

A DATA-DRIVEN DECISION SUPPORT SYSTEM FOR MOBILE TELEMATICS

by **Mohammad SiamiNamini**

Thesis submitted in fulfilment of the requirements for
the degree of

Doctor of Philosophy

under the supervision of Dist. Prof. Jie Lu, and Dr. Mohsen
Naderpour

University of Technology Sydney
Faculty of Faculty of Engineering and Information Technology

October 2020

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Mohammad SiamiNamini declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Computer Science, Faculty of Engineering and Information Technology 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.

Signature: Production Note:
Signature removed prior to publication.

Date: 4/11/20

DEDICATION

*To my darling wife for her passion and patience
and to my beloved parents for their
encouragement which enabled my dreams to
come true.*

And

*To my lovely little son Ariyan, who made me
stronger, happier and more fulfilled than I
could have ever imagined. I love you to the
moon and back.*

ACKNOWLEDGEMENT

I would like to express my appreciation to Distinguished Professor Jie Lu, who supervised this Ph.D. research, for her knowledgeable comments, valuable support and guidance during hard times, and constructive suggestions along the way. I want to thank Jie for her wonderful scientific and spiritual support during this journey and her willingness to allow my research to follow my interests. I am also grateful to Dr Mohsen Naderpour, my co-supervisor, for his knowledgeable suggestions and valuable technical advice throughout my PhD study.

Looking back over my PhD journey, I see my wife, Anahita, truly standing shoulder to shoulder with me. Thank you very much for being so supportive in all circumstances throughout the four years of this Ph.D., for understanding the stress I was subjected to, for sacrificing your time for me while you were also completing a Ph.D. and for caring for our little son, Ariyan, to give me the freedom to follow my scientific interests. I could not have accomplished this without your constant love and support.

I would like to thank my parents, the first ones who taught me, for their encouragement and support despite the geographical distance. Pursuing a PhD would not have been possible without their love and assistance. I express my gratitude to all my friends and colleagues in the Decision Systems & e-Service Intelligence (DeSI) Laboratory for their help and valuable comments during my study.

I would also like to express my appreciation to the Faculty of Engineering and Information Technology and the Centre for Artificial Intelligence (CAI) at the University of Technology Sydney (UTS) for their provision of conference registration and travel funds during this research. This research was also supported by

an International Research Scholarship (IRS) and the UTS President's Scholarship (UTSP) funded by UTS.

Last, but not least, a special thank you goes to Ms Sue Felix, Ms Jemima Moore, and Ms Michele Mooney for helping me to identify and correct grammar and syntax problems in my publications.

ABSTRACT

Mobile telematics is an emerging technology that collects data on human behaviour using smartphones. All smartphones have internal sensors with the capability to record and transmit data to an external server. This emerging technology is easy to use, the initial cost is very low, and generates a massive amount of data which are noisy, complex, and uncertain. This opens many opportunities for data-driven decision making such as driving behaviour risk analysis, usage-based insurance, remote sensing, and fleet management. Traditional decision-making techniques are not able to work with this type of unstructured and complex data. Thus, new techniques are needed based on advanced analytics to analyze mobile telematics streams.

This research develops a big data-driven decision support system (DSS) for mobile telematics. The research relies on the capabilities of advanced analytics techniques, machine learning, and fuzzy logic. The research presents an innovative analytical system for mobile telematics which consists of four major components: 1) a data preparation component that prepares a trajectory dataset to a new and ready-for-analysis format; 2) a driving style pattern recognition that extracts hidden human patterns in mobile telematics using unsupervised learning and unlabelled data; 3) a fuzzy risk assessment is proposed to assess risk of drivers by fuzzy logic using extracted patterns by unsupervised learning; and 4) a missing data imputation component which is a novel Choquet Fuzzy Integral Vertical Bagging (CFIVB) algorithm to classify large labelled mobile telematics stream datasets.

The proposed models were evaluated on two real-world mobile telematics datasets, namely an unlabelled dataset collected by a usage-based insurance company

containing 500,000 journeys of 2500 drivers, and an anonymized driving behaviour dataset consisting of streaming data of 408 trips of 310 unique drivers. Various validation measures were used to evaluate the performance of the proposed models. The area under a curve (AUC) and accuracy are used to evaluate the classification algorithms and the Davis–Boulding index, the Calinski–Harabasz index, execution time, and mean square error are utilized to evaluate clustering algorithms and find the optimal number of clusters. The sensitivity analysis results show the proposed model is consistent across different variations of the model.

The proposed DSS can be applied on all stream data risk assessments. Moreover, 29 unique driving styles were extracted from mobile telematics data and these patterns can be applied as labels for supervised learning modelling. In addition, performance measures depict the CFIVB algorithm performs well in this domain, and it can be applied for similar problems.

TABLE OF CONTENTS

Abstract.....	V
List of Figures	XI
List of Tables	XII
List of Algorithms	XIII
Chapter 1:.....	1
Introduction	1
1.1 Background.....	1
1.2 Research Problems	3
1.3 Research Objectives.....	6
1.4 Research Contributions.....	8
1.5 Research Methodology	9
1.5.1 General Research Methodology	9
1.5.2 Thesis Research Process.....	11
1.6 Thesis Structure	13
1.7 Publications of This Research	14
Chapter 2:.....	16
Literature Review	16
2.1 Introduction.....	16
2.2 Mobile Telematics	16
2.3 Mobile Telematics And Usage-Based Insurance.....	18
2.4 Fuzzy Decision Support System.....	22
2.5 Driving Style Analytics.....	24
2.6 Change Detection Algorithm.....	27
2.7 Deep Auto-Encoder	28
2.8 Fuzzy Clustering.....	29
2.9 Fuzzy Sets And Fuzzy Logic Systems.....	30
2.10 Partitive Clustering.....	32
2.11 Self-Organizing Map.....	32
2.12 Dangerous Driving Behaviour	34

2.13	Summary	37
Chapter 3:	38
Big Data-Driven Decision Support System Framework		
38		
3.1	Introduction.....	38
3.2	Decision Support System Framework.....	40
3.2.1	Data Preparation	42
3.2.2	Driving Style Pattern Recognition.....	44
3.2.3	Fuzzy Risk Assessment.....	44
3.2.4	Missing Data Imputation	45
3.2.5	Information Pipeline.....	46
3.3	Summary	48
Chapter 4:	49
Data Preparation		
49		
4.1	Introduction.....	49
4.2	Mobile Telematics Data	50
4.3	Data Transformation.....	51
4.4	Data Pre-Processing.....	52
4.4.1	Change Detection.....	53
4.4.2	Feature Extraction	55
4.4.3	Feature Selection	58
4.5	Implementation.....	58
4.6	Summary	59
Chapter 5:	60
Driving Style Pattern Recognition		
60		
5.1	Introduction.....	60
5.2	Two-Stage Clustering.....	62
5.2.1	SOM.....	62
5.2.2	Deep Auto-encoder	63
5.2.3	Clustering	64
5.3	Implementation.....	65
5.3.1	Two-Stage Clustering	67

5.3.2	Performance Evaluation	68
5.3.3	Experimental Results.....	70
5.4	Extracted Driving Patterns.....	72
5.5	Summary	79
Chapter 6:	81
Fuzzy Risk Assessment.....		81
6.1	Introduction.....	81
6.2	Risk Factor Identification	83
6.2.1	Risk Factor Mining	83
6.3	Fuzzy Risk Modelling.....	85
6.4	Risk Assessment.....	86
6.4.1	Event Detection	86
6.4.2	Stream Risk Calculation	86
6.5	Implementation.....	87
6.5.1	Risk Factor Mining	87
6.5.2	Fuzzy Risk Modelling	89
6.5.3	Stream Risk Calculation	93
6.3	Evaluation.....	95
6.3.1	Sensitivity Analysis.....	96
6.4	Summary	98
Chapter 7:	100
Missing Data Imputation		100
7.1	Introduction.....	100
7.2	Data Preparation	101
7.3	Choquet Fuzzy Integral Vertical Bagging Classifier	102
7.3.1	Fuzzy Integral	104
7.3.2	Preliminary Fuzzy Densities:	105
7.3.3	Adaptive Fuzzy Measures:.....	106
7.3.4	Choquet Fuzzy Integral.....	108
7.4	Experiment Results.....	110
7.5	Summary	114

Chapter 8:	116
Conclusion and Future Work	116
8.1 Conclusions	116
8.2 Future Works	119
Bibliography	122
APPENDIX: ABBREVIATIONS	131

LIST OF FIGURES

Figure 1-1 - General Research Methodology	10
Figure 1-2- Thesis Structure.....	13
Figure 2-1- A Deep Auto-Encoder With Many Layers	29
Figure 2-2 -Visualizing μ_{\max} Calculation	36
Figure 2-3- Safe Driving Area	37
Figure 3-1- Smartphone-Based Vehicle Telematics (Wahlström, Skog & Händel 2017).....	40
Figure 3-2- Mobile Telematics Big data Decision Support System	41
Figure 3-3- Information Pipeline.....	46
Figure 3-4- Information Pipeline Cross Table.....	47
Figure 4-1- Data Preparation Component	49
Figure 4-2- A Sample Driver’s Trips	50
Figure 4-3- Two-Dimensional Vehicle Coordinate System (ISO 8855) ((ISO) 2011–12).....	51
Figure 4-4- Change Detection Sample	54
Figure 5-1- The Driving Style Pattern Detection Framework	61
Figure 5-2- Change Detection Scores	66
Figure 5-3- Two-Stage Clustering.....	67
Figure 5-4- Sum Of Square Error Per Number Of Clusters For SOM + K-Means Clustering.....	73
Figure 6-1- Fuzzy Risk Assessment Component	82
Figure 6-2- Fuzzy Membership Functions Of Probability, Severity, And Risk	91
Figure 6-3- Sensitivity Analysis Optimistic Strategy Risk	97
Figure 6-4- Sensitivity Analysis Pessimistic Strategy Risk	97
Figure 6-5- Sensitivity Analysis Neutral Strategy Risk	97
Figure 7-1- Missing Imputation Component (Choquet Fuzzy Integral Vertical Bagging).....	103
Figure 7-2 - 5-Fold Cross-Validation.....	112
Figure 7-3- Choquet Fuzzy Integral Vertical Bagging	114

LIST OF TABLES

Table 2-1- Usage-Based Insurance and Mobile telematics	19
Table 2-2- Attributes for Risk Assessment with Smartphone Data.....	22
Table 5-1- Selected Dataset.....	65
Table 5-2- Davis-Boulding Index Results.....	70
Table 5-3- Calinski-Harabasz Index Results.....	71
Table 5-4- Execution Time Results (Minutes).....	72
Table 5-5- Frequent Driving Behaviors - Part I.....	76
Table 5-6- Frequent Driving Behaviors - Part II.....	77
Table 5-7- Frequent Driving Behaviors -Part III.....	78
Table 5-8- Frequent Driving Behaviors - Part IV.....	79
Table 6-1- Fuzzy Clustering Result.....	88
Table 6-2- Probability Linguistic Variables.....	91
Table 6-3- Severity Linguistic Variables.....	92
Table 6-4- Risk Linguistic Variables.....	92
Table 6-5- Risk Matrix.....	92
Table 6-6- Mamdani Model (Mamdani 1977).....	93
Table 6-7- Calculated Risk Scores.....	94
Table 6-8- Trip Risk Calculation Process.....	95
Table 7-1- Stream Data Introduction.....	101
Table 7-2- Data Description.....	110
Table 7-3- Correlation Analysis Results.....	111
Table 7-4- Results.....	113

LIST OF ALGORITHMS

Algorithm 2-1- Partitive Clustering Algorithm (Xiao & Yu 2012).....	32
Algorithm 5-1- Performance Validation Algorithm.....	69
Algorithm 5-2- Finding optimal number of clusters algorithm	70
Algorithm 5-3- The Matching Method.....	75
Algorithm 6-1- Fuzzy Clustering Algorithm (Shen Et Al. 2019).....	84
Algorithm 7-1- Choquet Fuzzy Integral Vertical Bagging (CFIVB) Classifier	109