UNIVERSITY OF TECHNOLOGY SYDNEY
School of Electrical and Data Engineering

Modeling, Analysis and Application of Big Traffic Data for Intelligent Transportation Systems

by

Peibo Duan

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

Doctor of Philosophy

Sydney, Australia

2019

Certificate of Authorship/Originality

I, Peibo Duan declare that this thesis, is submitted in fulfillment of the requirements for the award of doctorate, in the School of Electrical and Data Engineering at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature: Signature removed prior to publication.

Date: 12th, January, 2020

ABSTRACT

Modeling, Analysis and Application of Big Traffic Data for Intelligent Transportation Systems

by

Peibo Duan

Intelligent Transportation System (ITS), an integrated system of people, roads, and vehicles by utilizing information and communications technology, has emerged as an efficient way of improving the performance of transportation systems, enhancing travel security, and providing more choice to travelers. Recently, it has been seen that the big data era for ITS is coming due to the wide use of traffic detectors like traffic cameras and GPSs. These traffic detectors can collect various types of traffic data that significantly contribute to the development of ITS, which has the benefit of the public with convenient and safe travel.

With big traffic data, data-driven methods provide powerful and theoretical support for data modeling, analysis, and applications. However, existing methods still suffer from some shortcomings. First, traffic predictors usually use black-box methods to capture the spatiotemporal correlation between traffic. As a result, it reduces the flexibility of predictors due to the time-varying spatial-temporal correlation caused by frequent variation of road conditions. Second, it is impossible to cover all urban areas with traffic detectors. Thus, data absence and data sparsity have an essential impact on the reliability of travel state monitoring in a large road network. Lastly, most big data applications are based on the centralized method for processing and analyzing data, which consume more time and computational resources, optimal decision making. These make research on big traffic data in ITS become both exciting and essential.

In this thesis, a physically intuitive approach is developed for short-term traffic flow prediction that captures the time-varying spatiotemporal correlation between traffic, mainly attributed to the road network topology, travel speed, and trip distribution. Experimental results demonstrate its superior accuracy and lower computational complexity compared with its counterparts. After that, a novel methodology is presented to estimate link travel time distributions (TTDs) using end-to-end (E2E) measurements detected by the limited traffic detectors. The experimental results show that the estimated results are in excellent agreement with the empirical distributions. Lastly, a distributed scheme is proposed for taxi cruising route recommendations based on taxi demands predicted by the proposed Graph Convolutional Network (GCN) based method. Experiment and simulation are both implemented. Experimental results validate the accuracy of the proposed taxi demand predictor. Simulation results indicate that our proposed taxi recommendation scheme is better than its counterparts in the aspects of minimizing the number of vacant taxis and maximizing the global revenue of taxi drivers.

Dissertation directed by Professor Guoqiang Mao
School of Electrical and Data Engineering

Acknowledgements

The completion of this dissertation has been possible with the inspiration and encouragement from many people, to whom I am greatly indebted.

First of all, I overwhelmingly pay my immeasurable appreciation and deepest gratitude to my supervisor, Professor Guoqiang Mao, for his persistent guidance, valuable recommendations, generous advice never-ending patience and ongoing support. I feel extremely fortunate to be mentored by him during my Ph.D. candidature.

I am also grateful to other members in the research team of Prof. Guoqiang, for their making a friendly working environment, and for their support, great assistance, and valuable advice to my research.

I sincerely acknowledge the deepest love and ongoing support from my beloved families during my studies. I am grateful to my mother, my grandparents, and my father in heaven, for their encouragement and blessing.

Finally, but mostly, I wish to express my deepest gratitude to my friends, Mencheng Gao, Haoqi Zhou, Yiyang Jiang, Ying Wang, Ying He, and Yang Yang, who were always standing by me in my hard times during this work, especially in the innumerable days and nights fulfilled with anxiety, self-denial and hopelessness. Thanks for all the joy and beautiful memories they brought to my life. Although for some of them, we may never see each other again, I consider my self extremely fortunate to have all of them in the journal of my life.

Peibo Duan

Sydney, Australia

July 2019

List of Publications

The following is a list of publications in refereed journals and conference proceedings produced during my Ph.D. candidature. In some cases, the journal papers contain material overlapping with the conference publications.

Journal Papers

J-1. **Peibo Duan**, Changsheng Zhang, Guoqiang Mao, and Bin Zhang, "Applying distributed constraint optimization approach to the user association problem in heterogeneous networks," *IEEE Transactions on Cybernetics*, vol. 48, no. 6, pp. 16961707, 2018.

J-2. **Peibo Duan**, Guoqiang Mao, Weifa Liang, Degan Zhang, "A Unified Spatio-temporal Model for Short-term Traffic Flow Prediction," **Accepted** by *IEEE Transactions on Intelligent Transportation System*, NOv. 2018.

J-3. **Peibo Duan**, Guoqiang Mao, Baoqi Huang, Jun Kang, "Estimation of Link Travel Times Distribution with Limited Traffic Detectors," **Accepted** by *IEEE Transactions on Intelligent Transportation System*, 2019.

Conference Papers

C-1. **Peibo Duan**, Guoqiang Mao, Shangbo Wang, Changsheng Zhang and Bin Zhang, "STARIMA-based Traffic Prediction with Time-varying Lags," *IEEE 19th Int. Conf. on Intelligent Transportation Systems*, pp. 1610-1615, 2016.

C-2. **Peibo Duan**,Guoqiang Mao, Wenwei Yue, and Shangbo Wang, "A Trade-off between Accuracy and Complexity: Short-term Traffic Flow Prediction with Spatio-temporal Correlations," *IEEE 21th Int. Conf. on Intelligent Transportation Systems*, pp. 1658-1663, 2018.

C-3. **Peibo Duan**, Guoqiang Mao, Changsheng Zhang, and Jun Kang, "A Unified STARIMA based Model for Short-term Traffic Flow Prediction," *IEEE 21th Int. Conf. on Intelligent Transportation Systems*, pp. 1652-1657, 2018.

C-4. **Peibo Duan**, Guoqiang Mao, Baoqi Huang, and Jun Kang, "Estimating Link Travel Time Distribution Using Network Tomography Technique," **Accepted** by *IEEE 22th Int. Conf. on Intelligent Transportation Systems*,Jun. 2019.

Contents

List of Figures

Abbreviation

ARIMA - AutoRegressive Integrated Moving Average

ANN - Artificial Neural Network

ARIMAX - ARIMA with EXogenous variables

BPNN - Back Propagation Neural Network

BN - Bayesian Network

BTMS - Bluetooth Traffic Monitoring System

BFS - Breadth First Search

CNN - Convolutional Neural Network

CDF - Cumulative Density Function

DTMC - Discrete Time Markov Chains

DCRNN - Diffusion Convolutional Recurrent Neural Network

EM - Expectation Maximization

E2E - End-to-End

FCL-Net - Fusion Cnvolutional Long short-term memory Network

GNN - Graph Neural Network

GMM - Gaussian Mixture Model

GCN - Graph Convolutional Network

GN - GirvanNewman

HPP - Homogeneous Poisson Process

HW - Holt Winters

ITS - Intelligent Transportation System

KARIMA - Kohonen ARIMA

KNN - K Nearest Neighbor

KDE - Kernel Density Estimator

LSTM - Long Short Term Memory

MAC - Media Access Control

MDP - Markov Decision Process

MIP - Mixed Integer Programming

POI - Position of Interest

PDF - Probability Density Function

RNN - Recursive Neural Network

STARIMA - Space-Time ARIMA

SVR - Support Vector Regression

SARIMA - Seasonal ARIMA

STW-KNN - Spatio-Temporal Weighted KNN

SVD - Singular Value Decomposition

STL - Seasonal and Trend decomposition using Loess

STGCNN - Spatial-Temporal Graph Convolutional Neural Network

STM-GCN - Spatio-Temporal multi-GCN

SACF - Spatial AutoCorrelation Function

SPACF - Spatial Partial AutoCorrelation Function

SMS - Space Mean Speed

TTP - Traffic Transition Probability

TMS - Time Mean Speed

TTD - Travel Time Distribution