ROAD TRAFFIC PREDICTION USING K-MEANS CLUSTERING AND OPTIMIZED MULTILAYER PERCEPTRON USING GENETIC ALGORITHM

Jappreet kaur, S.k Gupta, Mohit Marwaha
Student, Associate Professor, Assistant Professor
Computer science &Engineering,
Beant college of engineering and technology, Gurdaspur, India.

ABSTRACT

With the growth of population, numbers of vehicles have also grown exponentially but the traffic management system has not developed at the same pace. As a result, traffic has become a major issue in major cities and countries of the world, traffic jams have increased leading to more accidents and pollution. This paper presents a novel machine learning strategy which uses K-means clustering and Multilayer Neural Networks for the purpose of predicting traffic in better way. Genetic Algorithm has been used to optimize the results of MLP. The proposed system can solve the traffic problem up to great extent without use of any man-power. It also highlights the better accuracy of the model than using Multilayer Neural Network or any of these alone and gives the future scope of our model.

Keywords: Neural Networks, K-means Clustering, Traffic-flow Prediction.

I. INTRODUCTION

An intelligent Transport system is needed for trouble-free traffic prediction on roads and for better management of traffic. Increase in number of vehicles has proven to result in more traffic leading to more socio-environment problems. Resources are getting wasted, lanes are getting blocked, pollution is increasing and most importantly our time is getting wasted. Thus, researchers have been looking into depth to deal with this problem. Traffic prediction at different locations varies according to the population growth, specific event, weather conditions, day of the week and so on. So 100% guaranteed results cannot be produced by model practically. However one can definitely aim for the best results. Machine Learning is one of the leading technologies which helps to make predictions in less time and with better results.

This paper has been organized as follows:

Sect. 2 presents some of the existing works; Sect. 3 presents the proposed mathematical modeling of road transportation system; simulation and analysis results are given in Sect. 4; and conclusions are drawn in Sect. 5.

2. LITERATURE SURVEY

Till date, many models have been proposed and implemented for predicting and managing the road-traffic data. Those models can be differentiated as parametric, non-parametric and artificial intelligence models. Using an approach which is somewhat similar to what we have proposed in the paper, Krzysztof Halawa [1] used single multilayer perceptron model on major intersections of Poland, even reducing the curse of dimensionality but failed to gain the best results on some points.

References [2] and [3] have used variants of ARIMA model which are SARIMA and H-ARIMA respectively. In [4] MA, ES and ARIMA have been integrated using NN which is again a technique for short term prediction of traffic.

In [5], it has been mentioned that there is need for short term traffic prediction due to sudden change in traffic after short interval of time. It is based on non-parametric regression algorithm i.e. k-means algorithm. These proved useful for short term traffic prediction when input data is limited. Models definitely held an advantage over simple ARIMA model in terms of accuracy but are again limited to input small data and not for large datasets.

[6] shows how traffic prediction can be improved using tweet semantics. Here tweet semantics from San Francisco Bay area of California have been used but again author discussed how it proved a limitation to heterogeneous data.

Bei Pan [7] made an advanced approach to include the video database to solve the purpose. He included traffic from rush-hour too for image processing and used H-Arima+ model. Accuracy of 91% has been achieved by this approach but again vision of cameras during night hinders the goals for better accuracy.

Approach adopted by Ari Wibisono[8], using FIMT-DD is also worth noticing which is a simulation on real world map data.

Now we will discuss the literature presented by authors in the field of machine learning. Anuja Nagare [9] gave us three variants of neural network models and explained them precisely for traffic flow prediction. JIAN-ME1 XIAO[10] used RBF neural networks to solve the purpose along with gradient descent and fuzzy clustering

Adel Abdennour[11] presented the paper using neural networks using video streams and very accurately showed how their model is better than the previous works conducted in similar fashion.

Similarly, CORINNE LEDOUX [12] and Kranti Kumar[13] used neural networks for traffic flow prediction using neural networks.

At last Jason Hall [14] came forth with the limitations of neural networks which encouraged us to add on some extra effort for the problem.

Keeping in mind the optimization of our input parameters [15] gave in depth review of collaboration of Genetic Algorithm (GA) with neural networks.

3. PROPOSED SYSTEM

The proposed system uses KNN approach to form clustering and then filtering is performed using layers of multi layer perception. Optimization is applied in the form of Genetic algorithm to optimize the number of nodes within the input layer along with weight. Convergence becomes better as the number of iterations increases. The proposed methodology first of all uses KNN to form clusters of related items from within traffic dataset. This is a primary step leading to formation of clusters that allows searching within a particular place hence saving time.

3.1 KNN

K nearest approach identifies the nearest neighbors from within the dataset and group them together to form a clusters. The dataset used contains name of the lanes, traffic count along with time interval. Grouping of items is performed on the basis of name of lane at distinct time interval. Complexity reduction procedure arranges the overall data in the form of relevant order so that clusters corresponding to desire items can be formed. This two phase process is critical in order to predict the recommendations. Algorithm corresponding to KNN is given as under

- Input the item and categories it as $(x_i, y_i) = Item_i\ Recieved\ from\ decision\ tree$
- Move to the next level and obtain next item as $(x_{i+1}, y_{i+1}) = Item_{i+1}$
- Find Euclidean distance between items.
  $euc_i = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}$
- $if\ euc_i \le cluster_i$
- $Cluster_i = Item_i$
- $end\ of\ if$

3.2 MULTI LAYER PERCEPTRON APPROACH

In this approach number of nodes comparable with the traffic dataset tuples is taken. These nodes construct a network. Number of nodes varies with the iterations of optimization approach used. The optimization approach collaborated with the proposed system genetic in nature. Every nodes have a weight assigned to it. These weights are adjusted with back propagation mechanism. The process continues until generation expires or prescribed tolerance is achieved. The pseudo code associated with the proposed approach is ~\ref{mlp}

- Input layer definition

$$Nodes_i = Dataset_{Traffic_{tupples_i}}$$

$$where\ i\ define\ number\ of\ tuples\ within\ the\ traffic\ dataset$$

- Hidden layer definition

Processing layer contains the neurons which are defined by normalization function

$$Normalization_i = \frac{Total_{Nodes}}{K}$$

Where, K is the parameter while value vary within the range of 2 to 4.

- Activation function definition

Activation function indicates the activation of weight function which is gradient descent in this case.

$$update_i = lerning_{rate} * gradient\_of_{parametes}$$

$$learning\ rate\ is\ constant\ for\ each\ iteration$$

- Backpropagation for weight adjustment

This is used to check prescribed tolerance. If it is not achieved then weight is adjusted by the random factor between [1-5].

If (Perscribed_tolerance)

Then list error rate

Else

Adjust weights $w_{i_j} = W_{i_j} + rand(1,5)$

This algorithm primarily used to obtain the least possible error in the prediction process.

3.3 GENETIC APPROACH FOR OPTIMIZATION

This approach is used to obtain optimized number of nodes and weights present within MLP. It takes number of tuples within the dataset as input and calls the MLP described in the second section. The iterations are repeated until optimal weights along with the number of nodes in the input layer are achieved. The pseudo code for the same is described in

- Initialization

Clusters data from the KNN is served as initial population.

- Fitness function

Fitness function involves minimization of errors obtained after MLP

If (Minimum_Error)

Convergence with prediction and error rate

Else

Move to the next phase

End of if

- Based on fitness function, selection of population
- Performing crossover

2 or 4 point crossover is performed and nomination is obtained

$$Normalization_i = \frac{Normalization_i}{K}$$

Normalization selects the parents for next phase

- Performing mutation

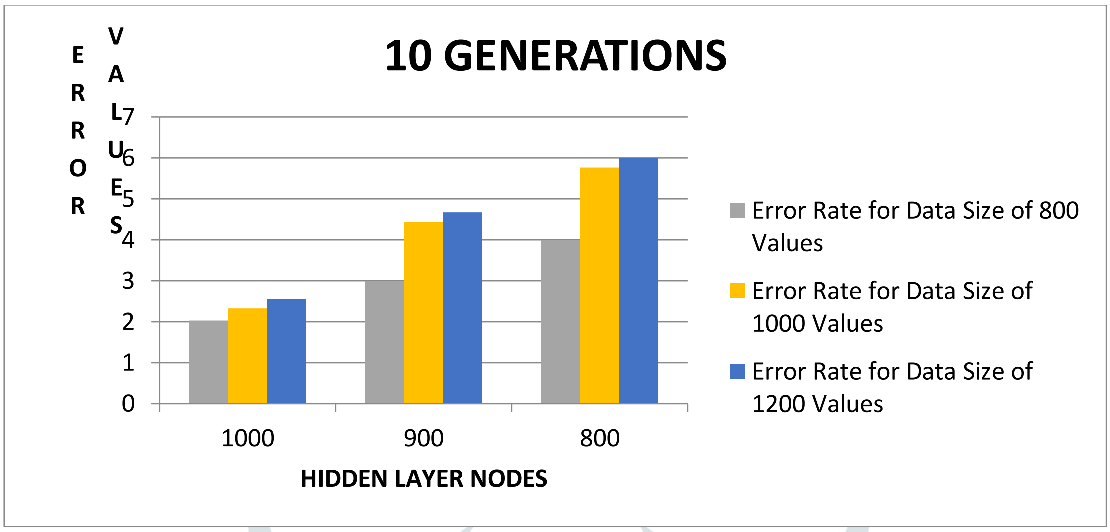
Mutation mutates the parents and new chromosomes are generated.

The result of this methodology is optimal predictions although time consumption could be high.

4. PERORMANCE ANALYSIS AND RESULTS:-

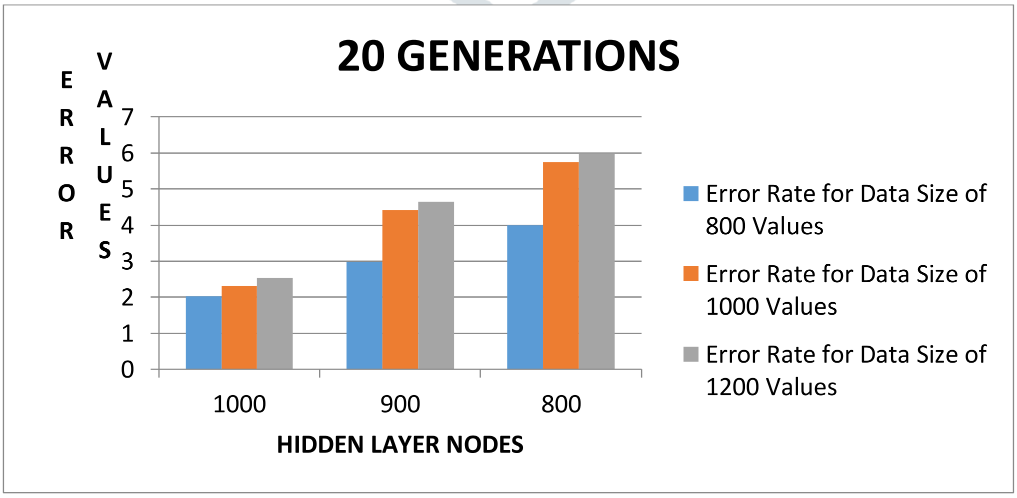
4.1 PERFORMANCE OF PROPOSED SYSTEM

As per the requirement of the system number of generations, hidden layer nodes and data size needed to be optimised. So these parameters were tuned and the following results were obtained.



4.1.1 Case 1- Number of Generations=10

In this case, hidden layer nodes can have three values i.e 1000, 900 and 800. Further, data size is also varied as 1000, 900 and 800 as per the need. The lowest error rate recorded in the experiments was found out to be when data size=800, hidden layer nodes=1000 and mutation probability=0.3. Next case will follow the increase in number of generations and we will observe the nature of results obtained



4.1.2 4.1.1 CASE 1- Number of Generations=20

In case 1, numbers of generations were increased to 20. Data size is being varied as 100,900 and 800 respectively. Optimised value for number of hidden layer neurons again came out to be 1000, error decreased was minor yet significant with a difference of 0.013. Thus, it was concluded that these are the best results which can be obtained by our system.Hence, this case is better than previous case and the best which can be obtained by our model.

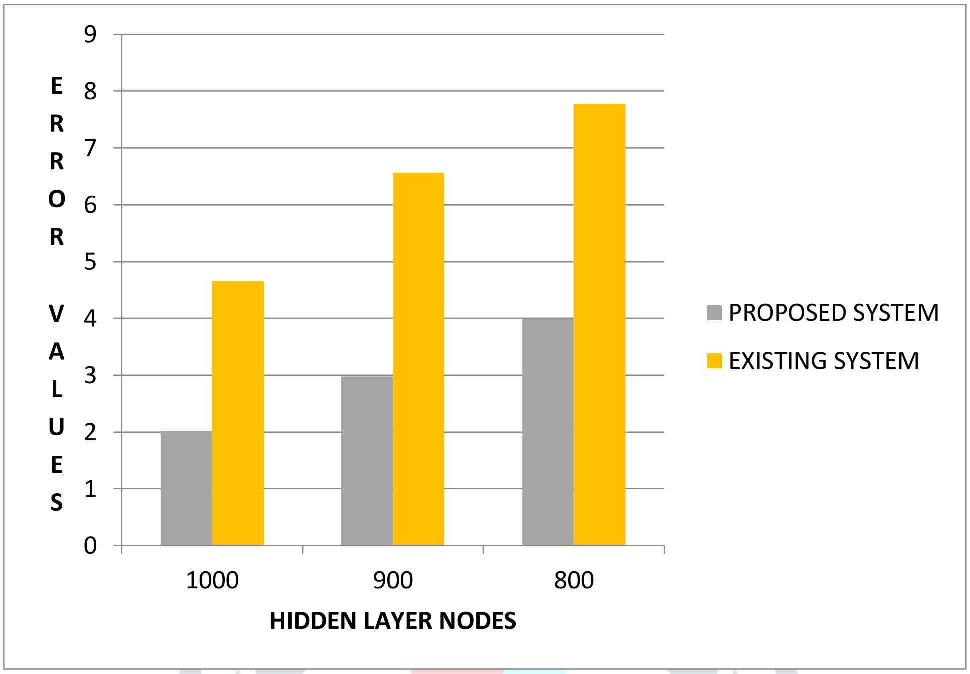


Fig 3: 4.1.2 COMPARISON OF EXISTING AND PROPOSED SYSTEM

Research is fruitful only if the result of proposed model outperforms the results of existing model. In the existing model, only multilayer perceptron model is used for predictions. In the Fig 3, we compared the best case of our proposed work with the existing model results. As observed from the graph, proposed model is better in every case when hidden layer node parameter varies. Hence, our model is valid and is better than its existing version in every way.

5. CONCLUSION AND FUTURE SCOPE

If we look into depth the present scenario of the condition of roads, machine learning is the only solution to help the world solve this problem. So, the goal of this paper is to design a simple and accurate system that is of utmost importance these days for traffic prediction. It uses a small data size of merely 1500 values and predicts accurate results which can be further improved for large datasets. In this method, road lanes are taken into consideration and traffic is predicted for given time interval. In the future, researchers can look into concepts such as weather, ongoing roadwork and traffic at intersections for better and more accurate results. Region identification such as plazas, hospital etc. can also be looked upon while calculating the results and then merged along with the proposed system.

6. REFERENCES

[1] Krzysztof Halawa, Marek Bazan, “Road traffic predictions across major city intersections using multilayer perceptrons and data from multiple intersections located in various places,” *Ophthalmology,* IET Intell. Transp. Syst., 2016, Vol. 10, Iss. 7, pp. 469–475.

[2] S. V. Kumar and L. Vanajakshi, “Short-term traffic flow prediction using seasonal ARIMA model with limited input data,” *Eur. Transp. Res. Rev.*, vol. 7, no. 3, pp. 1–9, 2015.

[3] B. Pan, U. Demiryurek, and C. Shahabi, “Utilizing Real-World Transportation Data for Accurate Traffic Prediction.”

[4] M. Tan, S. C. Wong, J. Xu, Z. Guan, and P. Zhang, “An Aggregation Approach to Short-Term Traffic Flow Prediction,” vol. 10, no. 1, pp. 60–69, 2009

[5] X. Pang, C. Wang, and G. Huang, “A Short-Term Traffic Flow Forecasting Method Based on a Three-Layer K-Nearest Neighbor Non-Parametric Regression Algorithm,” no. July, pp. 200–206, 2016.

[6] J. He, W. Shen, P. Divakaruni, L. Wynter, and R. Lawrence, “Improving Traffic Prediction with Tweet Semantics,” pp. 1387–1393.

[7] A. A. Shafie, M. H. Ali, F. Hafiz, and R. M. Ali, “Smart Video Surveillance System for Vehicle Detection and Traffic Flow Control,” vol. 6, no. 4, pp. 469–480, 2011.

[8] A. Wibisono, W. Jatmiko, H. A. Wisesa, B. Hardjono, and P. Mursanto, “Knowledge-Based Systems Traffic big data prediction and visualization using Fast Incremental Model Trees-Drift Detection (FIMT-DD),” Knowledge-Based Syst., vol. 93, pp. 33–46, 2016.

[9] Anuja Nagare , Shalini Bhatia ,"Traffic Flow Control using Neural Network", International Journal of Applied Information Systems (IJAIS), Volume 1– No.2, January 2012

[10] Jian-Me1 Xiao, Xi-Huai Wang,"Study On Trafific Flow Prediction Using Rbf Neural Network", 26-29 August 2004

[11] Adel Abdennour,"Evaluation of Neural Network Architectures for MPEG-4 Video Traffic Prediction", IEEE Transactions On Broadcasting, VOL. 52, NO. 2, JUNE 2006

[12] Corinne Ledoux," An Urban Traffic Flow Model Integrating Neural Networks", *TranspRes.- C,* Vol. 5, No. 5, pp. 287-300, 1997

[13] Kranti Kumara, Manoranjan Paridab & Vinod Kumar Katiyara,"Short term traffic flow prediction in heterogeneous condition using artificial neural network",2013

[14] Jason Hall and Philip Mars" The Limitations Of Artificial Neural Networks For Traffic Prediction"

[15] J. David Schaffer , Darrell Whitley , Larry J. Eshelman “Combinations of Genetic Algorithms and Neural Networks: A Survey of the State of the Art”,1992