Investigating Relationships Between Roads Based on Speed Performance Index of Road on Weekdays

Bagus Priambodo[1,2], Azlina Ahmad[1], Rabiah Abdul Kadir[1]

[1]Institute of Visual Informatics, Universiti Kebangsaan Malaysia,
Bangi, Malaysia
azlinaivi@ukm.edu.my
[2]Faculty of Computer Science, Universitas Mercu Buana, Jakarta, Indonesia
bagus.priambodo@mercubuana.ac.id

Abstract. Traffic congestion or traffic jam occurs as a ripple effect from a road congestion in the neighbouring area. Previous studies show that spatial correlation is exist between roads in neighbouring roads. There is similar traffic pattern observed between roads in a neighbouring area with respect to day and time. Nowadays, various machine learning model have been developed to predict traffic flow to provide traffic information. However, studies on relationships between road segments in a neighbouring area are still limited. It is important to investigate these relationships because they can assist drivers in avoiding roads which are impacted by road congestion or by a roadblock in a neighbouring area. Hence, this study investigates relationships of roads in a neighbouring area based on similarity of traffic condition. Traffic condition is influenced by number of vehicles and average speed of vehicles. In our study we determine traffic condition based on speed performance index of road in interval time. We used k-means clustering method to cluster condition of traffic flow on road segments. The experiments show that relationship roads can be revealed by clustering traffic condition in interval time.

Keywords: Relationship between roads, Roads Clustering, Speed Performance Index, K-Means Clustering

1 Introduction

The increasing number of vehicles in urban areas lead to worsen traffic congestion. Not only it causes environmental pollution, traffic congestion also causes stress and economic losses [1]. Thus, it is important to provide traffic information to drivers which can assist their driving plans and at the same time improve their driving habits [2]. Furthermore, the availability of traffic information can affect changes in flow of traffic in an area.

Traffic congestion is a situation where the vehicle exceeds capacity of the road. Some indications that a road is congested include low average speed, longer travel time than usual, and long queue of cars. Therefore, effective methods or models need to be developed to identify congested links, to analyse the relationship between the occurrence of congestion and increasing traffic flow, and to find congestion distribution in a road network. Some research used neural network for prediction of traffic flow based on speed of traffic in neighbouring links using all day's data [3]. A research by

H. Badioze Zaman et al. (Eds.): IVIC 2019 LNCS 11870, pp. 582-591, 2019.
https://doi.org/10.1007/978-3030-34032-2_51

Zhou & Huang [4] used neural network to predict traffic flow on road intersections using all day's data.

Other studies revealed that on the same day (working day or weekends), traffic on the road has similar pattern at the same time interval [5]. During working days or weekend, adjacent roads have similar road traffic condition based on historical data [6]. Investigating similar traffic conditions at adjacent roads can lead to traffic flow patterns of the neighbouring roads. Discovering relationships between roads in a neighbouring area can provide information on roads that are impacted by congestion thus can be a guide to drivers in avoiding congested roads.

2 Related Work

Neural network models [2][3][4] and time series [1], are commonly applied to forecast traffic flow and traffic congestion. Some studies used probability method such as Bayes classifier [5][6]. Others used k-NN as non-parametric model to predict short-term traffic flow [7][8][9][10]. However, studies on relationship between road segments in a neighbouring area are relatively new and certainly need to be explored further.

Research on relationships between road segments used extracted historical data of taxi GPS trajectory to study the correlation between two roads segments [11]. In their study, they 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, then congestion occurs at road segment B. Another research applied data from sensors using correlation method to find relationships between roads segments [12]. Visualization method was also used in several studies to investigate traffic flow patterns in neighbouring roads [13][14] [15].

3 Problem

One of the complexities related to predicting traffic congestion is unpredictability. Another related complexity is the behaviour of traffic congestion that is dynamic and interrelated. Many factors affecting traffic flow such as speed of vehicle, weather, special days or events and also occurrence of accidents. These complexities may influence the prediction of traffic flow on neighbouring roads that are affected by a congested road segment in the same area.

Congestion in an urban area, generally will spread through road networks due to the increase in traffic volume. Many types of research on traffic flow produced promising results based on historical traffic data. Unfortunately, spatial-temporal propagation of traffic congestions in an urban area is still unclear [16]. Monitoring and understanding traffic congestions are complex and unpredictable. Sometimes traffic congestions occur, and sometimes they do not. Furthermore, traffic congestions affect traffic flow in the same area because they will propagate to neighbouring roads in the same area [14][16].

Determining relationships between roads in a neighbouring area can provide information on the propagation of traffic congestion to traffic management offices. This information will assist them in performing traffic flow engineering. From Figure 1, we anticipate that traffic flow on road 158324 will be influenced by traffic flow on surrounding roads such as road 158446. This means that traffic flow on surrounding roads will be affected by the traffic condition on road 158324.

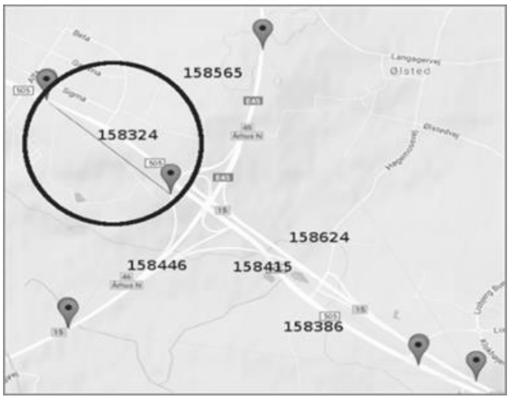

Fig. 1. Road 158324 and its surrounding roads.

Traffic condition is influenced by the number of vehicles and the average speed when passing by the roads (See Table 1). Previously, we have conducted experiments to find correlation value between neighbouring roads using the correlation method. However, using the correlation method, correlation value vary from one area to another area [4][12]. There exists a relationship between road A to road B if traffic condition in road A and road B have the same traffic condition from time t + T. Road A and road B are in the same area within a certain distanced d. K-means method can be used for clustering traffic flow in neighbouring roads. However, using an unsupervised method like k-means can produce results that vary for each road in the neighbouring area. Therefore, it would be difficult to determine if the traffic condition in road A at time t is similar to traffic condition in neighbouring roads. To address this issue, we calculate speed performance index of each road. In this experiment, we considered data on weekdays (Monday to Thursday) only. We cluster similarity of speed performance index of each road using k-Means clustering method. Then, we search the best k by using Bayes and Chi-square to evaluate the result of clustering by adjusting k value.

4 Dataset and Method

4.1 Dataset

In this study, the data set that we used are collected from IoT traffic sensor in Aarhus, Denmark [17][18][19]. A total of 449 sensors were installed, as shown in Fig 2.

Fig. 2. Map of the location of 449 IoT traffic sensors in the city of Aarhus, Denmark. For example, the name of the sensor at location A is 173225. This sensor is placed from Hinnerup Street Nordjyske Motorvej 0 Aarhus, Aarhus 15 Aarhus Denmark, the distance between the two sensors 3253 meters. We conduct experiment using average speed and time for calculating speed performance index. More details of this sensor are present in Table 1 and Table 2.

Table 1. Details of traffic data taken from sensor 173225.

Start	End	Point 1	Point 2	Observation
10/1/2014 1:45:00 AM	11/13/2014 10:40:00 AM	Coordinates(lat,long): 56.2348975970264, 10.125013142824, City: Hinnerup Street:Nordjyske Motorvej 0 Postal Code: 8382	Coordinates(lat,long): 56.2138565106957, 10.144907008598. City: Aarhus Street: 15 0 Postal Code: 8200	Distance between two sensors (meters): 3253 Duration (seconds): 100 Organization: COWI Road type: STREET NDT in KMH: 117 EXT ID: 192

Table 2. Details of traffic data taken from sensor 173225.

Avg Measured Time	Avg Speed	Median Measured Time	TIME STAMP	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

To determine the relationship between roads, we need to find similarity of traffic condition in neighbouring roads. Traffic flow condition is influenced by traffic speed of vehicles passing through and a number of vehicles. In this paper, traffic condition are determined based on the speed performance index [20]. We cluster neighbouring roads based on speed performance index using k-Means. We consider road A as a neighbouring road to road B if the distance between them is approximately four (4) kilometers. Details of our methodology are represented in Figure 3.

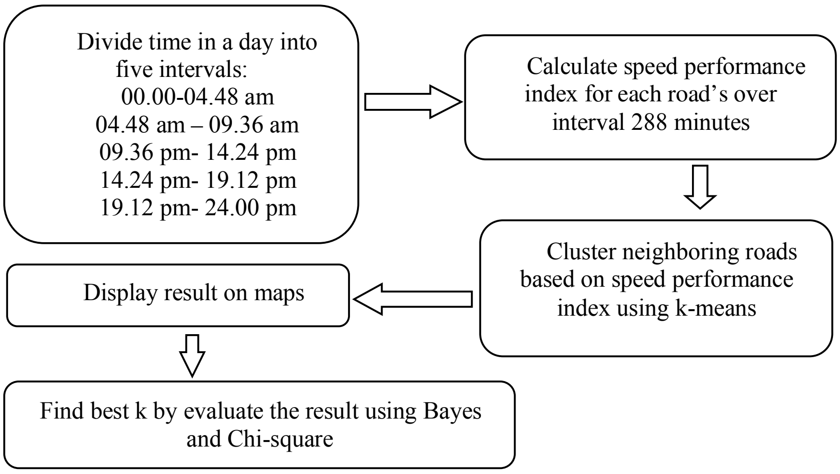

Fig. 3. Methodology to determine relationships of neighbouring roads based on speed performance index using k-means

We follow the speed performance index based on the travel speed from [20], with some adjustments. We define maximum speed limit as 160 km/hour. Instead of calculating speed performance index hourly, we calculate speed performance index for every 288 minutes defined in (0). BTMB has two thresholds (25, 50) as criterion of classification of urban traffic state. This study uses the speed performance index to measure the road traffic state but adopt one threshold only that is 0-33 for congestion. The speed performance index is calculated for neighbouring roads every day from February 2014 until October 2014.

$$R_v = \frac{v}{V_{max}} \times 100 \qquad (\mathbf{0})$$

Rv : the speed performance index;
v : the average travel speed, km/h;
Vmax : the maximum permissible road speed, km/h.

K-means is one of the best and simplest clustering algorithm [17] available today. It is a method of grouping data which divides the existing data into two or more cluster. The k-means method attempts to cluster the existing data into more than one cluster.

Data in one cluster has the same characteristics with each other but has different characteristics from the data in the other clusters. In other words, this method seeks to minimize variations between existing data within a cluster and maximize variation with existing data in other clusters [21]. The procedure used in performing optimization using k-means is as follows:

1. Specify k=n, as we would like to adjust similarity result based on k.
2. Randomly select distinct data points as initial cluster means.
3. Then, calculate the distance between each of cluster means and all other points using Euclidean distance formula.
4. Assign each point to the cluster having the closest mean.
5. 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.
6. Repeat steps 3 to 5 until the centroids do not change or the maximum number of repetitions is reached (we set the maximum number of repetitions as 1000).

The total within sum of square or the total within-cluster variation is defined (1) 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

5 Results and Discussion

We calculate the speed performance index from February 2014 until October 2014. In this study, we present and discuss the results of road 158324 and road 193294 with their neighbouring roads. For road 158324 we use interval II (04.48 am – 09.36 am), and for road 193294 we use interval III (09.36 am- 2.24 pm). This is because congestion occurred over these time intervals.

Results of clustering for neighbouring roads of 1584324 are shown in table 3 and the results of clustering neighbouring roads of 193294 are shown in table 4. There are different number of neighbouring roads for road 158324 and road 193294. For road 158324, there are 21 neighbouring roads, whereas for road 193294 there are 60. The number of k clusters are determined by the number of neighbouring roads. The higher the number of neighbouring roads, the higher the number of k clusters. Thus, road 193294 has higher k (k=15) compared to road 158324 (k = 5) since road 193294 has more neighbouring roads than road 158324. Adjusting k cluster influences the number of similar roads with destination roads namely road 158324 and road 193294. A higher number of k clusters will show roads which have higher relationships with destination roads. As seen in table 4, results of clustering for neighbouring roads of 1584324 with k=8 filtered only three (3) roads that have high Bayes and Chi-square result (p-value less than 0.05) compared with when k=3. Similarly, as seen in table 5, results of clustering for neighbouring roads of 193294 with k = 15 filtered five (5) roads that have high Bayes and chi-square result (p-value less than 0.05) compared with when k = 5.

Table 3. Clustering results of neighbouring roads for road 158324

K-3	Bayes	Chi Square P Value	K-8	Bayes	Chi Square P Value
158386	0.96	0.00000095	158386	0.96	0.00000095
158415	0.82	0.21	158415	0.82	0.21
171572	0.65	0.77	173011	0.70	0.01
173011	0.70	0.01			
173118	0.66	NA			
173225	0.66	NA			

Table 4. Clustering results of neighbouring roads for road 193294

K-5	Bayes	Chi Square P Value	K-15	Bayes	Chi Square P Value
182875	NA	NA	193322	0.62	0.31
193268	1.00	0.41	193430	0.69	0.012
193322	0.63	0.32	194878	0.82	0.0000114
193430	0.69	0.01	195015	0.62	0.68
194878	0.82	0.0000114	195150	0.70	0.012
194905	0.65	0.56			
194960	0.64	0.16			
194986	0.74	0.00			
195015	0.62	0.68			
195041	0.57	0.21			
195070	0.88	0.00			
195096	1.00	0.24			
195150	0.70	0.01			
195259	0.78	0.23			
195286	0.64	0.03			
195312	0.70	0.00			
195923	0.50	0.77			
197274	0.73	0.33			
197355	0.67	0.30			
197408	0.70	0.47			
197434	NA	NA			

For further observation, the results are shown on the map. Results for neighbouring roads of road 158324 are shown in figure 4 and results for neighbouring roads of road 193294 are shown in figure 5. From figure 4, the black line shows roads that are having similar traffic condition with road 158324 based on 288-minute interval of traffic flow. We can see the black lines are near and connected with road 158324. Different results are obtained for road 193294 as seen in figure 5. With k = 5, the result shows several black lines are disconnected from road 193294. But with k = 15, the black lines are filtered. Black lines that are disconnected from road 193294 are removed. Only high relationship roads with road 193294 are shown on maps.

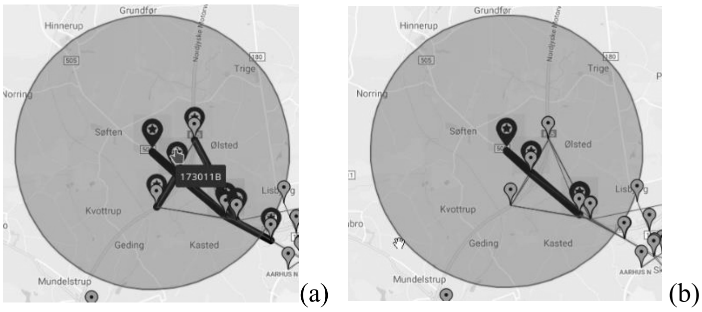

Fig. 4. Relationship of road 158324 with neighbouring roads using k-means (a) k=3 (b) k=8.

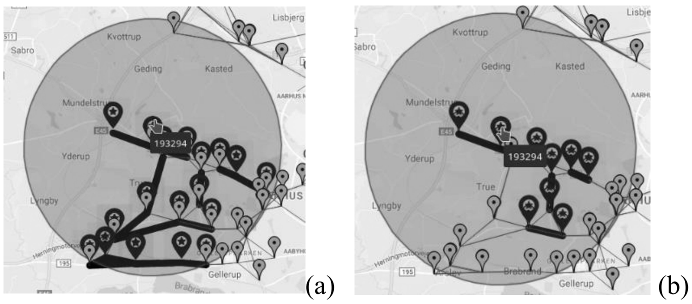

Fig. 5. Relationship of road 193294 with neighbouring roads using k-means (a) k = 5 (b) k = 15

6 Conclusion

Our main objective is to investigate the impact of road congestion on neighbouring roads. Our approach is to study the relationship between roads in a neighbouring area. The results of our experiments show that there exists a relationship between a road and its neighbouring roads. Neighbouring roads with high relationship with target road (road having a congestion), can be used to predict the propagation of traffic congestion. Our experiments show that relationship of roads can be revealed by clustering traffic condition during certain time intervals. The time interval can be adjusted for one hour, or more as need. Furthermore, by adjusting k to a higher value results in stronger relationship between roads in a neighbouring area. The experiment shows that by adjusting k to a higher value results in clustered neighbouring roads which have higher relationship with target road (using Bayes and chi-square) if compare to using smaller k value. Our next plan is to apply findings of the experiments with probability method to predict propagation impact of road congestion.

6 Acknowledgment

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.

References

1. Dong, H., Jia, L., Sun, X., Li, C., Qin, Y.: Road traffic flow prediction with a time-oriented ARIMA model. In: NCM 2009 - 5th International Joint Conference on INC, IMS, and IDC. pp. 1649–1652 (2009). https://doi.org/10.1109/NCM.2009.224.
2. Lee, E.-M., Kim, J.-H., Yoon, W.-S.: 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. pp. 626 –635 (2007). https://doi.org/10.1007/978-3-540-74171-8_62.
3. Kumar, K., Parida, M., Katiyar, V.K.: Short Term Traffic Flow Prediction for a Non Urban Highway Using Artificial Neural Network. In: Procedia - Social and Behavioral Sciences. pp. 755–764. Elsevier B.V. (2013). https://doi.org/10.1016/j.sbspro.2013.11.170.
4. Priambodo, B., Ahmad, A.: Traffic flow prediction model based on neighbouring roads using neural network and multiple regression. J. Inf. Commun. Technol. 17, 513–535 (2018).
5. Kim, J., Wang, G.: Diagnosis and Prediction of Traffic Congestion on Urban Road Networks Using Bayesian Networks. In: Australasian Transport Research Forum 2016. pp. 1–21 (2016). https://doi.org/10.3141/2595-12.
6. Aung, S.S., Naing, T.T.: Naïve Bayes Classifier Based Traffic Detection System on Cloud Infrastructure. 2015 6th Int. Conf. Intell. Syst. Model. Simul. 193–198 (2015). https://doi.org/10.1109/ISMS.2015.45.
7. Kindzerske, M.D., Ni, D.: Composite Nearest Neighbor Nonparametric Regression to Improve Traffic Prediction. Transp. Res. Rec. pp 30-35 (2007). https://doi.org/10.3141/1993-05.
8. Yu, B., Song, X., Guan, F., Yang, Z., Yao, B.: k-Nearest Neighbor Model for Multiple-Time-Step Prediction of Short-Term Traffic Condition. J. Transp. Eng. 142, 04016018 (2016). https://doi.org/10.1061/(ASCE)TE.1943-5436.0000816.
9. Pang, X., Wang, C., Huang, G.: A Short-Term Traffic Flow Forecasting Method Based on a Three-Layer K-Nearest Neighbor Non-Parametric Regression Algorithm. J. Transp. Technol. 6, 200–206 (2016).
10. Wu, Y., Tan, H., Jin, P., Shen, B., Ran, B.: Short-Term Traffic Flow Prediction Based on Multilinear Analysis and k- Nearest Neighbor Regression. 15th Int. Conf. Transp. Prof. 557–569 (2015). https://doi.org/10.1061/9780784479292.051.
11. Wang, Y., Cao, J., Li, W., Gu, T.: 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.
12. Priambodo, B., Ahmad, A.: Predicting traffic flow based on average speed of neighbouring road using multiple regression. In: Badioze Zaman, H., Robinson, P., Smeaton, A.F., Shih, T.K., Velastin, S., Terutoshi, T., Jaafar, A., and Mohamad Ali, N. (eds.) Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in

Bioinformatics). pp. 309–318. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-70010-6_29.
13. Petrovska, N., Stevanovic, A.: Traffic Congestion Analysis Visualisation Tool. IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC. 2015-Octob, 1489 –1494 (2015). https://doi.org/10.1109/ITSC.2015.243.
14. Wang, Z., Lu, M., Yuan, X., Zhang, J., Wetering, H. Van De: Visual traffic jam analysis based on trajectory data. IEEE Trans. Vis. Comput. Graph. 19, 2159–2168 (2013). https://doi.org/10.1109/TVCG.2013.228.
15. Anwar, A., Nagel, T., Ratti, C.: Traffic origins: A simple visualization technique to support traffic incident analysis. IEEE Pacific Vis. Symp. 316 –319 (2014). https://doi.org/10.1109/PacificVis.2014.35.
16. Jiang, Y., Kang, R., Li, D., Guo, S., Havlin, S.: Spatio-temporal propagation of traffic jams in urban traffic networks. Phys. Soc. (2017).
17. Syed Azimuddin, S., Desikan, K.: A Simple Density with Distance Based Initial Seed Selection Technique for K Means Algorithm. J. Comput. Inf. Technol. 25, 291–300 (2018). https://doi.org/10.20532/cit.2017.1003605.
18. Wang, G., Kim, J.: Prediction of Traffic Congestion and Incidents on Urban Road Networks Using Naive Bayes Classifier. (2016).
19. Bischof, S., Karapantelakis, A., Sheth, A., Mileo, A.: 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. pp. 1 –5 (2014).
20. He, F., Yan, X., Liu, Y., Ma, L.: A Traffic Congestion Assessment Method for Urban Road Networks Based on Speed Performance Index. In: Green Intelligent Transportation System and Safety. pp. 425–433 (2016). https://doi.org/10.1016/j.proeng.2016.01.277.
21. Kesumawati, A., Setianingsih, D.: 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. 8, 110–115 (2016).