Performance Improvement of Traffic Flow Prediction Model using Combination of Support Vector Machine and Rough Set

Minal Deshpande
Research Scholar
G H Raisoni College of Engineering
Nagpur, India.

Preeti Bajaj
Director
G H Raisoni College of Engineering
Nagpur, India.

ABSTRACT

Short term traffic flow prediction has become one of the important research fields in intelligent transportation system. The prediction of this traffic flow information quickly and accurately is important for traffic control and guidance to initiate the measuring steps well in advance. It makes the transport users better informed and makes the transport network smarter, safer and more coordinated. It plays a crucial role in individual dynamic route guidance, advance traffic information system (ATIS) and advance traffic management system (ATMS). This paper discusses the implementation of traffic flow prediction model using support vector machine. Rough set is used as a post processing tool to validate the prediction result. The objective is to improve traffic flow prediction performance. Data near Perungudi toll plaza in IT corridor in Chennai, India is used for the analysis. It is found that the use of rough set results in satisfactory performance improvement which is evaluated using mean square error as the performance measures.

General Terms

Intelligent Transportation Systems, Soft Computing Methods.

Keywords

Intelligent Transportation Systems (ITS), Rough Set Theory (RST), Short term traffic Flow Prediction, Support Vector Machine (SVM).

1. INTRODUCTION

With the rapidly increasing urbanization and traffic demand, transportation problems are becoming serious issue everywhere in the world. Infrastructure growth is limited because of the size, space constraints and because of lack of planning, technology. So the solution to this critical problem is to design intelligent systems to provide innovative and smarter services to the transport users. One such application is traffic flow prediction on short term basis which makes the transport users to be better informed and makes the transport network smarter, safer and more coordinated. In ITS, one of the basic component is traffic flow prediction which supports traffic monitoring and control systems [1].

Traffic flow prediction is supposed to be the key technology of ITS but at the same time a challenging problem. Short term prediction reflects rapid changes in traffic demand, for example, in flexible traffic control system. It forms an essential component of controlling traffic in real time and management system. It is used to predict the vehicle flow count in the next time interval which usually varies in the range of 5 minutes to about 30 minutes (half an hour).

As traffic flow exhibits nonlinear, time varying and stochastic characteristics, it is very difficult to predict the traffic flow accurately on all time periods of the day. Traffic flow gets affected by many nonlinear and uncertain elements like weather, road conditions, accidents, holidays, time and day of travel etc.

Support vector machine is one of the promising approaches in the traffic flow prediction problem. It is categorized into supervised learning method which is used to analyze the data and recognize patterns, used for classification and regression analysis. It can be used to approximate non-linear system with higher accuracy. It has good generalization and fast convergence characteristics. Hence it is suitable tool for traffic flow prediction problem. Application of rough set helps in evaluating the significance of data. It is defined as a formal approximation of the crisp set by a pair of sets. These two sets define the lower and the upper approximation obtained from the original set. In the proposed work, Rough set is used to perform the validation of result thereby used as a post processing tool.

The rest paper is organized into different sections as follows. Section II focuses on literature review. Section III discusses the data collection details. Section IV includes the description about the structure of SVM and Rough Set used for short term traffic flow prediction. Prediction experiment is performed and simulation results are discussed in section V. Conclusions are given in final section.

2. LITERATURE REVIEW

Various authors have proposed different techniques to design the traffic flow prediction model to get acceptable prediction accuracy. Artificial neural network (ANN), Moving average, Autoregressive Integrated Moving Average Method, Bayesian networks, Kalman filter and hybrid approaches were used for traffic flow prediction. H. Chang et al. proposed traffic flow prediction on multiple intervals using a non parametric K-nearest neighbour method [2]. This method requires large amount of historical data to produce satisfactory results. Qiangwei Li et al. proposed time dependent model of support vector machine for short term traffic volume prediction. Time varying characteristics of traffic flow are matched using this model [3]. Shiliang Sun et al. [4] proposed traffic flow forecasting using Bayesian network considering the traffic flows at the adjacent road links to analyze the variation of traffic flow at the current link. Also issue of incomplete data is handled. Wusheng Hu et al. [5] proposed the prediction model applying dynamic rolling prediction along with back propagation algorithm. But every method has some drawbacks.

Rough set theory is a mathematical tool which deals with vagueness. The concept of rough set originates assuming that some information is always associated with every object

existing in the universe of discourse. There are various applications where the concept of rough set theory has been applied like finance, banking and investment fields [6,7,8]. Bin-sheng Liu et al. proposed traffic flow prediction where rough set and genetic algorithm were applied for the selection of the relevant forecasting variable [9]. Zhenguo Zhou and Kun Huang reported rough fuzzy neural network model employed for prediction of traffic flow [10]. Pang Ming-bao and HE Guo-guang proposed recognition model of chaos using rough set and neural network where rough set theory was used to obtain reduced feature vector [11]. Xinrong Liang et al. proposed elman neural network based traffic flow prediction where original data is reduced using rough set theory [12].

After review of literature it was observed that all of the techniques used were as per the western road setting. In India, Traffic and road management are quite different from their western counterparts. In this work implementation of traffic flow prediction model is done using support vector machine. Rough set is used in a novel way. It is used as a post processing tool to validate the prediction result. The use of rough set results in considerable improvement in the performance of traffic flow prediction evaluated using performance measure.

3. DATA COLLECTION

Traffic count, velocity and density are the three fundamental measures of traffic flow. As we are concerned with traffic flow prediction problem, traffic flow count has given priority. Rest of the fields like speed and density doesn't affect the prediction performance. Data set employed in this work is acquired from IIT, Chennai. It is recorded at a location near Perungudi toll plaza in IT corridor in Chennai for 6 days continuously for 24 hours from Monday to Saturday, April 2014. The device used to measure the traffic flow count is TIRTL (The Infrared Traffic Logger) [13]. It is installed on opposite sides of the road perpendicular to the flow of traffic. It uses infrared light based technology. The data were reported at every time instant the vehicle passed through the detector section.

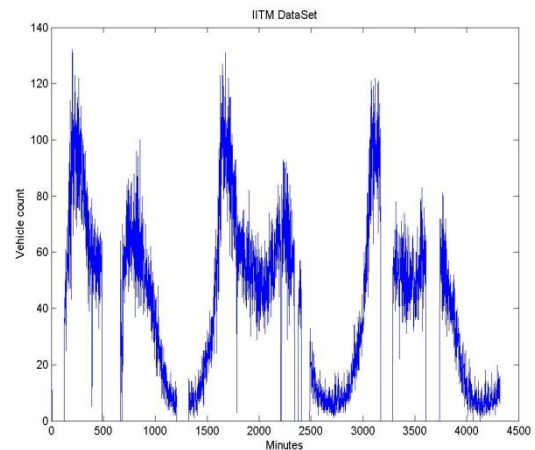**Figure 1: Location map of study area**

Figure 2: Traffic flow pattern of study area

Traffic flow pattern of recorded data is shown in the figure 2 which shows recurrent pattern of daily traffic profile. It is plotted by taking into account data of 3 days which is equal to 3 x 24 hours x 60 minutes=4320 minutes. It is observed that there are certain time intervals when no vehicles are recorded and peak traffic flow intervals also can be observed. Before preprocessing, data available in CSV format is converted into required time format with the help of script. The recorded data includes date, time, speed, velocity, vehicle classification, trigger class, trigger list and axle details etc. But some parameters are relevant only at some specific installations. As we are concerned with traffic flow count only, these fields do not contribute in the prediction performance. So pre processing is carried out in order to save the time in processing of invalid and redundant data so as to speed up the computation time. So data pre processing is done which includes deletion of null entries and to delete redundant entries and also to arrange the available data in 1 min time interval. After pre processing, the data were aggregated into 5 min, 10min and 15 min interval and processed.

4. METHODOLOGY USED

4.1 Support Vector Machine

Support vector machine is defined to be class of supervised learning algorithm. It is based on statistical learning theory. Both classification and regression problem can be solved with the help of support vector machine. As traffic flow prediction is a nonlinear problem equivalent to function approximation, it can be solved by using support vector machine for regression. The main concept is always the same: Obtaining optimized hyperplane so that margin maximization is achieved, to minimize error.

Let us define l number of training samples (x_1,y_1), $(x_2,y_2),....(x_l,y_l)$ Where $x_l \in R^n$ denotes the input samples and $y_l \in R$ represents the corresponding target value.

The regression function is defined by mathematical equation as

$$y(x) = \mathbf{w}.\varphi(x) + b$$

Where **w** is a standard vector and b represents the error. These are called the regression parameters required to found out,

$\varphi(x)$ denotes the higher dimensional mapping from original space R^n.

Always the main objective of the algorithm is to obtain minimized

$$\frac{1}{2}||w||^2 + C \sum_{i=1}^{N} (\xi_i + \xi_i^*)$$

Subjected to the constraints

$$y_i - \boldsymbol{w}x_i - b \le \epsilon + \xi_i$$

$$\boldsymbol{w}x_i + b - y_i \le \epsilon + \xi_i^*$$

Where both ξ_i and $\xi_i^* \ge 0$

The Kernel function plays a vital role in support vector machine. It transforms the training data into a new space which is of high dimensions so that it is possible to make them linearly separable in the transformed space.

$$y = \sum_{i=1}^{l} (\propto_i - \propto_i^*)\{\emptyset(x_i).\emptyset(x)\} + b$$

Using kernel function the equation becomes

$$y = \sum_{i=1}^{l} (\propto_i - \propto_i^*).K(x_i,x) + b$$

Where $K(x_i, x_j) = \emptyset(x_i)$

represents the inner product of the vector in the feature space of high dimension. There are different types of kernel functions available for use. But the most widely used kernel function is Gaussian Radial Basis Function because of the need to set very few parameters and excellent overall performance [14]. It is mathematically expressed as

$$K(x_i, x_j) = exp\left(-\frac{||x_i - x_j||^2}{2\sigma^2}\right)$$

4.2 Rough Set Theory

The Rough Set theory is defined as a formal approximation of a crisp set by a pair of sets. The main goal of the rough set analysis is defining the approximation of concept. Hence every rough set is associated with a pair of sets called lower and upper approximation. Rough set theory proposes a new mathematical approach to imperfect knowledge, ie to vagueness (or imprecision).In this approach, vagueness is expressed by a boundary region of a set.

In order to define rough set mathematically,

With every object in the universe some information is associated given by IS= (U, A) where U and A are finite and non-empty sets which represents data objects and attributes respectively.

For every $a \in A$, $a:U\text{->} V_a$ where V_a corresponds to value set of a.

Set B which is a subset of A determines a binary relation I (B) on U called indiscernibility relation.

The relation is defined as $(x,y) \in I(B)$ subject to the constraint that if a(x)=a(y) for every a in B.

Now A and B are related by

$$X \subseteq U \quad \text{and} \quad B \subseteq A.$$

The lower and upper approximations are represented mathematically by the following equations

$$\underline{B}X = \cup_{x \in U} \{ B(X): B(X) \subseteq X \}$$

$$\overline{B}X = \cup_{x \in U} \{ B(X): B(X) \cap X \ne \emptyset \}$$

Thus lower approximation consists of all the members that surely belong to the target set.
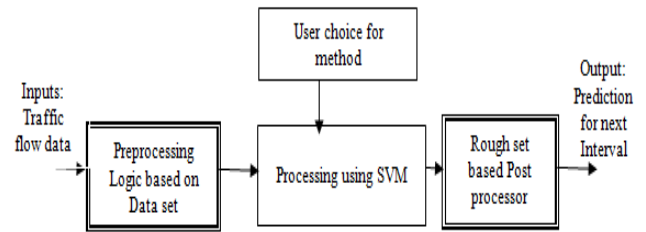

Figure 3: Rough set based model for traffic flow prediction

The vehicle count which is predicted has to be in accordance with the actual data which is available for training the system. If the predicted data deviates from the training set the system is going to have error in proportion to deviation. In order to reduce this error the rough set mechanism is used as a validator for prediction results by defining the lower approximation which is bound to be the member of target set.

If the prediction of the traffic count belongs to lower approximation set the error is minimal, but if it is beyond the lower approximation set then surely it will contribute to error thereby increasing it.

This work uses a correction methodology for the predicted traffic count to be corrected by choosing a traffic count from the training set value closest to the current predicted value and belongs to the lower approximation set.

5. EXPERIMENTS AND SIMULATION RESULTS

After pre processing of data is done, the traffic data is aggregated into different time intervals like 5 min, 10 min and 15 minutes. Before further processing time series data is normalized so as to make them in the range from 0 to 1 using the following formula:

$$x_{tNormal} = \frac{x_t - x_{min}}{x_{max} - x_{min}}$$

Where x_{max} and x_{min} represent the maximum and minimum value of each input sample. Data normalization is done to fasten the convergence speed during training of SVM. It is also to facilitate data processing to avoid numerical computational problems. So data sets were scaled to the range between 0 and 1.

RBF kernel function is used for the SVM model. Regularization parameter γ is defined equal to 10 and bandwidth for RBF kernel σ^2 is defined to be equal to 0.2.These values are finalized after trying different combinations and value which gives good results are selected for simulation.
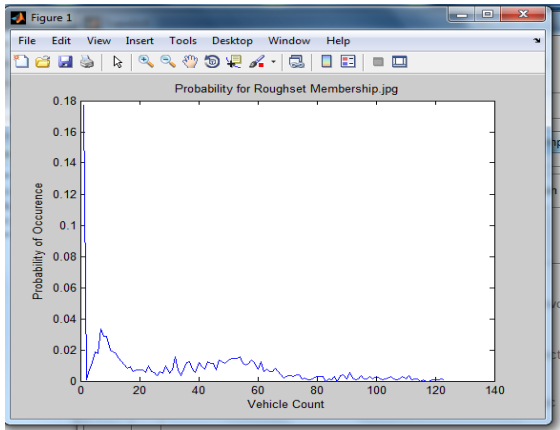
Figure 4: Probability for Rough set membership

Figure 4 shows the probability of occurrence of every vehicle count for defining rough set membership. In order to establish a rough set membership function we make use of the probabilities of the presence of a particular traffic count in the training set. As the training set is a bounded set with 0 on the lower side and a maximum value (say 250) on upper side, we compute the probability of occurrence in training set of all the numbers from 0 to 250. This gives a number which is from 0 and 1. As the counts which are not present in the training set will have zero probability these are considered as non members of the Rough set of the training data set. Values which are present have a non zero probability including the traffic count of 0 (when no vehicles arrive on the road during the aggregation time) and hence qualify to become the member of the rough set.
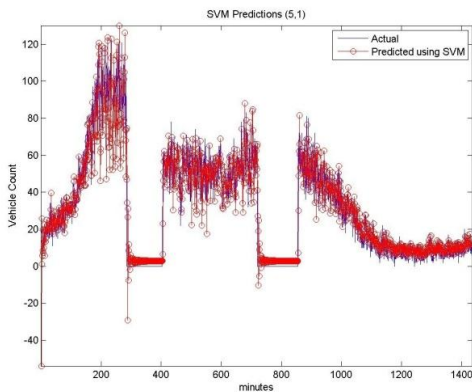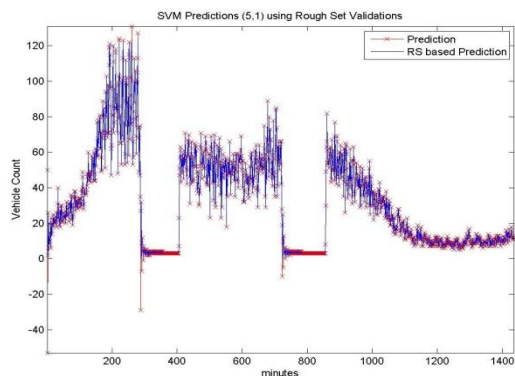

Figure 5: Actual Vs Predicted vehicle count using SVM

Figure 6: Rough set based validation of the prediction result

Figure 5 indicates the traffic flow prediction using SVM. It has been observed that the predicted traffic count matches nearly with the actual count. Simulation is performed for different test data. It also shows the validation of prediction result using rough set. In case of rough set we have negative numbers and numbers greater than maximum vehicle count all have a probability of zero as they are not in the initial data set. So we just replace these by nearest values which are part of the rough set. Figure 7 shows the performance validation using scatter plot. The ideal case should be that any actual value count should match with the forecasted value which is shown by the blue line. Any predictions/forecasted values which are matching with the corresponding actual values will be shown in alignment with the blue line. In this case we see that there are some forecasted values which deviate from the actual value.
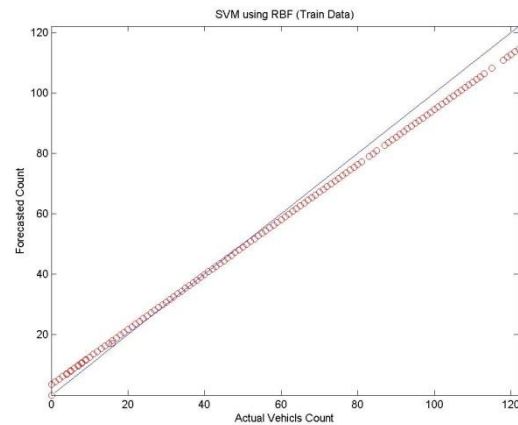

Figure 7: Scatter Plot for validation of the prediction result

Mean square error (MSE) is used to measure the prediction performance. As MSE is very sensitive measure it senses even small errors. MSE is defined as :-

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(y_t - \widehat{y_t})^2$$

Where y_t is the actual value of the traffic from the dataset and $\widehat{y_t}$ is the predicted value.

Table 1. Performance Measure

Method \ Measure	MSE
SVM (5 ,1)	1.42E-03
SVM RS App (5,1)	1.38E-03

Table 1 show the performance measure for SVM (5,1) system where 1 corresponds to aggregation interval and 5 represents the number of past samples used for prediction. Validation of prediction result using rough set results in improvement in mean square error.

6. CONCLUSION

Short term traffic flow prediction is one of the hot research topics in intelligent transportation system. Accurate and timely prediction can provide reliability for optimized traffic control and guidance. It has been observed that the variation obtained in the predicted and actual value using SVM is negligible for

not only the training data but also for test data. The prediction validation has been performed by plotting the predicted and actual values and checking them using a scatter plot. Rough set validates the result by defining the lower approximation which contains the members surely belong to be the part of target set and thus results in improvement in prediction performance.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Liang Zhao and Fei-Yue, " Short-term Fuzzy Traffic Flow Prediction Using Self-Organizing TSK-Type Fuzzy Neural Network", IEEE International Conference on Vehicular Electronics and Safety, ICVES, 2007.

[2] H. Chang, Y. Lee, B.Yoon and S. Baek," Dynamic near-term traffic flow prediction : system oriented approach based on past experiences", IET Intelligent Transportation System, 2012, Vol 6, Issue 3, pp 292-305.

[3] Qiangwei Li," Short-time Traffic Flow Volume Prediction Based on Support Vector Machine with Time-dependent Structure", Inetrnational Instrumentation and Measurement Technology Conference, I2MTC 2009, Singapore, 5-7 May 2009.

[4] Shiliang Sun, Changshui Zhang and Guoqiang Yu ," A Bayesian Network Approach to Traffic FlowForecasting", IEEE Transactions on Intelligent Transportation Systems, Vol 7, No. 1, March 2006.

[5] Wusheng HU et al.," The Short-Term Traffic Flow Prediction Based on Neural Network", 2nd International Conference on Future Computer and Communication, Volume 1, 2010, pp 293-296.

[6] S. Chen and P Wang, " Computational intelligence in economics and finance", Springer, ISBN: 3-540-44098-4, 2002, pp 15-22.

[7] L. Shen and H Loh, " Applying rough sets to market timing decisions", Decision Support Systems Journal , 2006, pp 65-72.

[8] K Ang and C Quek," Stock trading using RSPOP : A novel rough set based neuro-fuzzy approach", IEEE International conference on Trnsactions on Neural Networks, 2005, pp 105-112.

[9] Bin-Sheng Liu et al.," Research on Forecasting Model in Short Term Traffic Flow Based on Data Mining Technology", Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications, ISDA 2006.

[10] Zhenguo Zhou and Kun Huang," Study of Traffic Flow Prediction Model atv Intersection Based on R-FNN", International Seminar on Business and Information Management , 2008.

[11] PANG Ming-bao and HE Guo-guang ," Chaos Rapid Recognition of Traffic Flow by using Rough Set Neural Network", Inetrnational Symposiums on Information Processing, 2008.

[12] Xinrong Liang et al. ," Freeway Traffic Flow Model Based on Rough Sets and Elman Neural Network", IEEE 2009.

[13] http://www.ceos.com.au.

[14] Shi, D. Y. Lu, J, Lu,L.J.," A judge model of the impact of lane closure incident on individual vehicles on freeways based on RFID technology and FOA-GRNN method," Wuhan Univ. Technology, Vol 34, pp 63-68, 2012.

[15] Msizi Khoza and Tshilidzi Marwala," A Rough Set Theory Based Predictive Model for Stock Prices", 12th IEEE inetrnational Symposium on Computational Intelligence and Informatics, CINTI 2011, 21-22 November, 2011, Budapest, Hungary.