Prediction of Average Speed Based on Relationships between Neighbouring Roads Using K-NN and Neural Network

https://doi.org/10.3991/ijoe.v16i01.11671

Bagus Priambodo
Universiti Kebangsaan Malaysia, Bangi, Malaysia
Universitas Mercu Buana, Jakarta, Indonesia

Azlina Ahmad (✉), Rabiah Abdul Kadir
Universiti Kebangsaan Malaysia, Bangi, Malaysia
azlinaivi@ukm.edu.my

Abstract—For decades, various algorithms to predict traffic flow have been developed to address traffic congestion. Traffic congestion or traffic jam occurs as a ripple effect from a road congestion in the neighbouring area. Previous research shows that there is a spatial correlation between traffic flow in neighbouring roads. Similar traffic pattern is observed between roads in a neighbouring area with respect to day and time. Currently, time series models and neural network models are widely applied to predict traffic flow and traffic congestion based on historical data. However, studies on relationships between road segments in a neighbouring area are still limited. These studies can be used to improve the accuracy of prediction of traffic flow. Hence, this study investigates the relationships of roads in a neighbouring area based on similarity of traffic condition. In our study, clustering method is used to divide the speed of traffic into four (4) categories: very congested, congested, clear and very clear. We used k-means clustering method to cluster condition of traffic flow on-road segments. However, using an unsupervised method like k-means, results of clustering using k-means may vary for each road in the neighbouring area. To address this issue, instead of using only clustering method (k-means), we applied the k-Nearest Neighbour (k-NN) method to classify the traffic condition in neighbouring roads. From the classification of traffic condition in neighbouring roads, we then determine the relationship between road segments. Results show that combination of k-means and k-NN method produced better results than using both correlation method and using the k-means method only.

Keywords—K-nearest neighbour; k-means clustering; neural network; prediction of traffic speed; the relationship between roads

1 Introduction

Due to increase in population and number of private cars in this modern era, traffic congestion has become significantly worse, not only leading to economic losses but also causing human stress[1] and environmental damages such as pollution [2]. Therefore, it is important for drivers to have traffic information which can assist their driving plans which includes changing their driving habits and driving paths [3]. Availability of traffic information can lead to chain changes in traffic flow condition in upstream and downstream of road segments in the same area.

Traffic congestion is a circumstance in which road users exceed the capacity of the road. Characteristic features of road congestion are slow speed, long travel time, and a long queue of cars on the road. Therefore, effective methods or models need to be developed to identify congested links, to analyses the relationship between the occurrence of congestion and increasing traffic flow, and to find congestion distribution in the network. Many factors may influence traffic congestion. Some studies use vehicle speeds, weather, incident, and special days to predict traffic flow using linear regression[4] and neural network models[5][6]. Other factors that influence traffic congestion that they are dynamic and interrelated. Traffic congestion or traffic jams can propagate from one road to other roads in neighbouring roads [7].

In this study, our main study is to find the relationship between roads in neighbouring roads based on the similarity of average speed and volume of vehicle. Some research study that there is spatial relationship between roads in neighbouring roads, they used connected roads [5] and intersection [8] as parameters to predict traffic flow, another study used correlation method to select roads as input for prediction traffic flow [6]. The temporal relationship also considers as a factor that influences traffic flow. Some study shows that a road always has similar traffic pattern on same working day or weekends at the same time interval [5]. Adjacent roads have a similar history in terms of road traffic condition during workdays or weekend [6].

Investigating similar traffic conditions at adjacent roads can lead us to traffic flow pattern with its neighbouring roads. Discovering relationships between roads in a neighbouring area can provide information to guide drivers in avoiding congested roads and those that are impacted by the congestion. This result also can be used to improve the accuracy of prediction of traffic flow.

2 Related Work

Neural networks are successfully used to predict traffic flows [5][6], such as used successfully in other fields[9]–[11]. Some studies used a Bayes classifier to predict traffic congestion [12][13]. Others used non-parametric regression k-NN to predict short-term traffic flow[14][15][16][17]. However, studies on relationships between road segments in a neighbouring area are relatively new and certainly need to be explored further.

Determining relationships between roads in a neighbouring area can provide information on the propagation of traffic congestion. This result can be applied as a tool to predict traffic flow propagation. From Figure 1, we anticipate that traffic flow on the road 158324 will be influenced by traffic flow on surrounding roads such as road 158446. This means that traffic flow on neighbouring roads will affect the average speed of vehicles on road 158324.

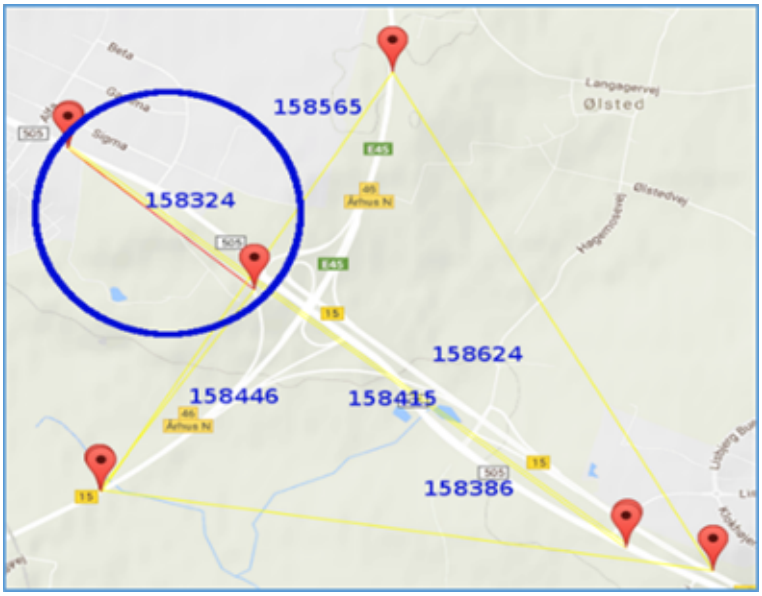

Fig. 1. Road 158324 and its neighbouring roads.

Research on relationships between roads [18], defined congestion correlation from road segment A to segment B with a certain distance d as: If congestion occurs in road segment A at time t and at time t + T, congestion occurs at road segment B. Another research applied data from sensors using correlation method based on mean of average speed at interval time to find relationships between roads segments [6] [19]. Other studies used connected road (upstream and downstream) using neural network [5][8], and k-NN [15]. Visualisation method was also used in several studies to investigate traffic flow patterns in neighbouring roads [1][7] [20].

3 Problem

Traffic condition is influenced by the number of vehicles and average speed of vehicles passing the roads (See Table 1). The number of vehicles and average speed are needed to determined traffic condition [21]. Previously we have conducted experiments to find correlation value between neighbouring roads using correlation method. However, correlation method allows the use of one variable, either average speed [19] or vehicle count. To deal with two or more variables, clustering is needed. Clustering is necessary to divide traffic conditions into four (4) categories:

- Very congested
- Congested
- Clear
- Very clear

There exists a relationship between road A to road B if traffic condition in road A is the same as road B, from time t + T. Road A and road B are within certain distance d. K-means method can be used for clustering traffic flow in neighbouring roads. However, using an unsupervised method like k-means, results of clustering using k-means may vary for each road in the neighbouring area. Therefore, it can be difficult to determine if a traffic condition in road A at time t is similar to traffic condition in neighbouring roads. To address this issue, instead of using only clustering method (k-means), we need to use classify method to classify neighbouring roads based on results from clustering in road A.

Previous research on prediction of short-term traffic flow has been conducted using k-NN, using k-NN with composite method [14], using three-layer k-NN [16], using k-NN with weight[22], using k-NN with tensor for calculating similarity traffic [17], and k-NN with connected link upstream or downstream[15]. However, their research did not focus on finding relationships between road segments.

Instead of using the k-NN method for regression, we proposed k-NN to classify traffic conditions in neighbouring roads based on results from clustering of traffic conditions of a particular road, since the k-NN classifier is considered as the most popular classifier for pattern recognition with its simplicity [23]. Relationships between roads were then obtained by calculating the number of similar traffic conditions at time t. Then, we observed the results by showing high relationship roads on the map. Next, the high relationship roads were used as factors for short-term prediction using neural network method. Finally, we compared our results that are using k-means and k-NN method with the results using k-means only and using correlation method.

4 Dataset and Method

4.1 Dataset

For our experiments, we used data set from IoT traffic sensor in Aarhus, Denmark [24][25][26]. The total number of IoT sensors is 449, and their location is as shown in Fig 2. However, in this study we only discuss road 158324 with its neighbouring roads (20 IoT sensors) and road 193294 with its neighbouring roads (60 IoT sensors). We define neighbouring roads as roads less than four (4) kilometres from target road.

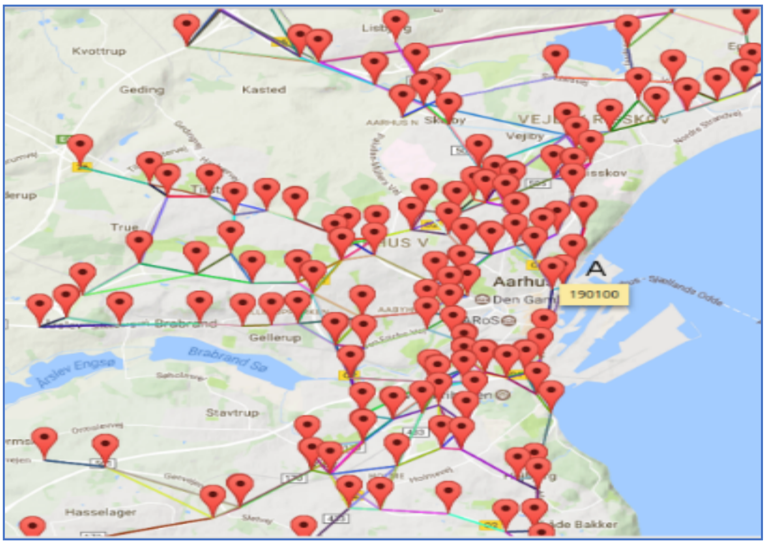

Fig. 2. Map of 449 IoT traffic sensors in the city of Aarhus, Denmark.

Table 1. Example of traffic data taken from sensor 190100

Time	Avg Speed	Median	TIME	Vehicle Count
376	14	376	2014-08-01T08:30:00	4
225	23	225	2014-08-01T08:40:00	3
285	18	285	2014-08-01T08:50:00	2

4.2 Methodology

Our methodology is described in fig 3.

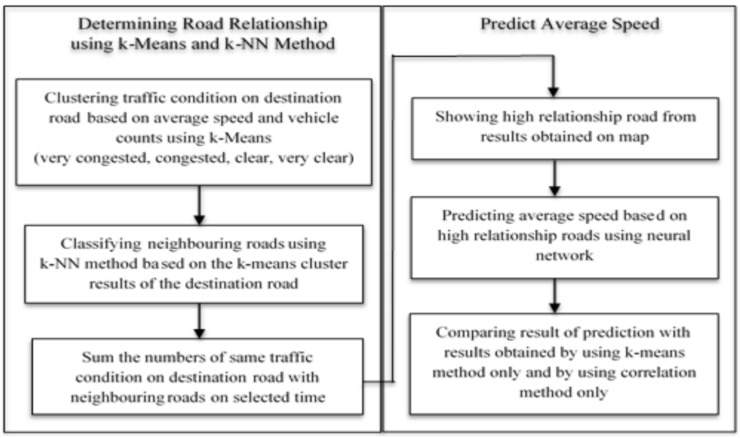

Fig. 3. Methodology to predict average speed based on the relationship between roads

Clustering road destination based on traffic condition: Traffic flow condition is influenced by traffic speed of vehicles passing through and the number of vehicles [21]. We proposed k-means clustering method to cluster traffic conditions into four (4) based on traffic speed and numbers of vehicles. The four (4) traffic conditions are very congested, congested, clear and very clear.

K-Means is an algorithm of grouping data which divides the existing data into k groups. It is one of the simplest clustering algorithm [27]. The k-means method attempts to cluster the existing data into k cluster, where data in one cluster have similar characteristics as each other and have different characteristics from the data in the other cluster. In other meanings, this method seeks to minimize variations between existing data within a cluster and maximize variation with existing data in other clusters [28]. The procedure used in performing optimization using k-means is as follows:

- Specify k=4, as we would like to cluster traffic conditions into four: Very Congested, Congested, Clear and Very Clear.
- Randomly select 4 distinct data points as initial cluster means.
- Then calculate the distance between each of cluster means and all other points using Euclidean distance formula.
- Assign each point to the cluster having the closest mean.
- For each of the k clusters recalculate the cluster centroid (means) by calculating the new mean value of all the data points in the cluster.
- Repeat steps 3 to 5 until the centroids do not change or the maximum number of repetitions, we set the maximum number of repetitions is 10.

The total within the sum of square or the total within-cluster variation is defined as:

$$\sum_{k=1}^{4} W(C_k) = \sum_{k=1}^{4} \sum_{x_i \in C_k} (x_i - \mu_k)^2 \quad (1)$$

Where:
x_i is data point belonging to the cluster C_k
μ_k is the mean value of the points assigned to the cluster C_k

In this paper, we conduct an experiment to find the relationship between road 158324 and road 193294 with its neighbouring roads. We started by clustering both roads. The results of clustering road 158324 and road 193294 are shown in Figure 4.

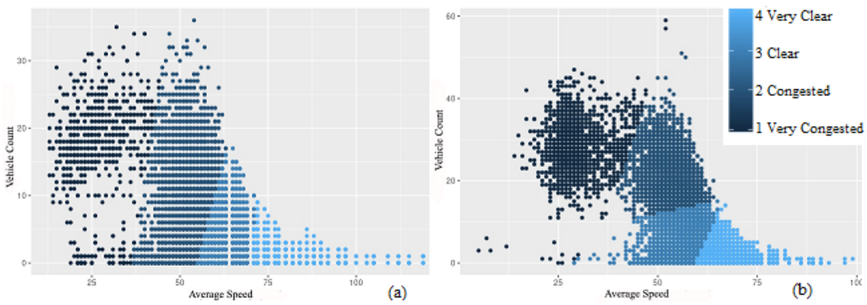

Fig. 4. Clustering result of (a) road 158324 traffic condition, (b) road 193294 traffic condition

Classifying neighbouring roads using k-NN: In this phase, we will classify neighbouring roads based on traffic conditions using results of clustering from the previous phase. Classification is a technique for grouping data into their appropriate classes [29]. After the clustering phase, the results obtained were used as the data training set (traffic condition on road 158324 and road 193294). K-NN then used to classify an unknown traffic condition in neighbouring roads using Euclidean distance using formula (2).

$$d = \sqrt{(avSpeed_a - avSpeed_b)^2 + (vCount_a - vCount_b)^2} \quad (2)$$

Where:
d is the distance between road A and road B
$avSpeed_a$ is the average speed of road A
$avSpeed_b$ is the average speed of road B
$vCount_a$ is vehicle count (volume) of road A
$vCount_b$ is vehicle count (volume) of road B

For example, we want to classify the state of traffic condition on road 158386 with given average speed value 73 and vehicle count value 2 based on traffic condition on road 158324. After normalization of data, we calculate the distance between given value with traffic condition cluster value in road 158324. This is to classify the road as a very congested, congested, clear or very clear. If we use the k-NN algorithm k = 1, the traffic condition is 3 or in clear state and the distance is 0,07. If we use the k-NN algorithm with k = 3 instead, it performs a vote among the three nearest neighbours: clear, clear, and very clear. Since the majority class among these neighbours is clear (two of the three votes), the traffic condition again is classified as clear, as seen in Table 2. The number K is typically chosen as the square root of N training data[30]. In this study, clustering result row is 32041 where the square root is 179, however we used k=135 to classify traffic condition. Some of the results obtained are displayed in Table 3.

Table 2. Example of classifying traffic condition using k-NN method

TIME	Average Speed	Vehicle Count	Cluster	Distance	Nearest Rank
3:55:00	82	0	4 (Very Clear)	0.33	
4:00:00	82	0	4 (Very Clear)	0.33	3
4:05:00	97	1	4 (Very Clear)	0.85	
4:10:00	71	2	3 (Clear)	0.07	1
4:15:00	53	2	2 (Congested)	0.71	
4:20:00	53	2	2 (Congested)	0.71	
4:25:00	54	1	2 (Congested)	0.67	
4:30:00	66	2	3 (Clear)	0.25	2

Table 3. The results of classifying traffic condition in neighbouring roads using k-NN with k=135.

TIME	Road 158324	Neighbouring Roads					
		158386	***158415***	***158624***	***158715***	***158895***	***173011***
2/18/2014 4:00	Clear	Very Clear	Clear	Clear	Clear	Clear	Congested
2/18/2014 4:05	Clear	Very Clear	Clear	Clear	Very Clear	Clear	Congested
2/18/2014 4:10	Clear	Very Clear	Clear	Very Clear	Very Clear	Clear	Congested

Calculating the same traffic condition in neighbouring roads: To obtain the relationship between roads, we calculate the number of roads in a neighbouring area which has same traffic condition (same cluster) as target road using formula (3). For example, between 06:00 am, and 07:00 am, there are 609 times road 158386 have same traffic condition with road 158324, out of 774 data. Then the relationship value that road 158324 and road 158386 when they have same traffic conditions is 609/774 = 0.7868. Table 4 shows the relationship values of neighbouring roads state when they have same traffic condition state with road 158324.

$$Relationship\ value = \frac{1}{n}\sum_{j=2}^{n}\left(Road_1 == Road_j\right) \tag{3}$$

Table 4. Relationship value of road 158324 with its neighbouring roads

With Road	Numbers of similar state at same time	Relationship Value
158386	609	0.786821705426357
158715	420	0.542635658914729
158595	404	0.521963824289406
173011	393	0.507751937984496

5 Result and Discussion

5.1 Results

We analysed the results obtained using both k-means and k-NN, using the correlation method and using the k-means method only. Firstly, we showed the highest relationship roads using both methods on the map. Secondly, we used the highest relationship road as input factor for predicting traffic flow using neural network method. The accuracy of prediction results was calculated.

We observed that traffic flow pattern on weekdays (Monday to Thursday) and weekends (Friday to Sunday) for roads 158324 and 193294 was different as. For road 158324, we calculated the prediction of traffic flow using data from 06.00 am to 07.00 am on Monday (duration of 50 minutes). Then, we calculated similar traffic condition based on results obtained using k-NN for the same time using weekdays data. For comparison, we calculated the correlation of all neighbouring roads based on mean value of average speed for interval 06.00 AM to 07.00 AM from 13-02-2014 to 02-06-2014. We chose interval 06.00 AM to 07:00 AM because at this time, congestion occurred (Refer to Figure 5). For road 193294, we performed prediction of traffic flow for 50 minutes starting from 15.00 PM to 16.00 PM on Monday. Then, we calculated similar traffic condition based on results obtained using k-NN at interval between 15.00 pm to 16.00 pm from weekday's data. For comparison, we calculated correlation based on mean value of average speed for interval between 15.00 PM to 16.00 PM because at this time, congestion occurred (Refer to Figure 6). The correlation results of road 158324 and its neighbouring roads are shown in Table 5. The correlation results of road 193294 and its neighbouring roads are shown in Table 6. The visualisation of the results are shown in Figure 8 and Figure 9. The stars in these figures represent six highest relationship values. In road 158324 relationship roads result using the correlation method, k-means only and combine k-means k-NN almost similar, both relationship roads are near to road 158324 and connected with road 158324 as seen in Fig 8. But there is a different result in road 193294. In road 193294, correlation method shows high relationship road with 193294 but there is some high relationship road location are distance away and not connected with road 193294. Also, when using k-means only method, there is some high relationship road location are distance away and not connected with road 193294. A different result is obtained using combine of k-means k-NN method, all high relationship road is connected to road 193294.

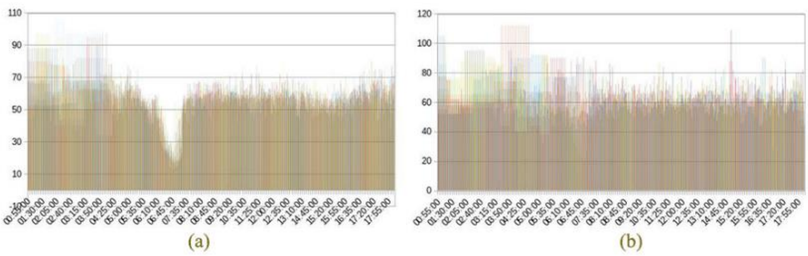

Fig. 5. The traffic pattern on road 158324, (a) pattern on weekdays, (b) pattern on all days

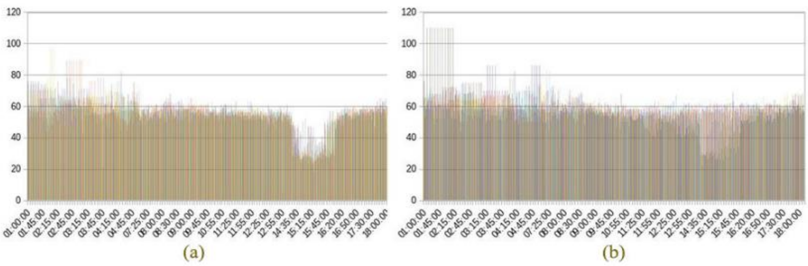

Fig. 6. The traffic pattern on road 193294, (a) pattern on weekdays, (b) pattern on all days

To find the relationship road with road 158324 we compared three methods:

- The correlation method
- K-means

- Combination of k-means and k-NN method

When using the correlation method, we selected six (6) roads with relationship value above 0.5. On the other hand, while using k-Means, we obtained three (3) roads with relationship value above 0.5. When we combine both methods; k-means and k-NN, we selected four (4) roads with relationship value 0.5 and above. The results for other roads using these three methods are shown in Table 5. We applied the same method to road 193294 and the results obtained are given in Table 6.

Table 5. Relationship value result of road 158324 and its neighbouring roads

Using Correlation		Using K-Means		Using k-NN	
Roads	***Correlation***	***Roads***	***Relationship***	***Roads***	***Relationship***
158386	0.79	158386	0.66	158386	0.78
158595	0.64	158895	0.54	158715	0.54
158536	0.61	173225	0.51	158595	0.52
158624	0.59	158415	0.49	173011	0.50
171969	0.58	158715	0.48	158415	0.47
173225	0.52	158536	0.43	158624	0.40

Table 6. Relationship value result of road 193294 and its neighbouring roads

Using Correlation		Using K-Means		Using k-NN	
Roads	***Correlation***	***Roads***	***Relationship***	***Roads***	***Relationship***
193348	0.8	193348	0.72	193348	0.59
194931	0.66	194931	0.67	194931	0.58
193268	0.61	194878	0.58	194878	0.50
193322	0.58	195312	0.55	195658	0.43
158595	0.56	193402	0.55	179444	0.42
195923	0.54	195737	0.54	193268	0.42

We plotted the results of each method on the map as shown in Figure 7. Road 158324 with high relationship values are represented with black icons and black lines. From the figure, we observed that the highest relationship roads generated using correlation method, k-means, and k-NN are at different locations but connected similarly. Most importantly, we observed that the relationship roads are connected to each other.

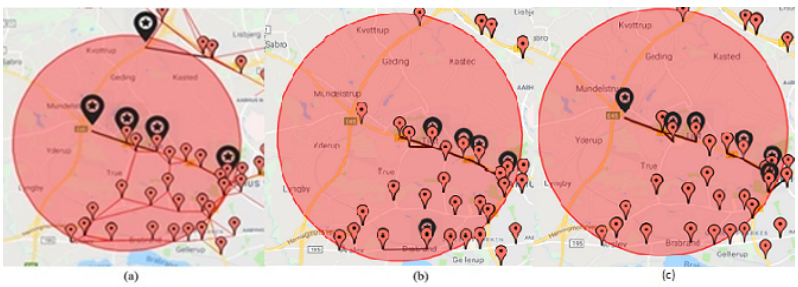

Fig. 7. Map plotting of results for road 158324 with its neighbouring roads (a) using correlation method (b) using k-means and (c) using k-NN.

Likewise, for road 193294, we plotted the results of roads with high relationship values on the map as shown in Figure 8. These roads are represented with black icons and black lines. We observed that when using the correlation method, there is a road with a high correlation value (black icon) that is not connected with road 193294 (See Figure 9 (a)). Similar results were obtained when using the k-means method. There is a road which is not connected to road 193294 (See Figure 8 (b)). However, when the k-NN method was used, all the roads with high relationship values are connected to road 193294 (See in Figure 8 (c)).

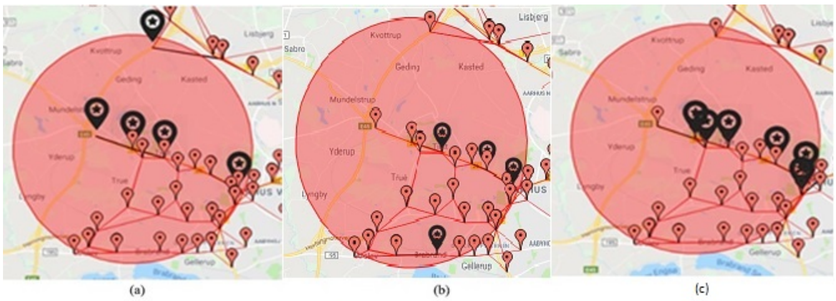

Fig. 8. Map plotting of results for road 193294 with its neighbouring roads (a) using correlation method (b) using k-means and (c) using k-NN

For further analysis of high relationship road for road 158324, we used results from all methods as input factors to perform prediction using neural network method. The results of prediction are shown in Table 7 and the errors of prediction in Table 9. We represented the results using a bar chart for better comparison as seen in Figure 9. From Figure 9, we observed that prediction using k-NN method and the highest relationship roads as input parameters has the lowest error when compared with using k-means or correlation method.

Table 7. Results of prediction using NN with correlation, k-Means, and k-NN on road 158324

TIME	Actual	Correlation	k-Means	k-NN
06:10:00	24	48.05	56.15	34.47
06:15:00	25	50.49	58.39	39.8
06:20:00	24	52.78	58.29	23.52
06:25:00	44	64.88	61.13	29.42
06:30:00	57	64.91	60.84	43.23
06:35:00	59	63.75	61.55	58.21
06:40:00	57	59.65	61.76	59.47
06:45:00	56	59.09	61.05	59.05
06:50:00	56	63.64	58.02	61.18
06:55:00	50	64.33	58.99	61.91
07:00:00	50	64.65	60.16	62.4

For further analysis of road 193294, we used the high relationship values obtained from all methods as input factors for prediction using neural network. The results of prediction are shown in Table 8 and the related errors are given in Table 9. We presented the results of prediction using bar chart as seen in Figure 9. From the figures, we see that the results of prediction based on highest relationship roads using k-NN has the lowest error if compared to k-means and correlation methods.

Table 8. Results of prediction using NN with correlation, k-Means and k-NN on road 193294

TIME	Actual	Correlation	k-Means	k-NN
15:10:00	36	60.33	53.85	53.69
15:15:00	45	60.39	50.83	50.12
15:20:00	52	59.76	45.38	49.80
15:25:00	50	59.34	50.15	53.31
15:30:00	50	59.95	54.53	54.72
15:35:00	55	60.84	56.96	55.79
15:40:00	54	61.59	55.63	55.28
15:45:00	54	61.67	54.41	54.63
15:50:00	56	61.14	55.20	55.00
15:55:00	59	61.85	57.11	56.45
16:00:00	55	61.75	57.49	57.06

Table 9. Errors of short-time prediction on road 158324 and road 193294

	Road 158324			Road 193294		
Error	Correlation	K-Means	k-NN	Correlation	k-Means	k-NN
MAD	14.02	9.11	6.44	9.33	4.01	3.76
RMSE	16.71	13.11	7.94	10.9	6.29	5.97
MAPE	43.02	31.37	18.13	20.04	9.2	8.69

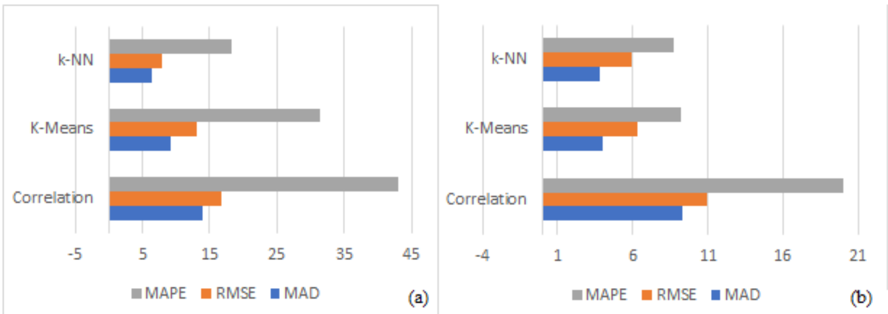

Fig. 9. Comparison of errors for short time prediction on (a) road 158324 and (b) road 193294

5.2 Discussion

From the results of all the experiments, we compared results using the correlation method, using the clustering k-means method and using combine of k-means and k-NN method. We found that varying results were obtained for road 158324 and road 193294. Roads with highest relationship values generated for road 158324 are connected, but for road 193294, one of the roads generated was located a distance away and not connected with road 193294. Better prediction results were obtained using combination of k-means and the k-NN methods. All roads with high relationship values generated using this combination were connected to each other. We further analysed choosing an input factor for neural network from the high relationship roads based on connection and location. As a conclusion, predictions of average speed based on roads with high relationship values using combination of k-means and the k-NN methods produce better results than using correlation and k-means methods only.

6 Conclusion

The main aim of our experiments in this study is to investigate and obtain the relationship between roads in a neighbouring area for predicting traffic flow. In this paper, we considered traffic congestion as being influenced by traffic speed and number of vehicles passing on the roads. We used k-means clustering method to cluster roads into four. Then, we classified neighbouring roads using k-NN and calculated the similarity of the traffic conditions to obtain the relationship values for roads in a neighbouring area. Our investigations showed that by using combination of k-means and k-NN methods to identify roads with high relationship values, results obtained were more consistent and accurate compared to using the correlation or the k-means methods.

7 Acknowledgement

The authors would like to thank the Ministry of Education (Malaysia) for funding this research work through Fundamental Research Grant Scheme (Project Code: FRGS/1/2018/ICT02/UKM/02/8). The authors would also like to extend the acknowledgement for the use of service and facilities of the Intelligent Visual Data Analytics Lab at Institute of Visual Informatics, Universiti Kebangsaan Malaysia.

8 References

[1] N. Petrovska and A. Stevanovic, “Traffic Congestion Analysis Visualisation Tool,” IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC, vol. 2015-Octob, pp. 1489–1494, 2015. https://doi.org/10.1109/itsc.2015.243

[2] Y. Jiang, R. Kang, D. Li, S. Guo, and S. Havlin, “Spatio-temporal propagation of traffic jams in urban traffic networks,” Phys. Soc., 2017.

[3] G. Zhu, K. Song, P. Zhang, and L. Wang, “A traffic flow state transition model for urban road network based on Hidden Markov Model,” Neurocomputing, vol. 214, pp. 567–574, 2016. https://doi.org/10.1016/j.neucom.2016.06.044

[4] H. Dong, L. Jia, X. Sun, C. Li, and Y. Qin, “Road traffic flow prediction with a time-oriented ARIMA model,” in NCM 2009 - 5th International Joint Conference on INC, IMS, and IDC, 2009, no. 1, pp. 1649–1652. https://doi.org/10.1109/ncm.2009.224

[5] E.-M. Lee, J.-H. Kim, and W.-S. Yoon, “Traffic Speed Prediction Under Weekday, Time, and Neighboring Links’ Speed: Back Propagation Neural Network Approach,” in Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and

Methodological Issues, 2007, no. Mic, pp. 626–635. https://doi.org/10.1007/978-3-540-74171-8_62

[6] B. Priambodo and A. Ahmad, "Traffic flow prediction model based on neighbouring roads using neural network and multiple regression," J. Inf. Commun. Technol., vol. 17, no. 4, pp. 513–535, 2018.

[7] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. Van De Wetering, "Visual traffic jam analysis based on trajectory data," IEEE Trans. Vis. Comput. Graph., vol. 19, no. 12, pp. 2159–2168, 2013. https://doi.org/10.1109/tvcg.2013.228

[8] Z. Zhou and K. Huang, "Study of Traffic Flow Prediction Model at Intersection Based on R-FNN," ISBIM 2008 Int. Semin. Bus. INFORMATION\nMANAGEMENT, VOL 1, pp. 531–534, 2009.

[9] W. Liu, "An Improved Back-Propagation Neural Network for the Prediction of College Students ' English Performance," Int. J. Emerg. Technol. Learn., pp. 130–142, 2019. https://doi.org/10.3991/ijet.v14i16.11187

[10] Y. Lan and X. Chen, "Application of immune feedback control algorithm based on BP network approximation in intelligent vehicle steering system," Int. J. online Biomed. Eng., vol. 15, no. 9, pp. 71–79, 2019. https://doi.org/10.3991/ijoe.v15i09.10584

[11] G. Zhou, Y. Ji, X. Chen, and F. Zhang, "Artificial neural networks and the mass appraisal of real estate," Int. J. Online Eng., vol. 14, no. 3, pp. 180–187, 2018. https://doi.org/10.3991/ijoe.v14i03.8420

[12] J. Kim and G. Wang, "Diagnosis and Prediction of Traffic Congestion on Urban Road Networks Using Bayesian Networks," in Australasian Transport Research Forum 2016, 2016, no. 2595, pp. 1–21. https://doi.org/10.3141/2595-12

[13] S. S. Aung and T. T. Naing, "Naïve Bayes Classifier Based Traffic Detection System on Cloud Infrastructure," 2015 6th Int. Conf. Intell. Syst. Model. Simul., pp. 193–198, 2015. https://doi.org/10.1109/isms.2015.45

[14] M. D. Kindzerske and D. Ni, "Composite Nearest Neighbor Nonparametric Regression to Improve Traffic Prediction," Transp. Res. Rec., no. 1993, p. pp 30-35, 2007. https://doi.org/10.3141/1993-05

[15] B. Yu, X. Song, F. Guan, Z. Yang, and B. Yao, "k-Nearest Neighbor Model for Multiple-Time-Step Prediction of Short-Term Traffic Condition," J. Transp. Eng., vol. 142, no. 6, p. 04016018, 2016. https://doi.org/10.1061/(asce)te.1943-5436.0000816

[16] 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," J. Transp. Technol., vol. 6, no. 4, pp. 200–206, 2016. https://doi.org/10.4236/jtts.2016.64020

[17] Y. Wu, H. Tan, P. Jin, B. Shen, and B. Ran, "Short-Term Traffic Flow Prediction Based on Multilinear Analysis and k- Nearest Neighbor Regression," in 15th International Conference on Transportation Professionals (CICTP-2015), 2015, pp. 557–569. https://doi.org/10.1061/9780784479292.051

[18] Y. Wang, J. Cao, W. Li, and T. Gu, "Mining Traffic Congestion Correlation between Road Segments on GPS Trajectories," 2016 IEEE Int. Conf. Smart Comput. SMARTCOMP 2016, 2016. https://doi.org/10.1109/smartcomp.2016.7501704

[19] B. Priambodo and A. Ahmad, "Predicting traffic flow based on average speed of neighbouring road using multiple regression," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2017, vol. 10645 LNCS, pp. 309–318. https://doi.org/10.1007/978-3-319-70010-6_29

[20] A. Anwar, T. Nagel, and C. Ratti, "Traffic origins: A simple visualization technique to support traffic incident analysis," IEEE Pacific Vis. Symp., pp. 316–319, 2014. https://doi.org/10.1109/pacificvis.2014.35

[21] W. X. Wang, R. J. Guo, and J. Yu, "Research on road traffic congestion index based on comprehensive parameters: Taking Dalian city as an example," Adv. Mech. Eng., vol. 10, no. 6, pp. 1–8, 2018.
[22] L. Zhang, Q. Liu, W. Yang, N. Wei, and D. Dong, "An Improved K-nearest Neighbor Model for Short-term Traffic Flow Prediction," Procedia - Soc. Behav. Sci., vol. 96, no. Cictp, pp. 653–662, 2013. https://doi.org/10.1016/j.sbspro.2013.08.076
[23] S. Hota and S. Pathak, "KNN classifier based approach for multi-class sentiment analysis of twitter data," Int. J. Eng. Technol., vol. 7, no. 3, pp. 1372–1375, 2018. https://doi.org/10.14419/ijet.v7i3.12656
[24] S. Kolozali, M. Bermudez-Edo, D. Puschmann, F. Ganz, and P. Barnaghi, "A knowledge-based approach for real-time IoT data stream annotation and processing," in Proceedings - 2014 IEEE International Conference on Internet of Things, iThings 2014, 2014 IEEE International Conference on Green Computing and Communications, GreenCom 2014 and 2014 IEEE International Conference on Cyber-Physical-Social Computing, CPS 20, 2014, no. iThings, pp. 215–222. https://doi.org/10.1109/ithings.2014.39
[25] S. Bischof, A. Karapantelakis, A. Sheth, and A. Mileo, "Semantic Modelling of Smart City Data Description of Smart City Data," in W3C Workshop on the Web of Things Enablers and services for an open Web of Devices, 2014, pp. 1–5.
[26] S. Bischof, C.-S. Karapantelakis, Athanasios Nechifor, A. Sheth, A. Mileo, and P. Barnaghi, "Real Time IoT Stream Processing and Large-scale Data Analytics for Smart City Applications," in Real Time IoT Stream Processing and Large-scale Data Analytics for Smart City Applications, 2014. https://doi.org/10.1016/b978-0-12-818014-3.00004-8
[27] S. Syed Azimuddin and K. Desikan, "A Simple Density with Distance Based Initial Seed Selection Technique for K Means Algorithm," J. Comput. Inf. Technol., vol. 25, no. 4, pp. 291–300, 2018. https://doi.org/10.20532/cit.2017.1003605
[28] A. Kesumawati and D. Setianingsih, "A segmentation group by Kohonen Self Organizing Maps (SOM) and K -means algorithms (case study: Malnutrition cases in Central Java of Indonesia)," Int. J. Adv. Soft Comput. its Appl., vol. 8, no. 3, pp. 110–115, 2016.
[29] D. Fitrianah, "Feature Exploration for Prediction of Potential Tuna Fishing Zones," Int. J. Inf. Electron. Eng., vol. 5, no. 4, pp. 270–274, 2015.
[30] P. Nadkarni, "Core Technologies: Data Mining and 'Big Data,'" Clin. Res. Comput., pp. 187–204, 2016.

9 Authors

Bagus Priambodo is a PhD student in Institute of Visual Informatics, and he is also a lecturer at Department of Information Systems, Faculty of Computer Science, Universitas Mercu Buana, Jakarta, Indonesia. bagus.priambodo@mercubuana.ac.id

Dr Azlina Ahmad is an Associate Professor at the Institute of Visual Informatics (IVI), Universiti Kebangsaan Malaysia (UKM) where currently she is one of the senior research fellows. Azlina received her Ph.D. in Multimedia from University of Malaya. Her research interests lie in the area of visualization, augmented and virtual reality, educational technology, and mathematics. She has been the principal investigator on research projects sponsored by Ministry of Science and Technology and Ministry of Education. She has experience in Project Management and is a certified Appraisal Team Members (ATM) for Capability Maturity Model Integration (CMMI) Level 2 from Carnegie-Melon University, USA. Azlina has collaborated actively with researchers in several other disciplines such as Islamic banking and finance, social science and humanities and engineering. E-mail: azlinaivi@ukm.edu.my

Dr Rabiah Abdul Kadir is an Associate Professor and Senior Research Fellow at the Institute of Visual Informatics, Universiti Kebangsaan Malaysia (UKM). She received her PhD (2007) in Computer Science from UKM. She was seconded as Assistant Professor at Najran University, Saudi Arabia for a year, from September 2012 to August 2013. In June 2014, she joined Institute of Visual Informatics, UKM as a Research Fellow until present. Her areas of research are in intelligent computing and computational linguistics. She is also active in research on big data analytics, semantic knowledge repository and extraction, and intelligent learning management system. E-mail: rabiahivi@ukm.edu.my

Article submitted 2019-09-13. Resubmitted 2019-10-22. Final acceptance 2019-10-22. Final version published as submitted by the authors.