DATA SCIENCE APPLICATION IN INTELLIGENT TRANSPORTATION SYSTEMS: AN INTEGRATIVE APPROACH FOR BORDER DELAY PREDICTION AND TRAFFIC ACCIDENT ANALYSIS

by

Lei Lin
January 5, 2015

A dissertation submitted to the
Faculty of the Graduate School of
the University at Buffalo, the State University of New York
in partial fulfillment of the requirements for the
degree of

Doctor of Philosophy

Department of Civil, Structural, and Environmental Engineering

ACKNOWLEDGEMENTS

First and foremost I would like to offer my greatest gratitude to my two supervisors, Dr. Adel W. Sadek and Dr. Qian Wang, who have supported me throughout my research studies with their patience and knowledge. Both of them have given me continuous encouragement and guidance on my research works in the past four and a half years. They always help me with useful suggestions to make research plans; they always respond very fast to my emails about any difficulties I met in the studies no matter how busy they are; and they always revise my paper sentence by sentence even a small typo in a figure. Without their help, this dissertation would not have been possible. I feel so fortunate to be one of their Ph.D. students. The moments that we have thought-provoking discussions, get promising experiment results, and work hard together on one paper will always be remembered by me.

Second, my special thanks to my family. My parents Jianzhou Lin and Jianrong Chu, who unconditionally support every decision I have made in my life, are my source of energy. My wife Yan Li, she always takes good care of me and also gives me huge help in getting my minor degree, the Master's Degree of Computer Science.

Finally, I would also like to thank my committee members, Dr. Panagiotis Ch. Anastasopoulos and Dr. Qing He for accepting the invitation to serve on my committee. I also feel very grateful for their precious advice and great ideas towards my dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ApEn	Approximate Entropy
AVI	Automatic Vehicle Identification
AVL	Automatic Vehicle Location
BMAP	Batch Markovian Arrival Process
CI	Conditional Independence
DTW	Dynamic Time Warping
ELC	Equally Likely Combination
ELV	Equally Likely Vehicles
EM	Expectation and Maximization
ETC	Electronic Toll Collection
FCM	Fuzzy C-means Clustering Method
FHWA	Federal Highway Administration
FP tree	Frequent Pattern Tree
k-NN	k Nearest Neighbor
HBDM	Hazard-based Duration Model
IAAFT	Iteratively Amplitude Adjusted Fourier Transform
ILD	Inductive Loop Detector
ITS	Intelligent Transportation System
LCC	Latent Class Clustering
LGP	Linear Gaussian Process

LRD	Long Range Dependence
MAPE	Mean Absolute Percentage Error
$M/E_K/n$	Queueing Model with Exponential inter-arrival times and Erlang service times
NITTEC	Niagara International Transportation Technology Coalition
OPR	Object Purity Ratio
PH	Phase Types
ROPR	Relative Object Purity Ratio
SARIMA	Seasonal Autoregressive Integrated Moving Average Model
SPN	Spinning Network Method
SPSS	Statistical Package for Social Sciences
SVR	Support Vector Regression

ABSTRACT

With the great progress in information and communications technologies in the past few decades, intelligent transportation systems (ITS) have accumulated vast amounts of data regarding the movement iof people and goods from one location to another. Besides the traditional fixed sensors and GPS devices, new emerging data sources and approaches such as social media and crowdsourcing can be used to extract travel-related data, especially given the wide popularity of mobile devices such as smartphones and tablets, along with their associated apps. To take advantage of all these data and to address the associated challenges, big data techniques, and a new emerging field called data science, are currently receiving more and more attention. Data science employs techniques and theories from many fields such as statistics, machine learning, data mining, analytical models and computer programming to solve the data analysis task. It is therefore timely and important to explore how data science may be best employed for transportation data analysis. In this doctoral study, an integrative approach is proposed for data science applications in ITS. The proposed approach constitutes to an integration of multiple steps in the data analysis process, or integration of different models to build a more powerful one. The integrative approach is applied and tested on two case studies: border crossing delay prediction and traffic accident data analysis.

For the first case study, a two-step border crossing delay prediction model is proposed, consisting of a short-term traffic volume prediction model and a