Traffic Prediction for Intelligent Transportation System Using Machine Learning

N Akhila, M Kavya, M Soumith Reddy and
Prasanta Kumar Pradhan

March 31, 2023

Traffic Prediction For Intelligent Transport System Using Machine Learning

N.Akhila[1*], M.Kavya[2], M.Soumith Reddy[3], Dr. Prasanta Kumar Pradhan[4]

Student, Department of Electronics and Communication Engineering, JB Institute of Engineering and Technology, Moinabad,500075 [1,2,3]

Associate Professor, Department of Electronics and Communication Engineering, JB Institute of Engineering and Technology, Moinabad,500075 [4]

* Corresponding Author **Email id:** nainiakhilagoud@gmail.com**.**

Abstract: Automobile manufacturers have developed various safety features to mitigate the risk of traffic accidents, but accidents continue to occur frequently in both urban and rural areas. To prevent accidents and improve safety measures, it is necessary to develop accurate prediction models that can identify patterns associated with different scenarios. By using these models, we can cluster accident scenarios and develop effective safety measures. We aim to achieve the maximum possible reduction in accidents using low-budget resources through scientific measures.

To achieve this goal, we need to collect and analyze a vast amount of data related to traffic accidents, such as accident location, time, weather conditions, and road features. Machine learning algorithms can be used to automatically identify patterns in the data and predict accident scenarios based on these patterns. These models can then be used to cluster accidents into different categories and develop safety measures tailored to each category.

By using this approach, we can develop cost-effective safety measures that can be implemented in a variety of settings. We believe that this approach has the potential to significantly reduce the number of traffic accidents and improve safety for drivers, passengers, and pedestrians alike.

Keywords:

Machine Learning, Random Forest, Decision Tree, Logistic Regression,
Support Vector Machine

1. **Introduction:**

The availability of precise traffic flow information is essential for various business sectors, government agencies, and individual travellers to make informed decisions regarding their travel routes[1]. The implementation of Intelligent Transportation Systems (ITS) is key to achieving accurate traffic flow prediction and improving traffic management efficiency while reducing carbon emissions[16].

Real-time traffic and historical data collected from diverse sensor sources, such as inductive loops, radars, cameras, mobile Global Positioning System (GPS), crowd sourcing, and social media, are used to predict traffic flow accurately[7]. The explosion of traffic data due to the extensive use of traditional sensors and new technologies has resulted in a vast volume of transportation data, making transportation control and management more data-driven[12].

Although many traffic flow prediction systems and models exist, most of them employ shallow traffic models and have limitations due to the high-dimensional nature of the dataset[6]. As a result, advancements in traffic flow prediction systems and models are crucial to meet the evolving needs of traffic management systems, public transportation systems, and traveler information systems[9].

Deep learning has emerged as a popular approach for solving complex problems in various domains, such as image classification, natural language processing, dimensionality reduction, and object detection[5]. This is due to its ability to learn intricate representations of data through the use of multi-layer neural networks, which can capture underlying patterns and structures in the data[10]. In particular, deep learning is being explored in the development of autonomous vehicles, which can potentially revolutionize transportation systems by reducing costs and improving safety[11]. Intelligent transportation systems (ITS) and researchers are working on driver assistance systems (DAS), autonomous vehicles (AV), and Traffic Sign Recognition (TSR) to provide timely and accurate information to ensure safe and efficient autonomous driving[4]. The use of deep learning in these areas is crucial for enabling the recognition and interpretation of complex visual and auditory cues, and the detection and prediction of objects and events in the environment[3]. Overall, deep learning holds great potential for advancing the capabilities of autonomous systems and enhancing their performance and safety[12].

The prediction of traffic flow information is a challenging task due to the large amount of data involved, making it difficult to achieve accurate predictions with low complexity[8]. While many algorithms have been developed for this purpose, their accuracy remains limited[2][15]. To address this issue, we propose using a combination of advanced techniques such as Genetic Algorithms, Deep Learning, Image Processing, Machine Learning, and Soft Computing[14]. These techniques have demonstrated strong performance in handling Big Data, as evidenced by numerous research papers and journals[13]. By leveraging these approaches, we aim to improve the accuracy of traffic flow predictions and overcome the challenges associated with large, complex datasets[17].

2.Purpose Of System

Traffic congestion is a common problem that can be predicted using actual traffic data. However, this data may not be readily available or accessible to all users, who often require advance knowledge of the best travel routes. To address this issue, it is necessary to predict real-time traffic based on past and recent data sets. Various factors contribute to traffic congestion, and a comparison of these data sets can help to identify patterns and trends. This analysis can then be used to predict congestion levels at different times of day, which can help drivers plan their journeys more effectively.

Fuel prices also play a significant role in traffic flow, and can cause congestion patterns to change rapidly. The objective of this prediction is to provide real-time information on gridlock and congestion, which is essential for intelligent transportation systems (ITS). However, traditional prediction methods may not be sufficient to manage the complex traffic patterns seen in modern cities. Therefore, ongoing research on traffic flow prediction is essential for the development of more effective ITS solutions.

3.Block Diagram

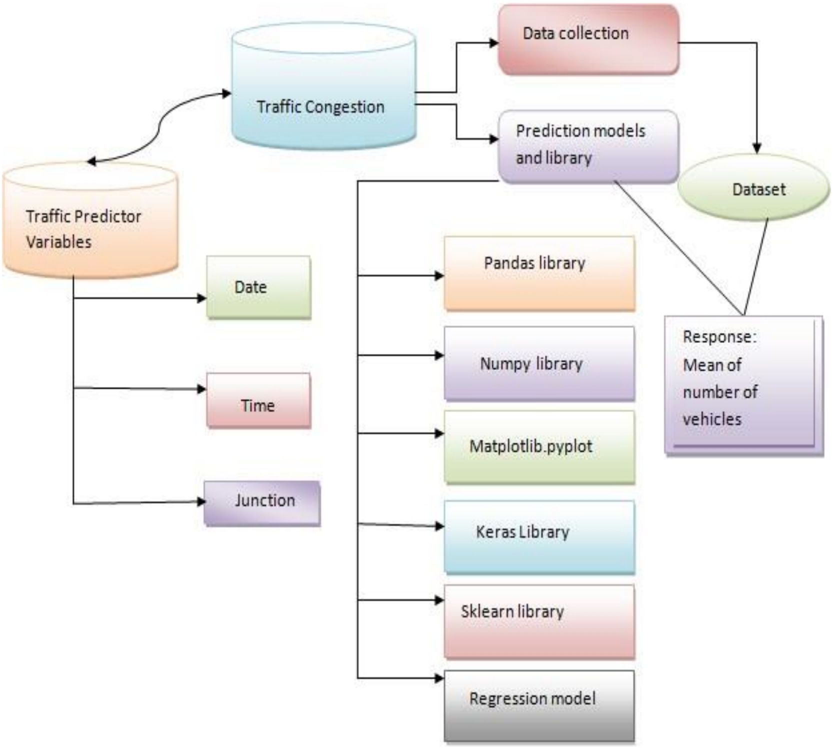


Fig 1 :Outline of the traffic prediction in this paper

4.Methodology

1. First, we take dataset.
2. Filter dataset according to requirements and create a new dataset which has attribute according to analysis to be done
3. Perform Pre-Processing on the dataset
4. Split the data into training and testing
5. Train the model with training data then analyze testing dataset over classification algorithm
6. Finally you will get results as accuracy metrics

Supervised learning

Supervised learning involves training a model on a dataset that contains labeled input and output parameters. The labeled dataset is used for both training and validation of the model. This approach is based on the principle of utilizing known output values to predict the corresponding input values accurately.

A. Classification:

This is a supervised learning task that involves predicting discrete values belonging to predefined classes. The output has a defined set of labels, such as 0 or 1, and the objective is to accurately predict the class to which an input belongs. The model's accuracy is evaluated based on its ability to correctly classify inputs into the correct class. This approach can be used for both binary and multi-class classification tasks. In binary classification, the model predicts a single class label, whereas in multi-class classification, the model predicts multiple class labels. For example, Gmail uses multi-class classification to categorize emails into categories like social, promotions, updates, and forums.

Example of Supervised Learning Algorithms:

- Gaussian Naive Bayes
- Decision Trees
- Support Vector Machine (SVM)
- Random Forest

ML | Types of Learning – Supervised Learning

Supervised learning is a machine learning paradigm in which the algorithm is trained on a dataset that has labeled data, where the target variable is already known. The objective of supervised learning is to train a function that can accurately predict the output variable based on the input variables. There are two primary types of supervised learning:

Classification: Classification is a supervised learning technique where the target variable is categorical, and the objective is to classify a new data point into one of the pre-defined categories. The algorithm learns from a labeled dataset to predict the class label of a new instance. Applications of classification problems include image classification, spam detection, sentiment analysis, and medical diagnosis. The performance of the classification model is typically evaluated based on metrics such as accuracy, precision, recall, and F1-score.

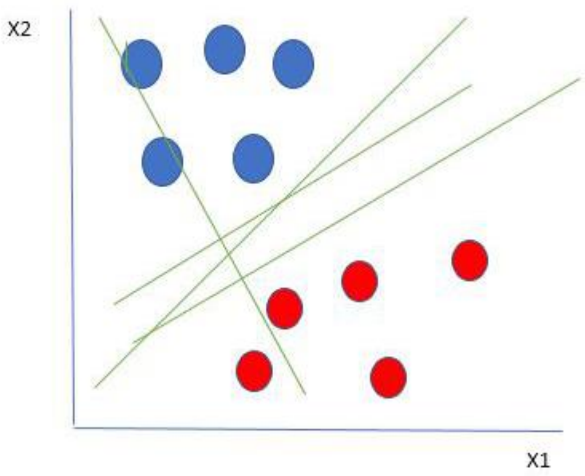
Regression is a supervised learning technique that deals with predicting a continuous output variable based on input variables. The primary objective of regression is to learn a function that can estimate the output value for a given input value. Examples of regression problems include stock price prediction, weather forecasting, and sales forecasting. In regression, the algorithm is trained on a labeled dataset with input and output variables. The performance of the regression model is typically evaluated based on metrics such as mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination (R-squared).

Supervised learning algorithms are extensively used in diverse domains such as natural language processing, computer vision, medical diagnosis, and speech recognition, among others. These algorithms are trained on labeled datasets to learn a mapping function that can predict the output variable for a given input variable accurately. Some of the most popular supervised learning algorithms include decision trees, random forest, and support vector machine (SVM). Decision trees construct a tree-like model to classify the data, while random forest uses multiple decision trees to improve the accuracy of the classification. SVM is a linear classifier that separates the data into different classes by finding the optimal hyperplane. These algorithms have been successfully used in various applications to achieve state-of-the-art results.

Supervised learning is a useful approach when the data is labeled. However, in some cases, the data may be unlabeled, or labeling the data may be too expensive. In such scenarios, unsupervised learning, semi-supervised learning, or self-supervised learningcould be more suitable. Unsupervised learning deals with learning from an unlabeled dataset, where the goal is to discover hidden patterns, structures, or relationships in the data. Semi-supervised learning, on the other hand, combines labeled and unlabeled data to improve the performance of the model. Self-supervised learning is a form of unsupervised learning where the algorithm learns from the data by generating supervisory signals automatically. These techniques have been applied to various domains such as natural language processing, computer vision, and robotics, among others, to address real-world problems where labeled data is limited or unavailable.

SVM

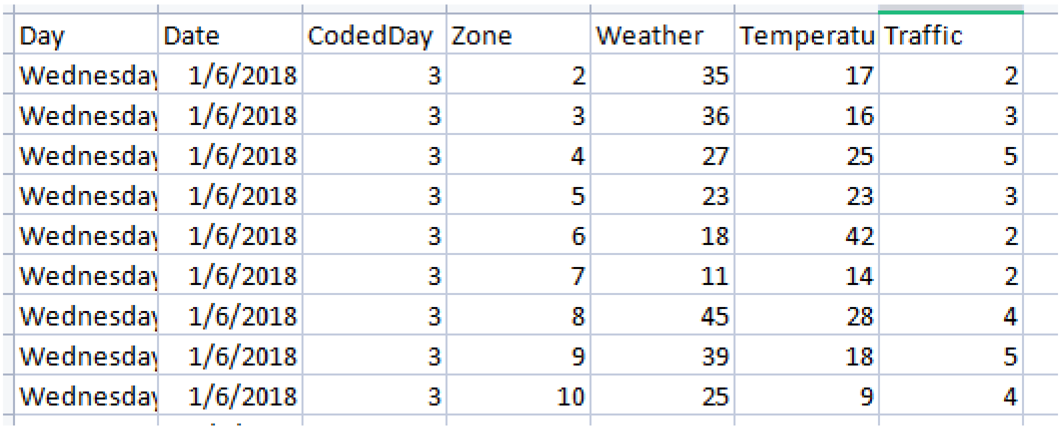
Support Vector Machine (SVM) is a supervised learning algorithm that can be used for classification as well as regression tasks. SVM aims to find a hyperplane in an N-dimensional space that can efficiently classify the data points. The number of dimensions of the hyperplane is determined by the number of features in the dataset. When there are only two input features, the hyperplane is simply a line, and for three input features, the hyperplane becomes a 2-D plane. However, when the number of features exceeds three, it becomes difficult to visualize the hyperplane. In such cases, SVM maximizes the margin between the two classes and adds a penalty each time a point crosses the margin. This is known as a soft margin, and the goal of SVM is to minimize (1/margin+∧(∑penalty)) to obtain the optimal hyperplane. Hinge loss is a commonly used penalty that is proportional to the distance of the violation. If there are no violations, there is no hinge loss. SVM has proven to be a powerful tool for classification tasks and has been successfully used in various applications such as image recognition, text classification, and bioinformatics, among others.



Linearly Separable Data points

4.1 Data Set

The dataset is the collection of about 1440 Open Data.world website .



Day	Date	CodedDay	Zone	Weather	Temperatu	Traffic
Wednesday	1/6/2018	3	2	35	17	2
Wednesday	1/6/2018	3	3	36	16	3
Wednesday	1/6/2018	3	4	27	25	5
Wednesday	1/6/2018	3	5	23	23	3
Wednesday	1/6/2018	3	6	18	42	2
Wednesday	1/6/2018	3	7	11	14	2
Wednesday	1/6/2018	3	8	45	28	4
Wednesday	1/6/2018	3	9	39	18	5
Wednesday	1/6/2018	3	10	25	9	4

Dataset consists of the following attributes (Dataset.csv)

1. **Day**
2. **Date**
3. **CodeDay**
4. **Zone**
5. **Weather**
6. **Temperature**
7. **Traffic**

Traffic congestion has become a major concern in urban areas due to various factors such as rapid population growth, uncoordinated traffic signal timing, and lack of real-time data. The increase in traffic congestion has resulted in significant economic losses, travel time delays, and environmental impacts. To address these issues, machine learning algorithms using Python 3 have been implemented to predict traffic flow patterns.

Data used in this study was collected from the Kaggle website, a platform for data science enthusiasts. The data includes two datasets, one from 2015 and the other from 2017. The datasets contain detailed information on traffic flow such as date, time, number of vehicles, and junction details. The datasets were collected to facilitate comparison between the two years and evaluate the effectiveness of the machine learning algorithms.

Pre-processing techniques were applied to the collected data to remove irrelevant information and aggregate the remaining data into 1-hour intervals. This allowed for accurate traffic flow prediction with each 1-hour interval. The pre-processing techniques included data cleaning, data normalization, and data transformation to ensure that the data used for training and prediction were of high quality and relevance.

The machine learning algorithms used in this study include various statistical models such as Random Forest, Gradient Boosting, and XGBoost. These models were trained using the pre-processed data and then used to predict future traffic flow patterns. The accuracy of the models was evaluated using statistical measures such as mean absolute error (MAE) and root mean square error (RMSE).

The implementation of machine learning algorithms has the potential to significantly improve traffic flow and reduce congestion in urban areas. The accurate prediction of traffic flow patterns can aid in the development of more effective traffic management strategies, including the optimization of traffic signal timing and the implementation of real-time traffic updates for commuters. Overall, the findings of this study provide valuable insights into the potential of machine learning algorithms in addressing the challenges of traffic congestion in urban areas.

4.2 Regression model

Regressor model analysis could even be a mathematical technique for resolvingthe connection in the middle of one dependent (criterion) variable and one or moreindependent (predictor) variables

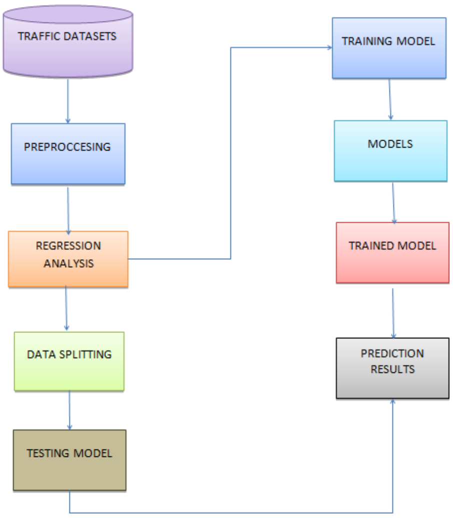


Fig 2: Regression model of traffic prediction in this paper

The evaluation yields a foretold value for the benchmark resulting from a sum of scalar vectors of the predictors. The accuracy is measured by computing mean square error. Thus obtaining the expected error from the observed value and also truth value which is equivalent to the standard deviation deployed within the statistical met

Results and discussion

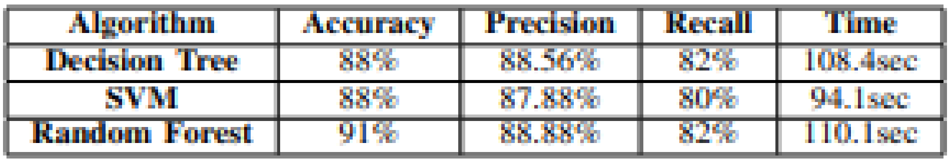
The movement of goods and humans is an integral part of existence. With the increase in the population and the necessity of social wellbeing of humans, travel is exponentially growing. As technology is evolving day by day so is the number of vehicles increasing. With this rapid rate of increase in vehicle, management of movement of vehicles is very critical. Vehicular management helps in optimizing the travel time and cost of the travel. For developing a precise vehicular management system, it is essential to have accurate background information. Traffic flow is one of the most important data which is required for developing a precise vehicular management system. This paper presents a review of recent deep learning approaches in the field of traffic flow prediction. Most of the contributions are application based while very few articles have a strong contribution to theory. Deep learning models for traffic forecasting have shown promising results to represent the non-linearity of traffic flow prediction. While there are several advantages to using the deep learning models to predict traffic flow individually, there are significant disadvantages also. Thus, recently, researchers are starting to move from deep learning architectures to hybrid and unsupervised methods. This review addressed the various existing deep learning architectures used for traffic flow prediction and the rising popularity of hybrid methods.

We have applied and tested different machine algorithms for achieving higher efficiency and accurate results. To identify classification and regression we have used a decision tree algorithm(DT).The goal of this method is to predict the value of target variables. Decision tree learning represents a function that takes as input a vector of attributes value and return a “Decision” a single output value.it falls under the category of supervised learning algorithm. It can be used to solve both regression and classification problem. DT identify its results by performing a set of tests on the training dataset.

Outliners detection is another critical step for an accurate result, and for this, we have used support vector machines(SVM’s),which is a set of supervised learning methods that can also be used for classification and regression .The SVM is beneficial for high dimensional spaces, and it helps in the condition where a number of samples are less than the number of dimensions.

The Random forest Algorithm is a robust machine learning algorithm. It is defined as bootstrap aggregation. The random forest algorithm is based on forecasting models, and it is mostly used to classify the data. The bootstrap algorithm is used to generate multiple models from a single training data sets. A bootstrap algorithm has also used a sample to estimate statistical quantities.

The results of performance of the models obtained through different machine learning algorithms that are discussed in this paper. In this table we defined various attributes like Accuracy, Precision, Recall and Time Taken.



Algorithm	Accuracy	Precision	Recall	Time
Decision Tree	88%	88.56%	82%	108.4sec
SVM	88%	87.88%	80%	94.1sec
Random Forest	91%	88.88%	82%	110.1sec

6 Conclusion

We plan to create a traffic flow prediction system using a machine learning algorithm that employs a regression model. This system will inform the public of current traffic conditions and predict traffic flow in the next hour. Users will also be able to learn about road conditions, such as the number of vehicles passing through a specific intersection. We recognize that traffic data is affected by changing weather conditions, fluctuating fuel costs, and variations in carpooling. Therefore, we will compare the prediction with traffic data collected over the past two years to provide accurate traffic flow information. The prediction will help users plan their route, make informed decisions, and avoid traffic congestion.

Our traffic flow prediction system will utilize a supervised learning algorithm to analyze past traffic data and create a regression model. This model will then be used to predict traffic flow in real-time. We will collect traffic data from various sources, including traffic cameras and sensors, to ensure accuracy. The system will also take into account weather conditions, such as rain or snow, which can impact traffic flow. Additionally, we will factor in fuel costs and carpooling data to provide a comprehensive analysis of traffic patterns. The prediction will be displayed on a user-friendly interface, allowing the public to easily access the information and make informed decisions about their travel plans. Overall, our system will help alleviate traffic congestion, reduce travel time, and enhance overall transportation efficiency.

7 Future Work

In the future, our traffic flow prediction system can be enhanced using advanced techniques such as deep learning, artificial neural networks, and big data. This will help us analyze more factors that affect traffic management and provide users with accurate suggestions for the easiest route to their destination. While many forecasting methods have already been applied, there is still scope for improving prediction accuracy. Using the increased availability of traffic data, we can develop new forecasting models to improve our predictions. Accurate traffic prediction is crucial for efficient transportation management, and our prediction method can help users plan ahead and avoid congestion. We aim to improve the accuracy of our prediction model in the future by developing user-friendly and accessible methods, such as integrating weather outlook and GPS data. Additionally, we will highlight accident-prone areas to ensure the safety of our users. We will achieve this through deep learning, big data, and artificial neural networks.

To further improve the accuracy of our traffic flow prediction system, we can also consider incorporating real-time data from social media and mobile applications. This data can provide additional insights into traffic patterns and help us make more accurate predictions. Additionally, we can explore the use of predictive analytics to anticipate changes in traffic flow and adjust our predictions accordingly. This will help us provide users with more accurate and reliable information about traffic conditions.

Furthermore, we can use advanced visualization techniques, such as heat maps, to provide users with a clear and easy-to-understand representation of traffic patterns. This will allow users to quickly identify areas of congestion and plan their route accordingly.

To ensure the scalability of our system, we can leverage cloud computing and distributed computing technologies to process large amounts of data in real-time. This will allow us to handle increasing volumes of traffic data and provide users with timely and accurate predictions.

Overall, our traffic flow prediction system will continue to evolve and improve as we incorporate new technologies and data sources. By providing users with accurate and timely information about traffic conditions, we can help reduce congestion and improve the efficiency of our transportation systems.

Acknowledgements:

The authors wish to thank the chief editor, the associate editor and the anonymous reviewers for their helpful feedback to improve the quality of this paper. We thank Jbiet, Department of Electronics and communication Engineering, Associate Professor Dr. Prasantha Kumar Pradhan Sir for his guidance and support in this project work.

Competing interests:

The authors declare there are no competing interests.

Authors’ contributions:

Soumith conceived the presented idea, and Akhila developed the theory and performed the computations. Kavya verified the analytical methods and provided guidance, while both Akhila and Kavya offered support and supervision to Soumith. All authors discussed the results and contributed to the final manuscript. Dr.Prasanta Kumar Pradhan also provided invaluable guidance with his expertise.

References:

1. Joaquín Abellán, Griselda López, and Juan De OñA. 2013. Analysisoftrafficaccidentseverityusingdecisionrulesviadecisiontrees. Expert Systems with Applications 40, 15 (2013), 6047–6054.

2. Mikhail Belkin and Partha Niyogi. 2001. Laplacian eigenmaps and spectral techniques for embedding and clustering. In NIPS, Vol. 14. 585–591.

3. Ruth Bergel-Hayat, Mohammed Debbarh, Constantinos Antoniou, and George Yannis. 2013. Explaining the road accident risk: Weather effects. Accident Analysis & Prevention 60 (2013), 456–465.

4. Ciro Caliendo, Maurizio Guida, and Alessandra Parisi. 2007. A crash-prediction model for multilane roads. Accident Analysis & Prevention 39, 4 (2007), 657–670.

5. Li-Yen Chang. 2005. Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network. Safety science 43, 8 (2005), 541–557.

6. Cheng Chen, Manchun Tan. Analysis of traffic flow in incident road section affected by traffic accident . Science Technology and Engineering, 2011; 28.

7. Heydecker B G钏Addison J D. Analysis and modeling of traffic flow under variable speed limits. Transportation Research Part C钏2011;19: 206—217.

8. MccreaJ钏MoutariS. A hybrid macroscopic-based model for traffic flow in road networks. European Journal of Operational Research钏2010; 207:676—684.

9. [US] the United States Traffic Research Committee.Ren Futian, Xiaoming Liu, Jian Rong translation.Road Capacity Manual . Beijing: People's Communications Press, 2007,12.

10. Jiang Liu. Mountain Two-lane road Capacity Research. Beijing: Beijing University of Technology, 2006.

11. Kuanmin Chen, Baojie Yan. Road capacity analysis . Beijing: People's Communications Press, 2003: 44-51,67-91.

12. Jin-chuan Chen, Xiao-ming Liu, Fu-tian Ren, et al.Advances in Operation Analysis of Road Interweaving Area . Highway Traffic Science and Technology.2000,17 (l): 46-50.

13. Jian Rong, Futian Ren, Xiaoming Liu.Study on Simulation Model of Basic Section of Expressway. Beijing: Beijing University of Technology, Ministry of Communications Highway Research Institute. 1999: 1-60.

14. Jinyu Duan, Design of Traffic Simulation Software System for Ramp Area of Freeway. Journal of Highway and Transportation Research and Development, 1999,16 (l): 31-35.

15. People's Republic of China Ministry of Housing and Urban-Rural Construction. Urban Road Engineering Design Code (CJJ37-2012) . Beijing: China Building Industry Press, 2012: 8-9.

16. Yang Weizhong, Zhang Tian. SPSS statistical analysis and industry application case. Beijing: Tsinghua University Press, 2011: 49-51. [13] Structural Equation Modeling. Chongqing: Chongqing University Press, 2009: 120-150.

17. Qisen Zhang, Yaping Zhang. Road capacity analysis. Beijing: People's Communications Press, 2002: 4-6,36-86 .