Time Series Traffic Speed Prediction Using k-Nearest Neighbour Based on Similar Traffic Data

Bagus Priambodo[1*], and *Yuwan* Jumaryadi[1]

[1]Universitas Mercu Buana, Faculty of Computer Science, Jakarta, Indonesia

Abstract. During the past few years, time series models and neural network models are widely used to predict traffic flow and traffic congestion based on historical data. Historical data traffic from sensors is often applied to time series prediction or various neural network predictions. Recent research shows that traffic flow pattern will be different on weekdays and weekends. We conducted a time series prediction of traffic flow on Monday, using data on weekdays and whole days data. Prediction of short time traffic flows on Monday based on weekdays data using k-NN methods shows a better result, compared to prediction based on all day's data. We compared the results of the experiment using k-NN and Neural Network methods. From this study, we observed that generally, using similar traffic data for time series prediction show a better result than using the whole data.

1 Introduction

Due to population increase and numbers of private cars in this modern era, traffic congestion has become significantly worse, not only leading to economic losses but also causes environment damages, human stress and environmental pollution [1]. Drivers need traffic information which can affect their driving to a certain option [2] and thus lead to chain changes in traffic flow state in the upstream and downstream of relative road section and other road segments in the network. Some studies show that a road always has a similar traffic state 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 work days or weekend [6]. Investigate similar traffic condition pattern can lead us to increase accuracy and reduce high computation in predicting traffic flow by using only similar traffic data rather use the whole data.

Related Work and Background

Time series [3], and neural network models[4][5], are often applied to prediction traffic flow and predict traffic congestion based on vehicle speeds, weather, incident, and special days of historical data. Some studies use a Bayes classifier to predict traffic congestion

---

* Corresponding author: bagus.priambodo@mercubuana.ac.id, yuwan.jumaryadi@mercubuana.ac.id

[6][7]. Other used non-parametric regression *k*-NN to predict short-term traffic flow[8][9][10][11].

Traffic flow is uncertain, nonlinear and complex, it is difficult to predict the traffic flow effectively and accurately by the predict method based on traditional mathematics and physics models [12], although the neural network can train based on historical data to modelling complex relationships of traffic flow, its unavoidable because its need lengthy amount of time required to train the network [8]. K-nearest neighbour non-parametric regression method [10] is a reliable method for short-term traffic flow forecasting, which can reflect the nonlinear characteristics.

Some studies have conducted predict short-term traffic flow using k-NN, using three-layer k-NN [10], using k-NN with composite method [8], using k-NN with weight[12], using k-NN with tensor for calculating similarity traffic [11], and k-NN with connected link upstream or downstream[9].

Previously, we have conducted a prediction of traffic flow in normal condition based on all day's data using multiple regression method[13]. Other research show there is a difference in congestion pattern between weekdays and weekend [14]and between regular and irregular days [15]. Based on this information, we found that traffic flow pattern on roads will be different on weekdays and weekends. We assumed time series prediction using the k-NN method on weekdays using only weekday's data will have a better result if compared to using all days' data.

2 Dataset and Method

2.1 Dataset

For our experiments, we used data set from IoT traffic sensor in Aarhus, Denmark [16][17]. The total number of IoT sensors is 449 and their location is shown in Figure 1.
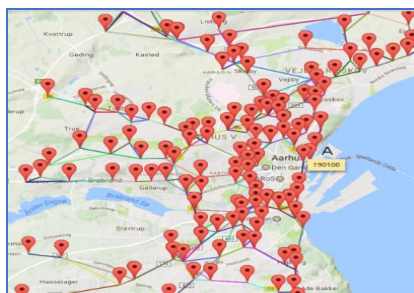

Fig. 1. Map of 449 IoT traffic sensors in the city of Aarhus, Denmark

2.2 Methodology

In order to achieve our main research objective that is to do time series traffic speed prediction using k-NN based on similar traffic data. We conduct our experiment as follows. In the beginning, we observe traffic average speed pattern on road 158415 and road 158595, on weekdays and in whole days. Second, we conduct time series prediction on road 158415 and road 158595 using week day's data and using whole days data. Third, we calculate an error using MAD, MAPE and RMSE.

2.3 Observe traffic average speed pattern

Our aim of the research is predicting average speed using a k-NN method based on similar traffic data. We observed traffic flow pattern on two roads and found there are different patterns of traffic flow between weekdays and the whole days. Weekdays' data is traffic data from Monday to Thursday. Whole week data is traffic data from Mondays to Sundays. We observed that traffic pattern on weekdays looks sharper than traffic pattern on whole days, as seen in figure 2 and figure 3 below.
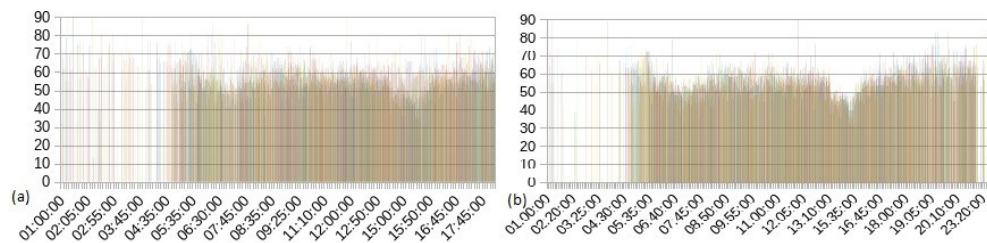

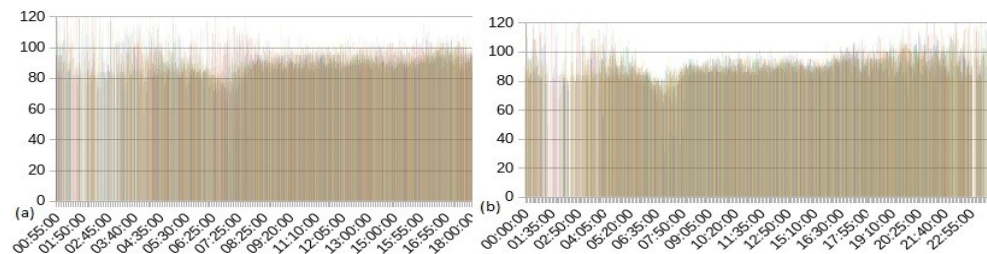**Fig. 2.** Traffic pattern on road 158415 (a) all day's data, (b) week day's data

Fig. 3. Traffic pattern on road 158595 (a) all day's data (b) week day's data.

2.4 Time series prediction

For time series prediction using the k-NN method, we set lags = 10 and k = 20. For time series prediction using elm neural network, we set lags = 4, layers = 3 (input, hidden and output), with 4 input neurons, 100 hidden neurons and one input neuron.

3 Result and Discussion

3.1 Dataset

Results of the experiment for short time prediction (25 minutes) on road 158415 are shown in Table 1 and the errors are displayed in Table 2. Results of the experiment for short time prediction (25 minutes) on road 158595 are shown in Table 3 and the errors are displayed in Table 4.

Table 1. Results of short time prediction using K-NN and ELM neural network on road 158415

TIME	Actual	K-NN Weekdays	NN Weekdays	K-NN All days	NN All days
15:05:00	45	50,15	47,91	49,50	48,24
15:10:00	44	54,2	48,84	46,85	49,49
15:15:00	53	45,35	49,60	41,05	50,25
15:20:00	48	50,7	49,59	44,85	50,27
15:25:00	52	51,8	49,31	47,75	50,11

Table 2. The error of short time prediction on road 158415

Error	K-NN Weekdays	NN Weekdays	K-NN All days	NN All days
MAD	5,180	3,087	5,340	3,129
RMSE	6,268	3,264	6,311	3,374
MAPE	11,01	6,48	10,75	6,65

Table 3. Results of short time series prediction using K-NN and ELM neural network on road 158595

TIME	Actual	K-NN Weekdays	NN Weekdays	K-NN All days	NN All days
06:05:00	73	69,2	74,69	70,95	74,46
06:10:00	69	72,9	75,79	64,55	75,45
06:15:00	74	67,55	77,29	62,7	77,16
06:20:00	78	73,45	78,03	68,8	78,08
06:25:00	71	69	78,58	70	78,74

Table 4. The error of short time series prediction on road 158595

Error	K-NN Weekdays	NN Weekdays	K-NN All days	NN All days
MAD	4,140	3,880	5,600	3,780
RMSE	4,381	4,848	6,890	4,770
MAPE	5,64	5,47	7,55	5,33

3.2 Discussion

From observation of traffic speed, we gathered information that traffic flow pattern on weekdays looks sharper than traffic flow in whole days. This show traffic flow on weekdays has a similar pattern. From experiment result, time series prediction using the k-NN method on Monday using weekdays data show a better result than prediction using whole days data. As seen in Figure 4 and Figure 5. This shows us that time series prediction using a k-NN method based on similar traffic flow data produced a better result than based on whole data.
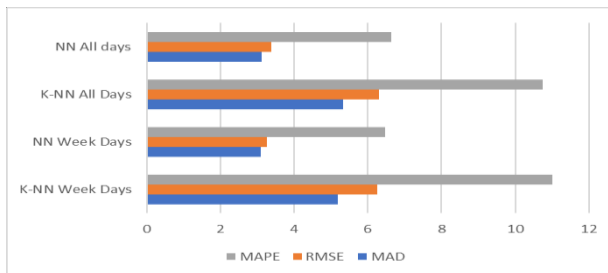

Fig. 4. Bar chart of error of short time series prediction on road 158415
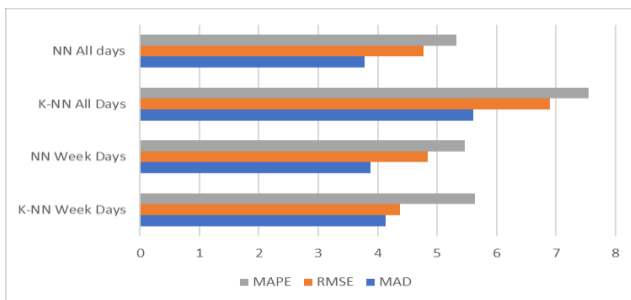
Fig. 5. Bar chart of error of short time series prediction on road 158595

4 Conclusions

Generally, our results show that time series prediction on Monday using a k-NN method based on weekdays' data or based on similar traffic data produced better results compared to prediction using whole day's data, also from our experiment, prediction using time series using neural network produced better results when compared with the k-NN method. Further investigation is needed to understand the results clearer. We propose to cluster traffic condition in each day on weekdays and weekends to obtain better results.

Acknowledgement

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).

2. 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).

3. 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*, no. 1, pp. 1649–1652 (2009).

4. E. E. E. Lee, J. J. J. Kim, and W. W. Yoon, "Traffic speed prediction under weekday, time, and neighboring links' speed: Back propagation neural network approach," in *… Theories and Applications. With Aspects of …*, no. Mic, pp. 626–635 (2007).

5. K. Kumar, M. Parida, and V. K. Katiyar, "Short Term Traffic Flow Prediction for a Non Urban Highway Using Artificial Neural Network," in *Procedia - Social and Behavioral Sciences*, vol. **104**, pp. 755–764 (2007).

6. J. Kim and G. Wang, "Diagnosis and Prediction of Traffic Congestion on Urban Road Networks Using Bayesian Networks," in *Australasian Transport Research Forum 2016*, no. 2595, pp. 1–21 (2016).

7. 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).

8.   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).

9.   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).

10. 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).

11. 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," *15th Int. Conf. Transp. Prof.*, pp. 557–569 (2015).

12. 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).

13. B. Priambodo and A. Ahmad, "Predicting Traffic Flow Based on Average Speed of Neighbouring Road Using Multiple Regression BT - Advances in Visual Informatics: 5th International Visual Informatics Conference, IVIC 2017, Bangi, Malaysia, November 28–30, 2017, Proceedings," pp. 309–318 (2017).

14. K. Lee, B. Hong, D. Jeong, and J. Lee, "Congestion pattern model for predicting short-term traffic decongestion times," in *2014 17th IEEE International Conference on Intelligent Transportation Systems, ITSC 2014*, pp. 2828–2833 (2014).

15. F. Rempe, G. Huber, and K. Bogenberger, "Spatio-Temporal Congestion Patterns in Urban Traffic Networks," *Transp. Res. Procedia*, vol. **15**, pp. 513–524 (2016).

16. G. Wang and J. Kim, "Prediction of Traffic Congestion and Incidents on Urban Road Networks Using Naive Bayes Classifier," no. February (2017, 2016).

17. S. Bischof, A. Karapantelakis, A. Sheth, A. Mileo, and P. Barnaghi, "Semantic Modelling 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).