predict long-term traffic volume with meteorological information. This study utilized a four-year traffic volume and meteorological information dataset in Chung-ju si to train and validate the models. The proposed model performed better than the other baseline scenarios with conventional and state-of-the-art deep learning techniques. Furthermore, the counterfactual scenarios analysis revealed the potential negative impacts of meteorological conditions on traffic volume. These findings will enable transportation planners predict hourly traffic volumes for different scenarios, such as harsh weather conditions or holidays. Furthermore, predicting the microscopic traffic simulation for different scenarios of weather conditions or holidays is useful.

Chen et al. [13] proposed in this study involves improvements to the Long Short-Term Memory (iLSTM) and Bidirectional Long Short-Term Memory (iBiLSTM) models, leading to the construction of the iBiLSTM-iLSTM-NN model. This model incorporates spatial data from surrounding intersections and employs data fitting techniques to establish the correlation between periodic queue length and traffic volume. Subsequently, a predictive model for periodic traffic volume is developed based on this correlation, enabling reliable forecasting of future traffic volumes within a given cycle. Additionally, actual intersection data is collected for simulation analysis. The results indicate that the prediction error of periodic traffic volume is influenced by different traffic flow characteristics such as peak, off-peak, and normal periods, as well as different inbound lanes. The research findings can be applied to rapidly predict future traffic volumes for several periods based on the instantaneous queue length at the end of the red signal phase, providing reliable, accurate, and timely data for urban traffic signal control.

Ma et al. [14] proposed a novel approach for network-wide traffic state prediction where the statistical time series model ARIMA is used to post-process the residuals out of the fundamental machine learning algorithm MLP. This approach is named as NN-ARIMA. Neural Network MLP is employed to capture network-scale co-movement pattern of all traffic flows, and ARIMA is used to further extract location-specific traffic features in the residual time series out of Neural Network. The experiment results show that the post-processing the residuals of Neural Network by the ARIMA analysis helps to significantly improve accuracy of traffic state prediction by 8.9–13.4% in term of mean squared error reduction. In order to verify the efficiency of the ARIMA analysis in the post-processing, Multidimensional Support Vector Regression (MSVR) model is also employed to replace the role of Neural Network in the comparative experiment. Two streams of comparisons, (1) NN vs. NN-ARIMA and (2) MSVR vs. MSVR-ARIMA, are performed and show consistent results. The proposed approach not only can capture network-wide co-movement pattern of traffic flows, but also seize location-specific traffic characteristics as well as sharp nonlinearity of macroscopic traffic variables. The case study indicates that the accuracy of prediction can be significantly improved when both network-scale traffic features and location-specific characteristics are taken into account.

Haghighat et al. [15] Intelligent Transportation Systems (ITS) have seen efficient and faster development by implementing deep learning techniques in problem domains which were previously addressed using analytical or statistical solutions and also in some areas that were untouched. These improvements have facilitated traffic management and traffic planning, increased safety and security in transit roads, decreased costs of maintenance, optimized public transportation and ride-sharing company's performance, and advanced driver-less vehicle development to a new stage. This papers primary objective was to provide a review and comprehensive insight into the applications of deep learning models on intelligent transportation systems accompanied by presenting the progress of ITS research due to deep learning. First, different techniques of deep learning and their state-of-the-art are discussed, followed by an in-depth analysis and explanation of the current applications of these techniques in transportation systems. This enumeration of deep learning on ITS highlights its significance in the domain. The applications are furthermore categorized based on the gap they are trying to address. Finally, different embedded systems for deployment of these techniques are investigated and their advantages and weaknesses over each other are discussed. Based on this systematic review, credible benefits of deep learning models on ITS are demonstrated and directions for future research are discussed.

Hydari et al. [16] proposed that latest technological improvements increased the quality of transportation. New data-driven approaches bring out a new research direction for all control-based systems, e.g., in transportation, robotics, IoT and power systems. Combining data-driven applications with transportation systems plays a key role in recent transportation applications. In this paper, the latest deep reinforcement learning (RL) based traffic control applications are surveyed. Specifically, traffic signal control (TSC) applications based on (deep) RL, which have been studied extensively in the literature, are discussed in detail. Different problem formulations, RL parameters, and simulation environments for TSC are discussed comprehensively. In the literature, there are also several autonomous driving applications studied with deep RL models. Our survey extensively summarizes existing works in this field by categorizing them with respect to application types, control models and studied algorithms. In the end, we discuss the challenges and open questions regarding deep RL-based transportation applications

Muthuramalingam et al. [17] proposed that IoT based Intelligent transportation system (IoT-ITS) helps in automating railways, roadways, airways and marine which enhance customer experience about the way goods are transported, tracked and delivered. A case study on Intelligent Traffic Management System based on IoT and big data, which will be a part of, smart traffic solutions for smarter cities. The ITS-IoT system itself forms an eco-system comprising of sensor systems, monitoring system and display system. There are several techniques and algorithms involved in full functioning of IoT-ITS. The proposed case study will examine and explain a complete design and implementation of a typical IoT-ITS system for a smart city scenario set on typical Indian subcontinent. This case study will also explain about several hardware and software components associated with the system. How concepts like Multiple regression analysis, Multiple discriminant analysis and logistic regression, Cojoint analysis, Cluster analysis and other big data analytics techniques will merge with IoT and help to build IoT-ITS will also be emphasized. The case study displays some big data analytics results and how the results are utilized in smart transportation systems.

Dazango et al. [18] have shown that the solution of every well-posed kinematic wave (KW) traffic problem with a concave flow-density relation is a set of least-cost (shortest) paths in space-time with a special metric. The equi-cost contours are the vehicle trajectories. If the flow-density relation is strictly concave the set of shortest paths is unique and matches the set of waves. Shocks, if they arise, are curves in the solution region where the shortest paths end. The new formulation extends the range of applications of kinematic wave theory and simplifies it considerably. For example, moving restrictions such as slow buses, which cannot be treated easily with existing methods, can be modeled as shortcuts in space-time. These shortcuts affect the nature of the solution but not the complexity of the solution process. Hybrid models of traffic flow where discrete vehicles (e.g., trucks) interact with a continuum KW stream can now be easily implemented.

Ma et al. [19] proposed a dynamic factor model to forecast traffic state for groups of locations. The model decomposes the grouped traffic time series into the latent common factor component and idiosyncratic component. It uses a few latent factor series to represent the comovement of the underlying dynamics of grouped traffic flows, and idiosyncratic component to represent location-specific traffic characteristics. The dynamic factor model is estimated by the maximum likelihood method via an iterative EM (expectation maximization) algorithm. The traffic state forecast for each location is a combination of the respective forecast from the common factor component and idiosyncratic component. The dynamic factor model exhibits four advantages. It provides an excellent way to (1) seamlessly incorporate spatial correlations among grouped traffic flows into forecast; (2) produce forecast simultaneously for group locations; (3) perform dimension reduction such that high-dimension grouped traffic time series can be modeled at a low-dimension space; (4) consider not only location-specific information but also global common dynamics in the forecast. Meanwhile, it also has capacity to accommodate typical characteristics of traffic flows including temporal correlation, seasonality, structural change in mean and/or covariance function, and cointegration. Forecast accuracy is significantly improved across highway network as well as urban road network in comparison with the Sparse VAR and ARIMA models. The proposed method is suitable for large-scale network traffic forecast in the context of big data environment.

Diao et al. [20] showed that prediction of short-term volatile traffic becomes increasingly critical for efficient traffic engineering in intelligent transportation systems. Accurate forecast results can assist in traffic management and pedestrian route selection, which will help alleviate the huge congestion problem in the system. This paper presents a novel hybrid DTMGP model to accurately forecast the volume of passenger flows multi-step ahead with the comprehensive consideration of factors from temporal, origin-destination spatial, and frequency and self-similarity perspectives. We first apply discrete wavelet transform to decompose the traffic volume series into an appropriation component and several detailed components. Then we propose a more efficient tracking model to forecast the appropriation component and a novel Gaussian process model to forecast the detailed components. The forecasting performance is evaluated with real-time passenger flow data in Chongqing, China. Simulation results demonstrate that our hybrid model can achieve on average 20%-50% accuracy improvement, especially during rush hours.

IV. Performance Metrics

The training is stopped based on the mean square error or mse given by [21]:

$$\boldsymbol{mse} = \frac{\sum_{i=1}^{n} e_i^2}{n} \quad \textbf{(11)}$$

The final computation of the performance metric is the mean absolute percentage error given by:

$$\boldsymbol{MAPE} = \frac{100}{M} \sum_{i=1}^{N} \frac{E - E_i|}{i} \quad \textbf{(12)}$$

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

It can be concluded from previous discussions that smart and intelligent traffic systems are extremely important for handling the large volumes of traffic emanating from semi-urban and urban environments worldwide due to mass exodus of populations. While several applications of smart and intelligent traffic monitoring may be considered such as Route Optimization & Navigation, Smart Parking, Lighting, Accident Detection, Road Anomalies and Infrastructure Management etc., one of the key attributes to be computed is the network wide traffic volume. This paper presents a systematic review on the existing statistical techniques employed off late for the prediction of network wide traffic volume along with salient points of previous work. The paper presents a foundation on the development of effective and accurate prediction models for the traffic volumes pertaining to intelligent traffic and transportation systems.

References:

[1] 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 2018, vol. 5, Issue 4, pp. 239-253.

[2] 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, 2018, vol. 19, no. 3, pp. 878-888.

[3] 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.

[4] 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.

[5] 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

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] H. Ibrahim and B. H. Far, "Data-oriented intelligent transportation systems," Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014), 2014, pp. 322-329.

[11] D.C. Han, "Prediction of Traffic Volume Based on Deep Learning Model for AADT Correction", Applied Sicences, MDPI, 2024, vol.14, no.20, pp.1-19.

[12] S Park, M Kim, J Kim, "Hourly Long-Term Traffic Volume Prediction with Meteorological Information Using Graph Convolutional Networks", Applied Sicences, MDPI, 2024, vol.16, no.6, pp.1-12.

[13] Y. Chen, J. Guo, H. Xu, J. Huang and L. Su, "Improved Long Short-Term Memory-Based Periodic Traffic Volume Prediction Method," in IEEE Access, vol. 11, pp. 103502-103510, 2023.

[14] 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.

[15] 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.

[16] 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.

[17] Muthuramalingam S., Bharathi A., Rakesh kumar S., Gayathri N., Sathiyaraj R., Balamurugan B. (2019) IoT Based Intelligent Transportation System (IoT-ITS) for Global Perspective: A Case Study, ntelligent Systems Reference Library, Springer 2020, vol 154, pp.279-300.

[18] CF Daganzo, A variational formulation of kinematic waves: basic theory and complex boundary conditions, Transportation Research Part B: Methodological, Elsevier 2005, vol.39, Issue 2, pp.187-196

[19] T Ma, Z Zhou, C Antoniou, "Dynamic factor model for network traffic state forecast", Transportation Research Part B: Methodological, Elsevier, 2018, vol.118, pp.281-317

[20] 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

[21] S Satunin, E Babkin, "A multi-agent approach to intelligent transportation systems modeling with combinatorial auctions", Expert Systems with Applications, Elsevier 2014, vol. 41, no. 15, pp.6622-6633.