C02029: Doctor of Philosophy
32903 PhD Thesis: Analytics
December 2021

Customer Behavior Analytics and Visualization

Md Rafiqul Islam

School of Computer Science
Faculty of Engineering and Information Technology
University of Technology Sydney
NSW - 2007, Australia

Customer Behavior Analytics and Visualization

by Md Rafiqul Islam

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Analytics

Under the supervision of Professor Guandong Xu and Dr. Xianzhi Wang

School of Computer Science
Faculty of Engineering and Information Technology
University of Technology Sydney
NSW - 2007, Australia

December 2021

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, **Md Rafiqul Islam** declare that this thesis, is submitted in fulfilment of the requirements for the award of **Doctor of Philosophy in Analytics**, in the **School of Computer Science**, **Faculty of Engineering and Information Technology at the University of Technology Sydney, Australia**.

This thesis is wholly my own work unless otherwise referenced 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.

[Md Rafiqul Islam]

DATE: 14th December, 2021

PLACE: Sydney, Australia

Dedication

To my beloved parents and teachers

ACKNOWLEDGMENTS

First of all, I owe my profound gratitude to my supervisor, Prof. Guandong Xu for providing me with all the resources, scholarly advice, and magnanimous technical, and financial support throughout my Ph.D. study. As a mentor, Prof. Guandong Xu showed me the quality of being devoted to my work through his contagious passion for research by being a role model.

I would also like to convey my appreciation and gratitude to my co-supervisor Dr. Xianzhi Wang for his insightful comments and encouragement. Besides my supervisors, my sincere appreciation goes to external collaborators Dr. Imran Razzak and Dr. Shaowu Liu who provided me essential support during my research work. Without their precious support, it would be difficult to achieve this goal.

I sincerely acknowledge my dear teacher Prof. Dr. Abu Raihan M. Kamal, Dr. Ashad Kabir, Dr. Md. Samiullah, Dr. Anwar Ulhaq, and friends Dr. Munshi Muhammad Abdul Kader Jilani, Dr. Mohammad Muntaseer Mahfuz, Dr. Md Shamsur Rahim, Dr. Shakil Ahmed Khan, Dr. Pejush Chandra Sarker, Shanjita Akter, Md. Ashraful Haque, and lab mates who have been extremely supportive during the entire course of my Ph.D. study.

I would like to acknowledge full financial support through the International Research Scholarship (IRS) and Faculty of Engineering and Information Technology (FEIT) Scholarship throughout my Ph.D. study.

Last but not the least, my heartfelt appreciation goes to all my family members, especially my father Md Abdul Kader, mother Most Tahmina Begum, and my uncle Md Fazlul Haque who have been extremely supportive during the entire course of my Ph.D. I would also extend my gratitude to my wife Aireen Rahman. Without her cooperation and support, it would not have been possible to accomplish this goal.

List of Publications

List of Journal Articles (Published/Accepted)

1. **Islam, Md Rafiqul**, Imran Razzak, Xianzhi Wang, Peter Tilocca, and Guandong Xu. "Natural language interactions enhanced by data visualization to explore insurance claims and manage risk." Annals of Operations Research (2022): 1-19. (**Refer to Chapter 1 & 5**)

2. **Islam, Md Rafiqul**, Shanjita Akter, Md Rakybuzzaman Ratan, Linta Islam, Imran Razzak, and Guandong Xu. “Strategies for evaluating visual interactive system: a systematic review and new perspectives." Journal of Visualization, (2022). (**Refer to Chapter 1 & 6**)

3. **Islam, Md Rafiqul**, Shaowu Liu, Rhys Biddle, Imran Razzak, Xianzhi Wang, Peter Tilocca, and Guandong Xu. “Discovering dynamic adverse behavior of policyholders in the life insurance industry." Technological Forecasting and Social Change 163 (2021): 120486. (**Refer to Chapter 1 & 3**)

4. **Islam, Md Rafiqul**, Shanjita Akter, Md Rakybuzzaman Ratan, Abu Raihan M. Kamal, and Guandong Xu. “Deep visual analytics (dva): applications, challenges and future directions." Hum-Centric Intell Syst 1, no. 1-2 (2021): 3-17. (**Refer to Chapter 1 & 6**)

5. Mosiur Rahman, **Md Rafiqul Islam**, Sharmin Akter, Shanjita Akter, Linta Islam, and Guandong Xu. “DiaVis: exploration and analysis of diabetes through visual interactive system." Hum-Centric Intell Syst (2021). (**Refer to Chapter 4**)

6. Sharif, Omar, **Md Rafiqul Islam**, Md Zobaer Hasan, Muhammad Ashad Kabir, Md Emran Hasan, Salman A. AlQahtani, and Guandong Xu. “Analyzing the impact of demographic variables on spreading and forecasting COVID-19." Journal of Healthcare Informatics Research (2021): 1-19. (**Refer to Chapter 2**)

7. **Islam, Md Rafiqul**, Shaowu Liu, Xianzhi Wang, and Guandong Xu. “Deep learning for misinformation detection on online social networks: a survey and new perspectives." Social Network Analysis and Mining 10, no. 1 (2020): 1-20. (**Refer to Chapter 2**)

LIST OF CONFERENCE ARTICLES (PUBLISHED)

1. **Islam, Md Rafiqul**, Imran Razzak, Xianzhi Wang, Peter Tilocca, and Guandong Xu. “UCBVis: understanding customer behavior sequences with visual interactive system." In 2021 International Joint Conference on Neural Networks (IJCNN), pp. 1-8. IEEE, 2021. (**Refer to Chapter 1 & 6**)

2. **Islam, Md Rafiqul**, Jiaming Zhang, Md. Hamjajul Ashmafee, Imran Razzak, Jianlong Zhou, Xianzhi Wang, and Guandong Xu. “ExVis: explainable visual decision support system for risk management." In 2021 8th International Conference on Behavioural and Social Computing (BESC), IEEE, 2021. (**Refer to Chapter 4**)

3. Zerafa, Joshua, **Md Rafiqul Islam**, Ashad Kabir, and Guandong Xu. “ExTraVis: exploration of traffic incidents using visual interactive system." In 25th International Conference Information Visualisation (IV 2021). IEEE, Institute of Electrical and Electronics Engineers, 2021. (**Refer to Chapter 2**)

4. **Islam, Md Rafiqul**, Shaowu Liu, Imran Razzak, Muhammad Ashad Kabir, Xianzhi Wang, Peter Tilocca, and Guandong Xu. “MHIVis: visual analytics for exploring mental illness of policyholders in life insurance industry." In 2020 7th International Conference on Behavioural and Social Computing (BESC), pp. 1-4. IEEE, 2020. (**Refer to Chapter 4**)

Table of Contents

Certificate of Original Authorship ii

Acknowledgments vii

List of Publications ix

Table of Contents xii

List of Figures xiii

List of Tables xv

Abstract xix

1 Introduction 1

1.1 Background and Motivation 1

1.1.1 Analytics for Decision Making 1

1.1.2 Customer Behavior Discovery 2

1.1.3 Customer Behavior Visualization 3

1.1.4 Summary 3

1.2 Research Objectives 4

1.3 Research Problems 5

1.4 Research Contributions 7

1.5 Thesis Organization 8

2 Related Work 11

2.1 Literature Search Methodology 11

2.2 Understanding Customer Behavior 12

2.2.1 Categorization of Customer Behavior 13

2.2.2 Current Customer Behavior Issues 16

2.2.3 Summary 18
2.3 Visualizing Customer Behavior 19
2.4 Customer Behavior Analysis using Data Mining Techniques 19
2.5 Summary 21

3 Pattern Mining for Discovering Customer Adverse Behavior 23
3.1 Background and Motivation 23
3.2 Preliminary on Adverse Behavior Selection 26
3.3 Methodology 28
3.3.1 Data Collection and Processing 28
3.3.2 Proposed Model 29
3.4 Experiment 33
3.4.1 Data Description 33
3.4.2 Interesting Pattern List Construction 34
3.5 Result and Discussion 35
3.5.1 Result Analysis 35
3.5.2 Discussion and Implications 41
3.6 Summary 43

4 Visual Analytics for Exploring Customer Behavior 45
4.1 Background and Motivation 45
4.2 Preliminary on Exploring Customer Behavior 50
4.2.1 Explainable Visualization System for Risk Management 50
4.2.2 Exploring Mental Illness of Customer with Visual Interactive System 52
4.2.3 Exploring Diabetes through Visual Interactive System 53
4.2.4 Summary 55
4.3 Methodology 56
4.3.1 Methodology of *ExVis* 56
4.3.2 Methodology of *MHIVis* 60
4.3.3 Methodology of *DiaVis* 63
4.4 Experimental Analysis 67
4.4.1 Experimental Analysis of *ExVis* 67
4.4.2 Experimental Analysis of *MHIVis* 67
4.4.3 Experimental Analysis of *DiaVis* 69
4.5 Result and Discussion 72
4.5.1 Result Analysis 73

4.5.2 User Study . . . 74
4.5.3 Discussion . . . 74
4.6 Summary . . . 79

5 NLI-driven-DV to Explore Customer Claim Behavior and Manage Risk 81
5.1 Background and Motivation . . . 81
5.2 Preliminary on Exploring Customer Claim Behavior and Manage Risk . . 84
5.2.1 Insurance Claims and Risk Management . . . 85
5.2.2 Visualization for Claim and Risk Management . . . 85
5.2.3 Natural Language Interfaces with Data Visualization . . . 86
5.3 Methodology . . . 87
5.3.1 Data Description . . . 87
5.3.2 Data Pre-processing . . . 88
5.3.3 Domain Characterization and Design Consideration . . . 88
5.3.4 Visual Analytics Solution . . . 90
5.4 Experimental Analysis . . . 91
5.4.1 Data Interpretation . . . 92
5.4.2 Query Analyzer . . . 93
5.4.3 Visualization Generation . . . 93
5.4.4 Implementation . . . 94
5.5 Result and Discussion . . . 95
5.5.1 Result Analysis . . . 95
5.5.2 User Study . . . 96
5.5.3 Discussion . . . 99
5.6 Summary . . . 101

6 Deep Visual Analytics for Understanding Customer Behavior 103
6.1 Background and Motivation . . . 103
6.2 Preliminary on Deep Visual Analytics for Understanding Customer Behavior 107
6.2.1 Exploring Customer Behavior with Visual Interactive System . . . 107
6.2.2 Deep Visual Analytics . . . 110
6.2.3 Strategies for Evaluating Visualization System . . . 113
6.3 Methodology . . . 114
6.3.1 Methodology of *UCBVis* . . . 114
6.3.2 Methodology of *Multi-DLMPVis* . . . 118
6.4 Experimental Analysis . . . 120

6.4.1 Experimental Analysis of *UCBVis* 120
6.4.2 Experimental Analysis of *Multi-DLMPVis* 121
6.5 Discussion 123
6.5.1 Evaluation 123
6.5.2 Challenges 123
6.5.3 Future Directions 126
6.6 Summary 127

7 Conclusion and Future Work **129**
7.1 Contributions 129
7.2 Future work 131

Bibliography **133**

List of Figures

Figure Name **Page**

2.1 Proposed review methodology for sample collection and analysis. 12

3.1 The structure of the *ARLAS* framework. 29
3.2 Identifying different rules with different support threshold. 36
3.3 Number of adverse policyholders per state. 37
3.4 Number of policyholders per state. 37
3.5 Number of adverse policyholder by occupation and gender. 39
3.6 Number of policyholders by age group. 40
3.7 Number of adverse policyholders by age group. 40

4.1 *ExVis*: exploring multiple views to explain visual decision support system. . 46
4.2 *MHIVis*: a visual interactive system for exploring mental health illness. . . . 47
4.3 *DiaVis*: a visual interactive system for exploring diabetes disease. 48
4.4 The framework of *ExVis* system. 57
4.5 The structure of bayesian network model. 58
4.6 The framework of *MHIVis* system. 61
4.7 The methodological framework of *DiaVis*. 64
4.8 The conceptual illustration of the study. 66
4.9 The mapping view of respondents. 71
4.10 The regional view of respondents. 71
4.11 The correlation between SBP and DBP. 72
4.12 The correlation between Height and Weight. 72
4.13 Similarity matrix of the attributes of *DiaVis*. 78

5.1 Summary of *InsCRMVis* interface components (Insurance claim and risk management with visual analytics). (a) query searching, (b) speech : allow users to freely express query to get visual insights, and (c) mouse/Touch/Pen: can be supported with visualization. 83
5.2 The data sources of the policyholders claims. 87
5.3 Proposed architecture of NLI based visualization. 91
5.4 System overview: key questions of the risk visualization framework. 92
5.5 Framework: overview of the interface functionalities such as input data, query analyzer, and visualization generation. 92
5.6 An illustration of a query analyzer while interpreting NL queries. 93
5.7 Sample questions with answers generated by the system. The answer in Q1(a), and Q1(b) is for query searching, Q2 is for speech and Q3(a) and Q3(b) is for touch/pen/mouse. 96
5.8 Post-study ranking for output visualization and opinion position. 99

6.1 Summary of *UCBVis* interface components. (a) behavior exploration workflow: allow user to identify potential behavior of customer, (b) frequent pattern view: allow user to choose sequential patterns for the focus, (c) attribute pair view: allow user to search attribute names with behavioral patterns, and (d) raw sequences view: can be supported with visualization. 105
6.2 *Multi-DLMPVis*: An interactive dashboard for multiple deep learning models performance visualizations (Figure courtesy [8]). 106
6.3 An interactive deep visual analytics system which is consisted of four major parts. A) distribution view, B) demographic chart, C) patient history, and D) knowledge graph view (Figure courtesy [166]). 110
6.4 The architecture of *UCBVis* system. 117
6.5 System architecture of *Multi-DLMPVis* System. 119
6.6 The input selection and performance ranking view of the *Multi-DLMPVis* System. 122
6.7 3D projection, confusion matrix, and misclassified instances view of the *Multi-DLMPVis* System. 123

List of Tables

	Table Name	Page
2.1	List of the key journals and conferences.	13
3.1	Key studies: different methods for adverse-selection detection in the insurance market.	27
3.2	Characteristics of demographic dataset.	34
3.3	Different interestingness measure of the different patterns.	35
3.4	The result of the experiment.	38
3.5	Advantages and limitations ML/DL techniques.	42
4.1	Characteristics of respondents	57
4.2	Conditional probability table (CPT) for bayesian network (BN).	58
4.3	Dataset description.	62
4.4	Diabetes dataset description.	63
4.5	Characteristics of respondents.	65
4.6	Feature comparison.	68
4.7	User study results.	75
5.1	Key questions identified in collaboration with domain experts.	89
5.2	Types of queries and visualization observed in this study.	94
5.3	User study response for output visualization to post-study questionnaires.	97
5.4	Key observations identified in collaboration with domain experts where expert responses 1, 2, 3, 4, and 5 indicate strongly disagree, disagree, neutral, agree, and strongly agree, respectively.	98
6.1	Overview of key representative works in visual and deep visual analytics.	111
6.2	Overview of various evaluation techniques and their applications.	115
6.3	Demographic characteristics of claim data set.	116
6.4	Domain specific requirements.	117

6.5 User study results. 124
6.6 Feature comparison analysis. 125

Abstract

Customer behavior refers to the study of customers and the procedure they use to pick, use, and dispose of products or services. The understanding of customer behavior analysis (CBA) is essential for improving business strategies. The existing studies have explored useful information to analyze customers' behaviors. However, they often fail to allow the analysts, including business management, development, decision-making, etc. Notably, the existing research on CBA is limited with four main challenges. First, the analysis of the absence of useful private information and the presence of asymmetric information of customers, e.g., discover adverse information in each cell rather than for each data instance. Second, exploring customer behavior with multi-dimensional and temporal data is necessary for any competitive and global business to improve its strategies. Third, the estimation of the correlation between claim analysis and risk management is key to avoiding fraud; Fourth, the lack of quantitative research necessitates performance analysis at the class, instance levels, and model visualization. Several approaches to addressing these issues were introduced that are inconsistent with models of rational choice. Due to the excellent ability to collect and classify valuable knowledge, data mining has become a standard support method for gaining interesting insight into customer behavior. Even though rapid and accurate identification of customer demands is critical to business management, it is not feasible to design all approaches to meet all criteria to be developed. Therefore, this thesis aims to exploit novel data mining techniques blending with visual analytics (VA) to explore customer behavior and provide valuable insight for decision-making support. Insurance data such as questionnaires, demographic, and claim data are used as a testbed to demonstrate our techniques. This thesis is categorized into four main themes: (1) pattern mining (PM) for discovering adverse behavior (AB); (2) visual analytics (VA) for exploring customer behavior; (3) natural language interaction driven data visualization (NLI-driven-DV) to analyze customer claim behavior and manage risk; (4) deep visual analytics (DVA) to provide a wide range of performance evaluations of different methods for understanding customer behavior (UCB). This is one of the first studies to utilize data mining techniques blending with visual analytics (VA) for exploring customer behavior from the insurance business aspect. The empirical results of this thesis show the advantages and effectiveness of the developed methods valuable for researchers and insurance managers (IMs). Moreover, various aspects of insurance data have been researched and integrated into sophisticated visual interactive systems (VIS) to gain a deeper understanding of customer behavior and to better business plans and make decisions.

Keywords: Behavior analysis, data analytics, data visualization, visual analytics, pattern mining, machine learning, risk management, business strategies.

List of Abbreviations

CBA	Customer Behaviour Analysis
AB	Adverse Behaviour
AS	Adverse Selection
IMs	Insurance Managers
PM	Pattern Mining
ASF	Adverse Selected Factors
BFP	Breaking Frequent Patterns
SEIFA	Socio Economic Indexes for Areas
RFA	Randomly Flipping Attribute
MHI	Mental Health Illness
MHIVis	Visual Analytics for Exploring Mental Health Illness
UCB	Understanding Customer Behaviour
UCBVis	Visual Interactive System for Understanding Customer Behavior
VA	Visual Analytics
VAS	Visual Analytics System
VIS/Vis	Visual Interactive System
D3	Data-Driven Documents
ExVis	Explainable Visual Interactive System
DAG	Directed Acyclic Graph
BN	Bayesian Networks
CPT	Conditional Probability Table
KLD	Kullback-Leibler Divergence
JSD	Jensen-Shannon Divergence
AI	Artificial Intelligence

ML	Machine Learning
DL	Deep Learning
ExTraVis	Exploration of Traffic Incidents Using a Visual Interactive System
ITD	Incident Trend Dashboard
FARS	Fatality Analysis Reporting System
FTS	Free Text Search
TIC	Traffic Incident Controllers
NLP	Natural Language Processing
NLQ	Natural Language Query
ExNLQVis	Explainable NLQ based Visual Interactive System
DVA	Deep Visual Analytics
DS	Data Science
IV	Information Visualization
LIWC	Linguistic Inquiry and Word Count
LR	Logistic Regression
LSTM	Long Short Term Memory
UI	User Interface
HCI	Human Computer Interaction
RM	Risk Management
DM	Diabetes Mellitus
DiaVis	Visual Interactive System for Exploring Diabetes Disease
ARM	Association Rule Mining
CRM	Customer Risk Management