Predicting Traffic Flow Based on Average Speed of Neighbouring Road Using Multiple Regression

Bagus Priambodo[1,2], Azlina Ahmad[1],

[1] Universiti Kebangsaan Malaysia, Institute of Visual Informatics, Bangi, Malaysia
[2] Universitas Mercu Buana, Information System, Jakarta, Indonesia

Abstract. The prediction of traffic flow is a challenge. There are many factors that can affect traffic flow. One of the factors is an inter path relationship between neighbouring roads. For example, an individual incidents (such as accidents) may cause ripple effects (a cascading failure) which then spreads and creates a sustained traffic jam the neighbouring area. To know the relationship between road segments we propose multiple regression method to predict the traffic based on the nearby surrounding roads. The prediction factor is chosen from a high-relation road with the path to be searched. To know the relationship between roads we calculate their correlation among neighbouring roads. The results are then displayed on the map for further observation. From this study, we demonstrate that multiple regression method can be used to predict impact of speed of vehicles on neighbouring roads on traffic flows.

Keywords: Traffic flow prediction · Multiple regressions · Traffic flow propagation.

1 Introduction

Traffic congestion is a condition in which road users exceed the capability of the road. Characteristic features of road congestion are slow speed, longer travel time, and the length of the queue of cars on the road. During past few years, numerous algorithms for traffic flow prediction have been proposed to predict traffic flow. One interesting research is about impact of road flow of neighbouring road. Previous research in analysis of traffic flow shows when there is a traffic jam or slow speed situation on a road, it will impact other roads [2] [3] [15]. This finding is important for road users as they can avoid impacted roads with slow traffic or traffic jam by providing alternative roads.

2 Related Work

Time series models are widely used to predict traffic flow and traffic congestion. Time series predictions are based on historical data on the same road location. The Arima method [1] [7] [16] is often used for time series prediction. The seasonal Arima model shows high performance in traffic flow forecasting [1], performance improvement using Arima method can also be done with addition of day classification [7]. In addition to Arima and regression [6], neural networks are also commonly used for traffic flow prediction [10] and traffic congestion based on time series, among others: [5] using linear fuzzy for short time prediction on toll roads to design a number of sensors on the highway, [8] based on the similarity of space and time. Considering many factors that can affect the flow of traffic than many traffic flow predictions is done using multivariate methods. Among them are predicting the traffic road safety level based on the licensed drivers, factor, Gross Domestic Product and accident using multilevel regression and predicting traffic [12] based on weather data using neural networks [11]. Short time prediction use neural networks based on speed data from various vehicles, days, and traffic density.

Aside from predictions using both linear and nonlinear methods, there are also other methods used to predict traffic flow. The road congestion prediction is based on the slices of the crossroads using the BML model [9]. Traffic prediction is based on similarity of traffic congestion pattern [13] but the result is not satisfactory.

In predicting traffic flow and traffic congestion on a road, many factors are involved which include previous data, [16] vehicle speed [11], weather [12], and accidents [3]. Another factor that affects traffic congestion on a road is congestion at neighbouring road. Previous research in this aspect of road congestion are the inter-road relationship extracted using of the 3D Markov model [2], visualizing and highlighting impact traffic seen as affecting adjacent roads [3], the visualization of traffic jam which reflects the spread of traffic flow [15], the detection of traffic jam based on the slices between the intersections which indicates an inter path relationship [9] and mining congestion between road segments [14].

3 Problem

Many factors influence traffic flow. Factors such as history [16], vehicle speeds [11], weathers [12], accidents [3], and special days or events can be used to predict traffic flows. In addition to the factors we have mentioned, the congestion level factor on neighbouring roads greatly affects road congestion. Previous research which involves extracting inter-road relationship [2], visualizing traffic accident impact with high lights [3], spreading of traffic flow when visualizing traffic jam [15], detection of traffic congestion based on intersection [9], and mining congestion between road links [14], have assumed that in the traffic flow there is a relationship between neighbouring roads.

The strong relationship between one road and the other roads around it makes the road a strong candidate for factor input in predicting speed on the road. This type of prediction can help in predicting road speed on roads that have damaged sensor or missing data. The surrounding road traffic can be used as a tool to predict traffic flow propagation. As we can see in Figure 1, road 158324 is influenced by the road that surrounds it. If there is traffic jam or road congestion on the neighbouring road it will affect the average speed on the road 158324.

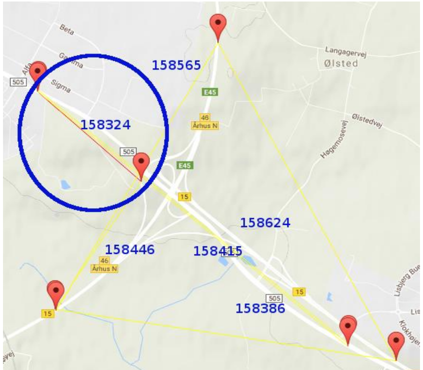

Fig. 1. Road 158324 and its neighbouring road.

4 Methodology

Our research methodology is described as follows:

1. Data set.
2. Determining relationship between road segments using correlation method between road segments among neighbouring roads.
3. Predicting the speed using multiple regression method.

4.1 Data set

Our data set is taken from IoT traffic sensor in Aarhus, Denmark [17] [18]. The total number of IoT sensors are 449 sensors as we can see in Figure 2.

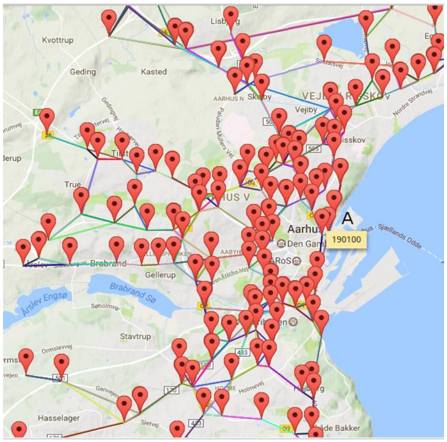

Fig. 2. Map of 449 Iot traffic sensors in city of Aarhus, Denmark.

An example sensors in A location, sensor name is 190100. This sensor are placed from Nørreport 93 Aarhus, Denmark to Spanien 63 Aarhus, Denmark, the distance between points is 1490 meters. In this experiment we only use average speed as representation of traffic flow, and timestamp data. Specific details of this sensor are described in Table 1 and Table 2.

Table 1. Example traffic data taken from sensor 190100.

Duration from	Duration to	Start Point	End Point	Cross Observation point data
2014-10-01 01:45:00	2014-11-13 10:40:00	City: Aarhus Street: Nørreport 93 Postal Code: 8000 Coordinates (lat,long): 56.161017815103236, 10.21197608217426	City: Aarhus Street: Spanien 63 Postal Code: 8000 Coordinates (lat, long): 56.14892750591274, 10.209599775463175	Distance between two points in meters: 1490 Durationof measurements in seconds: 202 NDT in KMH: 27 EXT ID: 359 Road type: MAJOR_ROAD

Table 2. Example traffic data taken from sensor 190100.

status	avgMeasuredTime	avgSpeed	extID	medianMeasuredTime	TIMESTAMP	vehicleCount	_id	REPORT_ID
OK	376	14	1051	376	2014-08-01T08:30:00	4	20749724	190100
OK	225	23	1051	225	2014-08-01T08:40:00	3	20750622	190100
OK	285	18	1051	285	2014-08-01T08:50:00	2	20751520	190100

4.2 Determining relationship between road segments among neighbouring roads

To predict average speed using neighbouring road, we need to find the highest correlation among all roads in neighbouring roads. We consider a neighboring road is a road that has a distance approximately four kilometers from the road location. We calculate correlation among all neighbouring roads using formula (0).

$$Cor_{xy} = \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \cdot (y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2 \cdot \sum_{i=1}^{n} (y_i - \overline{y})^2}} \qquad \textbf{(0)}$$

We calculate correlation of all neighbouring roads based on the median value of average speed at 11.00 am to 13.00 pm from 01-08-2014 until 28-09-2014. We choose 11:00 am to 13:00 pm because at the time vehicles are always passing through the sensor or vehicle count > 0. In this paper, we only discuss the results of two roads which is road 158324 and road 158744. The neighbouring road of correlation of road 158324 is shown in Figure 3 and the result of correlation of road 158324 is described in Table 3. The neighbouring road of correlation of correlation of road 158744 is shown in Figure 4 and the results of correlation of road 158744 is described in Table 4.

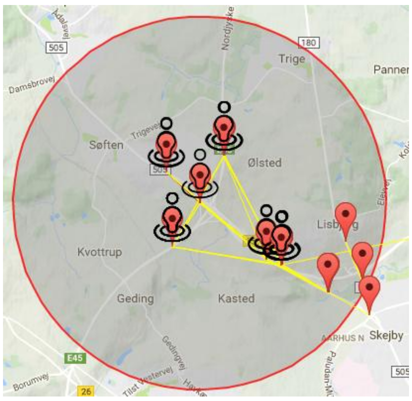

Fig. 3. Neighbouring road 158324 and correlation results.

Table 3. Correlation result of road 158324.

Roads	Correlation results	Distance (km)
158624	**0.789**	1.104
158386	**0.770**	1.867
158595	**0.757**	2.197
158415	**0.715**	0.975
158475	**0,701**	1.104
158446	**0.686**	1.143
158536	**0.637**	2.197
158355	**0.629**	1.104
171969	**0,641**	1.104
172329	0,514993548	1.867

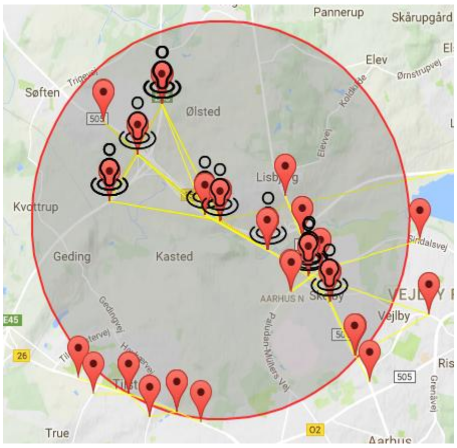

Fig. 4. Neighbouring road 158744 and correlation results.

Table 4. Correlation result of road 158744

Roads	Correlation results	Distance (km)
158954	**0,436660191**	1,157
158446	**0,425732191**	2,629
158505	**0,414264573**	2,196
158475	**0,393660707**	2,382
159014	**0,378868354**	2,826
158595	**0,378515879**	0,018
158565	0,34790664	2,629
158924	0,342129916	0,349

4.3 Predicting the speed using multiple regression method

First, we predict average speed on road 158324. We choose the road that has correlation values > 6, obtained nine roads to become independent variables. Nine independent variables are road 158624, road 158386, road 158595, road 158415, road 158475, road 158446, road 158536, road 158355 and road 171969. One dependent variable is road 158324 (Y). The data was taken from 01-08-2014 to 28-09-2014 15:00 pm. We obtained the coefficients of multiple regressions as follow:

$$Y = -1.412068+0.060048X_1 + 0.281535X_2 + 0.184692X_3 + 0.020628X_4 - 0.002242X_5 + 0.123846X_6 + 0.071511X_7 + 0.039800X_8 + 0.047805X_9 \quad (1)$$

Results of prediction and deviation of error rate are described in Table 5 below.

Table 5. Results of 158324 traffic flow prediction.

Time	Prediction	Actual	MAD	MSE	RMSE	MAPE
15:05	61,0715	60	1,07	1,15	1,07	1,75
15:10	63,0549	68	3,01	12,80	3,58	4,80
15:15	63,2426	68	3,59	16,08	4,01	5,71
15:20	64,3689	65	2,85	12,16	3,49	4,53
15:25	62,4896	60	2,78	10,97	3,31	4,42
15:30	61,0362	60	2,49	9,32	3,05	3,96

Second, we predict the average speed on road 158744. We choose the road that has correlation values > 3.5 since there is no road that has correlation value with road 158744 above 0.6, obtained six roads to become independent variables. Independent variables used to find dependent variable of road 158744 (Y), were road 158954, road 158446, road 158505, road 158475, road 159014 and road 158595. The data taken from 01-08-2014 to 28-09-2014 15:00 pm. We get the coefficients of multiple regressions as follows:

$$Y = 56.432470 + 0.115115X_1 + 0.005259X_2 +0.084683X_3 - 0.151605 X_4 + 0.285097X_5 - 0.004562X_6 \quad (2)$$

Results of prediction and deviation of error rate are described in Table 6 below.

Table 6. Results of road 158744 traffic flow prediction.

Time	Prediction	Actual	MAD	MSE	RMSE	MAPE
15:05	92,01	84	8,01	64,22	8,01	8,71
15:10	84,63	87	5,19	34,91	5,91	5,75

15:15	82,87	82	3,75	23,53	4,85	4,19
15:20	89,16	79	5,35	43,45	6,59	5,99
15:25	90,40	79	6,56	60,75	7,79	7,31
15:30	91,40	81	7,20	68,66	8,29	7,99

5 Results and Discussion

After conducting the experiments, we calculate the error of prediction using mean absolute deviation (MAD), mean square error (MSE), root mean square error (RMSE), and mean absolute percentage error (MAPE). Using result from Table 5, and Table 6. and using line chart shown in Figure 5 and Figure 6, results indicate that prediction average speed in road 158324 is more accurate than prediction in road 158744. Deviation error between predicted value and actual value in road 158744 is higher in road 158324.

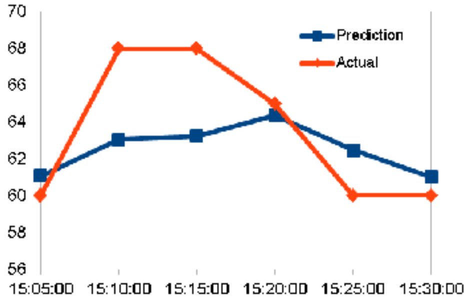

Fig. 5. Average speed prediction in 158324

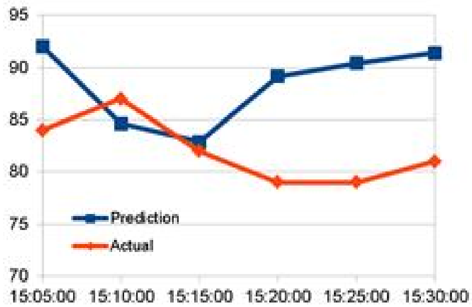

Fig. 6. Average speed prediction in 158744

A clearer result can be seen using line chart as shown in Figure 7 and Figure 8. Both figures show clearly that the results of predicted value in road 158744 is less accurate compared to predicted value in road 158324. The error trend on road 158744 is greater for long interval. While at road 158324, result is positive since the error trend is lower. From Table 3 and Table 4 it can be seen that the correlation neighbouring road 158324 has a higher value than the correlation value of neighbouring road 158744, since all road correlation values are above 0.6. This is different from the correlation values at road 158744, which all correlation values are below 0.6. The high correlation value indicates that the prediction of average speed at road 158324 is better or more accurate than prediction of average speed at road 158744. The question is what factors caused the correlation value in road 158744 lower than road 158324 since there are also correlation on other paths which value is below 0.6. This is probably due to location of the roads, the average speed on the roads, and the number of vehicles passing through the roads which leads to the differences in the value of correlation.

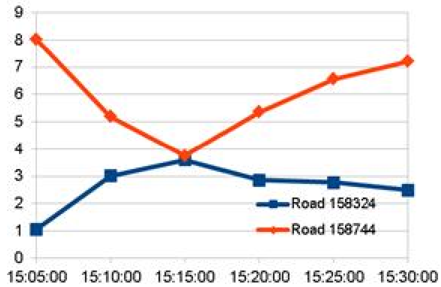

Fig. 7. Line chart MAD between road 158324 and 158744

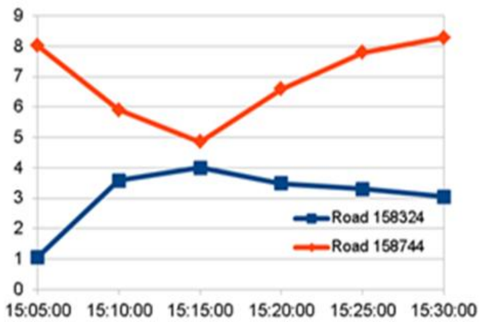

Fig. 8. Line chart RMSE between Road 158324 and 158744

6 Conclusion

Our research aims to investigate the impact of traffic flow of one road on traffic flow of neighbouring roads. The experiments conducted shows that there is a relationship between a road and its neighboring roads. The strong relationship between one road and the other road around it, makes it an important factor for predicting traffic flow. The experiments show that high correlation roads can be used to predict the value of average speed for short time prediction. However, not all roads have high correlation value with their neighbouring roads. The location of roads, the average speed on the roads, the number of vehicles passing through the road are possible factors that lead to the difference in correlation value of the road.

References

1. Abadi, A., Rajabioun, T., Ioannou, P.A.: Traffic Flow Prediction for Road Transportation Networks With Limited Traffic Data. IEEE Trans. Intell. Transp. Syst. 16, 653–662 (2015). doi:10.1109/TITS.2014.2337238
2. Ahn, J., Ko, E., Kim, E.Y.: Highway traffic flow prediction using support vector regressionand Bayesian classifier. 2016 Int. Conf. Big Data Smart Comput. 239–244 (2016). doi:10.1109/BIGCOMP.2016.7425919
3. Anwar, A., Nagel, T., Ratti, C.: Traffic origins: A simple visualization technique to supporttraffic incident analysis. IEEE Pacific Vis. Symp. 316–319 (2014). doi:10.1109/PacificVis.2014.35
4. Cai, H., Yan, L., Zhu, D.: Traffic Safety Level in China. 363–369 (2015)
5. Chan, K.Y., Dillon, T.S.: Traffic flow prediction using orthogonal arrays and Takagi-Sugenoneural fuzzy models. (2014)
6. Dai, H., Yang, Z.: Real-Time Traffic Volume Estimation with Fuzzy Linear Regression. 2006 6th World Congr. Intell. Control Autom. 3164–3167 (2006). doi:10.1109/WCICA.2006.1712950
7. Dong, H., Jia, L., Sun, X., Li, C., Qin, Y.: Road Traffic Flow Prediction with a Time-Oriented ARIMA Model. 1649–1652 (2009). doi:10.1109/NCM.2009.224
8. Hu, C., Xie, K., Song, G., Wu, T.: Hybrid Process Neural Network based on Spatio-Temporal Similarities for Short-Term Traffic Flow Prediction. 253–258 (2008)
9. Hu, W., Yan, L., Wang, H.: Traffic Jams Prediction Method based on Two-dimension Cellular Automata Model *. (2014)
10. Kumar, K., Parida, M., Katiyar, V.K.: Short Term Traffic Flow Prediction for a Non Urban Highway Using Artificial Neural Network. Procedia - Soc. Behav. Sci. 104, 755–764 (2013). doi:10.1016/j.sbspro.2013.11.170
11. Lee, J., Hong, B., Lee, K., Jang, Y.-J.: A prediction model of traffic congestion using weather data. 2015 IEEE Int. Conf. Data Sci. Data Intensive Syst. 81–88 (2015). doi:10.1109/DSDIS.2015.96
12. Lee, K., Hong, B., Jeong, D., Lee, J.: Congestion pattern model for predicting short-term traffic decongestion times. 2014 17th IEEE Int. Conf. Intell. Transp. Syst. ITSC 2014. 2828–2833 (2014). doi:10.1109/ITSC.2014.695814313
13. 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). doi:10.1109/SMARTCOMP.2016.7501704
14. Wang, Z., Lu, M., Yuan, X., Zhang, J., Wetering, H. Van De: Visual Traffic Jam Analysis Based on Trajectory Data a b c e d. 19, 2159–2168 (2013)
15. Xiong, W., Yu, Z., Eeckhout, L., Bei, Z., Zhang, F., Xu, C.: SZTS: A novel big data transportation system benchmark suite. Proc. Int. Conf. Parallel Process. 2015–Decem, 819–828 (2015). doi:10.1109/ICPP.2015.91
16. Zhang, R., Shu, Y., Yang, Z., Cheng, P., Chen, J.: Hybrid Traffic Speed Modeling and Prediction Using Real-World Data. Proc. - 2015 IEEE Int. Congr. Big Data, BigData Congr. 2015. 230–237 (2015). doi:10.1109/BigDataCongress.2015.40
17. R. Tönjes, P. Barnaghi, M. Ali, A. Mileo, M. Hauswirth, F. Ganz, S. Ganea, B. Kjærgaard, D. Kuemper, S. Nechifor, D. Puiu, A. Sheth, V. Tsiatsis, L. Vestergaard, "Real Time IoT Stream Processing and Large-scale Data Analytics for Smart City Applications", poster session, European Conference on Networks and Communications 2014
18. Stefan Bischof, Athanasios Karapantelakis, Cosmin-Septimiu Nechifor, Amit Sheth, Alessandra Mileo and Payam Barnaghi, "Semantic Modeling of Smart City Data", Position Paper in W3C Workshop on the Web of Things: Enablers and services for an open Web of Devices, 25-26 June 2014, Berlin, Germany