Machine Learning Techniques for Real-time Traffic Flow Prediction in Autonomous Driving

By Dr. Min-soo Kim

Professor of Computer Science, Pohang University of Science and Technology (POSTECH)

1. Introduction

The key insight introduced shows that traffic flow prediction is to predict the numerical values of traffic flow, such as speed, volume, and occupancy in the next several minutes in target road segments. Traffic flow prediction is not just a simple extension of the traditional time-series prediction problem but possesses some distinct features. Firstly, traffic flow data is of high-dimensional, spatiotemporal, and nonlinear. It is a big challenge to accurately model the correlation of traffic flow data in the past for the close to immediate future traffic condition forecasting. Secondly, traffic flow prediction exhibits a strong dependency on environmental conditions. Thus, it is important to quantify the effect of traffic incidents (such as accidents, special events, and weather events) on the traffic flow forecasting. Finally, actions taken by relevant agents can both directly and indirectly influence the traffic outcomes. Therefore, it is important to model the impact of relevant agents' actions in the short-term traffic flow forecasting.

[1] [2] With the rapid development of technologies like the Internet of Things (IoT), autonomous driving has gradually become a hot research topic. The core of autonomous driving is to realize intelligent perception and decision-making. Real-time prediction methods, by forecasting traffic flow through real-time traffic data (e.g., GPS logs, traffic camera images, and numerical data gathered by sensors) can greatly promote the safety, efficiency, and accessibility of autonomous driving. The research problem in traffic flow prediction is to forecast the following traffic flow, highway traffic flow, or lane-level traffic flow within a certain look-ahead period based on historical traffic data. It has significant theoretical and practical value for the smart transportation system, as well as the intelligent transportation system.

1.1. Background and Motivation

[3] [4] In recent years, the development of machine learning techniques has significantly advanced the real-time traffic flow prediction accuracy for many smart city services related to autonomous driving. It provides essential data for future optimization and prediction of vehicle routes, as well as pollutant and noise emissions management systems [6,7]. To reduce the additional dependencies, real-time traffic prediction systems should be equipped with onboard data acquisition and analysis. For this reason, an efficient deep learning data-based prediction system with low power consumption is required.[5] Traffic flow prediction is the process concerning the determination of future traffic states on the motorway based on the smallest amount of data. This is a challenging problem with many potential applications ranging from theory to practice, including driving assistance, traffic management, route guidance, ambient intelligence, management of goods transportation and environmental impact studies for citizens. Such predictions provide decision-support tools and may contribute to solving traffic issues which hamper modern (and future) society (e.g., electronic route guidance), and are also a basis for further studies (e.g., traffic psychology, and safety) [8,9]. Moreover, the European Commission suggests real-time traffic conditions monitoring using the vehicle-infrastructure cooperation of telematic systems through the powerful onboard data analysis and efficient intelligent information chain synthesis.

1.2. Research Objectives

The major objective of this work is to develop a Machine Learning-based real-time hybrid traffic flow prediction framework [6]. The novelty of the proposed work is; 1) to use the real-time Hidden Markov Model (HMM) to develop real-time traffic flow prediction and state prediction modules, which will accurately forecast future traffic flow pattern both in near and long term horizons [7]. The predicted traffic flow pattern will be used to generate future traffic flow and state data and to build a learner model to make a vehicle stop and go decision. In this work, a real-time Markov Model uses historical data to predict future traffic flow and state data using a forward-only model thereby forming a future trajectory data. The real-time Markov Model achieves real-time trajectory data since it uses traffic flow prediction and knowledge of the current state of the system. This knowledge combined with the developed spatial-temporal forward prediction model forms the basis of the key decision-making system in this work [8]. The key objective is to generate a real-time trajectory data and develop vehicle stopping and going decision learning models that will consider such critical information in real-time data before making any go and stop decision. The above three modules, i.e., traffic flow and link Markov prediction, traffic state prediction, and vehicle stop and go decision model, form a complete journey of the learner model. This model uses real-time traffic information, spatial-temporal relationships, and instant link state to predict and stop, and go in the road network. The major challenges, however, are the real-time decision making during stop signals and uncertainties in the traffic flow prediction. The temporal and spatial unknown dynamical behavior of incoming traffic, presence of time-varying and nonlinear parameters, the unknown dynamic structure of the input-output, the error in different dynamic behaviors of the system states, makes it a very difficult practical problem.

1.3. Scope and Organization of the Work

The essential characteristics of integrated decision-making in autonomous driving are time and space continuity. Integrated decision-making requires situational information about the surrounding environment, including predictions about the movement of other vehicles. Integrated decision-making also necessitates a capability of the vehicle's electronic control unit (ECU) to make autonomous decisions based on the EM. In this type of environment, the vehicle also requires perception models with prediction capabilities, the use of which could be implemented through various means: vehicle-vehicle (V2V) information sharing systems, vehicle-infrastructure (V2I) communication systems, or artificial intelligence (AI) enabled systems. The increased accuracy and confidence in predicting the future status of supervising vehicles and surrounding traffic flow can enhance real-time perception models in AD. As a data-driven method, although long short-term memory (LSTM) is sensitive to criteria such as sample selection and hyper parameter tuning, it is widely and successfully applied to traffic flow prediction tasks. LSTM models are known to be capable of long-term dependencies while learning in sequence data [9].

Traffic forecasting is proving to be a critical component in intelligent vehicle driving and smart transportation systems [10]. In the field of intelligent navigation, many studies have focused on short-term traffic flow forecasting methods [11]. In the field of machine learning (ML), various predictive models, such as autoregressive integrated moving average (ARIMA) and support vector machine (SVM), have been used for traffic flow forecasting. In the last decade, deep learning (DL) models have been increasingly used for time series forecasting and are shown to be generally more effective than traditional methods. The work in this paper is part of a more extensive study where modern ML methods are used in the development of a traffic forecasting prediction for an intelligent self-driving vehicle. More specifically, the scope of this study focuses on the variant of the long short-term memory (LSTM) model. In order to improve the LSTM performance, the input space is extended to include sensory data originating from exceptions related to the vehicle, its control system, and its perceptions and navigation devices, in addition to real-time traffic predictions.

2. Fundamentals of Traffic Flow Prediction

Modern traffic flow prediction methods can be categorized into two groups: model driven methods and data driven methods. Model driven methods describe the various traffic conditions using fundamental equations and statistical laws. They are based on mathematic models usually on the assumption of some traffic laws. Meanwhile, data driven methods construct models based on historical traffic data. Traffic flow prediction uses historical traffic flow, or other traffic information, to predict the future traffic situation. In technical description, traffic flow is represented in two forms: microscopic form and macroscopic form. The former depicts individual vehicle's motion and interactions, whereas the latter represents macroscopic variables such as traffic density, flow and speed. [12]

Traffic flow prediction is concerned with forecasting future traffic data based on historical traffic data. It is an important issue that has been extensively studied because accurate traffic flow prediction can provide commuters with guidance on traffic condition [8]. As emerging technologies affect various sectors, it is fundamentally changing mobility in smart cities. For example, the internet of things (IoT) and fifth generation (5G) communication technology make it easier to obtain real-time traffic data in real time. Meanwhile, smart phones can also collect traffic data and various methods for real-time traffic flow prediction can be improved [13]. Traffic flow prediction has already been used in many applications, including advanced traveler information systems and traffic management.

2.1. Concept of Traffic Flow Prediction

Traffic is mainly described by flow, a measure of the average number of vehicles passing a point of interest per unit time. The flow on the road network does not depend only on the traffic demands to move from origin to destinations, but also on traffic safety, the quality of life of road users, the pollution rate, etc. Thus, the measurement of traffic flow, traffic flow prediction, and traffic control are always the key research areas of the transportation community [2]. There are commonly two important traffic flow problems as follows: - Short-term flow forecasting: The prediction window is relatively short, such as five to sixty minutes. It is used for trip planning, congestion avoidance and the setting of ramp rates; long-term flow forecasting: The prediction window is relatively long, such as several hours to several days. It is used for urban planning, air quality management, traffic management, and the setting of road widening projects. For the two forecasting problems, both parametric and non-parametric approaches are used. Parametric approaches assume the dependences are structured by means of a mathematical model. Time series analysis is the main technique developing the predictive models used in the above approaches. Time series models like ARIMA, exponential smoothing methods, and state space models usually make a sound mathematical description of the data at hand by deterministically modeling the influences of past events, errors or exogenous input signals, and generate forecasts by iteratively applying causal relationships. Nevertheless, these models are generally based on the simplifying assumptions like the linear time-trending, the additive and constant-variance errors, and constant seasonal effects, which may have limits in effectively capturing complex system dynamics or the spiking value changes. Moreover, these models are built based on traditional statistical theories and become less powerful when it comes to dealing with high-dimensional, nonlinear, and noisy spatio-temporal attributes derived from the primitive traffic flow data [14]. Therefore, alternative methods for traffic flow prediction need to be studied.

2.2. Challenges and Importance

路The real-time traffic flow prediction is required for ensuring safety, security, comfort, convenience, and reliability in the modern transportation paradigm and agencies are nowadays relying on Data mining/Machine learning/Deep learning-based prediction mechanisms for solving these issues. Deep learning has obtained prominence to alleviate these limitations and difficulties due to its top performance in the field of modeling and inferring relations in chaotic data space containing a high degree of complicacy [10]. Various deep learning models and strategies are cogitated to fulfill real-time traffic flow forecasting requirements in multiple settings finding the importance of deep learning technique-adaptive, salient, robust, consistent, reliable, and applicable for different kinds of traffic datasets.

Traffic flow prediction plays a significant role in modern intelligent transportation systems (ITS) and smart cities [2]. Accurate and reliable traffic flow prediction is crucial for traffic management, guidance, and control that can help transportation sectors in saving time, fuel, and reducing emissions [15]. It can also assist traffic authorities, transportation research sectors, city planners, and the general public in major decision making and it also helps in recognizing heavy traffic, constricting long jams, and makes traveling and route planning convenient. Real-time traffic flow prediction has a high potential to minimize traffic jam events and travel time and it also provides alternative routes to minimize environmental impacts. However, real-time traffic flow prediction has its own challenges such as short-term data availability, data distortions, heterogeneity, prediction accuracy, privacy concerns, increasing variance, and missing/abrasive traffic data from various sensors. The main issues and challenges which inhibit traffic flow prediction are: - Short-term data availability (acquiring the traffic data from multiple discrete and distributed sources at a reasonable time interval); - Data distortions (incompleteness, misalignment, heterogeneity, offset, noise, bias, errors, missing or abrasive traffic data); - Privacy concerns (personal data privacy, privacy breach, threats, spam, spoofing or proofing of messages); changing topologies (dynamic and unpredictable changes in transportation networks, considering network modifications, traffic anomaly, and faults in routing paths); - Data heterogeneity (discrepancy, mismatches, inconsistency, variability and variances in features/attributes and behavioral patterns of data streams and database tables); - Prediction accuracy (type of traffic flow prediction, measurement of predicted accuracy, and it's data quality, integrity, reliability, and precision); - Increasing variance (many variabilities, supplying long-lines of variabilities, increased traffic loads on roads and limited road capacities); - Missing/Abrasive traffic data (losing, distorting, or collecting a disorder/chaos larger than noise/knowledge from traffic video feed, crowd video learning and video domain migration).

2.3. Traditional Approaches vs. Machine Learning

While robust forecasting methodology can assist in reducing transportation costs and positively impacts urban transportation planning, this kind of forecasting holds a prominent role for both traffic research and engineering. In contrast to the traditional methods such as exponential smoothing, autoregressive integrated moving average (ARIMA), and the combination auto-regressive integrated moving average (C-ARIMA), recurrent neural networks (RNNs), though if considering long-term real-time traffic forecasting, long short-term sequences and high real-time delay will be problematic. With the advent of advanced machine learning techniques, such as ensemble learning (EL), regression tree (RT) techniques, and traditional methods have become insufficient, due to the reason that urban congestion sources likely have different representations, especially regarding the growth of the wide aggregation of the urban spatial and temporal distribution datasets and diverse data sources, which is found widely in the latest academic literatures thus the dimensionality of features is incredibly high and will introduce high computational complexity and a reduction of prediction accuracy; nevertheless the previous are computationally costly in large-scale and have difficulty to deteriorate the dimensionality issue [16].

Traffic flow prediction is undoubtedly of paramount importance for the development of Intelligent Transportation Systems and Autonomous Driving and it only recently has grown in popularity due to the increasing use of deep learning models in traffic flow prediction. The most successful existing deep learning traffic flow models use Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) [17]. As an alternative, the present work combines optimized hyperparameters -- those that have been found most successful with LSTM and GRU in various domains of traffic flow prediction -- for nine models and are organized in straightforward categories. All hyperparameters combinations are derived from theoric systems with extensive hyperparameters, however the approach of using exhaustive hyperparameters is naïve. The goal of surveying a diverse collection of traffic flow models and selecting the five best-performing models for traffic flow is that traffic flow data are inherently stochastic and non-linear; these models are demonstrated for city-wise traffic flow prediction within Canada [13].

3. Machine Learning Algorithms for Traffic Flow Prediction

In recent years, data-driven methods have gained more attentions, such as the machine learning (ML) methods and the deep learning (DL) methods. The ML methods can capture the complex nonlinearity hidden in the traffic data with features based on are period term, time-based historical average, spatial dependence, time dependence, spatial-temporal influence, population of the density and the density of the population expressed in Fig. 3. It is worth to mention that the aforementioned features are extracted usually from the OT data and the features mentioned above are the trend import features for the traditional traffic flow prediction models. Also, the introduction of GT techniques has provided an alternative for the cities of Brazil in mobility planning. These features are complement to the present features, this study will verified the performance improvement that can be given from the GT. The connectivity of the Braint and SANDAG can potentially improve the prediction; generate the first results crude of his performance [8]. The principal contribution of this study is the fusion of the data coming from the GT of the BRAINT and the SANDAG by measuring the improvement of the prediction through the introduction of new time and spatial influences. This study focused on how to improve the accuracy of the traffic flow prediction based on very short-term prediction, by focusing on the discussion about the short-term traffic prediction mainly using the VDS data of the BRAINT (over which are aggregated the flows of the GT) and the SANDAG from San Diego. The focus is more on the correlation's structure of the network between the traffic and the GT, assuming that the latter can have lead, contemporaneous and lagged influences on the traffic. The research was performed using the VDS data of BRAINT (San Diego) and SANDAG (Brasilia), in which the first one is a first example of the introduction of the GT in the traffic prediction. For example, this paper is one of the first to use the global temperature energy consumption such as GT in the VDS data to predict a traffic flow where the few works about traffic prediction in which the GTs are considered are with the NARX GMDH and Cuckoo algorithm using the traffic speed. Consequently, we performed the same analysis by taking the flow of the GT of the SANDAG, the metropolitan zone of Brasília (DF), over the VDS data. The results of the experiments have shown as the contemporaneous influence together with the significant autoregressive features and influence features features the improvement in the prediction with respect to for architectural for all DMLa, FNNb, window and days and OML ones approaches;by comparison the showing the coverage source of the GTsignal is the end more between the distant one compared VDS with the BRAINT the with the contemporaneous influence features and feed forward neural network approach [18].

Traffic flow prediction is crucial to the Intelligent Transportation Systems (ITS) as it can support proactive traffic management for sustainable and efficient urban transportation. The techniques for traffic flow prediction can be roughly classified as (i) model-driven methods and (ii) data-driven methods. Model-driven methods mainly include the time series based methods [3] and the statistical methods like autoregressive integrated moving average (ARIMA) method. Though the model-driven methods have been widely used for short-term traffic flow prediction, these methods are unable to describe the high-dimensional data structure.

3.1. Supervised Learning

Recently, researchers have focused more attention on modeling the relationships among all the traffic flow variables spatio-temporally and then learning the temporal and spatial correlations shown by traffic flow data to achieve more accurate traffic flow prediction [8]. In this research direction, Traffic Flow Prediction (TFP) task is formulated as a spatio-temporal forecasting task, which concentrates on utilizing both time and spatially-temporally sequential information to improve spatio-temporal traffic forecasts by employing various deep learning techniques (Fig. 5). To the best of our knowledge, previous research only consider the task of TFP with utilizing time series information temporally, or exploring the problem of Spatio-Temporal Graph Learning (STGL) without applying the learned information to achieve traffic flow prediction task. Our contributions in this work can be summarized as: 1) Formulating the TFP task as STGC learning problem in order to represent, explore and utilize spatio-temporal correlations learned within the defined spatio-temporal causality graph framework; 2) We provide two different category of prediction models for the TFP task in the utilized spatio-temporal causality graph learning framework.

Traffic flow prediction is one of the most important tasks in intelligent transportation systems since traffic flow information contributes to traffic control, and traffic congestion mitigation and intelligent routing, alleviation of the environment pollution and urban air quality improvement [19]. In recent years, traffic flow prediction also becomes an essential task in the field of autonomous driving which is considered as one of the most dramatic application scenarios of intelligent transportation systems. Promising intelligent transportation systems aim to change our daily life in many aspects, these systems rely on real-time, accurate and reliable work of predictive models, and also require the efficient and effective algorithm capable of processing large-quantity time-series traffic data to forecast the future traffic flow in real-time [20].

3.2. Unsupervised Learning

Hsu et al. (2017) have realized that the pattern similarity detection, i.e. clustering of traffic flow, was a challenge to be addressed. They proposed a hybrid machine learning approach by fusing feature-based and shapelet-based clustering algorithms, named WARP (weighted auto-regression pattern) and DNDTW-C2E2 (dynamic time warping-based clustering and auto-regression prediction), respectively, to capture the non-linear synergistic and hidden patterns inside the time series traffic flow observations for different traffic analytics tasks including, urban area land use classification, real-time traffic flow pattern discovery, and the prediction of traffic flow and CO2 emission in urban areas [14]. The hybrid approach gained response rates of around 85%–93% in predicting 15-min, 30-min, 45-min, 1-h and 2-h patterns of case traffic series from Jakarta, Indonesia, and also presented higher accuracy scores than comparison machine learning methods including CART, MLP, XL model tree technique, and multi-core decision tree, and robustness improvements for different steps along the traffic analytics tasks. Though showing the applicability of the hybrid framework in other similar traffic analytics applications, the authors neither discussed datasets concurrent with the collected pollution resolution of relevant traffic situations nor fully validated its performance in real-time traffic flow prediction tasks.

In this strategy, traffic flow data are first generated, then unidentified patterns of such data are clustered into similar groups to identify hidden typical traffic characteristics, and then the prediction model are trained on the previously identified representative traffic patterns [21].

3.3. Reinforcement Learning

The estimation of travel times and traffic intensities play a crucial role on the development of autonomous driving [22]. In fact, a traffic prediction system that can return short-term accurate predictions of real-time traffic flows consists in a fundamental module if an autonomous vehicle aims to optimize its trajectory and control its status in a dynamic environment. The literature has demonstrated that in forecasting non-stationary traffic situations, no single model triumphs across all forecasting metrics. For this reason, combining forecasts has become one of the main focus of traffic forecasting research. Schemes for traffic data prediction can be divided into technologies requiring the exploitation of historical traffic data (offline schemes) and those using the information related to the traffic state at the time of forecast (online schemes).

Data prediction is crucial for navigation and control in the intelligent transportation field [23]. To forecast traffic flow, different algorithms (e.g. machine learning) can be employed for the layer and input data. Some scholars studied the trafficforecasting problem by employing reinforcement learning and improved the efficiency in both computational power and prediction accuracy. Reinforcement learning is a stochastic optimization methodology that learns the traffic information from the input data (a time-series traffic flow) in an automatic way, rather than empirical and calculated setting of the parameters. Reinforcement learning

(RL) algorithms are a set of abstract tools to enable computers to learn how to solve tasks without being explicitly programmed how to solve them [24]. The efficacy of these methods, however, is contingent upon the quality and representation of state space, i.e., the input traffic data chosen for prediction. For this reason, ensemble models and hybrid approaches have taken on great importance in traffic forecasting research.

4. Data Collection and Preprocessing

Data imputation, preprocessing the input features for real-time traffic flow prediction, and incorporating those new features into the model is crucial. For data imputation, the authors extend the use of data from more sources and use more features based on deep learning to encode the information. Data imputation is mainly processed with ENCODERDECODER LSTM following random forest. Citywide traffic accident risk (TAR) is presented to characterize the potential possibilities of traffic accidents at different urban locations in two-hour intervals. An informative map of TAR was produced, alongside known key factors in explained model, such as traffic flow and traffic accident density. After predicting the traffic flow stability index, the authors demonstrate with our historical data that the proposed model can further give drivers the most appropriate advice by calculating the shortest route and predicting the traffic flow on real-time.

Frequently used machine learning for time series traffic includes such data preprocessing steps as feature generation, normalization, and cleaning. The Great Recession 2007– 2009 had a profound effect on the motor carrier industry. This event brought about an economic slowdown in the United States, shook both the manufacturing and trucking industry to the core, and resulted in bankruptcy and liquidation of 10,000 trucking companies and 150,000 lost jobs [25]. These incidents tallied up to precipitating shipping demands and causes variations in market principles. The other main underlining factors that are viewed to bring about a slower recovery are over built inventories' in flatbed trailers coupled with a shift of freight lines to avoid van freight without mention of seasonal variations [26]. Two spate of researches were carried out first on the basis of the irregular arrangement of economic trens in two span of time, and of market rate state in pipelines respectively. No spatiotemporl model was proposed to examine the underlying space based pattern of midstream oil and gas prices.

4.1. Types of Data Sources

On one hand, urban road network traffic flow can be represented appropriately by spatial and temporal features. On the other hand, traffic prediction models are greatly affected by the sensors that are placed in the vicinity of traffic roads. In this study, the contribution of the investigation focuses on the data sources of real-time traffic prediction in virtual driving and autonomous driving scenarios. The first group is the data needed to predict the future traffic scene in urban areas, including regional road density installations, real-time traffic flow maps, and historical traffic flow maps. The second group should include road density distribution data and the position distribution of the vehicle at the current moment. The experimental results show that the traffic prediction results in each scenario are highly associated with the traffic data used as input. The influence of each data type and moisture content data on traffic prediction in virtual driving scenarios is also analysed, and the data showing the best improvement is spatial data in the environment with low moisture, effectively improving the prediction of left turn traffic volume. The third group includes weather data (which includes humidity information) and input reminders, as described in the proposed real-time traffic flow prediction. [27]

The development of this research was based on its use in the field of autonomous driving. Real-time traffic prediction is vital for the autonomous vehicle trajectory planning system in various traffic scenarios. Data sources play a key role in model training. Consequently, the study classified traffic data types into three groups, i.e., spatio-temporal sequence data, static spatial data, and external data [26]. Spatio-temporal sequence data refer to sensor data, GPS data, and camera data extracted from carlevant road position and status datasets, which are utilized in various autonomous driving scenarios. Spatio-temporal data also includes traffic accident data, which directly affects subsequent traffic flow. Static spatial data mainly contains location data, used as geographical information in the study. External data include arbitrary time-series data from social media and weather data.

4.2. Data Cleaning and Integration

Recently, the most immediate area where machine learning techniques can be successfully adopted is connected, autonomous, and cooperative vehicles, i.e., in the automotive domain, especially for personalized, operational, predictive, and cooperative driving paradigms [10]. Because of the increasing presence of big data in the automotive world, all these data are mostly processed by means of machine learning techniques that are able to predict human, political, social, environmental, and industrial behaviors with the simulated or real data originated from driving dynamics. In general, in a predictive and operative engine, it is highly feasible to act successfully if the engine is able to properly gather, preprocess, and exhaustively transform raw data into representative information. In summary, within this topic, we can find applications belonging to different modules such as (i) traffic conditions prediction and optimization, (ii) real traffic prediction and control, as well as (iii) advanced driver-assistance systems services in the case of car safety, infotainment, and comfort.

Accurate and efficient traffic speed prediction is an essential factor for autonomous driving and Intelligent Transportation System (ITS). In order to achieve optimal vehicle environment perception and path planning, we need to have a comprehensive understanding of the road traffic environment and predict the future road traffic speed at an accurate time interval [1]. Recently, traditional methods based on low-dimensional raw data and effective algorithms have been used for traffic prediction, but they often fail to achieve good results. In order to achieve effective traffic prediction, various machine learning techniques and various integrated data services have been used. This paper aims to provide these machine learning techniques used for traffic prediction with the focus on cloud-based data mining techniques.

4.3. Feature Engineering

Traffic flow forecasting is crucial for intelligent transportation systems, with short-term prediction being especially critical. Traffic data is characterized by high dimensionality, hidden data structures, and non-stationarity, which often poses challenges for traditional traffic prediction methods. Non-parametric prediction methods like wavelet-SVM have been preferred due to the strong periodic and regular behavior, and short-term uncertainty in the traffic flow data. Methods like ARMA, ARIMA, and GARCH have limited modeling ability in capturing both complex temporal characteristics of time series and dynamic transportation phenomena. For instance, wavelet singular vector machine can effectively reduce the influence of noise in traffic flow data and improve the prediction accuracy [11]. Deep learning, as a part of machine learning, is a new technique used in finding latent dependent relationships in traffic flow time series data. The application of stochastic biased term provides a global viewpoint when adjusting sense directions.

Real-time traffic prediction in a real traffic environment is complicated, due to the spatial and temporal diversity in traffic flow data. Techniques like ARIMA, Kalman filtering, and knearest neighbors do not perform well due to the high variability, non-stationary, and irregular patterns [19]. Temporal correlations and periodicity in traffic flow vary by time of day, posing challenges for prediction. In autonomous driving, the short-term period of traffic flow has a great impact, and the traffic state is updated within an hour. In a practical driverless car system, filtering methods like average filter, median filter, and polynomial filter have obvious limitations for only a few observation points, and advanced filters depend on a comprehensive assessment of the traffic flow parameters. We apply the wavelet singular vector machine (SVM) traffic flow prediction model [15] to address these problems. The process integrates data denoising, feature extraction, and traffic flow prediction into a single framework, which provides noise reduction strategy, reduces the multi-scale aspects of data from complex traffic systems, and extends the forecast horizon.

5. Evaluation Metrics for Traffic Flow Prediction

The proposed method for multimedia digital navigation system employing expert machine learning includes obtaining media data on multiple real-time traffic-related factors from multiple media servers; collecting the vehicle media data based on real-time traffic conditions; forecasting future traffic conditions based on historical traffic media data and current traffic media data; and providing navigation guidance on a digital map based on artificial intelligence. The primary decision support module includes expert AI engine, road AI engine, and other road conditions AI modules. The road conditions AI module includes traffic status, air quality, road condition, weather information, local services, events, off-road car counter, the car counter of the corresponding lane, the speed of the corresponding lane, and the number of cars on the road. The road AI engine includes voice assistant, map assistant, and road conditions to help users navigate.

It is a widely adopted practice to predict the traffic flow based on different models in the real-time traffic environment. The quality of traffic flow forecasting is a significant issue in the field of traffic prediction in autonomous driving [28]. There are several evaluation metrics which have been utilized in the road traffic flow forecasting literature to indicate the model's performance.

5.1. Mean Absolute Error (MAE)

Deep learning approaches have been found to outperform non-deep-learning-based methods for traffic flow prediction due to their exceptional capabilities in learning complex input-output relationships. The mean absolute error (MAE) is widely used to evaluate the predictive performance of various models. Compared to RMSE, MAE provides a better interpretation of the average error magnitude for ready reference. The MAE scores of different models under each dataset are detailed in Table 4, 5, 6, 7 which evaluates traffic flow prediction models for Shanghai East, Wujiaochang, Shanghai South, and Xingzhong Road, respectively. The spatial–temporal attention based fusion network (ST-AFN) is more accurate than other models by achieving the lowest MAE scores for all the datasets. Traffic flow prediction is important to autonomous driving and plays a significant role in Advanced Driver-Assistance Systems (ADAS). Accurate prediction of traffic flow is an indispensable requirement for autonomous vehicles, as prediction errors may result in inadequate traffic signal changes or inappropriate acceleration and braking of vehicles around the end of the current horizon. This is not only computationally expensive but also not practical in near real-time traffic flow prediction. Hence, real-time traffic flow prediction is a challenging task in traffic analysis. Furthermore, the use of a limited set of historical traffic flow information should inform real-time traffic flow prediction to be practical for deployment in road vehicles that traverse roads in real-time. The LSTM model takes as input spatial and temporal features extracted from traffic flow images by their multi-scale CNN, followed by multi-scale LSTMs with a spatial attention mechanism that allows focusing on the most informative regions or sequences.

5.2. Root Mean Square Error (RMSE)

To minimize this, it is of utmost importance to mitigate errors equally but also to resolve main error sources first. Furthermore, the traffic situation is very dynamic, it can fail to follow any trend, and if considered as a preceding noise, it can lead to wrong or falsely amplified conclusions. For this reason, we conclude that aside from using RMSE as a measure and focusing on achieving a more balanced prediction profile, during time validation the trending and average values should be also accounted for. If we have learned from previous examples, we hope to minimize negative impacts by avoiding misleading conclusions for future traffic situations [29].

[30] [9] Root Mean Square Error (RMSE) is another popular metric for evaluating a regression model's performance. The value shows the spread between the actual values and the

predicted values. Large errors generate large values, meaning that the error gets amplified, and therefore RMSE is better than the MAE for detecting it. However, it still does not inform on the exact distribution of the error spread. In our case, predicting traffic flow with large error values can be very costly or even life-threatening in the case of traffic light control or platooning. This is because for both cases unsafe driving can occur which could lead to traffic accidents. Therefore, the cost of neglecting false predictions can be far more serious for a safety critical AD system in an urban environment.

5.3. Mean Absolute Percentage Error (MAPE)

The uncertainty is also present, especially for forecasting values far from t ¯% where predictions are less accurate than for values c"t¯% due to unknown external and good/bad weather conditions. This issue is addressed in sMAPE, which results in less dependence of MAPE on the actual and forecasted values across the whole time series,path on inference and decisions under uncertainty">%One of the most frequently used measures to ascertain the quality of traffic prediction models is MAPE" % [31].

The mean absolute percentage error (MAPE) is a measure of prediction accuracy of a model for continuous variables % [32]. It measures the relative accuracy of predictions divided by the average actual value in the testing data. The less accurate the forecasted values, the higher the percentage error, imperfect predictions have a percentage error of 100%. The final value is averaged over all true and predicted results. It has the advantage of being simple to calculate, easy to understand and widely used because people prefer to understand the accuracy of their models in percentage terms % [16]. MAPE is widely used in transportation science when evaluating how good traffic flow models are">%sMAPE measuring the magnitude of error"", the accident model is considered to be successfully in providing accurate estimations for future values if = 30% for test data.

6. Real-time Implementation Considerations

Sustainable mobility, in particular, implies a significant re-orientation of current transportation systems, mainly being driven by technological advances. This transition is being supported by wireless communication technologies, advances in control algorithms, machine learning models providing precise demand predictions and novel system concepts such as shared automated electric shuttles [33].

Error in short-term traffic flow forecasting in congested area is due to sudden arrival of a large number of vehicles, changes in driver behaviour, and stop-start driving on traffic condition (Margi et al., 2014). An ARIMA model assumes that the time series is stationary, which is difficult to hold in non-stationary scenario and like Gaussian distribution, it has limitation over the Gaussian mixture model which cannot be used to predict non-Gaussian non-linear behaviour of different traffic phenomena. Consequently, the purpose of this study is to develop a hybrid artificial intelligent model to predict lane traffic flow in three typical scenarios involving congestion developing, recurrent congestion, and vanishing congestion [11].

CIM (referred to as SOT, '1The City of Innovative Mobility Act') [10] tells us that all these forecasting algorithms have excellent prediction accuracy after being trained with massive data from traffic sensors, but they either cannot handle speed prediction or face problems like kernel learning and data setting. The critical issue with this evaluation criteria is the fact that, the traffic responses may not strictly follow common linear or non-linear patterns during high congestion scenarios and sinusoidal patterns of GFP prediction (Greenshields, M.A. 1934).

6.1. Computational Efficiency

Building on traffic flow forecasting model comparison work, using different machine learning algorithms for short-term predictions, their computational cost was discussed. While all these approaches provide competitive accuracy, decision trees turned out to be the fastest, KNN the slowest, and linear models on a parallel computing system run at acceptable computational time. Moreover, Keras-TO and PyTorch-TO followed by PyTorch--TUTE_INF are faster than the other approaches for prediction times less than a second. There is a plateau of the computational time of prediction as the number of roads increases for large portions of the clocks, mainly due to data loading. This work suggests to pay attention to the computational cost of ML and deep learning models if they will be deployed in production, so that the deployment is feasible, and efficient, rather than optimal, methods should be preferred [15].

Traffic flow prediction is one of the essential tasks in most traffic control, especially in the context of intelligent transportation systems. Not only short-term forecasting, but for real-time applications such as autonomous driving or ride-sharing, very short-term prediction plays a significant role. However, the computational cost of running machine learning algorithms calls for the use of efficient methods that can handle millions of data points in a few milliseconds, a prerequisite for on-board deployment in autonomous vehicles. This section focuses on the computational efficiency of the machine learning models introduced and utilized for real-time traffic flow predictions in autonomous driving [14].

6.2. Model Deployment and Integration

In general, the model is often the theoretical results obtained by training and selecting signals in a certain scenario. Time and space data in different scenarios need to be analyzed and processed, and it is also significant for deploying into intelligent systems. In some networks such as UC, detection and control, the system is often deployed at the crossroads on the ground to guide the V2X to predict and diagnose. All information processing is done in the control center [26]. Therefore, ground processing and data transmission difficulties are very large. It is necessary to develop offline training, online prediction systems, and intelligent management systems that can process signals in real time based on some network-based traffic signals at the same time and manage system deployment to improve real-time performance efficiently. It has launched an adaptive LQ-like sorting reconciliation algorithm and an intelligent deep learning traffic flow prediction and detection technology for the background adaptively introduced in cross-road surveillance systems. In order to compare the prediction accuracy and real-time performance of different models effectively in driving simulation experiments, the simulator gets the information that needs to be predicted on-line in the simulation process and ensures that different traffic conditions are generated. Updates real-time images and the real-time prediction performance of detecting system accuracy.

These traffic prediction models are aimed at getting the traffic status after a full time step nT (usually nT - 5) [34]. Thus, the offline training process begins in the initial time period of several minutes. During this period, label data is available from historical data. After the training is finished, the prediction module is deployed for set predictions. Each time step uses nT - tw value to predict, and the result of each prediction unit is recorded and the latest v and q for the next control period are updated to ensure real-time prediction. The models need to be retrained every T seconds to predict the traffic status of the next time step. The online training process can use the label data of each T seconds in the actual prediction process. However, traffic needs to be predicted in real time from now on, rather than in the later offline prediction stage. Therefore, the prediction part of most models usually needs to be improved to ensure that the inversion operation in the process does not increase the real-time performance. During training, you also need to pay attention to whether to set the error at a certain length. If not, adding additional reset, normalization, and other network structures can speed up the inversion process in the face of prediction [33].

7. Case Studies and Applications

Different machine learning algorithms have been evaluated by Böök et al. [22] for modeling traffic flow, which have enabled real-time traffic flow prediction. Authors have used a mix of RNN, random forests, and feed forward neural networks as alternative regression models, using traffic data from San Francisco's Taxi, Limousine and now Vehicle status system. The authors used the dataset because of its intrinsic high spatiotemporal resolution, with events collected every 15 s. For prediction of traffic flow (in terms of occupancy levels) Recursive Neural Network (RNN) achieved superior results with accuracy close to 98 %. Lukči´c et al. [16] developed a new predictive model for traffic flow modeling using Ehrenfest ensemble modeling and deep learning. The model creates trajectories of vehicles and dynamically estimates road occupancy and traffic density using real-world traffic data. Proposed model achieved a 10%-min 1.05 average error and two-order prediction accuracy in some traffic states. Trend analysis showed that the model has the potential to accurately predict trends in traffic flow models through single step and multi-step prediction models, demonstrated by major test traffic spatiotemporal behavior.

Traffic scenario prediction can be divided into three different temporal scenarios: short- term, medium-term, and long-term prediction. In short-term scenario prediction, the aim is to extrapolate real-time data. For medium-term scenario prediction, it is useful to predict future scenarios for traffic network management, while in the long-term scenario, significant changes may have occurred in network models (e.g. road maps). In this chapter, the use of deep learning and machine learning for real time prediction was compared, highlighting their ability to describe and also model transient situations [14].

7.1. Urban Traffic Management

[24] [35]One of the fundamental components of intelligent transportation systems is traffic flow prediction, which can provide real-time information about the road conditions, thus enabling the development and execution of accurate route and schedule plans. Most of the existing traffic flow prediction methods have been designed for freeways and highways,

where the prediction results can be very sensitive to rapidly changing traffic conditions. In contrast, we do not have a well defined traffic flow prediction model specifically designed for urban traffic. Since urban traffic conditions are more unstable and uncertain, the non-homogeneous urban traffic models can have impacts on the more accurate traffic flow prediction. Ultimately, real-time traffic management in autonomous driving is a significant requirement in order to achieve full system autonomy [16]. Of particular interest, core urban traffic management issues can be reviewed in two primary aspects: i) real-time traffic efficiency and ii) management of shared road infrastructures. Real-time traffic efficiency involves effective traffic control to: i) balance road space allocations between different traffic performances, to maximise the use of road networks, and to avoid traffic local jams, and ii) facilitate a smooth overall traffic flow so that unnecessary traffic congestion can be eliminated. Management of shared road infrastructures focuses on addressing conflicts and service provision competitions among different transport vehicles on the main roads. Specifically, 'heterogeneity of unmanned vehicles' can lead to excessive competition on shared urban roads which can make traditional traffic management strategies difficult to implement and can result in chaos on these urban roads. The optimization of traffic control strategies through traffic flow prediction is a potent method of real-time traffic management. In this section, we introduce urban traffic transitivity of network and traffic efficients forecasting using urban traffic flow data in detail.

7.2. Highway Traffic Flow Optimization

CAVs have distinct advantages in real-time variable driving speed performance with designated CAV lanes. Traditional traffic optimization techniques mainly solve intersection control problems, whereas few methods focus on highway traffic, which has large-scale effects. It is important to study the long-term high-level control methods. In the future, extensive studies should be conducted to optimize the overall road network containing both intersections and highways. Efficient highway optimization techniques contribute to the reduction of congestion on main roads and are beneficial for improving the development of public transportation systems. The car-following model may give complicated simulation performance and underestimate the safety of the CAV convoy at practical level. It is an effective technique to deal with complicated congestion propagation characteristics by developing adaptive control strategies of the autonomous CAVs. [22] A feedback-based traffic flow control strategy within the context of mixed traffic including human-driven vehicles and connected automated vehicles is proposed to deal with various unexpected events at the larger scale, such as phantom, emergency brake, and capacity drop. It is also significant to exploit the potential capacity of the expressway. Deceleration striping by making the three-dimensional curbstone keep wig-waw-alternating changes can significantly enhance the capacity of a multi-lane highway. Suppose in the future more CAVs corresponding to a larger length become the main traffic flow or if the human-driven passenger cars and CAVs interact with learning-based strategy. The impact of environmental-related calibration, including parameters for the emergency micromanagement of unexpected events and emergency mutual aid in case of accidents, on the highway traffic flow is further considered.

[36] The traffic on the highways has unique characteristics. The arterial signal optimization focuses on the flow of the transportation network, while highway traffic flow is affected by interactions between vehicles, such as driver behavior, free flow capacity, traffic density, the blockage rate, and the propagation of the congestion front. This complicates the optimization of highway traffic flow. Traditional traffic optimization focuses on managing intersections and optimizing the signal timing. Traditional traffic optimization methods such as cycle times and coordination have problems such as high energy consumption, pollution, and overly high speeds [2]. A capital city in China uses fine-grained timing strategy large-scale traffic simulation methods, paying particular attention to the effects of multi-agent characteristics, and proposes the individual green adjustment strategy, which performs well under low-to-medium traffic. However, it fails under congestion conditions. CAV-related methods use the green signal-phase information as their exclusion and provide a smooth transition from free flow to congestion.

8. Challenges and Future Directions

Neural traffic forecasting model is a rapidly growing research with enormous potential for helping to facilitate the evolution of autonomous driving. It would provide significant benefits to traffic management and operation, including reduced travel time, improved safety, and energy efficiency. However, there are numerous open challenges in realizing these potential benefits. Deep learning algorithms are very sensitive to the training phase, i.e., the amount and quality of input data. Data collected from intelligent vehicles can be less reliable than other traffic sensors, which make the development of reliable architecture difficult [16]. Then, these adaptive algorithms are subject to overconfident predictions during abnormal traffic events such as car breakdowns, accidents, and traffic jams. Therefore, measures should be taken in the structure of deep learning and good data controlling for the prediction of traffic flow [11]. Third, as theoretical research indicates, combined prediction of multiple traffic data and the consideration of the actual operation of the traffic system can better predict traffic flow. It is necessary to improve the weighing of the loss function in new data sources, in order to improve the prediction accuracy for traffic control reference [2]. Lastly, the application of intelligence in assistance-drived to the policies of traffic rule and traffic flow in real-time will be challenging, and solid theory on all aspects should be developed to ensure the models could further be implemented in intelligent vehicle systems.

8.1. Data Privacy and Security Concerns

The work under the concerned privacy-preserving machine learning differs from the prior arts in that it aims to predict traffic conditions, such as road traffic speed and vehicular flow, at intersections, rather than raw vehicle status that may harm location privacy. At first, real-time traffic flow prediction is proposed to be problem P1, which is challenging because of the fluctuate data volume and rate in IoV. Then, a privacy-preserving task is transformed into a regularized learning problem P2 in the presence of location corroboration risk, and local differential privacy (LDP) is employed to tackle P2 by loss perturbation. In the end, experiments are conducted to verify the feasibility and reliability of our framework on real-world data. The main findings and contributions of this paper include that 1) the real-time IoV data can be well predicted with a better performance than the baseline and 2) LDP can effectively deal with the privacy concerns under reasonable assumptions. Furthermore, we will extend our study to a special case, namely smooth movement in the network, aiming to ease the traffic bottleneck by fully taking the traffic control mechanism into account [37].

Data-consuming services have been installed in vehicles to operate the autonomous driving system, which has significantly improved the driving safety and efficiency of vehicles. Using the collected traffic information in vehicles is likely to satisfy the requirement of efficient real-time traffic flow prediction in the internet of vehicle (IoV). A typical example is to forecast traffic flows regarding delay in vehicle-route optimization to reduce the computational burden. However, it is well known that location privacy can be easily compromised in the IoV. Also, if the raw data are directly utilized in the forecasting model, adversaries may use attack behaviors or events to tempt the IoV to join the traffic queue, leading to intentional traffic congestion. Thus, studies on privacy-preserving real-time traffic flow prediction are still missing [38].

8.2. Scalability and Adaptability in Dynamic Environments

The performance of the proposed technique has been monitored for several experiments with the NYC Taxi dataset and the comparative results demonstrated that the proposed T-GCN with ADBN considerably outperforms the state-of-the-art real-time traffic prediction method. In the [143] paper, authors proposed an end-to-end real-time traffic flow prediction method that models the spatial and temporal dependencies of various traffic information in a traffic network by using a time-aware graph convolutional unit and adversarially learns to balance the real unknown percentage class imbalance [39]. Predictive models built with this proposed method can generate accurate traffic flow predictions including future traffic speed, volume, road occupancy, and traffic pressure on vertices. In the publicly available taxi dataset within the real region of New York City, the proposed method shows a better prediction accuracy than the conventional state-of-art real-time traffic prediction models. The proposed tech also outperforms the latest Gated Graph Convolutional Network-based models which are closely related to our method. The proposed method does not need aggregation operations. Instead, behaviors in a spatio-temporal traffic network can be captured efficiently.

[40] The work of [143] introduced an efficient online real-time traffic flow prediction method that introduces two important contributions including a time-aware graph convolutional (T-GCN) network and a traffic-aware Adversarial Data Balancing Network (ADBN) to mitigate the class imbalance issue appearing in the real imbalanced scenarios. In this method, the temporal dependency of various traffic information embedded in a traffic network is captured by the T-GCN and a set of blocks named T-Gblock. The time dimension and spatial graph elements represented by road segments are embedded by using the block features which are fed into the encoder-decoder segment predictor to generate candidate traffic sequences. The load balancing problems in the class imbalance learning issues of sequence sets have been approached with the ADBN which is run consecutively with the aforementioned block predictor in order to increase traffic flow prediction performance.

9. Conclusion and Summary

After having chosen the most appropriate ML algorithms to employ in this paper, we developed training predictors capable of estimating traffic flow states with a linear interpolation approach. For this study, we focused on the arterial edges along the connection paths near future high urbanization areas [41]. To validate our proposed approach, we combined four case studies, and the numerical results were compared with the available anecdotal data from the traffic sensors.

While the majority of currently available traffic systems provide information obtained from traffic sensors, this typically includes only current density, traffic flow, and speed [11]. To tackle the shortcomings of traditional traffic management systems, we proposed a new GIS-based approach to assess future states of traffic volume by invoking ML algorithms [22]. Moreover, a comparative analysis of the similarities in the shape of the multi-temporal profiles among the different traffic flow signals is presented.

9.1. Key Findings and Contributions

State-of-the-art real-time traffic flow prediction techniques were systemized to help identify emerging trends, unveil potential shortcomings, and speed up future research [25]. In general, statistical, feature representation and machine learning, clustering and classification, and time series are the leading real-time traffic flow prediction techniques. The relevant literature shows mainly predictability and data characteristics and sizes for evaluation performance analysis, while there are no all-in-one traffic flow prediction platforms to compare and contrast time-wise, prediction scenario-, threshold-wise, and in difficult circumstantial conditions. Furthermore, significant differences in real-time traffic flow prediction results mostly reflect different testbeds, while additional benchmark tests, instantaneous conditions, overfitting, obtaining reliable and reusable predictions, and very fast prediction time must be highlighted.

The intersection of emerging technologies, domain expertise, and end-user's and provider's knowledge is the smart way to provide more accurate and usable technology [10]. The concept of big data forecasting has been first envisioned in 2011 when three parameters were deemed critical: data, methods, and knowledge. Although the scenario has changed, the parameters remain crucial in modern real-time traffic flow prediction models. The predictors identified in this research are designed to infer real-time traffic flow prediction by involving on-vehicle data, information collected at the vehicle and sub-vehicle levels, such as radar and image sensors, for new and earlier observed condition-specific data, such as LIDAR, ultrasonic sensors, and cameras, upside-down forward vehicle radar, camera images, side-view radar and camera images, air pollution as a sub-vehicle level predictor. Traffic volume, speed, density, and traffic flow direction predictions were also suggested as future indicators to test whether flow results could be improved with additional predictions.

9.2. Final Thoughts and Recommendations for Future Research

Of accelerating training and testing for the complex network with a massive node population, reducing the computation cost of traffic flow prediction is the key step. A method named LSTM (Long Short-Term Memory) is applied by [15] and is further modified to predict traffic flow by combining a set of carefully designed variational structured predictive networks called the variational LSTM. A novel time series (RF signal strength monitoring the vehicle moving and stopping behavior) is proposed as the hidden nodes of the variational LSTMs. The principle of the proposed variational LSTMs is rooted in neural networks and has the capability to learn the characteristics of the target real-time traffic flow with the use of vehicle-to-infrastructure (V2I) communication. Interpretability of the LSTM units, which is a fundamental problem in the deep learning technology, is resolved by using the RF signal strength. The RF signal strength is used as the hidden nodes of the variational structures to predict the flow traffic in wider areas and phase. It is discussed that the RF signal monitoring vehicle flow is a promising technique to predict the underlying traffic flow data in the intersection.

Traffic flow prediction is a key part of intelligent transportation in smart cities. Relevant research and methods are increasingly drawing attention in the context of studies in connected vehicles and shared economy. Urban planners hope to be able to make use of traffic parameters to enhance urban traffic prediction and control. It is essential to understand how traffic flow comes from and to be predicted, as it does and is also paramount to smoothly run in various transportation systems, including cars (opportunistic and automated).

Reference:

1. Tatineni, S., and A. Katari. "Advanced AI-Driven Techniques for Integrating DevOps and MLOps: Enhancing Continuous Integration, Deployment, and Monitoring in Machine Learning Projects". *Journal of Science & Technology*, vol. 2, no. 2, July 2021, pp. 68-98, https://thesciencebrigade.com/jst/article/view/243.
2. Prabhod, Kummaragunta Joel. "Advanced Techniques in Reinforcement Learning and Deep Learning for Autonomous Vehicle Navigation: Integrating Large Language Models for Real-Time Decision Making." *Journal of AI-Assisted Scientific Discovery* 3.1 (2023): 1-20.
3. Tatineni, Sumanth, and Sandeep Chinamanagonda. "Leveraging Artificial Intelligence for Predictive Analytics in DevOps: Enhancing Continuous Integration and Continuous Deployment Pipelines for Optimal Performance". Journal of Artificial Intelligence Research and Applications, vol. 1, no. 1, Feb. 2021, pp. 103-38, https://aimlstudies.co.uk/index.php/jaira/article/view/104.