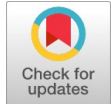


# Traffic Flow Prediction using Combination of Support Vector Machine and Rough Set

Minal Deshpande, Preeti Bajaj



**Abstract:** Congestion is the primary issue related to traffic flow. Avoiding congestion after getting into is not possible. So the only way is to make the informed decision by knowing the traffic situation in advance. This can be achieved with the help of traffic flow prediction. In the proposed work, short term traffic flow prediction is performed using support vector machine in combination with rough set. Traffic data used for analysis is collected from three adjacent intersections of Nagpur city and traffic flow is predicted at downstream junction. The work has attempted to study the effect of aggregation intervals and past samples on the prediction performance using MSE threshold variation. Rough set is used as a post processor to validate the prediction result. Accurate and timely prediction can provide reliability for optimized traffic control and guidance.

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

## I. INTRODUCTION

With the fast growth in the economic development and increased urbanization, transportation problems are becoming serious issue everywhere in the world. Infrastructure growth is limited because of the size, space constraints and also because of lack of planning, lack of technology and lack of integration of different services that falls under transportation. The only way to solve 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 as well as more coordinated. In ITS, traffic flow prediction is one of the basic and essential components 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 for predicting

the traffic flow value 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 found to be very difficult to predict the traffic flow accurately and that too 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 good accuracy. It has good generalization and fast convergence characteristics. Hence it is suitable tool for traffic flow prediction problem. Rough set 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 which are obtained from the original set. In the proposed work, Rough set is used to perform the validation of prediction result thereby used as a post processing tool. The rest paper is organized into different sections as follows. Section II covers literature review. Section III explains the data collection details. In section IV the description about the structure of SVM and Rough Set used for short term traffic flow prediction is included. Prediction experiment is performed and simulation results are discussed in section V. Conclusions are given in final section..

## II. 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. [2] proposed traffic flow prediction on multiple intervals using a non parametric K-nearest neighbour method. This method requires large amount of historical data to produce satisfactory results. Qiangwei Li et al. [3] proposed short term traffic volume prediction using time dependent model of support vector machine. Time varying characteristics of traffic flow are matched using this model. 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.

Manuscript published on 30 September 2019.

\*Correspondence Author(s)

**Minal Deshpande**, Research Scholar, G H Raisoni College of Engineering, Nagpur, India. Email:minalgade2002@gmail.com

**Preeti Bajaj**, Electronics Engineering, G H Raisoni College of Engineering, Nagpur, India. Email:preetib123@yahoo.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Rough set theory is a mathematical tool which deals with vagueness. The concept of rough set is originated assuming the fact that some information is always associated with every object that exists 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 [9] proposed traffic flow prediction where rough set and genetic algorithm were applied for the selection of the relevant forecasting variable. Zhenguang Zhou and Kun Huang [10] reported rough fuzzy neural network model employed for prediction of traffic flow. Pang Ming-bao and HE Guo-guang [11] suggested chaos recognition model using rough set and neural network where rough set theory was used to obtain reduced feature vector. Xinrong Liang et al. [12] proposed elman neural network based traffic flow prediction where original data is reduced using rough set theory.

After review of literature it was observed that use of support vector machine in prediction is explored but use of rough set along with this tool to improve its prediction performance was not touched upon which defines the direction for current work.

### III. DATA COLLECTION

The dataset collection is done by means of video recording. The cameras are installed at 4 corners of 3 adjacent intersections, one each at upstream section of three sides of intersection and one at the downstream section to collect the data from all the three locations at the same time. The data is recorded for 12 hours on working day starting from 8 am to 8 pm.

Data extraction from the videos is carried out by manual counting. Data flow and corresponding time are extracted as the aim is to predict traffic flow.



Fig. 1. Location map of study area

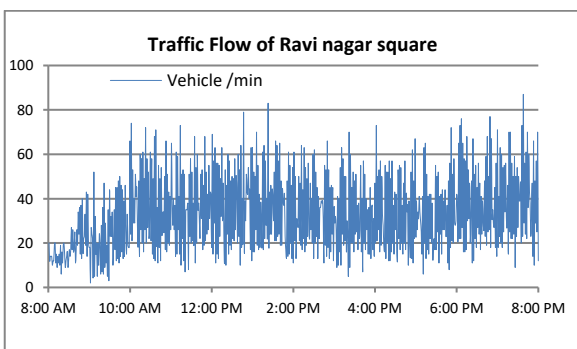


Fig. 2. Traffic flow pattern of study area

Nagpur is recorded as a tier-2 city of India. Fig. 2 shows the traffic pattern of the data for 1 day recorded for 12 hours calculated as 12(hours)\*60(minutes)=720 minutes. Data pre processing is done to arrange the available data in 1 min time interval. After pre processing, the data obtained is the aggregated data at different time interval of 5 min, 10min and 15 min and processed. Data set can be treated for temporal as well as spatial prediction as it is a multi-point data set with traffic recordings of same time slot.

### IV. METHODOLOGY USED

#### A. Support Vector Machine

Support vector machine is defined to be class of supervised learning algorithm. The theory is based on statistical learning. Both classification and regression problem can be solved with the help of support vector machine. As traffic flow prediction is entirely a nonlinear problem equivalent to function approximation, support vector machine can be used to solve it for regression. The main objective of LS-SVM is to apply the minimization of sum of squared errors to the objective function.

In Least-square SVM we have equality constraints.

$$\underset{w, \xi}{\text{Min}} \tau(w, \xi) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{k=1}^N e_k^2$$

Subject to:

$$Y_k [w^T \phi(x_k) + b] = 1 - e_k, \quad k = 1, \dots, N$$

Where  $W$  = normal vector to hyper plane

$b$  = bias term

$e$  = least square error

There are different types of kernel functions available for use. But the most often and widely used kernel function is Gaussian Radial Basis Function because it has excellent overall performance and very few parameters are required to be set [14]. It is mathematically expressed as

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

#### B. Rough Set Theory

The Rough Set theory is defined to be the 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, boundary region of a set is used to define vagueness.

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  $a \in A$ ,  $a:U \rightarrow V_a$  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

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

The lower and upper approximations are represented mathematically by the equations as follows

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

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

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

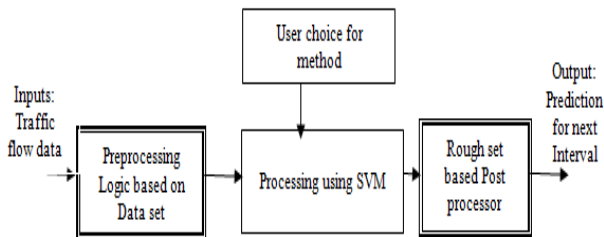


Fig. 3. Rough set based model for traffic flow prediction

Lower approximation is achieved by means of non-negativity of traffic flow count thereby ensuring that set under consideration does not remain empty. Upper approximation is achieved by considering maximum road capacity for flow. 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.

V. 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 to have same range of values 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 input sample.

1. RBF kernel function is used for the SVM model.
2. Regularization parameter  $\gamma$  is defined equal to 100.
3. Bandwidth for RBF kernel  $\sigma^2$  is defined to be equal to 2.

These values are finalized after trying different combinations and value which gives good results are selected for simulation.

Fig. 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.

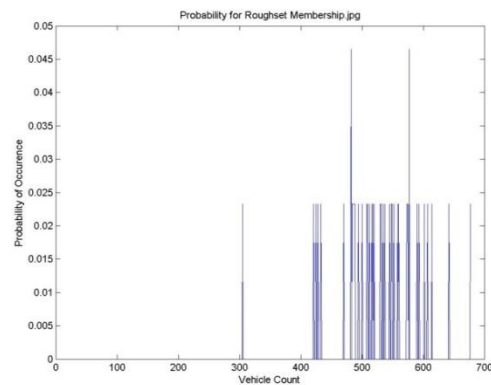


Fig. 4. Probability for Rough set membership

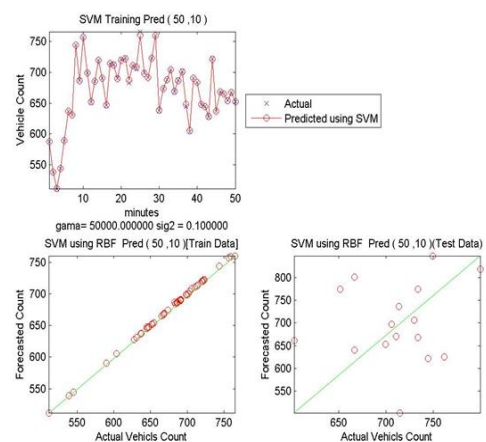


Fig. 5. Actual Vs Predicted vehicle count using SVM

Fig. 5 indicates the traffic flow prediction using SVM. It has been observed that the predicted traffic count matches exactly with the actual count. Simulation is performed for different test data. In this case we see that there are some forecasted values which deviate from the actual value.

Table- I.: Performance measures for different aggregation interval

Effect of Aggregation Interval with Prediction History =5					
Method	Aggregation Interval	MSE	RMSE	Error Mean	Error Std. deviation
SVM Test ( 5 , 1 ):	1	9.51E+01	9.75E+00	1.23E+00	9.70E+00



SVM Test ( 25 ,5 ):	5	7.88E+02	2.81E+01	6.84E+00	2.76E+01
SVM Test ( 50 ,10 ):	10	2.68E+03	5.18E+01	1.55E+00	5.33E+01
SVM Test ( 75 ,15 ):	15	2.51E+03	5.01E+01	2.44E+01	4.58E+01

SVM Test (4,1)	3.17E+02	1.78E+01	1.03E+00	1.78E+01
SVM RS (4,1)	1.11E+02	1.05E+01	1.21E+00	1.05E+01
SVM Test (40,10)	5.18E+02	2.28E+01	8.58E+00	2.12E+01
SVM RS (40,10)	5.10E+02	2.26E+01	8.49E+00	2.11E+01

Table I shows the effect of different aggregation interval on prediction performance where first number indicates history used and second number indicates aggregation interval in minute.

With an average of 33 vehicles per minute we have the following %MSE thresholds

Table- II . MSE thresholds validation for SVM

Method used	MSE Percentage threshold			
	2	5	8	10
SVM Test ( 5 ,1 ):	--	--	--	OK
SVM Test ( 25 ,5 ):	--	OK	OK	OK
SVM Test ( 50 ,10 ):	--	OK	OK	OK
SVM Test ( 75 ,15 ):	OK	OK	OK	OK

From table II it is observed that lower aggregation interval of 1 minute does not satisfy even 9% MSE threshold as compared to higher intervals of 15, 10 and 5. It performs well for higher % MSE threshold.

Table-III: Performance measures for different history samples

Effect of History with Aggregation Interval=5					
Method	History	MSE	RMSE	Error Mean	Error Std. deviation
SVM Test ( 10,5 ):	2	1.06E+03	3.26E+01	6.23E+00	3.25E+01
SVM Test ( 15,5 ):	3	7.81E+02	2.79E+01	4.32E+00	2.80E+01
SVM Test ( 20,5 ):	4	5.35E+02	2.31E+01	5.96E-01	2.35E+01
SVM Test ( 25,5 ):	5	7.88E+02	2.81E+01	6.84E+00	2.76E+01

Table-IV: MSE thresholds validation for history variation

Method Used	MSE Percentage threshold			
	7	8	10	20
SVM Test ( 10,5 ):	--	--	--	--
SVM Test ( 15,5 ):	--	OK	OK	OK
SVM Test ( 20,5 ):	OK	OK	OK	OK
SVM Test ( 25,5 ):	OK	OK	OK	OK

From the table IV it is observed that SVM with short history (=3) does not perform well for % MSE threshold below 8, also history of 2 does not work even with MSE threshold of 20%. This indicates shorter history is not advisable to be used with Nagpur dataset for SVM.

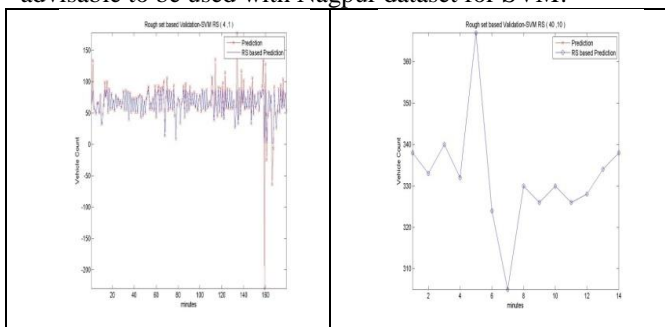


Fig. 6. Rough set based validation of prediction result

Fig. 6 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.

Table-V : Performance measures with and without rough set for prediction system

Method \ Measure	MSE	RMSE	Error Mean	Error STD
------------------	-----	------	------------	-----------

VI. CONCLUSION

Short term traffic flow prediction has achieved significant importance in the field of intelligent transportation system. 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 which is validated by using scatter plot. Rough set validates the prediction result by defining the lower approximation which contains the members that definitely belong to be the part of target set and thus results in considerable improvement in prediction performance. Prediction error is quite less where there are outliers. So use of rough set in traffic flow prediction application appears to be the promising field which needs to be researched further.

REFERENCES

- 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.
- 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.
- 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.
- 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.
- 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.
- S. Chen and P Wang, " Computational intelligence in economics and finance", Springer, ISBN: 3-540-44098-4, 2002, pp 15-22.
- L. Shen and H Loh, " Applying rough sets to market timing decisions", Decision Support Systems Journal , 2006, pp 65-72.
- 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.
- 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.
- 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.
- PANG Ming-bao and HE Guo-guang ," Chaos Rapid Recognition of Traffic Flow by using Rough Set Neural Network", International Symposiums on Information Processing, 2008.
- Xinrong Liang et al. ," Freeway Traffic Flow Model Based on Rough Sets and Elman Neural Network", IEEE 2009.
- <http://www.ceos.com.au>.
- 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 International Symposium on Computational Intelligence and Informatics, CINTI 2011, 21-22 November, 2011, Budapest, Hungary.
16. Minal Deshpande, Preeti Bajaj," Performance Improvement of Traffic Flow Prediction using Combination of Support Vector Machine and Rough Set", International Journal of Computer Applications(0975-8887),Volume 163 , Issue 2 , April 2017.

### AUTHORS PROFILE



**Minal Deshpande** is pursuing a PhD from the GH Raisoni College of Engineering, RTM Nagpur University, Nagpur, India. She has 15 years of teaching experience and currently working as an Assistant Professor in KCES's College of Engineering, Jalgaon. Her research interests include artificial intelligence, soft computing and rough set theory. She has published 6 papers in international conferences and journals. Her citations are 15 and h-index is 3.



**Preeti Bajaj** earned her Doctorate in Electronics Engineering in 2004. Having 25 years of experience, currently, she is the Director of the GHRCE, Nagpur. Her research interest includes intelligent transportation system, soft computing and hybrid intelligent systems. She is Fellow in the Institute of Engineers and IETE, a senior member in IEEE and an LM in ISTE and CSI. She is presently the Secretary in the IEEE India Council. She is also the Chair in IEEE Sub-section, Nagpur. Many students pursued PhD and masters under her supervision. She has authored and co-authored 100 plus reviewed publications and three book chapters with citations 800 plus and h index 14.