MULTI-AGENT BASED BIG DATA ANALYTICS USING ONTOLOGICAL MODELING AND FUZZY LOGIC FOR DECEPTIVE CONSUMER REVIEW DETECTION

Dushyanthi Udeshika Vidanagama

(168029P)

Degree of Doctor of Philosophy

Department of Computational Mathematics

University of Moratuwa
Sri Lanka

June 2023

MULTI-AGENT BASED BIG DATA ANALYTICS USING ONTOLOGICAL MODELING AND FUZZY LOGIC FOR DECEPTIVE CONSUMER REVIEW DETECTION

Dushyanthi Udeshika Vidanagama

(168029P)

Thesis submitted in partial fulfillment of the requirements for the degree
Doctor of Philosophy

Department of Computational Mathematics

University of Moratuwa
Sri Lanka

June 2023

DECLARATION

I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date: 19.06.2023

The above candidate has carried out research for the PhD thesis under my supervision.

Name of the supervisor: Dr (Mrs) A.T.P. Silva

Signature of the supervisor: Date : 19/06/2023

Name of the supervisor: Snr. Prof. A.S. Karunananda

Signature of the supervisor: Date : 19/06/2023

ABSTRACT

Customer reviews play a vital role in e-commerce. Most online customers depend on online reviews before making a purchasing decision, and the credibility of online reviews significantly impacts the company's reputation. Given the significance of reviews in generating revenue, some business owners reward reviewers who are dishonest. Consumer reviews are continuously produced at a high volume, velocity, and degree of unstructured. Hence, they can be treated as big data and need big data analysis methods to detect deceptive reviews. Due to the challenges highlighted among different classification processes, there is a research challenge in designing an effective deceptive review detection mechanism without focusing on labelled datasets in real-time. This research combines review-centric and reviewer-centric features in the feature selection stage and finds the deceptive level of each review based on those features without focusing on classification. The purpose of this research is to use multi-agent technology, a modern trend in Artificial Intelligence, to support the deceptive review detection process to automate complex tasks such as real-time data acquisition, feature selection, and calculating deceptive levels by ensuring high accuracy. The system follows a module architecture where all modules are incorporated with agents performing module tasks by communicating, coordinating, and negotiating with each other. The fuzzy agents in the credibility analysis module incorporate review content honesty, content quality, reviewer expertise, and reviewer trustworthiness for analysing the review credibility and reviewer credibility. The deceptive levels were calculated based on the credibility values. The human evaluated results were used to compare the results of the proposed model, k-means cluster results, and outlier-based deceptive identification method. The evaluation results indicated that the accuracy of detecting deceptive consumer reviews using multi-agent technology in big data analytics is 81% where the cluster model showed 73% and outlier-based model showed 63%. Also, the evaluation showed the importance of considering both review credibility and reviewer credibility when deciding on deceptive level. Therefore, the challenges encountered in existing deceptive review detection methodologies, such as scarcity of labelled data for model training, real-time data analysis, and uncertainty of credibility, were resolved by incorporating multi-agent technology in big data analytics. The ultimate goal of noticing the misleading level of reviews is to create an assured customer who will boost the business's revenue by expanding purchases because of the trustful and reliable reviews.

Keywords : e-commerce, review honesty, content quality, reviewer trustworthiness, reviewer expertise, credibility, fuzzy inferencing

DEDICATION

I dedicate this thesis

to

my beloved

father

and

husband

ACKNOWLEDGMENT

Foremost, I wish to express my sincere gratitude to my two supervisors, Snr. Prof. A.S. Karunananda and Dr (Mrs) A.T.P. Silva, for their continuous support for my PhD research, and their patience, motivation, enthusiasm, and immense knowledge. I am grateful to Snr. Prof. A.S. Karunananda for enlightening me the first glance of research. Their guidance helped me throughout the research work and during thesis writing. They were the best advisors and mentors for my PhD study. I would like to thank the other members of my progress review evaluation panel, Dr (Mrs) D.D.M. Ranasinghe, Dr L. Ranathunga, Mrs D.T. Kulawansa, and Dr (Mrs) T. Piyatilake, for their encouragement, insightful comments, and challenging questions.

I sincerely thank the Vice Chancellor, General Sir John Kotelawala Defence University, Major General Milinda Peiris RWP RSP VSV USP ndc psc, former Dean, Faculty of Computing, Cmdr. JU Gunaseela USP, psc., present Dean, Faculty of Computing, Dr A.D.A.I. Gunasekara, former Head of Department, Department of Information Technology, Mr R.P.S. Kathriarachchi, for encouraging me, offering resources, and releasing me from hard work during this challenging period. Further, I wish to thank my fellow staff members Dr (Mrs) N. Wedasinghe and Major R.M.M. Pradeep, for their advice, encouragement, and support bestowed upon me during this period. My heartfelt gratitude also goes to all staff members of Faculty of Computing, General Sir John Kotelawala Defence University, for being with me and supporting me in numerous ways.

Also, I wish to thank Ms Darshanie Premathilake, freelance copy editor of EDIT& READ Copy Editing and Proofreading Services, for providing me with proofreading service on time. Last but not least, I thank my beloved father, Dr W. Vidanagama, my beloved husband, Mr S.M.A. Kulathunga, and all the family members for supporting me spiritually throughout my journey to academic achievement.

CONTENTS

DECLARATION i

ABSTRACT ii

DEDICATION iii

ACKNOWLEDGMENT iv

TABLE OF CONTENT v

LIST OF FIGURES ix

LIST OF TABLES xii

LIST OF ABBREVIATIONS xiii

1 INTRODUCTION 1
- 1.1 Prolegomena . . . 1
- 1.2 Background and Motivation . . . 2
- 1.3 Problem in Brief . . . 5
- 1.4 Aim . . . 5
- 1.5 Objectives . . . 6
- 1.6 Proposed Solution . . . 6
- 1.7 Thesis Outline . . . 7
- 1.8 Summary . . . 8

2 LITERATURE REVIEW 9
- 2.1 Introduction . . . 9
- 2.2 Deceptive Reviews . . . 9
- 2.3 Deceptive Review Detection . . . 10
  - 2.3.1 Techniques Used in Deceptive Review Detection . . . 11
  - 2.3.2 Features Used in Deceptive Review Detection . . . 25
- 2.4 Big Data Analytics . . . 32
  - 2.4.1 Life Cycle of Big Data Analytics . . . 33
  - 2.4.2 Role of Big Data Analytics in Deceptive Review Detection . . . 36
- 2.5 Agent Technology for Big Data Analytics . . . 37
  - 2.5.1 Agents in Big Data Analytics . . . 38
  - 2.5.2 Intelligent Agents and Fuzzy Logic . . . 40
- 2.6 Ontological Modeling . . . 41
  - 2.6.1 Ontologies in Big Data Analytics . . . 41
  - 2.6.2 Ontologies in Agents . . . 42
- 2.7 Research Gap . . . 44
- 2.8 Problem Definition . . . 48

2.9 Summary 49

3 TECHNOLOGY ADOPTED **50**
3.1 Introduction 50
3.2 Agent Development Tools and Languages 50
3.3 Ontology Development Tools and Languages 51
3.4 Data Acquisition Tools and Languages 52
3.5 Pre-Processing and Feature Extraction Tools 52
3.6 Fuzzy Logic Modelling Tools and Languages 53
3.7 Database and Data Analysing Tools 53
3.8 Summary 53

4 METHODOLOGY **55**
4.1 Introduction 55
4.2 Proposed Approach 55
4.2.1 Hypothesis 55
4.2.2 The System Model 56
4.2.3 Features of the System 57
4.2.4 Users 58
4.3 Overall System Architecture 58
4.4 Module Architecture 60
4.5 Proposed Multi-Agent based Big Data Analytics Architecture 63
4.6 Ontological Modeling for Deceptive Review Detection 68
4.7 User Module Designing 70
4.8 Monitoring and Acquisition Module Designing 71
4.9 Pre-Processing Module Designing 71
4.10 Feature Engineering for Deceptive Review Identification 74
4.10.1 Review Honesty Features 81
4.10.2 Content Quality Features 82
4.10.3 Aspects-Opinion Extraction and Sentiment Analysis 84
4.10.4 Reviewer Expertise Features 89
4.10.5 Reviewer Trustworthiness Features 90
4.11 Fuzzy Logic Models Designing for Credibility Analysis 92
4.12 Summary 101

5 IMPLEMENTATION **102**
5.1 Introduction 102
5.2 Development of Ontologies 102
5.3 Monitoring and Acquisition 111
5.4 Data Pre-Processing 114
5.5 Feature Extraction for Deceptive Review Detection 121
5.5.1 Review-related Basic Feature Extraction 121
5.5.2 Reviewer and Product related Basic Feature Extraction 122
5.5.3 Content-related Feature Extraction 123
5.5.4 Reviewer-related Feature Extraction 127
5.6 Fuzzy Inferencing Models 128

5.6.1 Fuzzy Inference Model for Review Honesty (Model 1) 129
5.6.2 Fuzzy Inference Model for Content Quality (Model 2) 129
5.6.3 Fuzzy Inference Model for Reviewer Expertise (Model 3) . . 130
5.6.4 Fuzzy Inference Model for Abnormal Behaviour (Model 4) . . 130
5.6.5 Fuzzy Inference Model for Recent Reviewer Trustworthiness (Model 5) 134
5.6.6 Fuzzy Inference Model for Review Credibility (Model 6) . . . 134
5.6.7 Fuzzy Inference Model for Reviewer Credibility (Model 7) . . 137
5.7 Credibility Analysis 137
5.8 User Request Handling 140
5.9 Summary 141

6 EVALUATION 145
6.1 Introduction 145
6.2 Datasets Preparation 145
6.3 Preliminary Analysis 146
6.3.1 Feature Sets Anaysis and Cluster Model Preperation 146
6.3.2 Deceptive Review Identification from Outlier Detection . . . 148
6.4 Experimental Setup 149
6.5 Experimental Results 152
6.6 Results Analysis 153
6.7 Summary 154

7 CONCLUSION AND FUTURE WORK 156
7.1 Introduction 156
7.2 Concluding Remarks 156
7.3 Limitations 159
7.4 Future Work 160
7.5 Summary 160

REFERENCES 161

Appendix A: Slot Details 177

Appendix B: Responses Rate Summary Frequency 180

Appendix C: Antecedent Tables 181

Appendix D: Implemented source codes 186

Appendix E: Fuzzy Models 190
E.1 Fuzzy model of Review Honesty Calculation 190
E.2 Fuzzy model of Content Quality Calculation 192
E.3 Fuzzy Model for Reviewer Expertise Calculation 194
E.4 Fuzzy Model for Reviewer Expertise Calculation 196
E.5 Fuzzy Model for Reviewer Abnormal Behaviour Calculation 199
E.6 Fuzzy Model for Reviewer Recent Trustworthiness Calculation 201

E.7 Fuzzy Model for Review Credibility Calculation 204
E.8 Fuzzy Model for Reviewer Credibility Calculation 206

Appendix F: Questionnaire for Feature Weight Calculation **208**

LIST OF FIGURES

Figure 2.1 Fuzzy Inferencing System. 23
Figure 2.2 Life Cycle of Big Data Analytics. 33
Figure 2.3 Conventional Big Data Analytics Pipeline. 37
Figure 2.4 Communication with blackboard and Control agent. 38
Figure 2.5 Ontologies for Data Integration [147],[146]. 43
Figure 4.1 Overall process of the system. 57
Figure 4.2 Overall System Architecture. 59
Figure 4.3 Deceptive Analysis Module. 63
Figure 4.4 Proposed Multi-agent based Big Data Analytics Architecture. . 65
Figure 4.5 Top-level design of pre-processing ontology. 69
Figure 4.6 Top Level Design of Aspect Ontology. 70
Figure 4.7 Top Level Design of Application-specific Ontology. 71
Figure 4.8 Design of Agent Action Architecture. 72
Figure 4.9 Design of Agent Action Architecture (Continued). 73
Figure 4.10 Data flow of user request handling. 73
Figure 4.11 Agents comm. flow in monitoring and acquisition module. . . 74
Figure 4.12 Agent Communication flow in Pre-Processing Module. 75
Figure 4.13 Agent Comm. Flow in Feature Extraction Module. 76
Figure 4.14 Aspect-Opinion Generation. 85
Figure 4.15 Relevant Feature Identification. 85
Figure 4.16 New Aspect Generation. 86
Figure 4.17 Sentiment Expectation of New Aspects. 87
Figure 4.18 Sentiment-Score Calculation for Product Feature-Opinion Pairs. 89
Figure 4.19 Communication Flow of Agents in Credibility Analysis Module. 93
Figure 4.20 Overview of Fuzzy Model. 95
Figure 4.21 Weighted Averages of Features. 96
Figure 4.22 Input variable weights in fuzzy models. 97
Figure 5.1 Visualization of ErrorDetection class. 104
Figure 5.2 Visualization of CleaningMethod class. 105
Figure 5.3 SWRL Rules of Pre-Processing Ontology. 105
Figure 5.4 SPARQL query for cleaning actions of a review. 106
Figure 5.5 DeceptiveDetection Ontology. 107
Figure 5.6 ReviewVocabulary Interface. 108
Figure 5.7 RequestAcquisition agent action implementation. 108
Figure 5.8 DeceptiveDetection ontology class. 109
Figure 5.9 Ontology Registration within Monitoring Agent. 110
Figure 5.10 Aspect Ontology for Mobile Phone Domain. 110
Figure 5.11 Comm. Flow between Agents of Monitoring and Acqui. Module. 112
Figure 5.12 Monitoring and Acquisition of Monitoring Agent. 113
Figure 5.13 Data Crawling procedure. 115
Figure 5.14 YAML file for Product details extraction in Amazon. 115
Figure 5.15 YAML file for review extraction in Amazon. 116

Figure 5.16 Product details crawling in eBay. 116
Figure 5.17 eBay Reviewer details crawling. 117
Figure 5.18 eBay other reviews crawling. 117
Figure 5.19 Review attributes insert to Pre-Processing ontology. 118
Figure 5.20 Communication Flow between Agents of Pre-Processing Module.119
Figure 5.21 Querying the ontology for pre-processing actions. 119
Figure 5.22 Requesting Content Similarity Score values. 124
Figure 5.23 Comparison of received similarity score values. 124
Figure 5.24 Content Similarity Calculation. 124
Figure 5.25 Check for Existing aspects in Domain Ontology. 125
Figure 5.26 New aspect insertion to Domain Ontology. 126
Figure 5.27 Comm. Flow between Agents of Feature Extraction Module. . 128
Figure 5.28 MF Arrangements for Input and Output Variable of Model 1. . 129
Figure 5.29 Contour Graphs and Input-Output Surfaces of Model 1. 130
Figure 5.30 MF Arrangements for Input and Output Variable of Model 2. . 131
Figure 5.31 Contour Graphs and Input-Output Surfaces of Model 2. 132
Figure 5.32 MF Arrangements for Input and Output Variable of Model 3. . 132
Figure 5.33 Contour Graphs and Input-Output Surfaces of Model 3. 133
Figure 5.34 MF Arrangements for Input and Output Variable of Model 4. . 133
Figure 5.35 Contour Graphs and Input-Output Surfaces of Model 4. 134
Figure 5.36 MF Arrangements for Input and Output Variable of Model 5. . 135
Figure 5.37 Contour Graphs and Input-Output Surfaces of Model 5. 135
Figure 5.38 MF Arrangements for Input and Output Variable of Model 6. . 136
Figure 5.39 Contour Graphs and Input-Output Surfaces of Model 6. 136
Figure 5.40 MF Arrangements for Input and Output Variable of Model 7. . 137
Figure 5.41 Contour Graphs and Input-Output Surfaces of Model 7. 138
Figure 5.42 Comm. Flow between Agents of Credibility Analysis Module. 139
Figure 5.43 Summary Request Handling Form. 140
Figure 5.44 Algorithm for User Request Handling. 141
Figure 5.45 Implementation of User Module. 142
Figure 5.46 Summary Analysis for NOKIA 6101. 143
Figure 5.47 Detailed Deceptive Levels of Product B000BHB4XO. 143
Figure 5.48 MongoDB Query for Summary Retrieval. 144
Figure 6.1 Silhouette Index Analysis of K-means clustering 147
Figure 6.2 K-means Cluster Analysis Results for fuzzy features. 148
Figure 6.3 Outlier Distribution using Mahalanobis distance. 149
Figure 6.4 Threshold value determination for the proposed model. 152
Figure 6.5 Performance Measure Analysis 154
Figure D.1 Review Content lemmatization. 186
Figure D.2 Dependency parsing for review content. 186
Figure D.3 POS tags generation. 187
Figure D.4 Dependency relations extraction for review content. 187
Figure D.5 Receiving the sentiment value of the review content. 188
Figure D.6 Sentiment Calculation using SentiWordNet. 188
Figure D.7 Opinion count extraction. 189

Figure D.8 Readability score calculation. 189
Figure D.9 Linguistic Quality calculation. 189

LIST OF TABLES

Table 2.1 Research Challenges of Deceptive Review Detection Approaches. 45
Table 4.1 Possible Features. 62
Table 4.2 Cleaning Actions of Processing Agent. 75
Table 4.3 Description of Extracted Features. 76
Table 4.4 Feature Notations. 79
Table 4.5 Abnormal rate patterns of reviews in a product. 92
Table 4.6 Features of Fuzzy Logic Models. 96
Table 5.1 Deceptive Interpretation with Emoji Faces. 141
Table 6.1 Dataset Summary Statistics. 146
Table 6.2 Model Summary Results against Human Evaluation Results. . . 153
Table A.1 Slot details of RequestAquisition. 177
Table A.2 Slot details of RequestCredibilitySave. 177
Table A.3 Slot details of RequestDetailSummary. 177
Table A.4 Slot details of RequestEarlierTrust. 177
Table A.5 Slot details of RequestInfo. 177
Table A.6 Slot details of Review. 178
Table A.7 Slot details of ProcessedReview. 178
Table A.8 Slot details of InformProcess. 178
Table A.9 Slot details of RequestExtraction. 178
Table A.10 Slot details of RequestProcess. 179
Table A.11 Slot details of RequestCleaning. 179
Table A.12 Slot details of RequestSentiment. 179
Table A.13 Slot details of SendDetailSummary. 179
Table A.14 Slot details of SendSentimentScore. 179
Table A.15 Slot details of Opinion. 179
Table A.16 Slot details of SendOpinions. 179
Table A.17 Slot details of RequestEarlyTrust. 179
Table B.1 Summary rate frequency for features by responses. 180
Table C.1 Antecedent Table for fuzzy inference test Model 1. 181
Table C.2 Antecedent Table for fuzzy inference test Model 2. 182
Table C.3 Antecedent Table for fuzzy inference test Model 3. 182
Table C.4 Antecedent Table for fuzzy inference test Model 4. 183
Table C.5 Antecedent Table for fuzzy inference test Model 5. 184
Table C.6 Antecedent Table for fuzzy inference test Model 6. 184
Table C.7 Antecedent Table for fuzzy inference test Model 7. 185

LIST OF ABBREVIATIONS

ACL	Agent Communication Language
AID	Agent Identity
AMT	Amazon Mechanical Turk
CNN	Convolution Neural Network
CQ	Content Quality
ELM	Elaboration Likelihood Model
FCL	Fuzzy Control Language
FIPA	Foundation for Intelligent Physical Agents
GANs	Generative Adversarial Networks
HSM	Heuristic Systematic Model
HTML	Hypertext Markup Language
ICF	Iterative Computation Framework
IDE	Integrated Development Environment
JADE	Java Agent Development Framework
JSON	JavaScript Object Notation
KNN	Kth Nearest Neighbour
LDA	Latent Dirichlet Allocation
LR	Logistic Regression
MAS	Multi-Agent System
MF	Membership Function
MISO	Multiple-Input-Single-Output
NB	Naïve Bayes
NLP	Natural Language Processing
NLTK	Natural Language Tool Kit
NoSQL	Not-Only-Structured Query Language
OWL	Web Ontology Language
PMI	Point-wise Mutual Information
POS	Part-Of-Speech
RC	Review Credibility
RDF	Resource Description Framework
RE	Reviewer Expertise
RH	Review Honesty
RT	Reviewer Trustworthiness
RvC	Reviewer Credibility
SPARQL	SPARQL Protocol and RDF Query Language
SPSS	Statistical Package for Social Sciences
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TF-IDF	Term Frequency-Inverse Document Frequency
TISO	Two-Input-Single-Output
TMF	Triangular Membership Functions

TrMF	Trapezoidal Membership Functions
UH	Unsupervised Hedge Algorithm
URLs	Uniform Resource Locators
URSM	Unified Review Spamming Model
XML	Extensible Markup Language
YML	Yet Another Markup Language