An improved Long Short-Term Memory Networks with Takagi-Sugeno Fuzzy for Traffic Speed Prediction Considering Abnormal Traffic Situation

Journal:	*Transportmetrica B: Transport Dynamics*
Manuscript ID	TTRB-2019-0049
Manuscript Type:	Full Paper
Date Submitted by the Author:	30-Mar-2019
Complete List of Authors:	George, Shiju; VIT University, Santra , Ajit Kumar; VIT University
Keywords:	Traffic speed prediction, Fuzzy Optimized Long Short-Term Memory, Agglomerative Hierarchical K-mean, Gaussian bell-shaped membership function,, Whale Optimization Algorithm

An improved Long Short-Term Memory Networks with Takagi-Sugeno Fuzzy for Traffic Speed Prediction Considering Abnormal Traffic Situation

Shiju George[1], A K Santra[2]

[1]Research Scholar, School of Information Technology and Engineering, VIT University Vellore, Tamil Nadu, India

[2]Senior Professor, School of Information Technology and Engineering, VIT University Vellore, Tamil Nadu, India

Abstract

Traffic speed prediction is an emerging paradigm for achieving a better transportation system in smart cities and to improve the heavy traffic management in the Intelligent Transportation System (ITS). The accurate traffic speed prediction is affected by many factors such as abnormal traffic condition, traffic incident and lane closure, and traffic congestion. To overcome these problems, we propose a new method named Fuzzy Optimized Long Short-Term Memory (FOLSTM) Neural Network for long term traffic speed prediction. FOLSTM technique is the combination of Machine Learning (ML) and Computational Intelligence (CI), capable of predicting the speed for macroscopic traffic key parameters. At first, a hybrid unsupervised learning method known as Agglomerated Hierarchical K-means (AHK) Clustering is used to divide the input samples into group of clusters. Second, the Gaussian bell-shaped fuzzy membership function is designed for calculating the degree of membership (High, Low, and Medium) for clusters based on parameters by using Takagi-Sugeno fuzzy rules. Finally, the whale optimization algorithm (WOA) is used in LSTM to optimize the parameters in fuzzy rules and calculate the weight value. FOLSTM evaluate the accurate traffic speed from the abnormal traffic data to overcome the nonlinearity in traffic prediction. The predicted result shows that the proposed method is better as compared with existing methods for traffic prediction in terms of Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE) metrics.

Keywords: - Traffic speed prediction, Fuzzy Optimized Long Short-Term Memory (FOLSTM), Agglomerative Hierarchical K-mean Clustering (AHK), Gaussian bell-shaped membership function, Whale Optimization Algorithm (WOA).

1. Introduction

The accurate Traffic Speed Prediction (TSP) plays a vital role in the Intelligent Transport System (ITS) to improve the efficiency of traffic management [1]. TSP is a complex process due to the abnormal traffic situations (road accidents and traffic jam), applied areas (structure of the area and road curves), and traffic congestion. Furthermore, the Road Side Unit (RSU), sensors, inductive loops, and Global positioning system (GPS) are the detectors used to collect the traffic features (average speed, volume, and occupancy) for multi-lane simultaneously. Many prediction techniques were developed to predict the Traffic Speed (TS) in various environments [2-3].

In last few years, the researchers focus towards CI algorithms (Fuzzy logic [4], and Evolutionary algorithms [5-6]) and ML algorithms (Support Vector Machine (SVM) [7], Artificial Neural Network (ANN) [8], K- nearest neighbour (KNN) [9], and clustering algorithms [10]) are proposed for short term and long term TSP. Li *et al*. [11] have introduced the Niche Immune Genetic Algorithm-Support Vector Machine (NIGA-SVM) to overcome the correlation of velocity with its factors under various conditions based on road traffic data. It applied traffic highways for TSP. This system would increase the accuracy of TSP. The major drawback is that, it cannot face long-time forecasting. Hu *et al*. [12] proposed hybrid Particle Swarm Optimization-Support Vector Regression Model (PSO-SVR) to avoid the time series prediction problem in real-time. The drawback of PSO-SVR method is that it regulate only the direction of short-term TSP. Lu *et al.* [13] introduced GA method to track the most optimal Membership Functions (MF) for fuzzy rules based on traffic guidelines. It was suitable algorithm for TSP, but the major drawback of GA is that it cannot solve the optimization problem. Yao *et al.* [7] introduced an SVM model for predicting the speed of short-term traffic, the performance of this model is related to the spatial and temporal model. The benefit of this SVM forecast model is analysed by driving traffic, control and route analysis using GPS taxi data. The main problem with SVM forecast method is that it could only provide GPS data from taxis. But, this data does not represent the entire population in the traffic. Kang *et al.* [14] introduced Long Short-Term Memory Neural Network (LSTM NN) for short-term TSP by capturing nonlinear dynamic in an effective manner. The advantage of LSTM NN is that it avoids the Back Propagation (BP) and time series problem. Here the data were collected using a microwave detector, and it could not consider the stopped vehicles for precise TSP. Jiang *et al.* [15] have introduced the Hidden Markov Models (HMMs) for long and short-term TSP. In this method, the average traffic speed is predicted form the motorway traffic by using neural model. It individually perform the forward-backward algorithm for TSP. But it could never capture high time correlation for TSP. Tang *et al.* [16] proposed the Evolving -Fuzzy Neuro Neural Network (EFNN) for TSP with less predictable errors and a slow error propagation. The result of the EFNN forecast was better than the traditional model, but multi-step head processing is a complex process for accurate TSP in long term traffic networks. Lin *et al.* [17] introduced a Topic Enhanced Gaussian Process Aggregation Model (TEGPAM) to overcome the problem faced during multi-source data fusion. Their work takes the advantage of fusing cross-domain data (tweet sensor and trajectory sensor data) along with the conventional sensor data to address the insufficient speed data during an emergency like accidents. Even though they fused the multi-source data for emergency situations, it's a complex process with Natural Language Processing (NLP) to check the traffic prediction in large traffic networks.

According to the discussed literature review, the performance of the above algorithms are not perform well for TSP. To overcome the existing drawbacks, this paper proposes a hybrid combination of CI with ML algorithms to predict the best TSP result from the abnormal traffic situation. The ML techniques is designed to classify the inputs to the number of clusters (based on traffic features), to reduce the overlapping problems and avoid the cluster centroid problems. The CI techniques are more flexible for evaluating accurate traffic speed. The performance of proposed hybrid algorithm was better than the single CI and ML approaches in TSP. We also compare our method to some traditional approaches such as Long Short Term Memory (LSTM) [14], Back Propagation Neural Network (BPNN) [16], Evolving Fuzzy Neural Network (EFNN) [18], Auto Regressive Moving Average (ARIMA) [19], and Artificial Neural Network (ANN) [20] methods to validate our approach.

The significant contributions and of this paper is given below: -

- We propose a new AHK-FOLSTM model for TSP to provide better performance for TSP in long time dependencies.
- The AHK clustering is used to form a group of cluster for accurate 80% of TSP. The main objective of AHK method to avoid the centroid problems in k-mean clustering and thereby reduces the overlapping problems.
- Takagi-Sugeno Fuzzy Logic Controller (FLC) is the first layer of FOLSTM, designed to measure the degree of membership for each cluster and detect nonlinear interactions between the TSP variables by considering the Gaussian bell shape MF. The LSTM NN can able to capture the nonlinearity and randomness of the TSP. LSTM can overcome the problems of forward propagation error through the cell memory block.
- WOA is applied in LSTM to calculate and update the weight values, which avoid the TSP link problems and increase the performance speed in the LSTM. The FOLSTM Network

has some periodic features that can give low error by using BP in term of long term forecasting.

- We extract the periodic features (volume, and occupancy and average speed) from the TS data to predict an accurate traffic speed for long-term and the performance is evaluated based on the metrics such as MAPE, MAE, and RMSE.

The remaining section of the manuscript as follows: - Section 2 present the model description of proposed traffic prediction. In section 3 experimental result and analysis of the proposed method using FOLSTM and its corresponding dataset descriptions are explained, finally the conclusion of the work are given in section 4.

2. Proposed Traffic Speed Prediction Model

In this section, we first describe the basic concept of our proposed work by using AHK-FOLSTM method to forecast the traffic speed. Here, we combine both CI method (Fuzzy, WOA) process with ML method (LSTM) to overcome the drawback of the existing methods. First, the hybrid AHK clustering method is used to form different clusters by using intermediate clusters of AHC. Second, the Gaussian bell-shaped MF is designed to estimate the degree of membership for all the cluster based on the parameters of MF (high, medium and low) and the Takagi-Sugeno fuzzy rule generates the linear functions in the FLC. Finally, the WOA algorithm improves the parameters of the Takagi-Sugeno fuzzy rule and solve the optimization problems of LSTM. The LSTM can avoid the TSP link problems and predict the accurate traffic speed values. The architecture of AHK-FOLSTM method of traffic speed prediction is shown in figure 1.

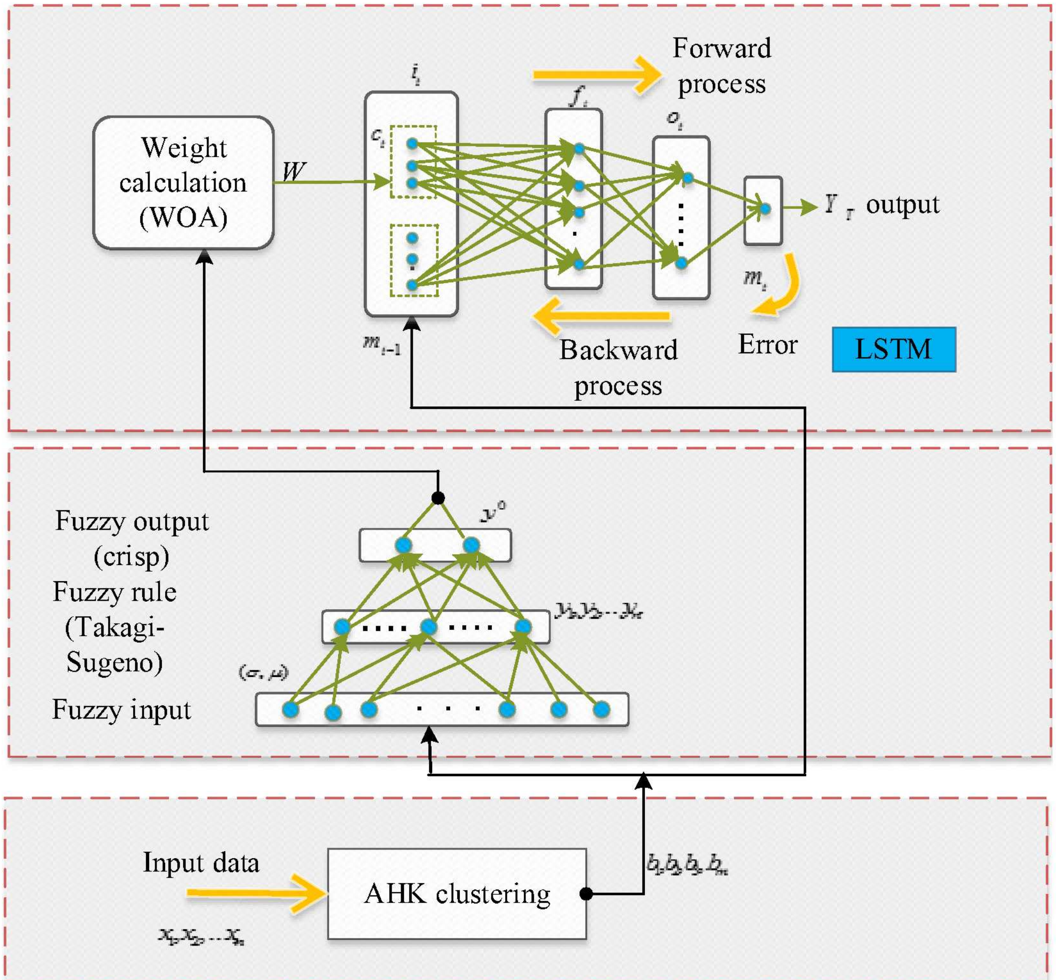


Figure 1: Architecture of Agglomerated Hierarchical K-means- Fuzzy Optimized Long Short - Term Memory (AHK-FOLSTM)

This FOLSTM method is mainly used for the long term TSP by capturing nonlinear dynamic in an effective manner. It reduces the drawback of forward propagation error issues and provides the long temporal result. Moreover, it can automatically determine the time lags in the traffic system, optimizes the time lag and enhances the accuracy of TSP.

2.1 Clustering Based on AHK Method

The AHK clustering is an unsupervised learning process. This hybrid clustering approach is a powerful model for hidden knowledge that can cluster the historical data to forecast the traffic speed. The aim of this hybrid method is to form the clusters based on the traffic environment key attributes without any overlapping problem. This model will reveal the similarity of unlabeled data units. Based on Euclidean distance (ED), the similar input values are merged into group of clusters. The ED is calculated by equation (1).

$$\text{ED} = \sqrt{(x_i - c_i)^2 + (x_j - c_j)^2} \tag{1}$$

Where x_{ij} represent the observed value of the variables, c_{ij} represent the centroid value of cluster for each variable. Let x_{ij} (i = 1, 2..., m and j=1, 2..., n), $x = \{x_1, x_2, ...x_n\}$ is the attributes, in which m and n denotes the rows and columns of the data. The centroid value of this cluster is created based on minimum value of input variables.

After calculating the distance, the clusters are formed based on average linkage and complete linkage clustering techniques, which avoids the overlapping problem in conventional algorithms.

The Agglomerative Hierarchical Clustering (AHC) method has certain process. First, the A_i , A_j considered as a separate cluster $\{A_{i1}\}, \{A_{i2}\}, ...\{A_{im}\}$ and $\{A_{j1}\}, \{A_{j2}\}, ...\{A_{jn}\}$. Then it analyze the distance between A_j and A_i cluster data point. The distance matrices of the clusters is considered by using D = $(d_{ij})_{m \times n}$, where $d_{ij} = ED(A_i, A_j)$. Where A_i, A_j represent the intermediate cluster. All the intermediate clusters $A_1, A_2, A_3, ..A_l$ are combined into single cluster. Finally, the K mean clustering algorithm divide the intermediate clusters into certain number of clusters and the k denote the number of cluster. The K mean clustering is to order the intermediate clusters from the AHC and is represented as:

$$A_k = \{A_1, A_2, ..., A_l\}, \tag{2}$$

Where k =1, 2….l, l indicate the number of intermediate clusters. Then the distance of the each cluster is evaluated by using the following equation:

$$d(A_k, c_k) = \sum_{k}^{l} |A_k - c_k| \tag{3}$$

Where, A_k denote the input variables, c_k is a center of the cluster and $|.|$ denotes the general ED. The objective function to form the final clusters is defined as:

$$b = \min\{d(A_k, c_k)\} \tag{4}$$

Where the b represent the cluster conditions. The AHK algorithm process flow is described as flowchart in Figure 2.

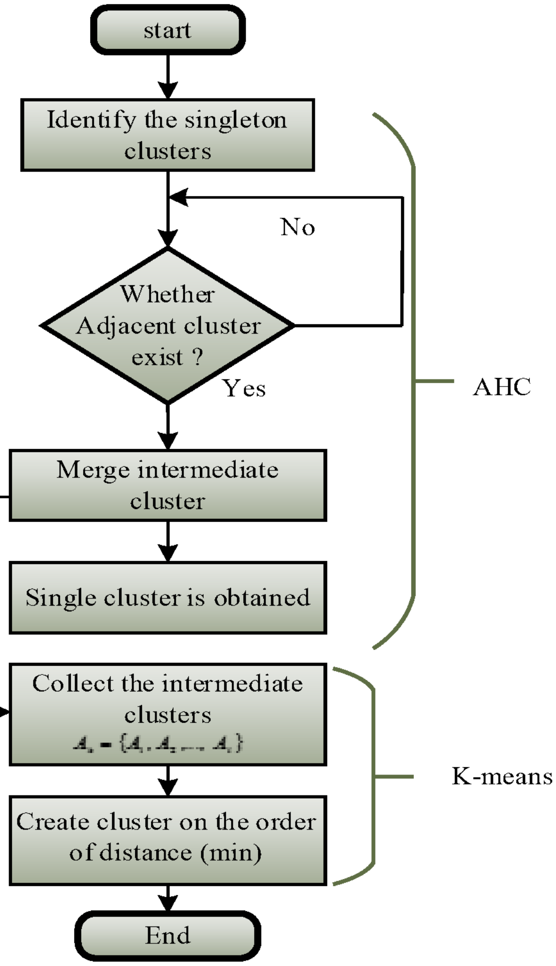


Figure 2: Flowchart of AHK clustering algorithm.

2.1 FOLSTM Method

2.2.1 Takagi-Sugeno fuzzy inference system

FLC is used to perform efficient and effective traffic prediction. This is the first layer of the FOLSTM. The operation of the fuzzy model is based on MF, it represents the elements of a fuzzy set. Here, the Gaussian bell shapes MFs is used to calculate the membership degree of each clusters. The bell shape MF perform based on three parameters: high, medium and low. This MF is more advanced than another type of MFs functions, the degree of membership is measured using equation (5).

$$mf(k,\sigma,\mu) = e^{\frac{-(d-\mu)^2}{2\sigma^2}} \tag{5}$$

Where *mf* is a membership function, μ denotes the mean value of clusters on *d* dimension, and σ denotes the standard deviation (SD) values of each cluster. The function of Gaussian bell shape MF depend on two parameters μ and σ. If the parameter σ is negative, the MF shape becomes an upside-down bell. The parameter μ is placed at the center of the curve. The Gaussian bell shape MFs is more advance than other generalized MF. Although the bell shape and Gaussian MFs can achieve evenness, they can reach the symmetric MF.

The Mean and SD of fuzzy is calculated by using equation (6) and (7).

$$\text{Mean } (\mu) = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{6}$$

$$\text{SD } (\sigma) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{7}$$

Where $\{y_{1,} y_2, ..., y_N\}$ are the clusters formed using AHK clustering, y_i is actual values and $\bar{y}$ denotes the average value of the input sample, *N* denotes the number of membership degree in vector $(y_1 - \mu, ..., y_N - \mu)$.

The Takagi-Sugeno fuzzy inference system in FLC introduces an *n* fuzzy set rules for computing output vectors with the set of input vectors. The Takagi-Sugeno rule is given below:

Rule 1: if P is A_{11} and Q is A_{12}, Then $y^0 = (y_1, y_2, ..., y_m)$

Rule 2: if P is A_{21} and Q is A_{22}, Then $y^0 = (y_1, y_2, ...y_m)$

Where A_{11}, A_{12} and, A_{21}, A_{22} are represent the fuzzy set of its MFs of P and Q antecedent variable, y indicate the consequence of variable Q, and $(y_1, y_2, ..., y_m)$ are the linear parameter in the If-Then function. A linear regression model indicates that the given operating system is valid as a result (mean and SD) of Takagi-Sugeno rule. The main advantage of this rule is to maintain the nonlinearity interaction between the TSP, so the TSP parameters are automatically adjust and makes the output either linear or constant.

The final output of mean function y_i $(y_1, y_2, ...y_M)$ is a linear function, where i =1, 2, 3….K. So, the input point is $y^0 = [\, y_1^0, y_2^0, ... y_M^0 \,]$, inferring results of the fuzzy interference system rule is represented by y^0.

2.22 Prediction Based On WOA-LSTM

LSTM is a type of Recurrent Neural Network (RNN), its structure is shown in figure 3. LSTM is one of the most successive method for TSP. This model contains four important layers such as cell weight state, input layer, hidden layer, and output layer. The prediction function of each layer is given below:

Cell state: The cell state is a kind of conveyor belt. This state crosses the complete chain, with some linear connections. The weighted value is given to the cell weight state.

Forget gate: It can remove the unwanted data's from the weighted state. The clustering output and previous hidden state (m_{t-1}) are the input value (x_t) of forget gate. The output of forget gate is compared with the cell states output is characterised by 0 and 1, it is related to each number of cell state. If the output is 1, it represents that the forget gate want to keep the information, while the output is 0, it represents "Absolutely this exemption".

Input gate: The input gate is used to store the new input variables has two sub layers such as sigmoid layer and *tan h* layer. It is used to create new values and added with the weighted value. This combined process is evaluated by equation (16).

Output gate: The function of output model is based on the cell weight state, here the sigmoid layer run first. It provide the data to run from the cell state through the *tan h* layer between -1 and 1, finally, the output of tan h is multiplied with the output of the sigmoid layer. The final output of this gate is calculated by using equation (20).

The hidden layer has a joining loop in its cell weight state with some fixed-weighted values. It is denoted by cell weight blocks, each block contains the number of weighted cells. The result of output layer (m_t) is return back to the hidden layer (m_{t-1}) is called BP process. These three gates are assisted by a small hole between each cells weight state. The structure of LSTM is shown in figure 3,

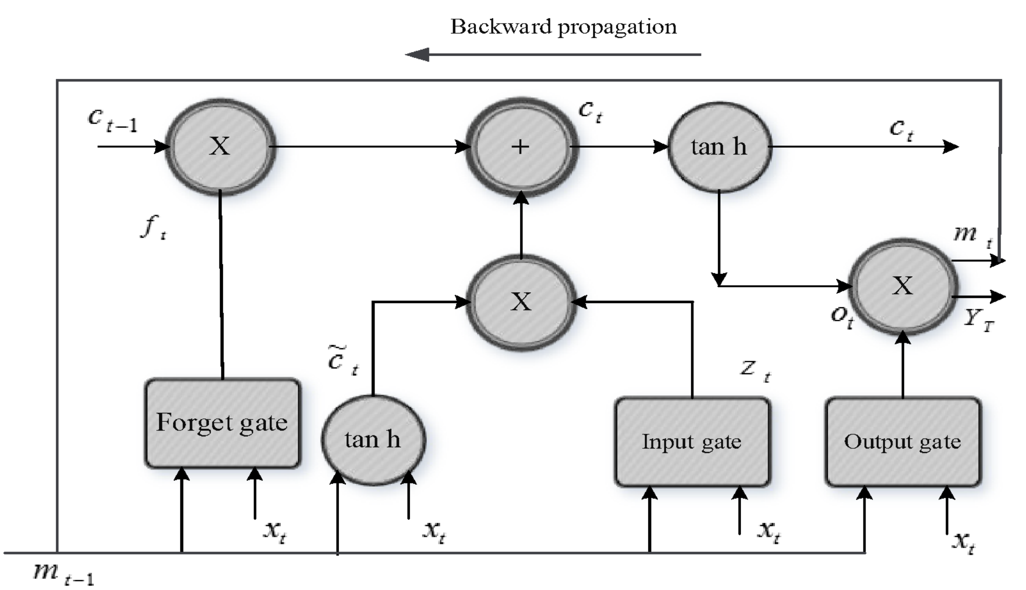


Figure 3: The structural design of the Long Short Term Memory Network with three layers.

Weight updating using WOA

The WOA algorithm function is related to the whale behaviors. It is used to optimize the best weight solution, and this value is used in the LSTM network layer. The fitness of this algorithm is to balance the number of selected values in each solution. The best solution of this method depend on the bubble-net feeding method. In WOA, the bubble-net nourishing winding is used to establish the mathematical model.

The model is used to find best solution by using Encircling prey, to detect the target. WOA assumes that the best current solution is the position of the damen optimal solution. As the number of imbalances increases from the starting end, it will change its position towards the right solution. After finding the best solution, the opposite search agents try to update their position according to position of the best solution using equation (8) and (9).

$$\vec{D} = \left|\vec{C}.\vec{X}_p(t) - \vec{X}(t)\right| \tag{8}$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A}.\vec{D} \tag{9}$$

Where $\vec{D}$ indicates the distance of the whale prey (obtained best solution), t symbolizes the current iteration, $\vec{A}$ and $\vec{C}$ is a coefficient vector, $\vec{X}_p$ denotes the position vector, $\vec{X}$ represent the best arrangement of the position vector. P denotes the multiplication of the elements and $|\,|$ is an absolute value. The equation of $\vec{A}$ and $\vec{C}$ is given below:

$$\vec{A} = 2\vec{a}.\vec{r} - \vec{a} \tag{10}$$

$$\vec{C} = 2.\vec{r} \tag{11}$$

Where $\vec{a}$ represent that the iteration process is linearly decreased from a higher to lower level and $\vec{r}$ is an uninformed vector. This technique has two type of mathematical models namely Spiral Updating Position and Shrinking Encircling Mechanism used to calculating the weighted value.

a) Shrinking Encircling Mechanism:

The best solution obtained indicate larger inertia weight as for global search and smaller weight as local search. In order to improve the local search capacity, convergence accuracy and speed an inertia weight is introduced in the WOA to improve the capacity of algorithm. The expressed form of the inertia weight is given as.

$$W = W' - (W' - W'')\left(\frac{t}{T_{\max}}\right)^{\frac{1}{t}} \tag{12}$$

Where W is an inertia weight of the WOA algorithm, W' represents the maximum value of inertia weight, W'' is a minimum value of inertia weight, t is a current iteration number, $T_{\max}$ denotes the maximum number of iterations.

b) Spiral Updating Position:

In this process, the rate of a spiral is formed between prey and whale to reproduce the spiral-shape movement, it is used to update the data's. Then the scientific formulation of search agent position updating during the optimization procedure is given by equation (13).

$$\vec{X}(t+1) = \vec{D}^t \cdot e^{jl} \cdot \cos(2\pi l) + \vec{X}_p(t) \tag{13}$$

Where $\vec{D}^t = \left| X_p(t) - X(t) \right|$, $\vec{D}^t$ denotes the distance among the best solution and X_p denotes the position vector.

Thus, the fitness function of individuals of WOA is as follows:

$$\vec{X}(t+1) = \vec{D}^t \cdot e^{jl} \cdot \cos(2\pi l) + \vec{X}^*(t); \quad \text{If } p \le 0.5 \tag{14}$$

Where j is a continuous logarithmic shape, l denotes the random value between the [-1, 1], and (.) is an element by element augmentation, p is a random solution in [0, 1]. The concrete steps of the WOA are given below:

Step 1: In the first step, the WOA is start from the set of random population $X_i (i = 1,2,...,n)$, and set of maximum iterations (t=1).

Step 2: Next, calculate the weighted value is depend on the fitness function of $X_i (i = 1,2,...,n)$ to evaluate the accurate solution X^*

Step 3: Update the current condition by using the equation (13). From this position, select the random solution. If the random value $p \le 0.5$, again update the random solution by using the equation (14).

Step 4: Check the random values and calculate the fitness value $X_i (i = 1,2,...,n)$. Finally, search if any best solution occur X^* by using equation (8) and (9). Based on this the weight value is calculated by equation (12).

Step 5: The overall optimization process is repeated until termination fitness criteria (weight) satisfied.

The LSTM network can perform both read and writing process and predict the traffic speed based on the prior information. The TSP equations are given below:

$$z^t = g(W_z x^t + W_i y^t + p_i \; \Theta \, b_i) \tag{15}$$

$$i_t = \sigma(W_{ix} + W_{im} m_{t-1} + W_{ic} c_{t-1} + b_i) \tag{16}$$

$$f_t = \sigma(W_{fx} x_t + W_{fm} m_{t-1} + W_{fc} c_{t-1} + b_f) \tag{17}$$

$$\widetilde{c}_t = \tan h\,(W_{cx} x_t + W_{cm} m_{t-1} + b_c) \tag{18}$$

$$c_t = f_t \;\Theta\, c_{t-1} + i_t \;\Theta\; g(W_{cx} x_t + W_{cm} m_{t-1} + b_c) \tag{19}$$

$$o_t = \sigma(W_{ox} x_t + W_{om} m_{t-1} + W_{oc} c_t + b_o) \tag{20}$$

$$m_t = o_t \;\Theta\; h(c_t) \tag{21}$$

$$Y_T = [m_{t-1}, \ldots .m_{t-n}] \tag{22}$$

Where the z^t is an input block, i_t is a input gate, f_t is a forget gate, $\widetilde{c}_t$ is a input of cell state, c_t is a output of cell state, o_t denotes the output gate, m_t denotes the hidden block, the W represents the rectangular input weighted matrices, x_t is an input vector at time t, b denotes the bias vectors, Θ indicates the scalar product of vector and $\sigma(\cdot)$ represents the average logistics sigmoid function. $W_{ox}, W_{om}, W_{oc}, W_{cx}, W_{cm}$ represents the weight matrices are connected to the x_t of three gates and cell state, $W_{ix}, W_{im}, W_{ic}, W_{fx}, W_{fm}, W_{fc}$ represent the weight matrices are connected to the c_{t-1} of cell input and three gates. The final output of this proposed system is represented by Y_T and m_{t-1} is a previous hidden state.

The BP error calculation includes that the error backward propagate over time and the error propagate backward to previous layer from the current layer. The BP of error calculation is calculated by using equation (15) to (22). The final output of the backward layer is represented by Y_T . The following procedure for the calculating the real traffic speed depends on:

Step 1: Consider a raw speed with periodic and residual components

$$S_t = P_t - R_t \tag{23}$$

Where S_t represent the speed time, P_t represent the periodic component and R_t represent the residual components

Step 2: Residual error are used to optimize the parameter in the FOLSTM to train dataset. Residual error are obtain by subtracting the P_t with S_t. The predicted residual error (PR_E) are captured by the predicted data Y_T and observed data X_N.

Step 3: Finally calculate the predicted value for real traffic speed data by combing the PR_E and P_t.

The pseudo code of the proposed AHK-FOLSTM model for accurate speed prediction is given in Table 1.

Table 1: Pseudo Code of proposed AHK-FOLSTM method

Algorithm: Pseudo code of AHK-FOLSTM
AHK Clustering ***Input:*** C: *the initial centroids* *X: n objects of original data's* ***Output:*** *A set of* ***b clusters.*** **Stage 1: Merge stage** 1. **Initialization:** Set initial value (x, and c_{ij}). **2. For** each c $\in C$ **do** 3. // agglomerative hierarchical clustering 4. Set the input features ($c \in C$); **5. While** $\|C\|$>A **do** 6. c_{ij} for next pair; 7. // update the distance of the cluster (C_i *and* C_j) 8. Merge (C_i *and* C_j); 9. $A' = A' - 1$; 10. Repeat step (6,8,9), until $A' = A$; **11. End while** **12. End for** **Stage 2: K mean** List the intermediate clusters '*A*'

13. For each c $\in C$ **do**

14. Tem dist = distance evaluation

15. End

16. **For each** *d* in *D*

17. **C=** argmin (centroid, distance D);

18. design (D, cluster *k*);

19. End for

20. **For** i=1, j=1 to *k*

21. C_i *and* C_j = mean (cluster *k*);

22. Evaluate the objective function Eq. (4)

23. End for

24. // repeat step (16, 20), until the centroid no change.

FOLSTM Prediction

Input: *No of clusters* $[k_1, k_2, \ldots k_n]$

y^0 *fuzzy interference system rule*

x=$[x_1, x_{2,\ldots,} x_n]$

Input weighted matrices $[W = W_1, W_{2,\ldots,} W_n]$

Output: $Y_T = [Y_1, Y_2, \ldots Y_n]$

Stage 1: Fuzzy base Takagi-Sugeno

Input: *No of clusters* $[k_1, k_{2,\ldots,} k_n]$

Output: *Fuzzy interference rule* $y^0 = [y_1^0, y_2^0, \ldots y_M^0]$.

1. **Begin;**

   Initialization: Set **n** cluster value (*b*).

2. **Population = {** b_{ij}} set of clusters;

3. **While (**True and false condition) **do**

4. Update the MF using Eq. (5)

5. Calculate MF values

6. Update the fuzzy interference rule

7. **End While**

8. **End**

Stage 2: LSTM

Input: x=[$x_1, x_{2,....} x_n$]

Input weighted matrices $[W = W_1, W_{2,....} W_n]$

Output: *TSP output* $Y_T = [Y_1, Y_2, \ldots Y_n]$

1. **Initialization:** The parameter of LSTM ($i_t, f_t, \tilde{c}_t, c_t, w$ and o_t)
2. Create_ weight value (WHO)
3. Evaluate the weight value of each agent by using Eq. (12)
4. **Update** the weight value
5. **For** t=n **do**
6. **Evaluate** initial gates f_t Eq. (17), i_t Eq. (16), $\tilde{c}_t$ Eq.(18)
7. **Update** c_t using Eq.(19)
8. **Evaluate** output gate (o_t) Eq. (20), hidden state (m_t) Eq.(21)
9. **End for**

Repeat same process in BP // final result denote in Eq. (22).

3. Experimental Result and Discussion

3.1 Dataset Description

San Diego: The travel speed data used in the study is collected from San Diego in USA for one year (Jan-Dec 2010). This dataset used 3,000 detectors in the road side to collect the cleaned and raw data for traffic process. The road segment is selected from 1,250 lane miles of 1-5 in San Diego and its total length is 2011.68 kilometre. The dataset contain every 30 second and 5 minutes information and hourly, daily road incidents and lane closures for two section along between northbound interstate and southbound interstate (5 mile post).

Beijing: In this dataset, the trajectory data is collected from the $PM_{2.5}$ data from the US Embassy in Beijing and the hourly metrological measurement from the Beijing Capital International Airport (BCIA). These two factors are collected at the time period from1st March to 31st July, 2016. There are about 24 weekends and 12 holidays among the dataset. In this, the weather condition is classified with ten types such as snow, sunny, rainy with the temperature ranges from -4 C -36 C. The time interval were used from the previous 5 days and 3 weeks information to extract the periodicity data. Here, the speed of the wind is divided into four levels. The first four month data are used as the training set and the remaining information collected at the last month is used as testing set.

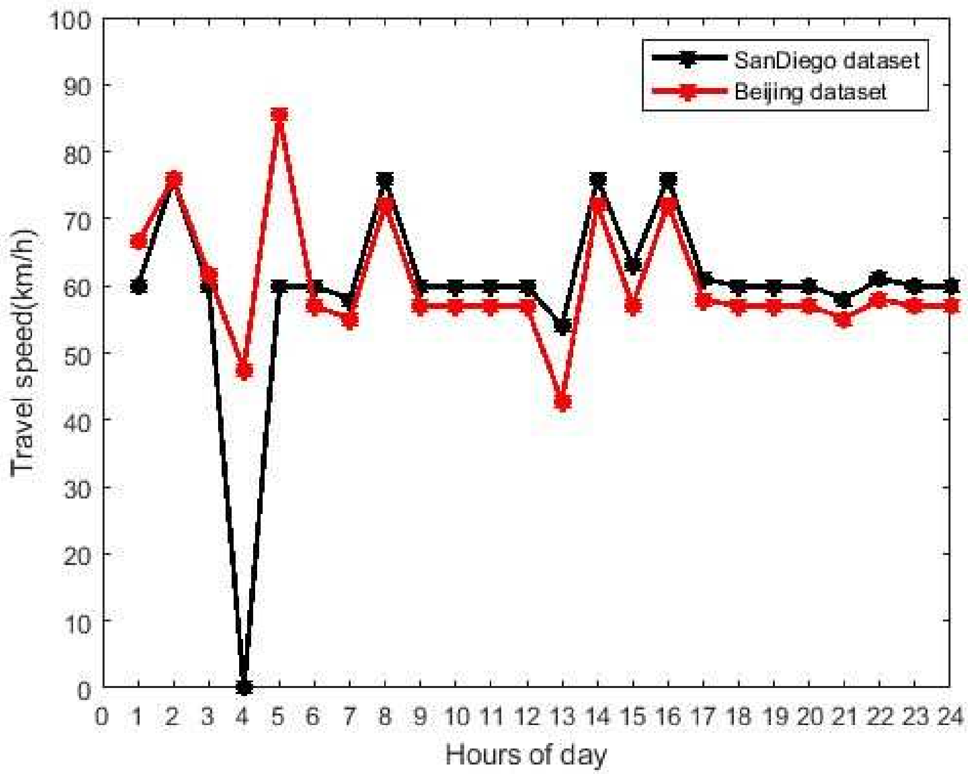


Figure 4: The traffic speed distribution in one day data from San Diego dataset and Beijing dataset.

San Diego and Beijing are the most commonly used dataset for traffic speed estimation. In this, we use the first 80% of the data as a training set and the rest of 20% of the data is used as testing test. Figure 4 show the traffic speed data obtained for one day based on the observed data from the San Diego dataset and Beijing dataset. It clearly shows a strong reduction in speed during peak hours and the speed values observed during the day oscillate more significantly than the observed speed values at night.

The proposed AHK-FOLSTM model is implemented in the user-friendly R programming language, which is more suitable for data analysis without getting into too much of details. The key features (speed, occupancy and volume) of this study is to determine the TSP. To evaluate the computational performance of the proposed system, different performance parameters, namely the Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE) are used. The unit of RMSE and MAE is km per hour. The equations of RMSE, MAE, and MAPE are shown as follows:

$$\text{MAPE}=\frac{1}{N}\sum_{i=1}^{N}\frac{\left|\hat{S}-S_i\right|}{S_i}\times 100\% \tag{24}$$

$$\text{MAE}=\frac{\sum_{i=1}^{N}\left|\hat{s}_i-S_i\right|}{N} \tag{25}$$

$$\text{RMSE}=\sqrt{\frac{\sum_{i=1}^{N}\left(\hat{s}_i-S_i\right)^2}{N}} \tag{26}$$

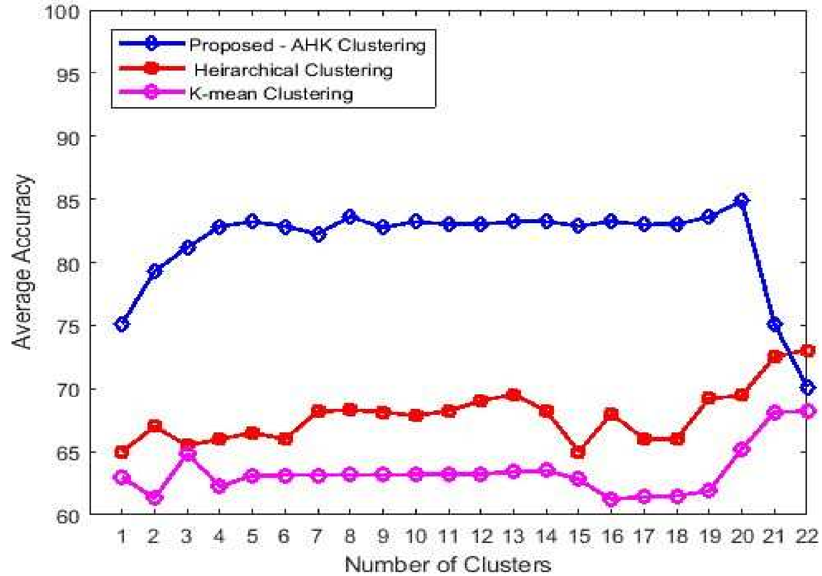
Where, S_i denotes the actual speed at the time i, number of observation is denoted by N and $\hat{S}_i$ is the predicted speed interval t.

3.2 Benchmarks

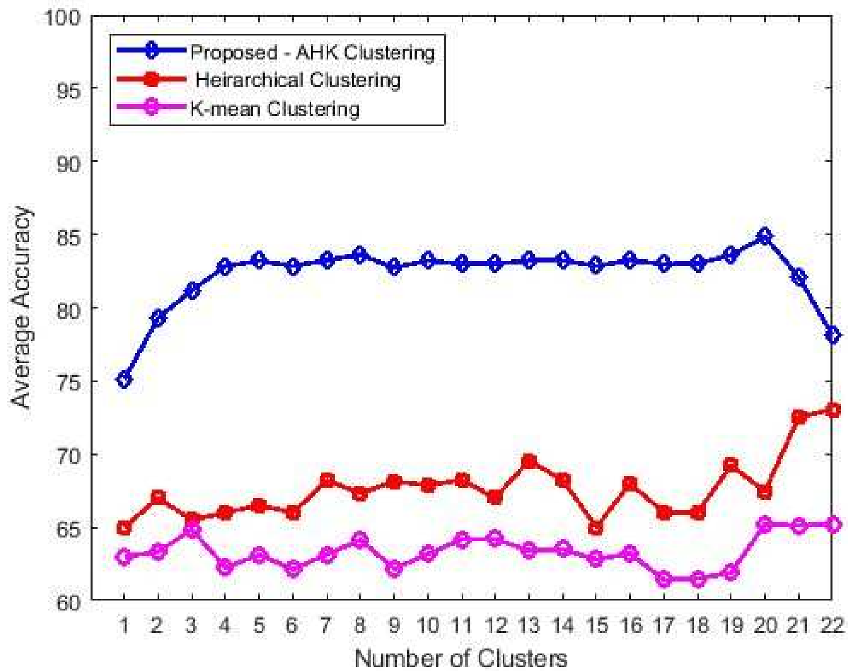
The comparison is performed with some other traditional approaches such as Auto Regressive Moving Average (ARIMA), Back Propagation Neural Network (BPNN), Artificial Neural Network (ANN), Evolving Fuzzy Neural Network (EFNN) and Long Short Term Memory (LSTM) methods. The BPNN is one of the method in artificial neural network that have one hidden layer associated with 50 neurons. The ARMA that uses the maximum likelihood estimation to predict the traffic speed condition. Artificial Intelligence (AI) based methods such as ANN are the widely used that handle the traffic prediction issues. However, the ARIMA and the ANN model lacks in prediction due to the non-linearity and the irregularity nature of the traffic speed condition. EFNN is the extension of the fuzzy network that uses five layers such as input, fuzzy input, rule layer, fuzzy output and the output layer. Here, the fuzzy layer represents the membership degree obtained from the input values can be modified during the learning process. LSTM model is designed from Recurrent Neural Network (RNN) to estimate the traffic flow prediction that learns the optimal time lags in an automatic manner.

3.3 Result Analysis from San Diego and Beijing datasets

The performance of AHK-FOLSTM prediction is measured by means of four metrics such as Mean Square Error (MSE), MAPE, MAE and RMSE. In this, the parameters of data clustering is evaluated by using MSE based on cross-validation method. The MAPE error prediction process depends on the fractional variation between predicted result and observed values. MAE measures the average magnitude of the error to analyse the variation between predicted and observed value. The RMSE provides the prediction error based on the historical record of the TSP value.



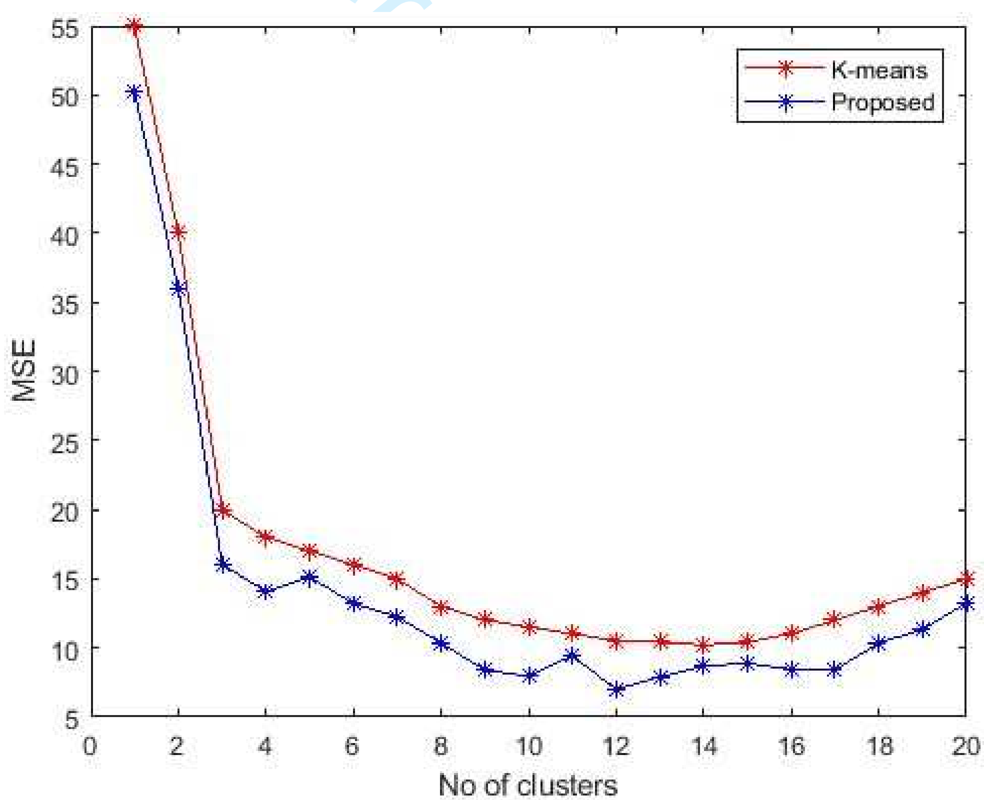
(a)



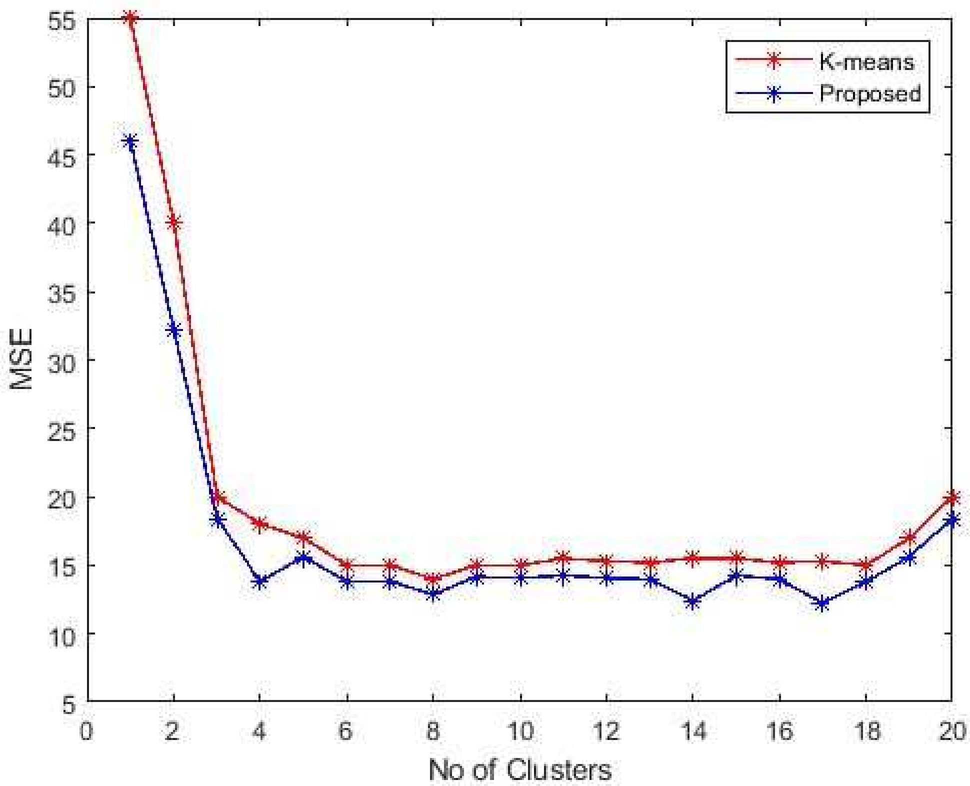
(b)

Figure 5: The average accuracy of each cluster for both proposed and existing method (a) San Diego dataset, and (b) Beijing dataset.

To develop the FOLSTM prediction method, the traffic information were divided into twenty two group of clusters using AHK clustering. Figure 5 (a) and (b) illustrate the average accuracy for different clusters using San Diego dataset and Beijing dataset. From fig 5(a), the result show that the proposed AHK method achieves a maximum accuracy of about 85% while the existing hierarchical and the K mean clustering reaches only about 72.12%, and 62.21% from San Diego dataset. Also in fig 5(b), the average accuracy of the proposed method achieves 84% accuracy measure but the existing hierarchical and the existing K mean clustering reaches only 73.52% and 66.11% from the Beijing dataset. Here, the result obtained from both datasets achieved a maximum range within 20 clusters. This is due to the hybrid form of K means and hierarchical clustering (AHK) that achieves a better performance than the individual clustering method. After the completion of 20 cluster, the accuracy level is automatically decreased. In this, we consider twenty clusters for traffic prediction because there is a less overlapping problems and high speed performance within the clusters. So, it's a best choice to get the good accuracy prediction than the other state-of art methods.



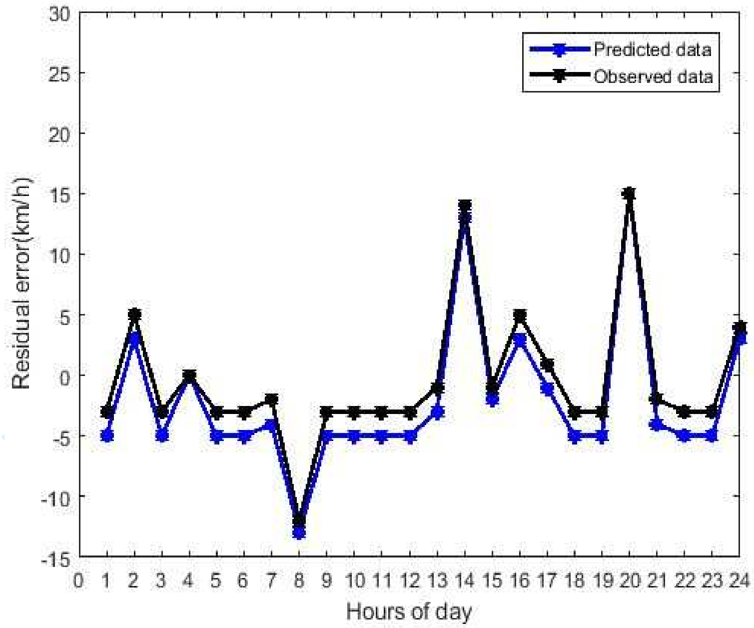
(a)



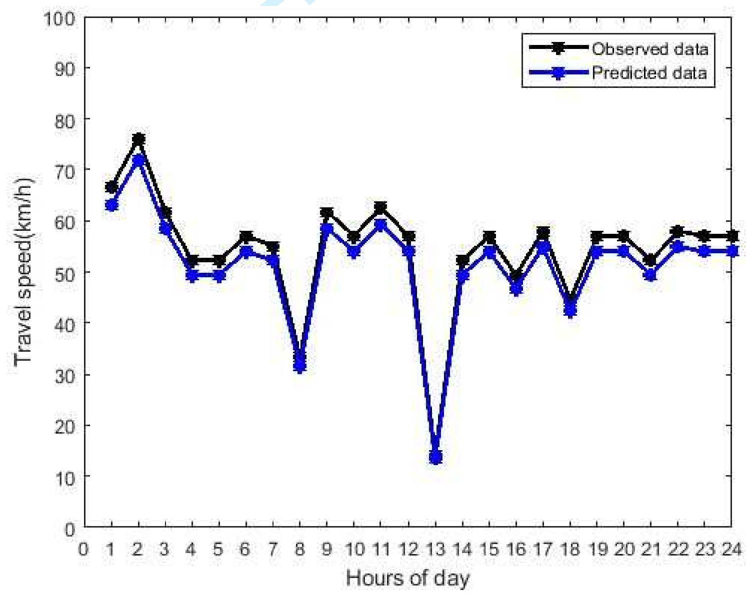
(b)

Figure 6: MSE analysis of AHK model for one day traffic speed prediction data (a) San Diego dataset, (b) Beijing dataset.

The MSE estimation is evaluated based on cross-validation in the training part to make the fairness comparison of TSP. The accurate forecasting performance of the MSE measure with the group of clusters from 1 to 20 of AHK model obtained for both San Diego and Beijing dataset is shown in figure 6 (a) and (b). The result show that, when number of clusters increases, the MSE value for the proposed method linearly decreased than the K mean method. This is due to the high prediction of cluster accuracy in the proposed approach of TSP. But the K-means obtains poor learning effect due to the classification of the input samples into different models with the reduced number of clusters [16].



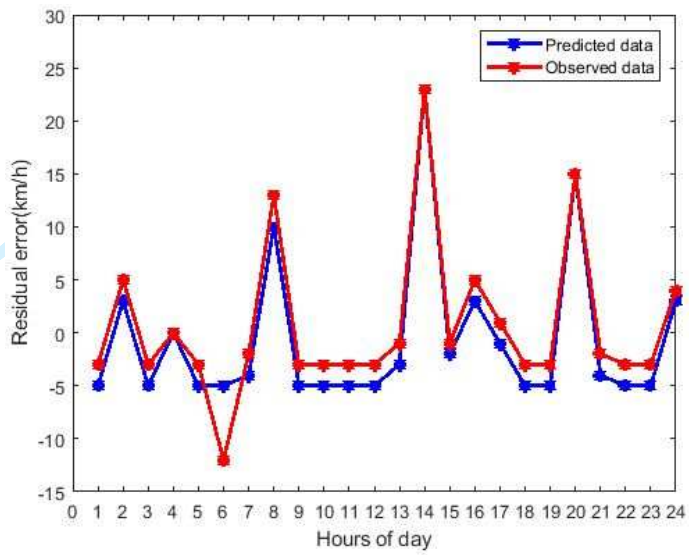
(a)



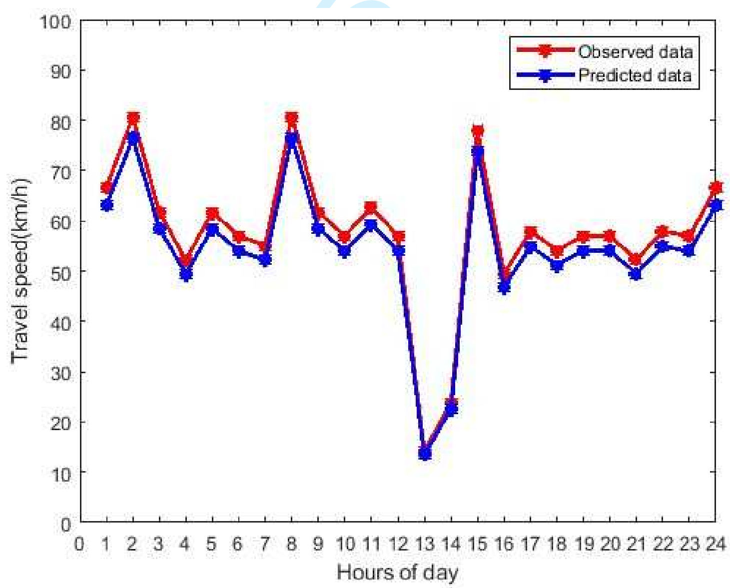
(b)

Figure 7: Prediction result analysis for speed using AHK-FOLSTM method using San Diego dataset (a) Residual error (b) Actual travel speed.

The residual error is calculated by the difference between observed data and predicted data to determine the predicted error on traffic speed. The residual error analysis performed with periodic component and observed data is shown in the fig 7(a). The actual predicted speed calculated based on residual error and periodic function is shown in fig 7(b). The result show that the predicted error is lower than the observed data indicates there is an accurate traffic speed prediction by our proposed AHK-FOLSTM method.



(a)



(b)

Figure 8: Prediction result analysis for speed using AHK-FOLSTM method (a) Residual error (b) Actual travel speed using Beijing dataset.

In traffic speed estimation, the residual error is calculated to determine the real time traffic speed. The residual error and actual travel speed of Beijing dataset is shown in figure 8 (a) and (b). The result show that the proposed method analysis is 7% lower than the observed data by AHK-FOLSTM approach.

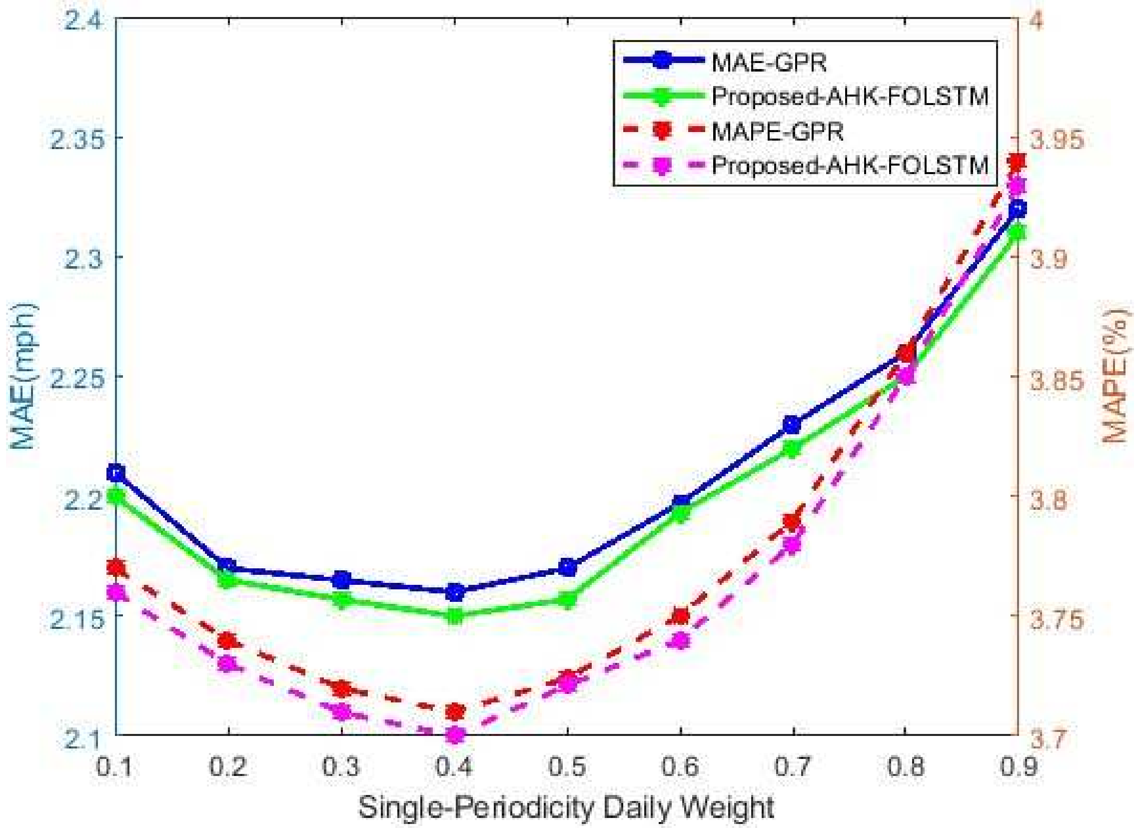


Figure 9: The MAPE and MAE value on AHK-FOLSTM model with the daily periodicities using San Diego dataset.

In traffic speed prediction, the MAE measure is used to estimate the error we can expect from an average prediction based on the difference between two continuous variable. The MAPE measure subtracts the actual value with the forecast value and divides it with the actual value to measure the accuracy of the forecast. This represents the high error from low error speed variation that used to show the accuracy of the prediction model. Hence, both this model is used to find the mean absolute error in percentage terms and show the infrequent errors. Figure 9 shows that the result of the daily single periodicity with different weights are combined together for predicting the MAE and MAPE errors obtained from San Diego dataset. In this, we use multiple weighted values in FOLSTM method to get the best result. To predict the accurate mean error, the weight of the daily range is taken from 0.1 to 0.9. In this, MAE and MAPE measure of the proposed AHK-FOLSTM method obtains the 2.33 *mph* and 3.93% but the existing GPR achieves of only 2.34 *mph* and 3.94% errors. In addition, by testing with different approach, the proposed method obtains 0.4 as a best weight. This is due to both MAE and MAPE measure of the proposed method only takes the magnitude of deviations of predicted values from those of the observed data [31].

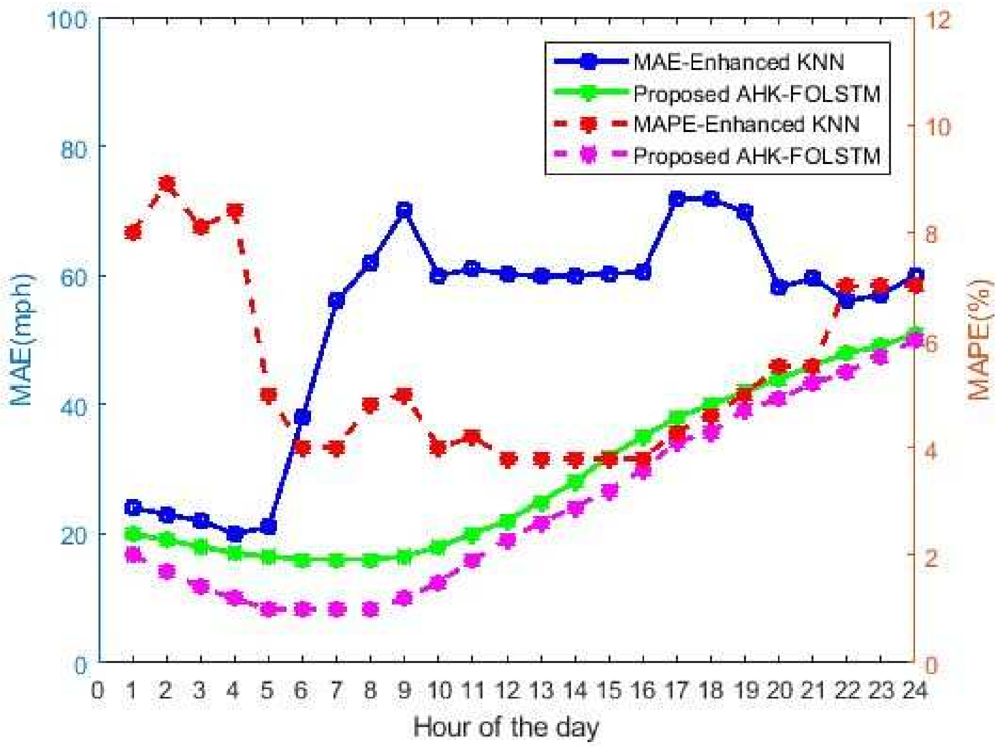
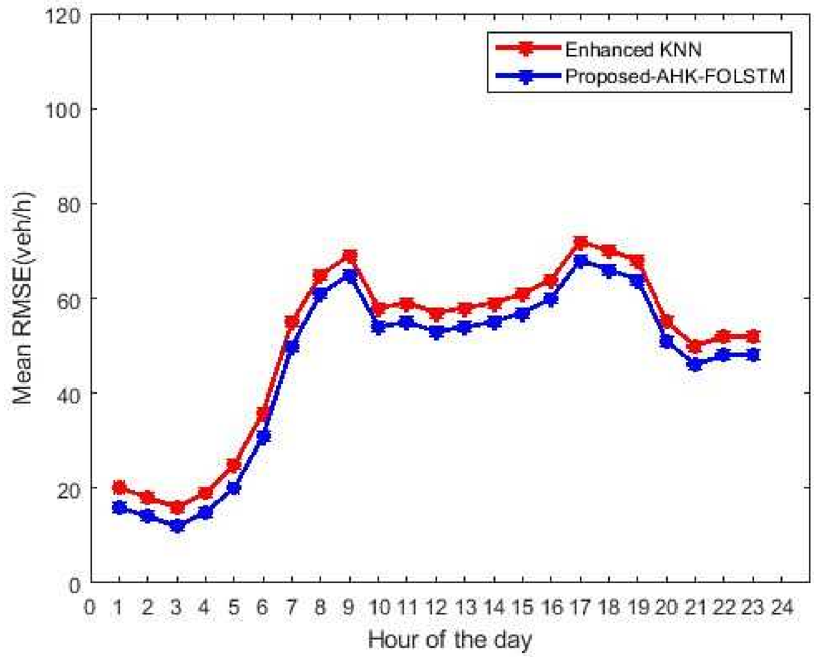
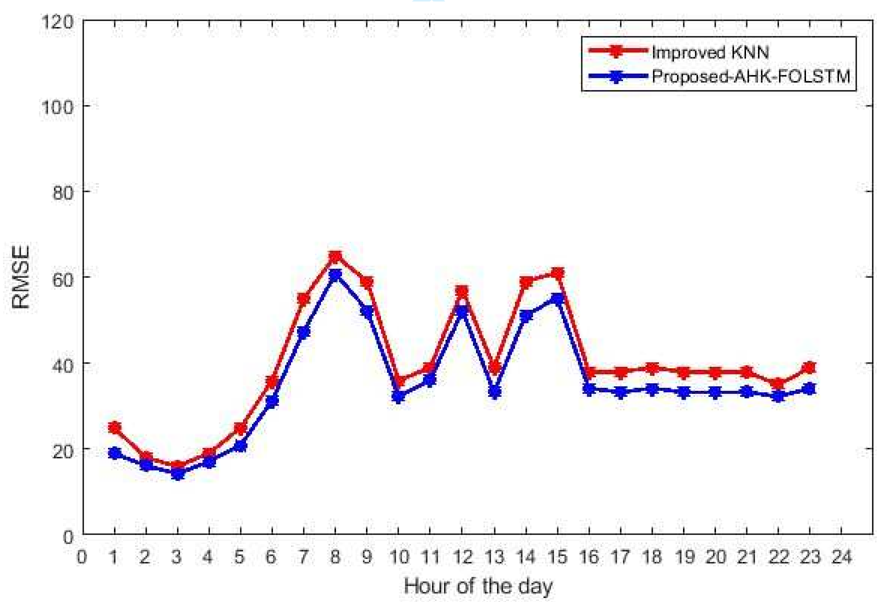


Figure 10: The MAPE and MAE value on AHK-FOLSTM model with the single day periodicities using Beijing dataset.

The forecasting result of MAPE and MAE for Beijing dataset using Enhanced KNN and the proposed AHK-FOLSTM method is shown in Figure 10. The result show that, our proposed method achieves better result than the existing KNN Enhanced method. In this, our hybrid technique reaches the low error in MAE of 4.9 *mph* and MAPE of 6 % than the KNN method. This is because of both MAE and MAPE consider only the magnitude of deviations predicted values from those observed data's from the Beijing dataset [32].



(a)



(b)

Figure 11: RMSE comparison for both proposed method and existing method (a) San Diego (b) Beijing dataset

RMSE is one of the error metrics that calculates the error by means of the deviation between the actual and expected values. Figure 11 (a) (b), shows the forecast errors in RMSE of Enhanced KNN and the proposed AHK-FOLSTM method for both San Diego and Beijing dataset. Here, we use the traffic level and the time of day to estimate which model has the biggest error in TSP. In comparison, both forecasting errors is a statistically increase in mean prediction errors between the AHK-FOLSTM and the Enhanced KNN approach in San Diego and with Improved KNN in Beijing dataset. This is due to the appropriate selection of clusters in description of periodic pattern in the traffic speed data and thus improves the performance of the AHK-FOLSTM method [32, 33].

3.4 Comparison Result

This section designs comparative methods based on the decomposition to validate the performance of our proposed approach by using two data sources (San Diego dataset and Beijing dataset), in particular, to explore the impact of each method.

The table (2) shows the comparison of various benchmarks on two dataset. From the comparison table, the result of the proposed method has obviously improved in TSP on both San Diego and Beijing dataset.

Table 2: Comparison of various benchmarks and proposed method on San Diego and Beijing dataset.

Methods	**San Diego dataset** (5 min time interval)			**Beijing dataset** (5 min time interval)		
	MAPE (%)	**RMSE** (mph)	**MAE** (mph)	**MAPE** (%)	**RMSE** (mph)	**MAE** (mph)
LSTM [14]	4.55	7.28	3.25	4.85	7.5	3.65
BPNN [16]	7.72	7.6	3.76	7.55	7.8	3.66
EFNN [18]	7.21	6.88	3.54	5.3	6.9	3.58
ARIMA [19]	5.78	6.9	4.26	5.86	6.5	4.56
ANN [20]	7.86	7.60	5.66	6.55	7.59	5.68
FOLSTM	**3.9**	**6.5**	**2.33**	**6**	**6.33**	**4.9**

Based on the above table, several interesting details are discussed below,

1. The table shows the MAPE, MAE and RMSE value for one step forecasting are significantly larger than the result of multi-step forecasting. The performance of

FOLSTM obtain better result than various benchmark methods (ARIMA, BPNN, ANN, EFNN and LSTM).

2. The proposed model can develop the TSP result by 19% and 22% when compared to ML model (BPNN) statistical model (ARIMA). The comparison result of these methods have complex learning ability and complex structure.
3. For the prediction performance of ML model (ANN) is higher than those of EFNN model, but the performance of these prediction methods are not used for long term prediction. The comparison result of TSP increases by average of 4% in our proposed method. Furthermore, deeply discuss about the function of benchmark comparison, the accuracy of TSP in proposed AHK-FOLSTM method has relatively better accuracy than other benchmarks.

4. Conclusion

In this paper, the TSP is designed based on the FOLSTM by using traffic speed data from San Diego in USA. In the training process, the AHK clustering method group the traffic data in to non-hierarchical clusters. The FOLSTM forecasting model is built based on the Takagi-Sugeno fuzzy interference rule. Gaussian bell shape membership function were designed to calculate the membership degree of each clusters. WAO algorithm update the weight of LSTM. The output of this method is compared with some existing methods (GPR and KNN). The performance of proposed TSP method is calculated based on three parameters MAPE, MAE and RMSE criteria. Based on this result, the number of interesting conclusions are obtained.

- Traffic speed as an input variable can provide acceptable performance for TSP, while the use of TS together with residence/speed as input can produce better results on the speed of traffic.
- The TSP is analyzed based on three characteristics: road occupancy, traffic volume and average traffic speed. Furthermore, these three variables expressed a different correlation between different states. In this research, the traffic occupation, average traffic speed and volume are used as an input variables of the proposed model to improve the accuracy of inaccurate traffic.

Reference

[1]. Lin, Y., Wang, P., & Ma, M. Intelligent Transportation System (ITS): Concept, Challenge and Opportunity. In Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International

Conference on Intelligent Data and Security (IDS), 2017 IEEE 3rd International Conference on IEEE, (2017) pp. 167-172.

[2]. Blythe, P., Rackliff, T., Holland, R., & Mageean, J. ITS Applications in Public Transport: Improving the service to the transport system. Journal of advanced transportation, 34(3) (2000) 325-345.

[3]. Asif, M. T., Dauwels, J., Goh, C. Y., Oran, A., Fathi, E., Xu, M, & Jaillet, P. Spatiotemporal patterns in large-scale traffic speed prediction. IEEE Transactions on Intelligent Transportation Systems, 15(2) (2014) 794-804.

[4]. Li, R., Jiang, C., Zhu, F., & Chen, X. Traffic flow data forecasting based on interval type-2 fuzzy sets theory. IEEE/CAA Journal of Automatica Sinica, 3(2) (2016) 141-148.

[5]. Huang, M. L. Intersection traffic flow forecasting based on ν-GSVR with a new hybrid evolutionary algorithm. Neurocomputing, 147 (2015) 343-349.

[6]. Lau, H. C., Chan, T. M., Tsui, W. T., Chan, F. T., Ho, G. T., & Choy, K. L. A fuzzy guided multi-objective evolutionary algorithm model for solving transportation problem. Expert Systems with Applications, 36(4) (2009) 8255-8268.

[7]. Yao, B., Chen, C., Cao, Q., Jin, L., Zhang, M., Zhu, H., & Yu, B. Short-term traffic speed prediction for an urban corridor. Computer-Aided Civil and Infrastructure Engineering, 32(2) (2017) 154-169.

[8]. Vanajakshi, L., & Rilett, L. R. A comparison of the performance of artificial neural networks and support vector machines for the prediction of traffic speed. In Intelligent Vehicles Symposium, 2004 IEEE, (2004) pp. 194-199.

[9]. Cai, P., Wang, Y., Lu, G., Chen, P., Ding, C., & Sun, J. A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. Transportation Research Part C: Emerging Technologies, 62 (2016) 21-34.

[10]. Nath, R. P. D., Lee, H. J., Chowdhury, N. K., & Chang, J. W. Modified K-means clustering for travel time prediction based on historical traffic data. In International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, Springer, Berlin, Heidelberg. (2010) pp. 511-521.

[11]. Li, Y., Chen, M., Lu, X., & Zhao, W. Research on optimized GA-SVM vehicle speed prediction model based on driver-vehicle-road-traffic system. Science China Technological Sciences, 61(5) (2018) 782-790.

[12]. Hu, W., Yan, L., Liu, K., & Wang, H. A short-term traffic flow forecasting method based on the hybrid PSO-SVR. Neural Processing Letters, 43(1) (2016) 155-172.

[13]. Lu, B., & Huang, M. Traffic flow prediction based on wavelet analysis, genetic algorithm and artificial neural network. In Information Engineering and Computer Science, 2009. ICIECS 2009. International Conference on IEEE, (2009) pp. 1-4.

[14]. Kang, D., Lv, Y., & Chen, Y. Y. Short-term traffic flow prediction with LSTM recurrent neural network. In Intelligent Transportation Systems (ITSC), 2017 IEEE 20th International Conference on IEEE, (2017) pp. 1-6.

[15]. Jiang, B., & Fei, Y. Vehicle speed prediction by two-level data driven models in vehicular networks. IEEE Transactions on Intelligent Transportation Systems, 18(7) (2017) 1793-1801.

[16]. Tang, J., Liu, F., Zou, Y., Zhang, W., & Wang, Y. An improved fuzzy neural network for traffic speed prediction considering periodic characteristic. IEEE Transactions on Intelligent Transportation Systems, 18(9) (2017) 2340-2350.

[17]. Lin, L., Li, J., Chen, F., Ye, J., & Huai, J. Road traffic speed prediction: a probabilistic model fusing multi-source data. IEEE Transactions on Knowledge and Data Engineering, 30(7) (2018) 1310-1323.

[18]. Kasabov, N. Evolving fuzzy neural networks-algorithms, applications and biological motivation. Methodologies for the conception, design and application of soft computing, World Scientific, 1 (1998) 271-274.

[19]. Okolobah, V., & Ismail, A. Forecasting peak load demand by ARIMA and EMD method. Arch. Des. Sci, 66 (2013) 2-15.

[20]. Kumar, M., & Thenmozhi, M. A hybrid ARIMA-EGARCH and Artificial Neural Network model in stock market forecasting: evidence for India and the USA. International Journal of Business and Emerging Markets, 4(2) (2012) 160-178.

[21]. Wattana, K., & Nishio, M. Application of a regression model for predicting traffic volume from dynamic monitoring data to the bridge safety evaluation. Journal of Civil Structural Health Monitoring, 7(4) (2017) 429-443.

[22]. Xu, J., Deng, D., Demiryurek, U., Shahabi, C., & Van Der Schaar, M. Mining the situation: Spatiotemporal traffic prediction with big data. IEEE Journal of Selected Topics in Signal Processing, 9(4) (2015) 702-715.

[23]. Su, L., Yao, Y., Li, N., Liu, J., Lu, Z., & Liu, B. Hierarchical Clustering Based Network Traffic Data Reduction for Improving Suspicious Flow Detection. In 2018 17th IEEE

International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE) IEEE. (2018) 744-753.

[24]. Fan, S. K. S., Su, C. J., Nien, H. T., Tsai, P. F., & Cheng, C. Y. Using machine learning and big data approaches to predict travel time based on historical and real-time data from Taiwan electronic toll collection. Soft Computing, 22(17) (2018) 5707-5718.

[25]. Tang, J., Zhang, G., Wang, Y., Wang, H., & Liu, F. A hybrid approach to integrate fuzzy C-means based imputation method with genetic algorithm for missing traffic volume data estimation. Transportation Research Part C: Emerging Technologies, 51 (2015) 29-40.

[26]. Zhang, X., Onieva, E., Perallos, A., Osaba, E., & Lee, V. C. Hierarchical fuzzy rule-based system optimized with genetic algorithms for short term traffic congestion prediction. Transportation Research Part C: Emerging Technologies, 43 (2014) 127-142.

[27]. Rewadkar, D., & Doye, D. Multi-objective auto-regressive whale optimisation for traffic-aware routing in urban VANET. IET Information Security, 12(4) (2018) 293-304.

[28]. Fu, R., Zhang, Z., & Li, L. Using LSTM and GRU neural network methods for traffic flow prediction. In Chinese Association of Automation (YAC), Youth Academic Annual Conference of IEEE. (2016) 324-328.

[29]. Chen, Y. Y., Lv, Y., Li, Z., & Wang, F. Y. Long short-term memory model for traffic congestion prediction with online open data. In Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on IEEE. (2016) 132-137.

[30]. Taamneh, M., Taamneh, S., & Alkheder, S. Clustering-based classification of road traffic accidents using hierarchical clustering and artificial neural networks. International journal of injury control and safety promotion, 24(3) (2017) 388-395.

[31]. Chen A, Pao HK, Lee YJ. Online Traffic Speed Forecasting Considering Multiple Periodicities and Complex Patterns.

[32]. Habtemichael, F. G., & Cetin, M. Short-term traffic flow rate forecasting based on identifying similar traffic patterns. Transportation Research Part C: Emerging Technologies, 66 (2016) 61-78.

[33]. Cai, P., Wang, Y., Lu, G., Chen, P., Ding, C., & Sun, J. A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting. Transportation Research Part C: Emerging Technologies, 62 (2016) 21-34.