Blekinge Institute of Technology
Blekinge Institute of Technology Doctoral Dissertation Series No. 2018:10
ISSN 1653-2090
ISBN 978-91-7295-360-4

AUTOMATED TRAFFIC TIME SERIES PREDICTION

Bin Sun

Akademisk avhandling
som för avläggande av Teknologie doktorsexamen vid
Blekinge Tekniska Högskola kommer att offentligen försvaras
i J1650, Valhallav. 1, Karlskrona, 2018-11-30 , kl. 13.30

Handledare:
Professor Wei Cheng, Kunming University of Science and Technology

Prashant Goswami, Institutionen för kreativa teknologier

Guohua Bai, Institutionen för kreativa teknologier

Opponent:
Professor Slawomir Nowaczyk, Halmstad University

Blekinge Tekniska Högskola
Institutionen för kreativa teknologier
Karlskrona

Abstract

Intelligent transportation systems (ITS) are becoming more and more effective. Robust and accurate short-term traffic prediction plays a key role in modern ITS and demands continuous improvement. Benefiting from better data collection and storage strategies, a huge amount of traffic data is archived which can be used for this purpose especially by using machine learning.

For the data preprocessing stage, despite the amount of data available, missing data records and their messy labels are two problems that prevent many prediction algorithms in ITS from working effectively and smoothly. For the prediction stage, though there are many prediction algorithms, higher accuracy and more automated procedures are needed.

Considering both preprocessing and prediction studies, one widely used algorithm is k-nearest neighbours (kNN) which has shown high accuracy and efficiency. However, the general kNN is designed for matrix instead of time series which lacks the use of time series characteristics. Choosing the right parameter values for kNN is problematic due to dynamic traffic characteristics. This thesis analyses kNN based algorithms and improves the prediction accuracy with better parameter handling using time series characteristics.

Specifically, for the data preprocessing stage, this work introduces gap-sensitive windowed kNN (GSW-kNN) imputation. Besides, a Mahalanobis distance-based algorithm is improved to support correcting and complementing label information. Later, several automated and dynamic procedures are proposed and different strategies for making use of data and parameters are also compared.

Two real-world datasets are used to conduct experiments in different papers. The results show that GSW-kNN imputation is 34% on average more accurate than benchmarking methods, and it is still robust even if the missing ratio increases to 90%. The Mahalanobis distance-based models efficiently correct and complement label information which is then used to fairly compare performance of algorithms. The proposed dynamic procedure (DP) performs better than manually adjusted kNN and other benchmarking methods in terms of accuracy on average. What is better, weighted parameter tuples (WPT) gives more accurate results than any human tuned parameters which cannot be achieved manually in practice. The experiments indicate that the relations among parameters are compound and the flow-aware strategy performs better than the time-aware one. Thus, it is suggested to consider all parameter strategies simultaneously as ensemble strategies especially by including window in flow-aware strategies.

In summary, this thesis improves the accuracy and automation level of short-term traffic prediction with proposed high-speed algorithms.

Keywords: Machine Learning, Time Series, Traffic Engineering