Traffic Prediction System Using Machine Learning Algorithms

Nazirkar Reshma Ramchandra [1], Dr. C. Rajabhushanam [2]
{reshma174@gmail.com[1], rajabhushanamc.cse@bharathuniv.ac.in[2]}

Research Scholar, Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India.[1],
Professor, Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India.[2]

Abstract. Traffic congestion is defined as the state on transport which is characterized by slower speeds of vehicles this is also because of the bad condition of the roads, weather, concern zone, temperature, etc. This traffic flow prediction is mainly based on the real-time dataset which is collected with the help of various cameras and sensors. In recent day the deep learning concepts has dragged the attention for the detection of traffic flow predictions. In this paper, some of the common and familiar machine learning concepts like Deep Autoencoder (DAN), Deep Belief Network (DBN), and Random Forest (RF) are applied on the online dataset for the traffic flow predictions. The important attributes of weather, temperature, zone name, and day are used to predict the traffic flow of the particular zone. The performance of the proposed system can be evaluated by using accuracy, precision, and RMSE, and MSE value. Among the three methods, the DT technique produces a better result.

Keywords: Traffic Flow, Machine Learning, Prediction, Accuracy, Recall, Performance.

1 Introduction

This world is very busy today. Each and everyone in this world has something to do always. This leads to many difficulties in society. One of the main problems is traffic congestion. This is due to overpopulation. Because of traffic congestion, there are many economic problems over society, health problems, and a heavier impact on the environment. So this condition should be changed. Decreasing the level of traffic congestion is an important research area. Proper traffic prediction is one of the challenging tasks because the traffic data is dynamically and complex. The traffic of the zone depends on the capacity of the road, type of road users, time, weather, events, and traffic policies, time of the day, etc. This research work uses machine learning concepts to predict traffic flow.

Section two elaborates on the view of various authors regarding machine learning concepts used to predict the traffic flow. Section three explains the methods used in this prediction system. Section four deals with the experiment result. Finally, section five concludes with the proposed system.

I3CAC 2021, June 07-08, Chennai, India

DOI 10.4108/eai.7-6-2021.2308860

2 Literature Review

The main aim of the Smart transport system is to distribute new services to various modes of transportation and managing traffic. The flow of traffic forecasting is the key attribute used to managing traffic. Due to the growth of modern technology in real-time, various new techniques and types of equipment are used in the traffic prediction system. From the various new techniques, deep learning is one of the important concepts used to retrieve important characteristics effectively from the amount of raw data with the help of unseen layers. Traffic data with nonlinear features is one of the important reasons for producing less accuracy result in traffic prediction. Shiju George et al., 2020 proposes a new bioinspired technique with a fuzzy concept. Technological indicators issue the flow characteristics of an input. Unseen layers of the Deep learning framework continuously learn the characteristics and transmit to the next level layer. Membership degree is measured with the help of membership methods. Finding optimized weight value with the help of the Dolphin Echolocation concept to set the model for data with nonlinear features. Experiments were conducted on two various datasets and display the output for the new proposed deep learning-based framework. Produced results show that the key significance of the traffic jam prediction [5].

Vehicle identification is an important technique in the transport system. By identifying the vehicle, the number of automobiles is known and the presences of the vehicles on the path are important factors. High dimensional data can be used to denote the automobiles. Feature retrieval and classifying features are the important processes used to identify the automobiles. High dimension data takes more computation time during the feature extraction process. D. M. S. Arsa et al., 2017, proposes the DBN (Deep Belief Network) technique for reducing the dimension of the data to detect the vehicles. In this research, the authors try to identify motorbikes and cars. Here DBN technique is used to reduce the dimension of the data and the SVM concept is used to classify the data. The proposed method applied to the UIUC dataset and the outcome of the current technique is compared with the PCA method. The experiment outcomes show that DBN concept provides a better result than traditional PCA method in the identification of automobiles [6]

In the transport system traffic flow forecasting is a major issue. Various existing techniques produce unsuccessful output due to various reasons like thin framework, engineering manually, and learning separately. W. Huang et al., 2014 proposes a new deep framework that contains two basic parts. At the base part contains DBN and the top portion contains the regression layer. DBN technique is applied for the purposed of unsupervised learning and it learns efficient attributes for traffic forecasting an unsupervised manner. It produces a better result for many places like audio and image classification. The regression layer is used for supervised forecasting. The experiment outcomes describe that the proposed method increases the performance of existing systems. The positive outcomes say that multitask regression and deep learning are important technologies in the transport system research [7].

Sheik Mohammed Ali et al., 2012 presented the best algorithm to categorize automobiles recognized using many inductive systems. RF (Random Forest) algorithm can be used for classification purposes. This proposed system is used to categorize the vehicles and count them based on the traffic situation. The output of the proposed method compared with other techniques depends on signature and threshold data. The outcome from the proposed system shows improved accuracy compared to signature and threshold-based techniques [8].

According to Zhenbo Lu et al., 2019 recognizing and identifying the mode of travel and passenger travel pattern are the major issues in the transportation system. MSD (Mobile-

phone Signaling Data) technique have various merits like wide area coverage and less acquisition amount, reliability and stability of data, and better performance in real-time. Here the authors develop a travel mode identification system using MSD integrated with travel data. GIS data and navigation type data. The proposed system applied to the Kunshan data set in China and the model produces better accuracy 90%. This accuracy level is suitable for all kinds of transport modes except for buses [9].

3 Proposed Methodology

Traffic is one of the major issues in an urban area. Due to the unemployment in the village, the people move to urban areas. The population rate is increased day by day in cities. Various techniques are followed to control the traffic flow in major areas. Machine learning concepts are playing a major role in all domains. Traffic flow is also predicted with the help of machine learning techniques. This research work uses three kinds of machine learning approaches like DAN, DBN, and RF.

DAN

DAN is one of the neural network technique using feed-forward concepts. It accepts a value x and converted to an unseen symbol or space (h). Autoencoders reduce the input into a low dimension and produce the output data from the given input representation. It contains three important elements. They are encoder, code, and decoder. The encoder process is represented by using the following equation (1).

$$h = \sigma(wX + b) - - - - - - - - - - - - - - - - - - (1)$$

From the above equation, W represents the unseen weight value and b represents the components and σ stands for the sigmoid method. The decoder element changes the resulting unseen representation into feature space y with the help of the following equation (2).

$$y = \sigma(W'\, h + \, b') - (2)$$

Here W represents weight and b states the weight value. The succeeding layer is an unseen layer of the preceding one. Each layer is trained with the help of the gradient descent method with an optimization technique, which is J (reconstruction error) of every layer. This concept is represented by using the equation (3).

$$argmin_{\ w1\ \ b1w2b2\ \ } [j] = argmin_{\ w1\ \ b1w2b2\ \ } \left[\frac{\sum_{i=1}^{m}\|x_i - x_i'\| + J_{wd} + J_{sp}}{2}\right] - - - - - (3)$$

From equation (3) J denotes the squared reconstruction error of the autoencoder layer $x_i \ and \ x_i'$ denotes the ith data of input.

DBN (Deep Belief Network)

DBN is one of the most familiar and efficient techniques among the entire deep learning techniques. It is used to identify, group, and produce pictures, video streams, and motion data. DBN is a mixture of RBMs (Restricted Boltzmann Machines). RBM contains a two-level layer graphical system that can be made up of two layers are hidden and visible layers represented by the following figure 1.

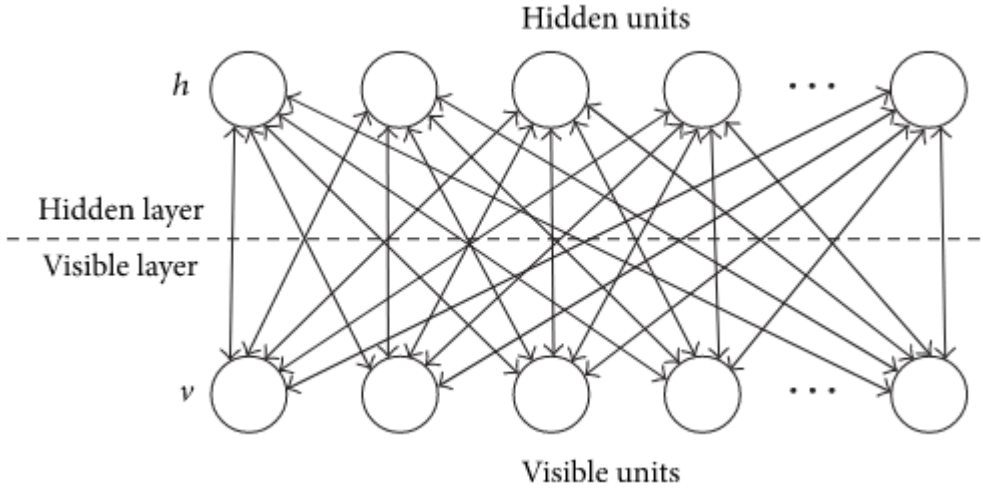


Figure 1 Restricted Boltzmann Machine [11]

Every element of the layer is linked with all other elements by using edge values. The elements available in a similar layer are disjoined with other elements. DBN approach is like a stack of different RBMs. All units' values are random identifiers. Noticeable and unseen elements are obeyed Gaussian and Bernoulli distribution. It is described as follows:

$$P(v_i|h) = N\left(b_i + \sum_{j=1}^{j} w_{i,j}\, h_j, 1\right), P\left(h_j = 1\middle|v\right) = \text{sigm}\left(a_j + \sum_{i=1}^{I} w_{i,j}\, v_i\right) - - - -(4)$$

From the equation (4) $N\left(b_i + \sum_{j=1}^{j} w_{i,j}\, h_j, 1\right)$ represents Gaussian distribution, mean value is $b_i + \sum_{j=1}^{j} w_{i,j}\, h_j$ and variance value is 1. The equation $\text{sigm}(z) = \exp(z)/(1 + \exp(z))$ is used to describes the sigmoid method. The identifiers I and J used to denote noticeable and unseen units. The values of b_i and a_j represent the bias value of the noticeable and unseen elements. The value of $w_{i,j}$ denotes the association between noticeable element v_j and unseen element h_j. Probability distribution method over noticeable and unseen elements can be represented by

$$p(v, h) = \frac{\exp\left(-E(v, h)\right)}{\sum_{v,h} \exp\left(-E(v, h)\right)} - - - - - \quad (5)$$

From the above equation $E(v, h)$ known as energy method described as

$$E(v, h) = -\frac{1}{2}\sum_{i=1}^{I} (b_i - v_i)^{\ 2} - \sum_{j=1}^{J} a_j h_j - \sum_{i=1}^{I}\sum_{j=1}^{J} w_{i,j}\, v_i h_j - - - - - (6)$$

A greedy method is used to train DBN.

RF (RANDOM FOREST)

According to Juan Cheng et al., 2019, RF is a combined learning technique based on DT. It is a high precision technique in machine learning it avoids the issues of single forecasting or classification system. RF is an ensemble technique that is used to manage continuous reaction and label values. The main benefit of the RF method is it does not affect the issue like

overfitting. Here the value of the error rate is less when more CARTs are further added with the existing one. The major idea used to create RF is to develop a major collection of weak models, which will generate the final powerful model. RF is the huge collection of un-pruned DTs (Decision Tree) with predictors. Un- pruned DTs are trained with the help of a bootstrap model from the given dataset. CART is a machine learning approach used to create RF. CART follows the greedy technique partition of the data recursively by using a top-down partitioning approach that segregates the attribute into a group of disjoint sections. RF consists of many DTs. The following equation (7) denotes the meaning of RF.

$$\{h(x, \theta_t), t = 1,2 \ldots . T\} - - - - - - - - - - - - (7)$$

Here $h(x, \theta_t)$ is represented tree classifier and $\{\theta_t\}$ denotes random vectors, x is known as an independent identifier, θ_t denotes random identifier, T indicates the number of DTs.

The average value of every DT as a prediction output, represented in equation (8)

$$\bar{h} = \frac{1}{T}\sum_{i=1}^{T}\{h(x, \theta_t)\} - - - - - - - - - - - - (8)$$

From equation (8) $h(x, \theta_t)$ is the result based on the values xand θ.

Bagging and subspace are introduced to increase the accuracy value and to avoid the overfitting in DT.

Performance Evaluation

The performance of the proposed system is evaluated for the various metrics like accuracy, precision, root mean squared error (RMSE), and MSE.

The metric RMSE is measured by using the following equation (9).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y_i})^2} \text{ ---------- (9)}$$

From the above equation, y_i is the real value collected during the time of travel, $\bar{y_i}$ denotes the predicted time when the traveling, and n indicates the total number of observations.

The MSE (Mean Square error) can be computed by using the equation(10)

$$\boldsymbol{MSE = \frac{1}{N}\sum(Y - \hat{Y})^2 - - - - - - - - (10)}$$

From the equation $(Y - \hat{Y})^2$ represents the square difference between normal value and forecasted value.

The accuracy value of the proposed system calculated by using the formula (11)

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} - - - - - - - - (11)$$

The recall metrics can be measured by using the equation (12)

$$Recall = \frac{TP}{TP + FN} - - - - - - - - - - - - (12)$$

TP denotes True Positive value. It means the observation value is positive and the predicted value also positive. FN indicates False Negative. FN represents the observation data is a positive value and the predicted value is negative. TN means True Negative. It means the observed value is negative and the predicted value also negative. FP denotes False Positive. It represents the observed data is a negative value but the predicted value is positive.

4 Result And Discussion

Traffic congestion is considered as one of the most important issues faced in urban areas. Due to the growth of the computing technologies various advanced concepts are used to predict the traffic congestion and traffic flow. This research uses three machine learning concepts such as DAN, DBN, and RF are used to predict the traffic flow. The traffic data set was collected from the online and initial done the preprocessing steps. After preprocessing apply machine learning concepts to predict the traffic flow. The performance of the proposed system is evaluated by using the metrics accuracy, precision, RMSE, and MSE.

Table 1 demonstrates the value of accuracy, precision, RMSE, and MSE.

Metrics/Method	ADN	DBN	RF
Accuracy	88.6%	90.43	92.6
MSE	3340.25	1709.71	346.35
RMSE	57.794895968	41.3486396	18.6104809

Figure 2 shows the performance comparison of various metrics like accuracy, recall, RMSE and MSE.

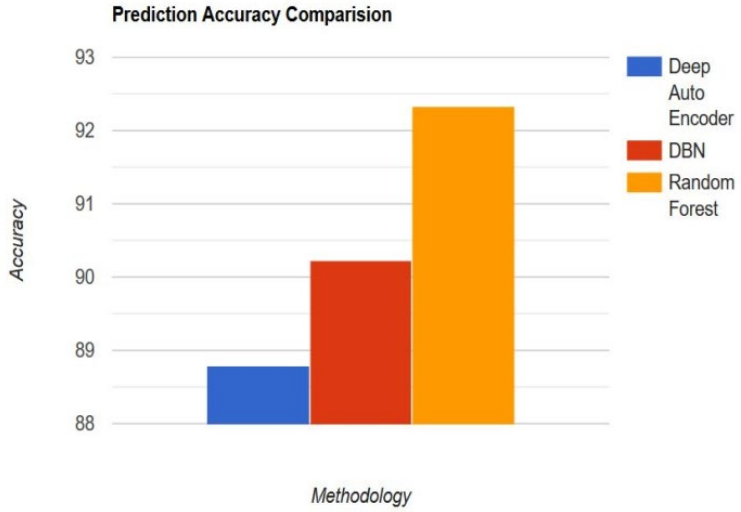
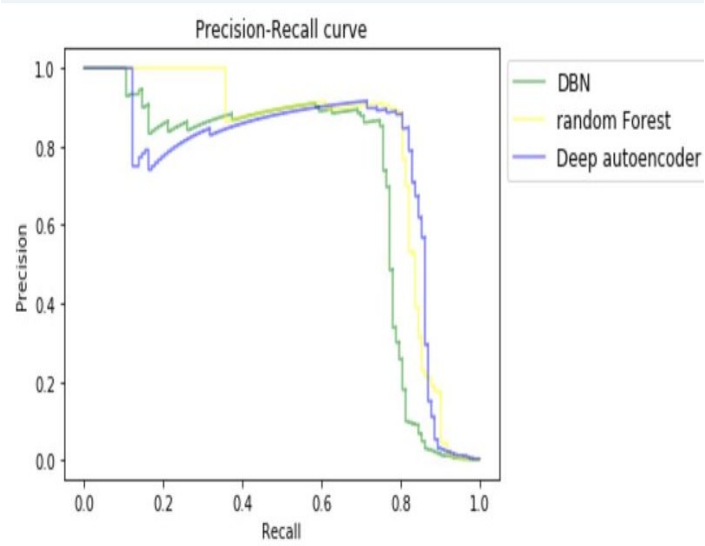
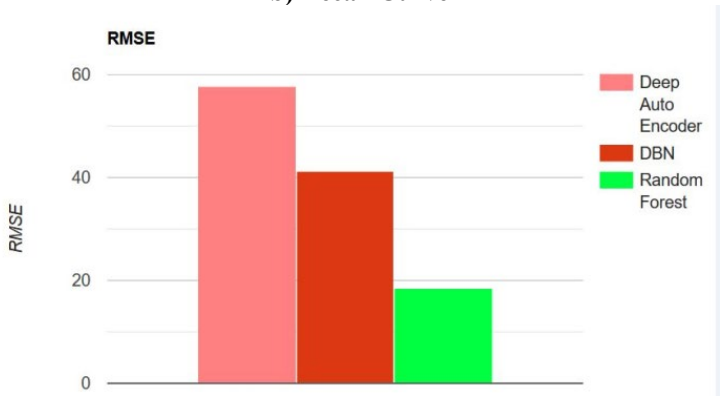


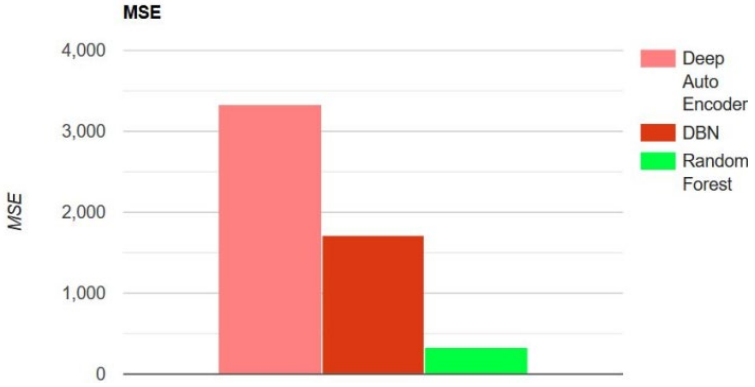
Figure 2 a) Accuracy Comparison



2b) Recall Curve



2c) RMSE value Comparison



2 d) MSE Value Comparison

Conclusion

Due to the huge population traffic is one of the important problems in most of cities. Computing techniques are used to predict the output based on given data in various domains. In smart transport system also uses machine learning concepts to predict the traffic flow. In this research work DAN, DBN, and RF approaches are used to measure the traffic flow. Here the data collected from an online website and are initially preprocessed. During this process, the unwanted and noisy data are removed from the original data. Then machine learning concepts are applied to the preprocessed data to predict the flow. The experimental results show that the DF approach produces a better result than the other two methods. This proposed work is implemented using Python programming

References

[1] Sen Zhang, Yong Yao , Jie Hu , Yong Zhao , Shaobo Li &Jianjun Hu (2019), "Deep Autoencoder Neural Networks for Short-Term Traffic Congestion Prediction of Transportation Networks" , Sensors, Vol. 19, 2229, pp. 1-19.

[2] A. Dairi, F. Harrou, Y. Sun & M. Senouci (2018), "Obstacle Detection for Intelligent Transportation Systems Using Deep Stacked Autoencoder and k -Nearest Neighbor Scheme," IEEE Sensors Journal, vol. 18, no. 12, pp. 5122-5132.

[3] S. Uma Devi & S. Nirmala SugirthaRajini (2019), " Detection of Traffic Violation Crime Using Data Mining Algorithms", Jour of Adv Research in Dynamical & Control Systems, Vol. 11, No.9, pp. 982-987

[4] Arya KetabchiHaghighat, VarshaRavichandra-Mouli, Pranamesh Chakraborty, YasamanEsfandiari, Saeed Arabi& Anuj Sharma (2020), Applications of Deep Learning in Intelligent Transportation Systems, Journal of Big Data Analytics in Transportation.

[5] Shiju George &Ajit Kumar Santra (2020), "Fuzzy Inspired Deep Belief Network for the Traffic Flow Prediction in Intelligent Transportation System Using Flow Strength Indicators", Big Data, Vol. No. 4.

[6] D. M. S. Arsa, G. Jati, M. Soleh& W. Jatmiko(2017), "Vehicle Detection using Dimensionality Reduction based on Deep Belief Network for Intelligent Transportation System," 6th IEEE International Conference on Advanced Logistics and Transport (ICALT), pp. 199-204.

[7] W. Huang, G. Song, H. Hong & K. Xie (2014), "Deep Architecture for Traffic Flow Prediction: Deep Belief Networks With Multitask Learning," IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 5, pp. 2191-2201.

[8] Sheik Mohammed Ali, Niranjan Joshi, Boby George & L. Vanajakshi (2012), "Application of random forest algorithm to classify vehicles detected by a multiple inductive loop system", 15th International IEEE Conference on Intelligent Transportation Systems

[9] Zhenbo Lu , Zhen Long, Jingxin Xia &Chengchuan An (2019), A Random Forest Model for Travel Mode Identification Based on Mobile Phone Signaling Data, Sustainability, 11, 5950, pp. 1-21.

[10] Juan Cheng , Gen Li &Xianhua Chen (2019), "Developing a Travel Time Estimation Method of Freeway Based on Floating Car Using Random Forests", Hindawi Journal of Advanced Transportation, pp. 1-13.

[11] LaisenNie,XiaojieWang,Liangtian Wan, Shui Yu, Houbing Song and Dingde Jiang(2018), "Network Traffic Prediction Based on Deep Belief Network and Spatiotemporal Compressive Sensing in Wireless Mesh Backbone Networks", Crowdsourcing for Mobile Networks and IoT.