AN IMPROVED MACHINE LEARNING MODEL OF MASSIVE FLOATING CAR DATA (FCD) BASED ON FUZZY-MDL AND LSTM-C FOR TRAFFIC SPEED ESTIMATION AND PREDICTION

By

FATEMEH AHANIN

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

February 2023

FSKTM 2023 5

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

AN IMPROVED MACHINE LEARNING MODEL OF MASSIVE FLOATING CAR DATA (FCD) BASED ON FUZZY-MDL AND LSTM-C FOR TRAFFIC SPEED ESTIMATION AND PREDICTION

By

FATEMEH AHANIN

February 2023

Chair : Associate Professor NORWATI MUSTAPHA, PhD
Faculty : Computer Science and Information Technology

In today's world, traffic congestion is a major problem in almost all metropolitans. There has been much previous research developing new methods to improve accuracy of Traffic State Prediction (TSP) which are designed according to its advantage for static sensors such as video cameras, inductive loop detectors and other static sensors. However, static sensors are not able to store longer traffic flow patterns and capture the dynamics of traffic flow and their instalment is too expensive. Floating Car Data (FCD) is a convenient and cost-effective method to gather traffic condition information. It is regarded as GPS sensors which can probe a large scale of traffic flows in real time. Although FCD can cover more road segments across the road network compared to static sensors, GPS data are prone to missing data because of urban canyons and tall buildings that will affect the traffic prediction accuracy. Existence of missing data (known as data sparsity) have made the traffic prediction tasks even more sophisticated. There are two techniques used by the existing methods of TSP which are either with Traffic State Estimation (Traffic State Estimation) or without TSE. While TSE estimates the missing data in traffic states, such as speed and density to reduce data sparsity, TSP uses the traffic data to forecast the traffic state within a certain time period in future. When there are missing data in the dataset, TSP may use TSE for estimation of missing data and then performs prediction.

The aim of this thesis is to improve accuracy of TSP with TSE as well as without TSE by the improvement of LSTM. There are three (3) methods are proposed in this study. In the first method, a new algorithm called LSTM-C (Long Short-Term Memory (LSTM) with Contrast) is proposed to improve prediction of traffic speed without TSE. The existing research in traffic speed prediction used LSTM with single variable (traffic speed) and multi variables (traffic speed and vehicle headway). However, multivariate LSTM does not add any significant contribution to adequately predict traffic speed compared to single variate LSTM. This signifies that LSTM model requires improvement in term of identification of traffic speed changes within a certain time period. Thus, this study improved the traffic speed prediction using LSTM with Contrast Measure which detects the decreasing and increasing patterns in traffic speed. Speed

prediction accuracy of the proposed method LSTM-C and previous work LSTM achieved 96.67% and 94.86 respectively. In the second method, a new traffic estimation method is proposed using Fuzzy C-Mean (FCM) clustering and Minimum Description Length (MDL). MDL uses patterns to express the repeated presence in the data of particular items or clusters. Spectral clustering and Hidden Markov Model (HMM) has been used in detecting patterns by the existing research to estimate traffic speed. HMMs are well-suited for capturing first-order dependencies, also known as Markov dependencies. In an HMM, the future state (or observation) depends only on the current state and is independent of the past states. This behaviour of HMM makes it less effective in estimation of traffic data, because it might be necessary to consider several previous states when estimating a missing state. This thesis uses Fuzzy C-Mean and concept of MDL to constitute patterns and estimate the missing traffic state based on n previous states. The implementation results demonstrate proposed Fuzzy-MDL method has achieved accuracy of 96.46% which outperform the HMM-based model that achieved 93.14%. In the third method, a hybrid algorithm called LSTM-C-EST, which is a combination of Fuzzy-MDL and LSTM-C is proposed. The idea to propose this method is that estimating the value of missing traffic speed can improve the traffic prediction results. In this model, the Fuzzy-MDL is applied as the pre-processing step to estimate the missing traffic speeds. Then this new estimated data is used for prediction to predict the traffic speed in the next 5 minutes. The results of this model is compared with LSTM-C as well as the study by (shuming Sun et al., (2019) which performed traffic estimation and traffic prediction using HMM and a single variant LSTM. The accuracy of LSTM-C-EST, LSTM-C, LSTM are 98.05%, 96.69%, 94.90% respectively, which proves the LSTM-C-EST outperform the other two algorithms.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

MODEL PEMBELAJARAN MESIN YANG DITINGKATKAN DARI DATA KERETA TERAPUNG MASSIF (FCD) BERDASARKAN FUZZY-MDL DAN LSTM-C UNTUK ANGGAARAN DAN RAMALAN KELAJUAN LALU LINTAS

Oleh

FATEMEH AHANIN

Februari 2023

Dalam dunia hari ini, kesesakan lalu lintas adalah masalah utama di hampir semua metropolitan. Terdapat banyak penyelidikan terdahulu yang membangunkan kaedah baharu untuk meningkatkan ketepatan Ramalan Keadaan Trafik (TSP) yang direka bentuk mengikut kelebihannya untuk penderia statik seperti kamera video, pengesan gelung induktif dan penderia statik lain. Walau bagaimanapun, penderia statik tidak dapat menyimpan corak aliran trafik yang lebih panjang dan tidak dapat menangkap aliran trafik yang dinamik serta instalasinya terlalu mahal. Floating Car Data (FCD) ialah kaedah yang mudah dan kos efektif untuk mengumpulkan maklumat keadaan trafik. Ia dianggap sebagai penderia GPS yang boleh menyiasat aliran trafik berskala besar dalam masa nyata. Walaupun FCD boleh meliputi lebih banyak segmen jalan merentasi rangkaian jalan berbanding dengan penderia statik, data GPS terdedah kepada kehilangan data disebabkan lembah bandar dan bangunan tinggi yang akan menjejaskan ketepatan ramalan trafik. Kewujudan data yang hilang (dikenali sebagai data sparsity) telah menjadikan tugas ramalan trafik lebih canggih. Terdapat dua teknik yang digunakan oleh kaedah TSP sedia ada iaitu sama ada dengan Anggaran Keadaan Trafik (TSE) atau tanpa TSE. Manakala TSE menganggarkan data yang hilang bagi keadaan trafik, seperti kelajuan dan ketumpatan untuk mengurangkan kelompongan data, TSP pula menggunakan data trafik untuk meramalkan keadaan trafik pada masa hadapan dalam tempoh masa tertentu. Apabila terdapat data yang hilang dalam set data, TSP boleh menggunakan TSE untuk menganggarkan data yang hilang dan kemudian melakukan ramalan.

Matlamat tesis ini adalah untuk meningkatkan ketepatan TSP dengan TSE dan juga tanpa TSE dengan penambahbaikan LSTM. Terdapat tiga (3) kaedah dicadangkan dalam kajian ini. Dalam kaedah pertama, algoritma baharu yang dipanggil LSTM-C (Long Short-Term Memory (LSTM) with Contrast) dicadangkan untuk meningkatkan ramalan kelajuan trafik tanpa TSE. Penyelidikan sedia ada dalam ramalan kelajuan lalu lintas menggunakan LSTM dengan pembolehubah tunggal (kelajuan trafik) dan pembolehubah berbilang (kelajuan lalu lintas dan hala tuju kenderaan). Walau bagaimanapun, LSTM dengan pembolehubah berbilang tidak menambah apa-apa sumbangan penting untuk meramalkan kelajuan trafik dengan secukupnya berbanding LSTM dengan pembolehubah tunggal. Ini menandakan model LSTM memerlukan penambahbaikan dari segi mengenal pasti perubahan kelajuan trafik dalam tempoh masa tertentu. Justeru,

kajian ini menambah baik ramalan kelajuan trafik menggunakan LSTM dengan Contrast Measure yang mengesan corak penurunan dan peningkatan dalam kelajuan trafik. Ketepatan ramalan kelajuan kaedah LSTM-C yang dicadangkan berbanding LSTM sedia ada adalah 96.67% dan 94.86. Dalam kaedah kedua, kaedah anggaran trafik baharu dicadangkan menggunakan pengelompokan Fuzzy C-Mean (FCM) dan Minimum Description Length (MDL). MDL menggunakan corak untuk menyatakan kehadiran berulang dalam data item atau kelompok tertentu. Pengelompokan spektrum dan Hidden Markov Model (HMM) telah digunakan untuk dengan mengesan corak dalam penyelidikan sedia ada bagi menganggar kelajuan lalu lintas. HMM sangat sesuai untuk menangkap kebergantungan tertib pertama, juga dikenali sebagai kebergantungan Markov. Dalam HMM, keadaan masa depan (atau pemerhatian) hanya bergantung pada keadaan semasa dan bebas daripada keadaan masa lalu. Tingkah laku HMM ini menjadikannya kurang berkesan dalam menganggar data trafik, kerana ianya perlu mempertimbangkan beberapa keadaan sebelumnya apabila menganggarkan keadaan yang hilang. Tesis ini menggunakan Fuzzy C-Mean dan konsep MDL untuk membentuk corak dan menganggar keadaan trafik yang hilang berdasarkan n keadaan sebelumnya. Keputusan pelaksanaan menunjukkan kaedah Fuzzy-MDL yang dicadangkan telah mencapai ketepatan 96.46% yang mengatasi prestasi model berasaskan HMM yang mencapai 93.14%. Dalam kaedah ketiga, algoritma hibrid yang dipanggil LSTM-C-EST, yang merupakan gabungan Fuzzy-MDL dan LSTM-C dicadangkan. Idea untuk mencadangkan kaedah ini adalah penganggaran nilai kelajuan trafik yang hilang boleh meningkatkan hasil ramalan trafik. Dalam model ini, Fuzzy-MDL digunakan sebagai langkah pra-pemprosesan untuk menganggarkan kelajuan trafik yang hilang. Kemudian data anggaran baharu ini digunakan untuk ramalan bagi meramalkan kelajuan trafik dalam 5 minit seterusnya. Keputusan model ini dibandingkan dengan LSTM-C serta kajian sebelum yang melakukan anggaran trafik dan ramalan trafik menggunakan HMM dan LSTM varian tunggal. Ketepatan LSTM-C-EST , LSTM-C, LSTM masing-masing adalah 98.05%, 96.69%, 94.90%, yang membuktikan LSTM-C-EST mengatasi dua algoritma yang lain.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank God Almighty for giving me the opportunity and guidance to achieving my goal.

I would like to express my special thanks of gratitude to my supervisor Associate Prof. Datin Dr. Norwati Mustapha and my committee members Dr. Maslina Zolkepli and Dr. Nor Azura Husin who provided invaluable advice and helped in difficult periods. They have been the backbone of this thesis, providing invaluable suggestions and constructive criticism that have breathed life into this study. Their insightful contributions have added depth and significance to the research.

I would also like to give special thanks to my parents, Fereidoun and Kobra, and my sisters Zahra and Elaheh, for their continuous support and understanding when undertaking my research and writing my project. Your prayer for me was what sustained me this far.

Last but not least, I give many thanks to the faculty of Computer Science and Information Technology, administration staff and lecturers for their kindness and cooperation with me during my studies at the University. I am also grateful to my friends for being there for me.

I certify that a Thesis Examination Committee has met on 8th February 2023 to conduct the final examination of Fatemeh Ahanin on her thesis entitled “An Improved Machine Learning Model of Massive Floating Car Data (FCD) Based on Fuzzy-MDL And LSTM-C for Traffic Speed Estimation and Prediction” in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the (insert the name of relevant degree).

Members of the Thesis Examination Committee were as follows:

Razali bin Yaakob, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)Name of Faculty

Name of Examiner 1, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Name of Examiner 2, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Prof. Dr. Xiaofang Zhou, PhD
Professor
Computer Science and Engineering (CSE)
The Hong Kong University of Science and Technology (HKUST)
China
(External Examiner)

____
SITI SALWA BINTI ABD GANI, PhD
Associate Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Norwati Mustapha, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Maslina Zulkepli, PhD
Senior Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Name of Member 2, PhD (omit `PhD' if not applicable)
Senior Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

SITI SALWA BINTI ABD GANI, PhD
Associate Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

Declaration by the Graduate Student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any institutions;
- intellectual property from the thesis and the copyright of the thesis are fully-owned by Universiti Putra Malaysia, as stipulated in the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from the supervisor and the office of the Deputy Vice-Chancellor (Research and innovation) before the thesis is published in any written, printed or electronic form (including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials) as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld in accordance with the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2015-2016) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software

Signature: ________________________ Date: __________________

Name and Matric No.: _______________________________________

Declaration by Members of the Supervisory Committee

This is to confirm that:

- the research and the writing of this thesis were done under our supervision;
- supervisory responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2015-2016) are adhered to.

Signature: ______________________________

Name of Chairman of Supervisory Committee: ______________________________

Signature: ______________________________

Name of Member of Supervisory Committee: ______________________________

Signature: ______________________________

Name of Member of Supervisory Committee: ______________________________

Signature: ______________________________

Name of Member of Supervisory Committee: ______________________________

TABLE OF CONTENTS

	Page
ABSTRACT	i
ABSTRAK	iii
ACKNOWLEDGMENTS	v
APPROVAL	vi
DECLARATION	viii
LIST OF TABLES	xii
LIST OF FIGURES	xiv
LIST OF ABBREVIATIONS	xvii

CHAPTER

1. INTRODUCTION	**1**
1.1 Background	1
1.2 Problem Statement	2
1.3 Research Objective	4
1.4 Research Scope	4
1.5 Thesis Organization	5
2. LITERATURE REVIEW	**7**
2.1 Introduction	7
2.2 Intelligent Transportation System (ITS)	7
2.2.3 Road Network	8
2.2.4 Traffic Data	8
2.2.4.1 Types of Traffic Data	8
2.2.4.2 Macroscopic vs Microscopic Traffic Data Analysis	10
2.3 Traffic State Estimation (TSE)	10
2.3.1 Model-based approach	11
2.3.2 Data-driven approach	12
2.3.2.1 Statistical methods	13
2.3.2.2 Machine Learning Methods	17
2.5.3 Streaming-data-driven	20
2.4 Traffic State Prediction (TSP)	20
2.4.1 Probabilistic Reasoning	21
2.4.1.1 Fuzzy logic	21
2.4.1.2 Hidden Markov Model (HMM)	21
2.4.1.3 Gaussian Processes (GPs)	22
2.4.1.4 Bayesian Network (BN)	23
2.4.2 Shallow Machine Learning (SML)	26
2.4.2.1 Artificial Neural Network (ANN)	26
2.4.2.2 Regression Model	27
2.4.2.3 Decision Tree (DT)	28
2.4.2.4 Support Vector Machine (SVM)	29
2.6.3 Deep Learning (DL)	33
2.6.3.1 Convolutional neural network (CNN)	33
2.6.3.2 Recurrent Neural Networks (RNNs)	34
2.5 Traffic Congestion	40
2.6 Gap Analysis	40

2.6 Summary 41
3. **RESEARCH METHODOLOGY** **43**
3.1 Introduction 43
3.2 Research Methodology 43
3.3 Problem Formulation 45
3.4 System Requirements 46
3.5 Dataset 46
3.5.1 Dataset Description 47
3.6 Data Pre-processing 48
3.6.1 Data preparation and Data Normalization 49
3.6.2 Data Cleaning 49
3.6.3 Parameter Determination and Data Formation 50
3.6.4 Splitting Dataset 51
3.8 Performance Measure 51
3.9 Summary 52
4. **SPEED PREDICTION USING LSTM-C** **53**
4.1 Introduction 53
4.2 The Basic of LSTM Model 53
4.3 Definition of Contrast Measure 55
4.4 Proposed LSTM-C Model 57
4.4.1 Usage of Average Speed and Contrast in LSTM 57
4.5 Model Implementation 60
4.5.1 Test Environment 60
4.5.2 Dataset 60
4.5.2.1 Data of speed prediction in a road segment 61
4.5.3 Hyperparameter 63
4.6 Results and discussion 67
4.6.1 Speed Prediction 67
4.6.2 Identification of Traffic congestion 73
4.6.3 Speed prediction in a road segment 74
4.7 Summary 80
5. **SPEED ESTIMATION USING FUZZY-MDL** **81**
5.1 Introduction 81
5.2 Proposed Fuzzy-MDL Model 81
5.3 Parameter Tuning 87
5.4 Performance Measurement 88
5.5 Summary 90
6. **COMBINING FUZZY-MDL AND LSTM-C** **93**
6.1 Introduction 93
6.2 Proposed LSTM-C-EST 93
6.3 Model implementation 97
6.4 Results and Discussion 98
6.4.1 Speed Prediction 98
6.4.2 Identification of Traffic Congestion Level 103
6.4.3 Speed prediction in a road segment 103
6.5 Summary 109
7. **CONCLUSION** **111**
7.1 Summary 111
7.2 Future Work 112
REFERENCES **115**

LIST OF TABLES

Table		Page
2.1	Review Of Statistical Methods In Tse	15
2.2	Review Of Ml Methods In Tse	18
2.3	Review Of Probabilistic Reasoning Algorithms In Tsp	24
2.4	Review Of Shallow Machine Learning Algorithms In Tsp	30
2.5	Review Of Deep Learning Algorithms In Tsp	37
2.6	The Traffic Congestion Level Based On Average Speed (V), By Chengdu Transportation Department	40
2.7	The Traffic Congestion Level Based On Average Speed (V), By Chengdu Transportation Department	41
3.1	Dataset Used In Experiments	46
3.2	The Features Of The Data Trajectory (Beijing)	47
3.3	Examples Of The Data Trajectory (Beijing)	47
3.4	Description Of Fcd Of Xuancheng City	48
3.5	The Input Features Of The Fcd (Xuancheng City)	48
3.6	Traffic Index And Speed Ranges	49
4.1	Number Of Test And Train Samples In Each Dataset	61
4.2	Descriptive Statistics Of The Speed Feature In The Fcd Of Beijing And Xuancheng	61
4.3	Range Of The Parameters Used To Find Optimal Parameters	64
4.4	Using Different Optimizers On The Model. The Optimal Structure Refers To The Number Of Neurons At Each Layer.	65
4.5	Lstm-C, Lstm, And Gru Hyperparameter	66
4.6	Performance Comparison Of The Proposed Model With Lstm Model On Fcd Of Beijing (Majumdar Et Al, 2021)	67
4.7	Performance Comparison Of The Proposed Model With Lstm Model Based On Xuancheng Fcd (Majumdar Et Al, 2021)	72

4.8 The Classification Of Traffic Congestion Level In Beijing Arterial Roads (Chen Et Al., 2020) 73

4.9 Congestion Level Identification Of The Proposed Model With Lstm Model (Majumdar Et Al, 2021) 73

5.1 The Effect Of Timeslot Support Value On The Mae 87

5.2 The Fuzzy-Mdl Parameters 88

5.2 Traffic Flow State Estimation Accuracy Comparison 89

6.1 Performance Comparison Of Lstm-C-Est, Lstm, Gru-C & Lstm 98

LIST OF FIGURES

Figure		Page
2.1	Static Traffic Detector (Radwan Et Al., 2023)	9
2.2	GPS Signals Communication Between GPS Enabled Devices and Satellites Receives (Family1st, 2022)	9
2.3	Traffic State Estimation with Assumptions and Inputs (Seo Et Al., 2017)	11
2.4	Classification of Artificial Intelligence Techniques for Traffic State Prediction (Akhtar & Moridpour, 2021)	20
2.5	Basic Structure of ASNN	27
2.6	Deep Learning Models	33
2.7	Basic Structure of Lstm Model	35
3.1	Research Methodology	44
4.1	The Structure of The Hidden Layer	54
4.2	GLCM Matrix For a Speed Sequence. In The Speed Sequence, 1 Represents a Speed Range Between 0 To 5 Km/H, 2 Represents Speed Range Between 5 to 10 Km/H, and so on. The Values of Cells Are The Total Occurrence of The Two Consecutive Speeds (Pair).	56
4.3	LSTM-C Network Model	58
4.4	The Workflow of The Proposed LSTM-C Model	59
4.5	A Sample of Time Series Data Sliding Window (Rolling Window Size of 6 And 8).	62
4.6	Average of Traffic Speed Of All Road Segments in FCD (Beijing) From Timeslot 73 (6:00 Am) to Timeslot 181 (3:00 Pm)	62
4.7	A Sample of Traffic Speed Data of Random 100 Road Segments Between Timeslot 73 to 181 (6:00 Am To 3:00 Pm) in Two Fcd Datasets (A) FCD of Beijing, (B) FCD of Xuancheng	63
4.8	The Average Speed Prediction For Rolling Window Size 6, 8 and 12 for LSTM-C, LSTM (Majumdar Et Al, 2021), GRU-C and GRU	71
4.9	Comparison of Predicted Speed And Rmse Between Proposed Model and Benchmark. 18 Timesteps are Used to Predict The Speed at The Next Timestep (Rolling Window Size = 18). The Left Part Shows Road Segment #1 with Large and Frequent Speed Fluctuation, and The Right Part Shows Road Segment #1000 with Relatively Low Speed Changes.	76

4.10 Comparison of Predicted Speed and RMSE Between Proposed Model and Benchmark. 6 Timesteps Are Used to Predict The Next Timestep (Rolling Window Size = 6). The Left Part Shows Road Segment #1 with Large and Frequent Speed Fluctuation, and The Right Part Shows Road Segment #1000 with Relatively Low Speed Changes. 79

5.1 Fuzzy C-Mean And MDL Model 81

5.2 Illustration of The GPS Noise In The FCD 82

5.3 The Algorithm To Converge A Local Minimum of J_m. 84

5.4 Traffic State Estimation Algorithm 86

5.5 Traffic Flow Distribution Over Time Using Fuzzy C-Mean Clustering And MDL 86

5.6 Illustrates The Decreasing Gradient of Mae as The Number of Timeslot Support Value Increases. 87

5.7 Comparison of Accuracies of The Fuzzy-MDL And Hmm-Based Models 89

5.8 Traffic States of Arterials in Beijing At The 50th Timeslot (2:20 Pm – 2:20 Pm) Before Estimation, and Figure 8(B) Shows The Traffic Condition of The Arterials After Estimation. 90

6.1 Average Speed of 6 Sample Road Segments, Before Missing Data Estimation, from Timeslot 85 To 104 (7 Am To 8:30 Am). Data Points with Value 0 Implies The Unobserved Values. 94

6.2 Average Speed of 6 Sample Road Segments, After Missing Data Estimation, from Timeslot 85 To 104 (7 Am To 8 30 Am). Data Points With Value 0 Implies The Unobserved Values. 94

6.3 The Number of Missing Speed Data in Each Timeslot 95

6.4 A Sample of Missing Speed Data 95

Figure 6.5: Steps of Identifying and Estimating Missing Speed Data 95

6.6 The Number of Missing Speed Data Before and After Using Proposed Speed Estimation Method (Fuzzy-MDL) 96

6.7 The Workflow of The Proposed LSTM-C-EST Model 97

6.8 The Average Speed Prediction for Rolling Window Size 6, 8, 12 and 18 for LSTM-C-EST, LSTM (Shuming Sun et al., 2019), GRU-C-EST, and GRU (Shuming Sun et al., 2019) 102

6.9 Comparison of Predicted Speed and RMSE Between Proposed Model and Benchmark. 18 Timesteps Are Used to Predict The Next Timestep (Rolling Window Size = 18). The Left Part Shows Road Segment #1 With Large and Frequent Speed Fluctuation, and The Right Part Shows Road Segment #1000 With Relatively Low Speed Changes. 105

6.10 Comparison of Predicted Speed and RMSE Between Proposed Model and Benchmark. 6 Timesteps Are Used to Predict The Next Timestep (Rolling Window Size = 6). The Left Part Shows Road Segment #1 with Large and Frequent Speed Fluctuation, and The Right Part Shows Road Segment #1000 with Relatively Low Speed Changes. 108

LIST OF ABBREVIATIONS

ACC	Accuracy
AI	Artificial Intelligence
ANN	Artificial Neural Network
BP	Back Propagation
BN	Bayesian Network
CNN	Convolutional Neural Network
DL	Deep Learning
DM	Data Mining
DML	Deep Machine Learning
DT	Decision Tree
FD	Fundamental Diagram
FNN	Feed-Forward Neural Network
FS	Feature Selection
FCD	Floating Car Data
FCM	Fuzzy C-Mean
Faster R-CNN	Faster Regional Convolutional Neural Network
GA	Genetic Algorithm
GPS	Global Positioning System
GRU	Gated Recurrent Units
ITS	Intelligent Transportation Systems

K-NN	K-Nearest Neighbour
LSTM	Long short-term memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage
ML	Machine Learning
MLP	Multilayer Perceptron
MSE	Mean Squared Error
PSO	Particle Swarm Optimization
RBF	Radial Basis Function
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SRMF	Sparsity Regularized Matrix Factorization
SVM	Support Vector Machine
TSE	Traffic State Estimation
TSP	Traffic State Prediction

CHAPTER 1

INTRODUCTION

1.1 Background

In today's world, traffic congestion is a major problem in almost all metropolitans. This phenomenon has affected several aspects from people's daily life to economy. Mobility of people, travel time duration, quality of life, transportation planning systems and traffic management are examples which bear the effects of traffic congestion. This problem is even becoming more crucial due to increasing numbers of vehicles especially in arterial roads. Governments, universities and Research and Development (R&D) sectors have tackled this problem and tries to reduce traffic congestion using traffic monitoring and management technologies. These technologies require sufficient and accurate traffic data.

Due to pervasive technologies in telecommunication and transportation systems, there are massive amount of traffic data available. In order to determine traffic state (e.g., flow velocity and traffic density), video cameras, inductive loop detectors, and other static sensors can be deployed at fixed positions on roads. While data collected from these devices are accurate and sufficient in order to be used for traffic management, these traditional approaches cannot cover all roads because they need considerable infrastructure development and high maintenance costs (Akhtar & Moridpour, 2021).

On the other hand, Floating Car Data (FCD) is a convenient and cost-effective method to gather traffic condition information. It is regarded as GPS sensors and does not need any specific device and offers good coverage across road networks with defined penetration rate. However, FCD are prone to have missing data because of urban canyons and tall buildings which affected the traffic estimation and prediction results (Newson and Krumm, 2009). Existence of these errors has made the estimation and prediction tasks even more sophisticated. Traffic management applications, trip recommendations and any other applications require accurate data to have precise. There are methods to remove the noises, however the amount of data is very scare, it is not suitable for the aforementioned applications. Moreover, in some cases the GPS devices might update their locations once in few minutes which exact speed and location remains unknown, and the vehicle might pass several road segments having diverse speed without updating its traveling data. The latter issue is known as data sparsity or missing data. Furthermore, in these data only some of the vehicles in the roads report their location which the real traffic characteristics of traffic such as flow and density remain unknown. Consequently, there will be a difference between the actual numbers of vehicles travelling on a road segment with the reported vehicles.

Machine Learning offers various methods and models which aid to discover knowledge, recognize hidden pattern, estimate and predict traffic states. These methods and models aid to overcome data sparseness to provide more accurate and sufficient traffic data. Traffic State Estimation (TSE) and prediction technologies have a challenging task to control and monitor traffic and aiming to improve traffic management and monitoring. TSE refers to the procedure of inferring traffic state (i.e., flow, density, speed, and other similar variables) from partially observed traffic data on road segments (Seo et al., 2017). In this manner, TSE can estimate the missing traffic states caused by removal of GPS errors or low sampling rates. Traffic State Prediction task is to forecast the traffic state variable within a certain time period. The accuracy of the short-term traffic speed prediction model significantly influences the performance of real-time traffic control and management in Intelligent Transportation Systems (ITS).

1.2 Problem Statement

Traffic State Estimation (TSE) and Traffic State Prediction (TSP) both aim to help traffic management using different methods. While TSE is more on the preprocessing step, TSP is a more prominent stage in traffic management which predicts traffic states such as speed. Here are two techniques used by the existing methods of TSP which is either with TSE or without TSE. While TSE estimates the missing data in traffic states, such as speed and density to reduce data sparsity, TSP uses the traffic data to forecast the traffic state within a certain time period in future. When there are missing data in the dataset, TSP may use TSE for estimation of missing data and then performs prediction.

In TSP, Speed is an important traffic states which can reveal the severity of congestion in a road segment. Therefore, the accuracy of traffic speed prediction is essential to identify traffic congestion earlier. In this reason simple and traditional ML algorithms such as Support Vector Machine (SVM), Artificial Neural Network (ANN), Regression Model, and Decision Trees (DT) (Tseng and Hsueh, 2018; Nadeem and fowdur, 2018; Feng et al., 2019; Xu et al., 2019; X. Yang et al., 2019; Navarro-Espinoza et al., 2022a) were used but some problems still remain. Failing to deliver effective results and accuracy drop when the amount of missing data increases are common issues in Shallow Machine Learning (SLM) methods. Another group which has powerful methods for traffic prediction is Deep Learning (DL). One of the advantages of DL over ANN is that it does not require prior knowledge for extracting features from the input data. Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and LSTM are DL algorithms which are utilized in traffic prediction (Ma et al., 2017; M. Chen et al., 2018). RNN and its branch LSTM are also well-known Deep learning algorithms that performed good on traffic state prediction (Ma et al., 2015; Zhang et al., 2019; Shuming Sun et al., 2019; Majumdar et al., 2021). One of the advantages of RNN algorithms is their short-term memory which allows them to model data nonlinearity in time series. Besides its short-term memory advantage, might interface difficulty in the training phase. This issue is smoothened in the LSTM but for LSTM-based models, the prediction always depends on the data. The existing research in traffic speed prediction used LSTM with single variable (traffic speed) and multi variables (traffic speed and vehicle headway) (Majumdar et al., 2021). However, they concluded multivariate LSTM does not add any significant contribution to adequately predict traffic speed compared to single variate LSTM. This signifies that the model's ability to learn meaningful patterns (increasing and decreasing speed changes) from the data requires improvement. Accurate identification of changes in speed data at road segments is still a challenge which needs to be addressed.

Moreover, many studies have used loop detectors data and other static sensors data (M. Chen et al., 2018; Zhang et al., 2019; Majumdar et al., 2021). However, loop detectors data are not able to store longer traffic flow patterns and capture the dynamics of traffic flow and their instalment is too expensive. FCD has more coverage of the road segments across the road network (compare to loop detector data) and can capture the dynamic nature of traffic trends. However, FCD data may suffer from data sparsity (missing data), which affects the accuracy of the LSTM. Hence, dealing with this problem of missing data is crucial for such models (Selim Reza et al.,2022). To deal with data sparsity, TSE methods play an important role. When the amount of available data is too sparse, the TSE algorithms face difficulty to identify traffic characteristics in the road segments and the accuracy of estimation would drop drastically. Some researchers have used tensor decomposition-based model to perform traffic state estimation (Chen et al., 2019; Xu et al., 2020). It is important to note that spatiotemporal correlations are the main component in many estimation algorithms. Hence, it is essential for TSE methods to incorporate spatiotemporal features properly, in a way that reflects their reliance in accurate estimation. X. Wang et al., (2015) proposed a method to estimate traffic speed by detecting patterns using spectral clustering and HMM. HMMs are well-suited for capturing first-order dependencies, also known as Markov dependencies. In an HMM, the future state (or observation) depends only on the current state and is independent of the past states. This behaviour of HMM makes it less effective in estimation of traffic data, because it might be necessary to consider several previous states when estimating a missing state.

This study addresses the following issues based on the above limitations:

1. Most of the existing research on traffic speed prediction lack effective identification in changes in traffic trends.

2. The accuracy of estimating missing traffic state must be improved.

3. While Deep Learning methods are the most recent technologies for TSP, they still require improving accuracy when there is missing data in the dataset.

1.3 Research Objective

The main purpose of the study is to propose a novel traffic state estimation and prediction model using FCD. The objectives of this research are as follows:

1. To propose Contrast Measure and combine with LSTM (LSTM-C) to identify changes in traffic speed within a time period to improve traffic speed prediction using FCD.

2. To propose a novel traffic state estimation method using Fuzzy C-Mean algorithm and Minimum Description Length (MDL) concept called Fuzzy-MDL, to estimate the missing traffic speed with diverse missing rate percentages and improve the accuracy of speed estimation.

3. To combine the proposed speed estimation (Fuzzy-MDL) and speed prediction (LSTM-C) methods called LSTM-C-EST to solve data sparsity in FCD dataset and enhance data in order to improve accuracy of speed prediction.

In the first objective, the proposed LSTM-C is evaluated using FCD. It is important to note that FCD includes missing data due to the aforementioned issues and the actual value of this missing data is unknown.

In the second objective, random data values form FCD are removed, and the proposed Fuzzy-MDL estimated the removed data values. Fuzzy-MDL is evaluated by comparing the estimated data values with the removed data values.

The third objective aims to combine the first and second proposed methods (objective one and two) as one method called LSTM-C-EST to estimate the actual missing data exists in the FCD dataset and fed it to the prediction layer and improve the accuracy, in comparison with the first method (objective one, LSTM-C). The purpose of the combination is to prove that effectiveness of the proposed estimation method (Fuzzy-MDL) and its capability in increasing prediction accuracy.

1.4 Research Scope

Traffic management and controlling is a huge research scope. There are many research paths in the area. Some of these studies have been carried out by traffic engineering and civil engineering departments which require certain kind of data or equipment from government or private sectors (e.g., accident detection, traffic light management).

The focus of the research is on traffic state estimation and prediction through ML (Fuzzy and LSTM) techniques using FCD. This study aims to estimate missing traffic speed and performs short-term traffic speed prediction using data in arterial roads. The goal is to discover patterns and trends in traffic data and improve speed prediction accuracy. Traffic theory and other fundamental concepts are not in this research scope and can be found in traffic engineering and civil engineering studies. The main FCD dataset used in this research is from Beijing. Another FCD dataset from Xuancheng is applied in Chapter 4. The benchmarks used in this study have similar background and are using Machine Learning and data-driven techniques only rather than employing physical traffic models and Traffic Theory. The performance measures used by benchmarks are MAE, MSE, and RMSE.

1.5 Thesis Organization

The thesis is organized as follows:

Chapter 1 explains the background of research, motivation, problem statement, research objectives and the scope of the research.

Chapter 2 provides in-depth research and examines relevant literature of the subject matter including traffic state estimation and prediction using traffic data. Moreover, this chapter includes a summary of ML techniques and its significance in improving traffic state estimation and prediction.

Chapter 3 discusses the research methodologies which have been used in this study. The research methodology provides a detailed guidance to the reader to understand this thesis.

Chapter 4 provides the design and evaluation of the proposed prediction method using the proposed LSTM-C.

Chapter 5 provides the design and evaluation of the proposed state estimation method using Fuzzy-MDL.

Chapter 6 describes the design and evaluation of the combined method using Fuzzy-MDL and LSTM-C called LSTM-C-EST.

Chapter 7 conclude the research and recommended some promising direction for future research.

REFERENCES

Adetiloye, T., & Awasthi, A. (2019). Multimodal Big Data Fusion for Traffic Congestion Prediction. Multimodal Analytics for Next-Generation Big Data Technologies and Applications, 319–335.

Akhtar, M., & Moridpour, S. (2021). A Review of Traffic Congestion Prediction Using Artificial Intelligence. Journal of Advanced Transportation, 2021.

Alghamdi, T., Elgazzar, K., Bayoumi, M., Sharaf, T., & Shah, S. (2019). Forecasting traffic congestion using ARIMA modeling. 2019 15th International Wireless Communications and Mobile Computing Conference, IWCMC 2019, 1227–1232.

An, S. H., Lee, B. H., & Shin, D. R. (2011). A survey of intelligent transportation systems. Proceedings - 3rd International Conference on Computational Intelligence, Communication Systems and Networks, CICSyN 2011, 332–337.

Antoniou, C., Ben-Akiva, M., Koutsopoulos, H.N., 2007. Nonlinear kalman filtering algorithms for on-line calibration of dynamic traffic assignment models. IEEE Trans.Intell. Transp. Syst. 8, 661–670.

Asencio-Cortés, G., Florido, E., Troncoso, A., & Martínez-Álvarez, F. (2016). A novel methodology to predict urban traffic congestion with ensemble learning. Soft Computing 2016 20:11, 20(11), 4205–4216.

Ban, X., Guo, C., & Li, G. (2016). Application of Extreme Learning Machine on Large Scale Traffic Congestion Prediction. 293–305.

Bogaerts, T., Masegosa, A. D., Angarita-Zapata, J. S., Onieva, E., & Hellinckx, P. (2020). A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data. Transportation Research Part C: Emerging Technologies, 112, 62–77.

Bonifay, W. (2022). Increasing generalizability via the principle of minimum description length. In Behavioral and Brain Sciences (Vol. 45). Cambridge University Press.

Cao, W., & Wang, J. (2019). Research on traffic flow congestion based on Mamdani fuzzy system. AIP Conference Proceedings, 2073(1), 020101.https://doi.org/10.1063/1.5090755

Chan, R. K. C., Lim, J. M. Y., & Parthiban, R. (2021). A neural network approach for traffic prediction and routing with missing data imputation for intelligent transportation system. Expert Systems with Applications, 171, 114573.

Chen, B., Hu, K., Li, Y., & Miao, L. (2022). Hybrid Spatio-Temporal Graph Convolution Network For Short-Term Traffic Forecasting. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 2022-October, 2128–2133. https://doi.org/10.1109/ITSC55140.2022.9921809

Chen, D., Yan, X., Liu, X., Li, S., Wang, L., & Tian, X. (2021). A Multiscale-Grid-Based Stacked Bidirectional GRU Neural Network Model for Predicting Traffic Speeds of Urban Expressways. IEEE Access, 9, 1321–1337.

Chen, M., Yu, X., & Liu, Y. (2018). PCNN: Deep Convolutional Networks for Short-Term Traffic Congestion Prediction. IEEE Transactions on Intelligent Transportation Systems, 19(11), 3550–3559.

Chen, Y. Y., Lv, Y., Li, Z., & Wang, F. Y. (2016). Long short-Term memory model for traffic congestion prediction with online open data. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 132–137.

Chen, Z., Jiang, Y., & Sun, D. (2020). Discrimination and Prediction of Traffic Congestion States of Urban Road Network Based on Spatio-Temporal Correlation. IEEE Access, 8(2), 3330–3342.

D'Andrea, E., & Marcelloni, F. (2017). Detection of traffic congestion and incidents from GPS trace analysis. Expert Systems with Applications, 73, 43–56.

Dimitrakopoulos, G., & Demestichas, P. (2010). Intelligent transportation systems: Systems based on cognitive networking principles and management functionality. IEEE Vehicular Technology Magazine, 5(1), 77–84.

Family1st. (2022, September 30). Everything you need to know about satellite GPS tracking system. Family1st. https://family1st.io/what-is-a-satellite-gps-tracking-system/

Fan, S., Herty, M., & Seibold, B. (2014). Comparative model accuracy of a data-fitted generalized Aw-Rascle-Zhang model. Networks & Heterogeneous Media. 2014, Volume 9, Pages 239-268, 9(2), 239. https://doi.org/10.3934/NHM.2014.9.239

Fazekas, Z., Balázs, G., Gerencsér, L., & Gáspár, P. (2017). Inferring the actual urban road environment from traffic sign data using a minimum description length approach. Transportation Research Procedia, 27, 516–523.

Feng, X., Ling, X., Zheng, H., Chen, Z., & Xu, Y. (2019). Adaptive multi-kernel SVM with spatial-temporal correlation for short-term traffic flow prediction. IEEE Transactions on Intelligent Transportation Systems, 20(6), 2001–2013.https://doi.org/10.1109/TITS.2018.2854913

Fulari, S., Vanajakshi, L., & Subramanian, S. C. (2017). Artificial Neural Network–Based Traffic State Estimation Using Erroneous Automated Sensor Data. Journal of Transportation Engineering, Part A: Systems, 143(8), 05017003.

Hara, Y., Suzuki, J., & Kuwahara, M. (2018). Network-wide traffic state estimation using a mixture Gaussian graphical model and graphical lasso. Transportation Research Part C: Emerging Technologies, 86, 622–638.

Holden, H., & Risebro, N. H. (2017). Follow-the-Leader models can be viewed as a numerical approximation to the Lighthill-Whitham-Richards model for traffic flow. Networks and Heterogeneous Media, 13(3), 409–421.

Huang, Z., Xia, J., Li, F., Li, Z., & Li, Q. (2019). A Peak Traffic Congestion Prediction Method Based on Bus Driving Time. Entropy 2019, Vol. 21, Page 709, 21(7), 709. https://doi.org/10.3390/E21070709

Jain, A., Koppula, H. S., Raghavan, B., Soh, S., & Saxena, A. (2015). Car that knows before you do: Anticipating maneuvers via learning temporal driving models. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 3182-3190).

Jin, J., & Ma, X. (2019). A non-parametric Bayesian framework for traffic-state estimation at signalized intersections. *Information Sciences*, *498*, 21-40.

Ke, R., Li, W., Cui, Z., & Wang, Y. (2020). Two-stream multi-channel convolutional neural network for multi-lane traffic speed prediction considering traffic volume impact. *Transportation Research Record*, *2674*(4), 459-470.

Kim, J., & Wang, G. (2016). Diagnosis and Prediction of Traffic Congestion on Urban Road Networks Using Bayesian Networks: Https://Doi.Org/10.3141/2595-12, 2595, 108–118.

Kong, X., Xu, Z., Shen, G., Wang, J., Yang, Q., & Zhang, B. (2016). Urban traffic congestion estimation and prediction based on floating car trajectory data. Future Generation Computer Systems, 61, 97–107.

Kumar, P. K., & Gopal, A. V. (2019). Developing big data frame works for efficient intelligent transport system for safety and security. Journal of Theoretical and Applied Information Technology, 97(6), 1833–1842.

Laña, I., Lobo, J.L., Capecci, E., Del Ser, J. and Kasabov, N., 2019. Adaptive long-term traffic state estimation with evolving spiking neural networks. Transportation Research Part C: Emerging Technologies, 101, pp.126-144.

Lee, C., Kim, Y., Jin, S., Kim, D., Maciejewski, R., Ebert, D., & Ko, S. (2020). A Visual Analytics System for Exploring, Monitoring, and Forecasting Road Traffic Congestion. IEEE Transactions on Visualization and Computer Graphics, 26(11), 3133–3146.

Lee, J., Hong, B., Lee, K., & Jang, Y. J. (2015, December). A prediction model of traffic congestion using weather data. In 2015 IEEE International Conference on Data Science and Data Intensive Systems (pp. 81-88). IEEE.

Liu, Yi, Feng, X., Wang, Q., Zhang, H., & Wang, X. (2014). Prediction of Urban Road Congestion Using a Bayesian Network Approach. Procedia - Social and Behavioral Sciences, 138, 671–678.

Liu, Yunxiang, & Wu, H. (2018). Prediction of road traffic congestion based on random forest. Proceedings - 2017 10th International Symposium on Computational Intelligence and Design, ISCID 2017, 2, 361–364. https://doi.org/10.1109/ISCID.2017.216

Lonare, S., & Bhramaramba, R. (2020). Traffic Flow Prediction Using Regression and Deep Learning Approach. In New Trends in Computational Vision and Bio-inspired Computing (pp. 641–648). Springer International Publishing.

Lopez-Garcia, P., Onieva, E., Osaba, E., Masegosa, A. D., & Perallos, A. (2016). A Hybrid Method for Short-Term Traffic Congestion Forecasting Using Genetic

Algorithms and Cross Entropy. IEEE Transactions on Intelligent Transportation Systems, 17(2), 557–569.

Lu, S., & Liu, Y. (2018). Evaluation system for the sustainable development of urban transportation and ecological environment based on SVM. Journal of Intelligent & Fuzzy Systems, 34(2), 831–838.

Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., & Wang, Y. (2017). Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction. Sensors 2017, Vol. 17, Page 818, 17(4), 818.

Ma, X., Yu, H., Wang, Y., & Wang, Y. (2015). Large-Scale Transportation Network Congestion Evolution Prediction Using Deep Learning Theory. PLOS ONE, 10(3), e0119044.

Majumdar, S., Subhani, M. M., Roullier, B., Anjum, A., & Zhu, R. (2021). Congestion prediction for smart sustainable cities using IoT and machine learning approaches. Sustainable Cities and Society, 64(September 2020).

Mihaylova, L., Boel, R., 2004. A particle filter for freeway traffic estimation. In: 2004 43rd IEEE Conference on Decision and Control (CDC)(IEEE Cat. No. 04CH37601), IEEE. pp. 2106–2111.

Mishra, P., Hadfi, R., & Ito, T. (2016). Adaptive Model for Traffic Congestion Prediction. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9799, 782–793.

Nadeem, K. M., & Fowdur, T. P. (2018). PERFORMANCE ANALYSIS OF A REAL-TIME ADAPTIVE PREDICTION ALGORITHM FOR TRAFFIC CONGESTION. Journal of Information and Communication Technology, 17(3), 493–511.

Nanthawichit, C., Nakatsuji, T., Suzuki, H., 2003. Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway. Transp. Res. Rec. 1855, 49–59

Navarro-Espinoza, A., López-Bonilla, O. R., García-Guerrero, E. E., Tlelo-Cuautle, E., López-Mancilla, D., Hernández-Mejía, C., & Inzunza-González, E. (2022). Traffic Flow Prediction for Smart Traffic Lights Using Machine Learning Algorithms. Technologies, 10(1), 5.

Nazemi, A., & Dehghan, M. (2015). A neural network method for solving support vector classification problems. Neurocomputing, 152, 369–376.

Newson, P., & Krumm, J. (2009). Hidden Markov map matching through noise and sparseness. Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS ’09, 336–343.

Onieva, E., Lopez-Garcia, P., Masegosa, A. D., Osaba, E., & Perallos, A. (2016). A Comparative Study on the Performance of Evolutionary Fuzzy and Crisp Rule Based Classification Methods in Congestion Prediction. Transportation Research Procedia, 14, 4458–4467.

Onieva, E., Milanés, V., Villagrá, J., Pérez, J., & Godoy, J. (2012). Genetic optimization of a vehicle fuzzy decision system for intersections. https://doi.org/10.1016/j.eswa.2012.05.087

Pueboobpaphan, R., Nakatsuji, T., Suzuki, H., 2007. Unscented Kalman filter-based real-time traffic state estimation. Technical Report.

Rahmani, M., Jenelius, E., & Koutsopoulos, H. N. (2015). Non-parametric estimation of route travel time distributions from low-frequency floating car data. Transportation Research Part C: Emerging Technologies, 58, 343–362. https://doi.org/10.1016/J.TRC.2015.01.015

Radwan, E., Harb, R., Ramasamy, S., & Putcha, S. (2023). Evaluation Of Safety And Operational Effectiveness Of Dynamic Lane Merge System In Florida.

Ran, B., Song, L., Zhang, J., Cheng, Y., & Tan, H. (2016). Using tensor completion method to achieving better coverage of traffic state estimation from sparse floating car data. PLoS ONE, 11(7), 1–16.

Reisi Gahrooei, M., & Work, D. B. (2015). Inferring traffic signal phases from turning movement counters using hidden Markov models. IEEE Transactions on Intelligent Transportation Systems, 16(1), 91–101.

Reza, S., Ferreira, M. C., Machado, J. J. M., & Tavares, J. M. R. S. (2022). A multi-head attention-based transformer model for traffic flow forecasting with a comparative analysis to recurrent neural networks. Expert Systems with Applications, 202, 117275. https://doi.org/10.1016/J.ESWA.2022.117275

Riyadh, M., Mustapha, N., Sulaiman, N., & Sharef, N. B. M. (2017). ONF-TRS: On-line noise filtering algorithm for trajectory segmentation based on MDL threshold. Journal of Artificial Intelligence, 10(1), 42–48.

Seo, T., Bayen, A. M., Kusakabe, T., & Asakura, Y. (2017). Traffic state estimation on highway: A comprehensive survey. Annual Reviews in Control, 43, 128–151. https://doi.org/10.1016/j.arcontrol.2017.03.005

Shen, Q., Ban, X., & Guo, C. (2017). Urban Traffic Congestion Evaluation Based on Kernel the Semi-Supervised Extreme Learning Machine. Symmetry 2017, Vol. 9, Page 70, 9(5), 70.

Shen, Q., Ban, X., Guo, C., & Wang, C. (2016). Kernel Based Semi-supervised Extreme Learning Machine and the Application in Traffic Congestion Evaluation. 227–236. https://doi.org/10.1007/978-3-319-28397-5_18

Shi, Q., & Abdel-Aty, M. (2015). Big Data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. Transportation Research Part C: Emerging Technologies, 58, 380–394. https://doi.org/10.1016/j.trc.2015.02.022

Shi, Z. (2011). *Advanced Artificial Intelligence*. WORLD SCIENTIFIC. https://doi.org/10.1142/7547

Simoni, M. D., & Claudel, C. G. (2020). A fast Lax–Hopf algorithm to solve the Lighthill–Whitham–Richards traffic flow model on networks. *Transportation Science*, *54*(6), 1516-1534.

Sun, Shiliang, Zhang, C., & Yu, G. (2006). A Bayesian network approach to traffic flow forecasting. IEEE Transactions on Intelligent Transportation Systems, 7(1), 124–133.

Sun, Shuming, Chen, J., & Sun, J. (2019). Traffic congestion prediction based on GPS trajectory data. International Journal of Distributed Sensor Networks, 15(5). https://doi.org/10.1177/1550147719847440

Sutarto, H.Y., Boel, R.K. and Joelianto, E., 2015. Parameter estimation for stochastic hybrid model applied to urban traffic flow estimation. IET Control Theory & Applications, 9(11), pp.1683-1691.

Takenouchi, A., Kawai, K., & Kuwahara, M. (2019). Traffic state estimation and its sensitivity utilizing measurements from the opposite lane. In Transportation Research Part C: Emerging Technologies (Vol. 104, pp. 95–109).

Tseng, F. H., Hsueh, J. H., Tseng, C. W., Yang, Y. T., Chao, H. C., & Chou, L. Der. (2018). Congestion prediction with big data for real-time highway traffic. IEEE Access, 6, 57311–57323. https://doi.org/10.1109/ACCESS.2018.2873569

Wang, J., Mao, Y., Li, J., Xiong, Z., & Wang, W.-X. (2015). Predictability of Road Traffic and Congestion in Urban Areas. PLOS ONE, 10(4), e0121825. https://doi.org/10.1371/JOURNAL.PONE.0121825

Wang, R., Work, D. B., & Sowers, R. (2016). Multiple model particle filter for traffic estimation and incident detection. IEEE Transactions on Intelligent Transportation Systems, 17(12), 3461–3470.

Wang, S., Li, F., Stenneth, L., & Yu, P. S. (2016). Enhancing Traffic Congestion Estimation with Social Media by Coupled Hidden Markov Model. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9852 LNAI, 247–264.

Wang, W. X., Guo, R. J., & Yu, J. (2018). Research on road traffic congestion index based on comprehensive parameters: Taking Dalian city as an example. Advances in Mechanical Engineering, 10(6), 1–8.

Wang, Xiang, An, K., Tang, L., & Chen, X. (2015). Short Term Prediction of Freeway Exiting Volume Based on SVM and KNN. International Journal of Transportation Science and Technology, 4(3), 337–352.

Wang, Xiaomeng, Peng, L., Chi, T., Li, M., Yao, X., & Shao, J. (2015). A hidden markov model for urban-scale traffic estimation using floating car data. PLoS ONE, 10(12), 1–20. https://doi.org/10.1371/journal.pone.0145348

Wang, Y., Papageorgiou, M., Messmer, A., 2007. Real-time freeway traffic state estimation based on extended kalman filter: a case study. Transport. Sci. 41, 167–181.

Wang, Y., Chen, Y., Li, G., Lu, Y., He, Z., Yu, Z., & Sun, W. (2023). City-scale holographic traffic flow data based on vehicular trajectory resampling. Scientific Data, 10(1). https://doi.org/10.1038/s41597-022-01850-0

Wu, Y., Tan, H., Li, Y., Zhang, J., Chen, X., 2018. A fused cp factorization method for incomplete tensors. IEEE Trans. Neural Netw. Learn. Syst. 30, 751–764.

Xing, Y., Ban, X., Liu, X., & Shen, Q. (2019). Large-Scale Traffic Congestion Prediction Based on the Symmetric Extreme Learning Machine Cluster Fast Learning Method. Symmetry 2019, Vol. 11, Page 730, 11(6), 730. https://doi.org/10.3390/SYM11060730

Xu, D., Dai, H., Wang, Y., Peng, P., Xuan, Q., Guo, H., 2019. Road traffic state prediction based on a graph embedding recurrent neural network under the scats. Chaos: Interdisciplinary J. Nonlinear Sci. 29.

Xu, D., Wei, C., Peng, P., Xuan, Q., & Guo, H. (2020). GE-GAN: A novel deep learning framework for road traffic state estimation. Transportation Research Part C: Emerging Technologies, 117, 102635.

Xu, D.W., Dong, H.H., Jia, L.M., Tian, Y., 2014. Road traffic states estimation algorithm based on matching of regional traffic attracters. J. Central South Univ. 21,2100–2107

Yang, Q., Wang, J., Song, X., Kong, X., Xu, Z., & Zhang, B. (2015). Urban traffic congestion prediction using floating car trajectory data. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9529, 18–30.

Yang, S. (2013). On feature selection for traffic congestion prediction. Transportation Research Part C: Emerging Technologies, 26, 160–169.https://doi.org/10.1016/J.TRC.2012.08.005

Yang, S. (2013). On feature selection for traffic congestion prediction. Transportation Research Part C: Emerging Technologies, 26, 160–169.https://doi.org/10.1016/J.TRC.2012.08.005

Yang, X., Luo, S., Gao, K., Qiao, T., & Chen, X. (2019). Application of Data Science Technologies in Intelligent Prediction of Traffic Congestion. Journal of Advanced Transportation, 2019.

Yu, J., Stettler, M. E. J., Angeloudis, P., Hu, S., & Chen, X. (Michael). (2020). Urban network-wide traffic speed estimation with massive ride-sourcing GPS traces. Transportation Research Part C: Emerging Technologies, 112(January), 136–152.https://doi.org/10.1016/j.trc.2020.01.023

Zaki, John F., Ali-Eldin, A., Hussein, S. E., Saraya, S. F., & Areed, F. F. (2020). Traffic congestion prediction based on Hidden Markov Models and contrast measure. Ain Shams Engineering Journal, 11(3), 535–551.

Zaki, John Fw, Ali-Eldin, A. M. T., Hussein, S. E., Saraya, S. F., & Areed, F. F. (2019). Time Aware Hybrid Hidden Markov Models for Traffic Congestion Prediction. International Journal on Electrical Engineering and Informatics, 11(1).

Zammit, L. C., Fabri, S. G., & Scerri, K. (2019). Real-Time Parametric Modeling and Estimation of Urban Traffic Junctions. IEEE Transactions on Intelligent Transportation Systems, 20(12), 4579–4589.

Zhan, X., Li, R., & Ukkusuri, S. V. (2020). Link-based traffic state estimation and prediction for arterial networks using license-plate recognition data. Transportation Research Part C: Emerging Technologies, 117, 102660.

Zhang, P., & Qian, Z. (Sean). (2018). User-centric interdependent urban systems: Using time-of-day electricity usage data to predict morning roadway congestion. Transportation Research Part C: Emerging Technologies, 92, 392–411.https://doi.org/10.1016/J.TRC.2018.05.008

Zhang, S., Yao, Y., Hu, J., Zhao, Y., Li, S., & Hu, J. (2019). Deep Autoencoder Neural Networks for Short-Term Traffic Congestion Prediction of Transportation Networks. Sensors 2019, Vol. 19, Page 2229, 19(10), 2229.

Zhang, T., Liu, Y., Cui, Z., Leng, J., Xie, W., & Zhang, L. (2019). Short-Term Traffic Congestion Forecasting Using Attention-Based Long Short-Term Memory Recurrent Neural Network. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11538 LNCS, 304–314. https://doi.org/10.1007/978-3-030-22744-9_24

Zhang, W., Yu, Y., Qi, Y., Shu, F., & Wang, Y. (2019). Short-term traffic flow prediction based on spatio-temporal analysis and CNN deep learning. *Transportmetrica A: Transport Science*, *15*(2), 1688-1711.

Zhang, X., Onieva, E., Perallos, A., Osaba, E., & Lee, V. C. S. (2014). Hierarchical fuzzy rule-based system optimized with genetic algorithms for short term traffic congestion prediction. Transportation Research Part C: Emerging Technologies, 43, 127–142. https://doi.org/10.1016/J.TRC.2014.02.013

Zhang, Y., & Ye, Z. (2008). Short-Term Traffic Flow Forecasting Using Fuzzy Logic System Methods. Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, 12(3), 102–112.

Zhao, J. (2015). Research on Prediction of Traffic Congestion State. MATEC Web of Conferences, 22, 01059.

Zheng, Y., Li, Y., Own, C. M., Meng, Z., & Gao, M. (2018). Real-time predication and navigation on traffic congestion model with equilibrium Markov chain. *International journal of distributed sensor networks*, *14*(4), 1550147718769784.

Zhang, Z., & Yang, X. (2020). Freeway Traffic Speed Estimation by Regression Machine-Learning Techniques Using Probe Vehicle and Sensor Detector Data. Journal of Transportation Engineering, Part A: Systems, 146(12).

Zhu, Li, Yu, F. R., Wang, Y., Ning, B., & Tang, T. (2019). Big Data Analytics in Intelligent Transportation Systems: A Survey. In IEEE Transactions on Intelligent Transportation Systems (Vol. 20, Issue 1, pp. 383–398). Institute of Electrical and Electronics Engineers Inc.

Zhu, Lin, Krishnan, R., Guo, F., Polak, J. W., & Sivakumar, A. (2019). Early Identification of Recurrent Congestion in Heterogeneous Urban Traffic. 2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019, 4392–4397.https://doi.org/10.1109/ITSC.2019.8916966

LIST OF PUBLICATIONS

Ahanin, F., Mustapha, N., & Zolkepli, M. (2020). An Efficient Traffic State Estimation Model Based on FUZZY C-MEAN Clustering and MDL Using FCD. Journal of Theoretical and Applied Information Technology, 31(14).

Ahanin, F., Mustapha, N., Zolkepli, M., & Husin, N. A. (2023). A Review of Traffic State Prediction (TSP) Methods in Intelligent Transportation Systems (ITS). International Journal of Academic Research in Business and Social Sciences, 13(3).