Short-Term Traffic Forecasting Using Self-Adjusting k-Nearest Neighbours

Bin Sun[1], Wei Cheng[2,1,*], Prashant Goswami[1], Guohua Bai[1]

[1]Blekinge Institute of Technology, Karlskrona 37179, Sweden
[2]Kunming University of Science and Technology, Kunming 650093, China
[*]Corresponding author, email: wei.cheng@bth.se

Abstract: Short-term traffic forecasting is becoming more important in intelligent transportation systems. The k-nearest neighbours (kNN) method is widely used for short-term traffic forecasting. However, the self-adjustment of kNN parameters has been a problem due to dynamic traffic characteristics. This paper proposes a fully automatic dynamic procedure kNN (DP-kNN) that makes the kNN parameters self-adjustable and robust without predefined models or training for the parameters. A real-world dataset with more than one year traffic records is used to conduct experiments. The results show that DP-kNN can perform better than manually adjusted kNN and other benchmarking methods in terms of accuracy on average. This study also discusses the difference between holiday and workday traffic prediction as well as the usage of neighbour distance measurement.

1.   Introduction

The paper full-text is available on [IET Digital Library](http://dx.doi.org/10.1049/iet-its.2016.0263).

The code is available on GitHub: https://github.com/SunnyBingoMe/sun2018shortterm-github

First author's web: http://ABOUT.DMML.NU

9. References

[1] An, S., Zhang, T., Zhang, X., *et al*.: 'Unrecorded Accidents Detection on Highways Based on Temporal Data Mining', *Math. Probl. Eng.*, 2014, **852**, (495).

[2] Zou, Y., Shi, G., Shi, H., *et al*.: 'Traffic incident classification at intersections based on image sequences by HMM/SVM classifiers', *Multimed. Tools Appl.*, 2011, **52**, (1), pp. 133–145.

[3] Clark, S.: 'Traffic prediction using multivariate nonparametric regression', *J. Transp. Eng.-ASCE*, 2003, **129**, (2), pp. 161–168.

[4] Guo, J., Huang, W., Williams, B. M.: 'Real time traffic flow outlier detection using short-term traffic conditional variance prediction', *Transp. Res. Pt. C-Emerg. Technol.*, 2015, **50**, pp. 160–172.

[5] Roess, R. P., Prassas, E. S., McShane, W. R.: 'Traffic Engineering', (Prentice Hall, 2010, 4 ed. Upper Saddle River, NJ.)

[6] Oh, S., Byon, Y.-J., Jang, K., Yeo, H.: 'Short-term Travel-time Prediction on Highway: A Review of the Data-driven Approach', *Transport Rev.*, 2015, **35**, (1), pp. 4–32.

[7] Davis, G. A., Nihan, N. L.: 'Nonparametric regression and short-term freeway traffic forecasting', *J. Transp. Eng.*, 1991, **117**, (2), pp. 178–188.

[8] Castro-Neto, M., Han, L.D., Jeong, Y., *et al*.: 'Toward Training-Free Automatic Detection of Freeway Incidents Simple Algorithm with One Parameter', *Transp. Res. Record*, 2012, **2278**, (1), pp. 42–49.

[9] Bernas, M., Placzek, B., Porwik, P., *et al*.: 'Segmentation of vehicle detector data for improved k-nearest neighbours-based traffic flow prediction', *IET Intell. Transp. Syst.*, 2015, **9**, (3), pp. 264–274.

[10] Bhatia, N., Ashev, V.: 'Survey of Nearest Neighbor Techniques', *International Journal of Computer Science and Information Security*, 2010, **8**, (2).

[11] Sun, B., Cheng, W., Goswami, P. *et al*.: 'An Overview of Parameter and Data Strategies for k-Nearest Neighbours Based Short-Term Traffic Prediction'. International Conference on Intelligent Traffic and Transportation (ICITT), Zurich, Switzerland, September 2017.

[12] Lin, L., Wang, Q., Sadek, A. W.: 'Short-Term Forecasting of Traffic Volume - Evaluating Models Based on Multiple Data Sets and Data Diagnosis Measures', *Transp. Res. Record*, 2013, **23**, (92), pp. 40–47.

[13] Vlahogianni, E. I., Karlaftis, M. G., Golias, J. C.: 'Short-term traffic forecasting: Where we are and where we're going', *Transp. Res. Pt. C-Emerg. Technol.*, 2014, **43**, (1), pp. 3–19.

[14] Hyndman, R. J., Athanasopoulos, G: 'Seasonal ARIMA models', in 'Forecasting: Principles and Practice'. (oTexts, 2015).

[15] Hyndman, R. J., Khandakar, Y.: 'Automatic time series forecasting: The forecast package for R', *J. Stat. Softw.*, 2008, **27**, (3), pp. 1–22.

[16] Yoon, B., Chang, H.: 'Potentialities of Data-Driven Nonparametric Regression in Urban Signalized Traffic Flow Forecasting', *J. Transp. Eng.*, 2014, **4**, (14), p. 27.

[17] Markovic, H., Basic, B. D., Gold, H. *et al.*: 'GPS Data-based Non-parametric Regression for Predicting Travel Times in Urban Traffic Networks', *PROMET-Traffic & Transportation*, 2010, **22**, (1).

[18] Hong, H., Huang, W., Xing, X., *et al.*: 'Hybrid Multi-metric K-Nearest Neighbor Regression for Traffic Flow Prediction'. 18th IEEE International Conference on Intelligent Transportation Systems, Gran Canaria Canary Islands, Spain, September 2015, pp. 2262–2267.

[19] Zheng, Z., Su, D.: 'Short-term traffic volume forecasting: A k-nearest neighbor approach enhanced by constrained linearly sewing principle component algorithm', *Transp. Res. Pt. C-Emerg. Technol.*, 2014, **43**, pp. 143–157.

[20] Dell'Acqua, P., Bellotti, F., Berta, R. *et al.*: 'Time-Aware Multivariate Nearest Neighbor Regression Methods for Traffic Flow Prediction', *IEEE T. Intell. Transp.*, 2015, **16**, (6), pp. 3393–3402.

[21] Boto-Giralda, D., Diaz-Pernas, F. J., Gonzalez-Ortega, *et al.*: 'Wavelet-Based Denoising for Traffic Volume Time Series Forecasting with Self-Organizing Neural Networks', *Comput.-Aided Civil Infrastruct. Eng.*, 2010, **25**, (7), pp. 530–545.

[22] Byon, Y.-J., Liang, S.: 'Real-Time Transportation Mode Detection Using Smartphones and Artificial Neural Networks: Performance Comparisons Between Smartphones and Conventional Global Positioning System Sensors', *J. Intell. Transport. S.*, 2014, **18**, (3), pp. 264–272.

[23] Vlahogianni, E. I., Karlaftis, M. G., Golias, J. C.: 'Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach', *Transp. Res. Pt. C-Emerg. Technol.*, 2005, **13**, (3), pp. 211–234.

[24] Zheng, W. Z., Lee, D. H., Shi, Q. X.: 'Short-term freeway traffic flow prediction: Bayesian combined neural network approach', *J. Transp. Eng.-ASCE*, 2006, **132**, (2), pp. 114–121.

[25] Moghaddam, S. S., Hellinga, B.: 'Real-Time Prediction of Arterial Roadway Travel Times Using Data Collected by Bluetooth Detectors', *Transp. Res. Record*, 2014, **2442**, pp. 117–128.

[26] Wu, S., Yang, Z., Zhu, X. *et al.*: 'Improved k-nn for Short-Term Traffic Forecasting Using Temporal and Spatial Information', *J. Transp. Eng.*, 2014, **14**, (7), p. 26.

[27] Kohn, R., Smith, M., Chan, D.: 'Nonparametric regression using linear combinations of basis functions', *Statistics and Computing*, 2001, **11**, (4), pp. 313–322.

[28] Ohba, Y., Ueno, H., Kuwahara, M.: 'Travel time calculation method for expressway using toll collection system data'. Proceedings of IEEE/IEEJ/JSAI International Conference on Intelligent Transportation Systems, Tokyo, Japan, 1999, pp. 471–475.

[29] Bajwa, S., Chung, E., Kuwahara, M.: 'Performance evaluation of an adaptive travel time prediction model'. Proceedings of IEEE Intelligent Transportation Systems, Vienna, Austria, September 2005, pp. 1000–1005.

[30] Ruiz Aguilar, J. J., Turias, I. J., Moscoso Lopez, J. A. *et al.*: 'Forecasting of short-term flow freight congestion: A study case of Algeciras Bay Port (Spain)', *DYNA*, 2016, **83**, (195), pp. 163–172.

[31] Ling, P., Gao, D., Zhou, X. *et al.*: 'Improve the diagnosis of atrial hypertrophy with the local discriminative support vector machine', *Bio-Med. Mater. Eng.*, 2015, **26**, pp. 1813–1820.

[32] He, Z., Lee, B. S., Snapp, R.: 'Self-tuning cost modeling of user-defined functions in an object-relational DBMS', *ACM T. Database Syst.*, 2005, **30**, (3), pp. 812–853.

[33] Zhang, X., Song, Q.: 'Predicting the number of nearest neighbors for the k-NN classification algorithm', *Intell. Data Anal.*, 2014, **18**, (3), pp. 449–464.

[34] Hong, H., Zhou, X., Huang, W. *et al.*: 'Learning Common Metrics for Homogenous Tasks in Traffic Flow Prediction'. 14th IEEE International Conference on Machine Learning and Applications (ICMLA), Miami, FL, December 2015, pp. 1007–1012.

[35] Singh, H., Desai, A., Pudi, V.: 'PAGER: Parameter less, Accurate, Generic, Efficient kNN-Based Regression'. in Bringas P. G., Hameurlain A., Quirchmayr, G.: *et al.*: 'Database and Expert Systems Applications, Pt 2' (Springer-Verlag Berlin, 2010.), pp. 168–176.

[36] Oh, S., Byon, Y. J., Yeo, H.: 'Improvement of Search Strategy With K-Nearest Neighbors Approach for Traffic State Prediction', *IEEE T. Intell. Transp.*, 2016, **17**, (4), pp. 1146–1156.

[37] Dubey, H., Pudi, V.: 'CLUEKR: CLUstering Based Efficient kNN Regression'. Pacific-Asia Conference on Knowledge Discovery and Data Mining, Gold Coast, Australia, April 2013, pp. 450–458.

[38] Takens, F.: 'Detecting strange attractors in turbulence'. in Warwick: 'Dynamical Systems and Turbulence: Proceedings of a Symposium Held at the University of Warwick' (University of Warwick Press, 1981), pp. 366–381.

[39] Packard, N. H., Crutchfield, J. P., Farmer, J. D. *et al.*: 'Geometry from a Time Series', *Phys. Rev. Lett.*, 1980, **45**, (9), pp. 712–716.

[40] Chang, H., Lee, Y., Yoon, B. *et al.*: 'Dynamic near-term traffic flow prediction: System-oriented approach based on past experiences', *IET Intell. Transp. Syst.*, 2012, **6**, (3), pp. 292–305.

[41] Kim, T., Kim, H., Lovell, D. J.: 'Traffic flow forecasting: Overcoming memoryless property in nearest neighbor non-parametric regression'. 8th IEEE International Conference on Intelligent Transportation Systems (ITSC), Vienna, Austria, September 2005, pp. 965–969.

[42] Everitt, B.: 'Euclidean distance'. in Cambridge: 'The Cambridge Dictionary of Statistics' (Cambridge University Press, 2002), p. 134.

[43] Sun, B., Cheng, W., Goswami, P. *et al.*: 'Flow-Aware WPT k-Nearest Neighbours Regression for Short-Term Traffic Prediction'. Proceedings of IEEE Symposium on Computers and Communication (ISCC), Heraklion, Greece, July 2017, pp. 48–53.

[44] Innamaa, S.: 'Self-adapting traffic flow status forecasts using clustering', *IET Intell. Transp. Syst.*, 2009, **3**, (1), pp. 67–76.

[45] Qi, Y., Smith, B. L.: 'Identifying nearest neighbors in a large-scale incident data archive', *Transp. Res. Rec.*, 2004, **1879**, (1), pp. 89–98.

[46] Maurer, C. R., Qi, R. S., Raghavan, V.: 'A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2003, **25**, (2), pp. 265–270.

[47] Laurikkala, J., Juhola, M., Kentala, E. *et al.*: 'Informal identification of outliers in medical data'. Fifth International Workshop on Intelligent Data Analysis in Medicine and Pharmacology, Berlin, Germany, May 2000, pp. 20–24.

[48] De Maesschalck, R., Jouan-Rimbaud, D., Massart, D. L.: 'The Mahalanobis distance', *Chemometrics and Intelligent Laboratory Systems*, 2000, **50**, (1), pp. 1–18.

[49] Filzmoser, P., Garrett, R. G., Reimann, C.: 'Multivariate outlier detection in exploration geochemistry', *Comput. Geosci.*, 2005, **31**, (5), pp. 579–587.

[50] Smith, B. L., Oswald, R. K.: 'Meeting Real–Time Traffic Flow Forecasting Requirements with Imprecise Computations', *Computer-Aided Civil and Infrastructure Engineering*, 2003, **18**, (3), p. 5.

[51] Sun, B., Cheng, W., Bai, G. *et al.*: 'Correcting and Complementing Freeway Traffic Accident Data Using Mahalanobis Distance Based Outlier Detection', *Teh. Vjesn.*, 2017, **24**, (5).

[52] Kehagias, D., Salamanis, A., Tzovaras, D.: 'Speed pattern recognition technique for short-term traffic forecasting based on traffic dynamics', *IET Intell. Transp. Syst.*, 2015, **9**, (6), pp. 646–653.

[53] Garcia, V., Debreuve, E., Barlaud, M.: 'Fast k nearest neighbor search using GPU'. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops CVPRW, Anchorage, AK, June 2008, pp. 1–6.

[54] Nvidia: 'CUDA Programming Guide 8.0'. (Nvidia Ltd., 2016).

[55] Sun, B., Ma, L., Cheng, W., *et al.*: 'An Improved k-Nearest Neighbours Method for Traffic Time Series Imputation', Chinese Automation Congress (CAC), Jinan, China, October 2017.

[56] Smith, B. L., Demetsky, M. J.: 'Traffic Flow Forecasting: Comparison of Modeling Approaches', *J. Transp. Eng.-ASCE*, 1997, **123**, (4), pp. 261–266.