ASPECT DETECTION IN SPORTSWEAR APPAREL REVIEWS FOR OPINION MINING

Rajapaksha Wasala Mudiyanselage Polwatte Gedara Sampath Rajapaksha

(179345X)

Degree of Master of Science in Computer Science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

May 2021

ASPECT DETECTION IN SPORTSWEAR APPAREL REVIEWS FOR OPINION MINING

Rajapaksha Wasala Mudiyanselage Polwatte Gedara Sampath Rajapaksha

(179345X)

This Dissertation submitted in partial fulfillment of the requirements for the Degree of Master of Science in Computer Science Specializing in Data Science, Analytics and Engineering

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

May 2021

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for 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 dissertation, 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).

Name: R.W.M.P.G.S Rajapaksha

Signature: UOM Verified Signature Date: 28/05/2021

The above candidate has carried out research for the Masters dissertation under my Supervision.

Name of the Supervisor: Dr. Surangika Ranathunga

Signature of the Supervisor: UOM Verified Signature Date: 28/05/2021

ABSTRACT

As a result of the growth of social media sites and e-commerce websites, most of these websites provide platforms for people to express their opinion about their products or services. Main purpose of these platforms is to improve customer shopping experience. Moreover, these websites can use customer reviews to improve their products or services. In the sportswear apparel industry, almost all e-commerce websites provide these platforms for customers to leave their feedback. Since manual analysis of huge number of reviews is practically impossible, the automated approach of sentiment analysis/opinion mining has got the attention.

Sentiment analysis can be classified into 3 categories such as document-level sentiment analysis, sentence-level sentiment analysis and aspect-level sentiment analysis. Document-level or sentence-level sentiment analysis does not give the complete information as reviews consist with multiple entities and may have different opinions for different entities. This issue has inspired the aspect level opinion mining.

There are two core tasks involve with aspect level opinion mining. Those are aspect extraction and aspect sentiment analysis. This research aim at the first task of aspect level opinion mining which is aspect extraction task for sportswear apparel reviews as none of pervious works consider a clothing review dataset. A new data set will be produced with manual annotations by domain experts. This study used different deep learning models and achieved state-of-the-art performance for sportswear apparel reviews. It serves as the baseline for future research.

Keywords

Sentiment Analysis, Opinion Mining, Multi-label classification, Aspect Based Opinion Mining, Aspect Extraction, BERT, RoBERTa, Sentence Pair Classification, Multi-label classification.

ACKNOWLEDGEMENT

My sincere appreciation goes to my family for the continuous support and motivation given to make this thesis a success. I also express my heartfelt gratitude to Dr. Surangika Ranathunga, my supervisor, for the supervision and advice given throughout to make this research a success. I also thank Mr. Hansa Perera from MAS Holdings who provided the dataset for this research. Last but not least I also thank my friends who supported me in this whole effort.

Table of Contents

DECLARATION i

ABSTRACT ii

ACKNOWLEDGEMENT iii

Table of Contents iv

List of Figures viii

List of Tables ix

CHAPTER 1 : INTRODUCTION 1

1.1 Background 1

1.2 Opinion Mining 2

1.3 Problem and Motivation 2

1.4 Objectives 3

1.5 Contribution 3

1.6 Report Organization 4

CHAPTER 2 : LITERATURE REVIEW 5

2.1 Opinion Mining 5

2.2 Opinion mining at different levels 6

2.2.1 Document-level opinion mining 6

2.2.2 Sentence-level opinion mining 6

2.2.3 Aspect-level opinion mining 6

2.3 Aspect Extraction 8

2.3.1 Unsupervised Methods 9

2.3.1.1 Frequency or Statistical 9
2.3.1.2 Bootstrapping (Unsupervised) 10
2.3.1.3 Heuristic or Rule-based 10
2.3.1.4 Topic Modeling 11
2.3.2 Semi -supervised Methods 12
2.3.2.1 Bootstrapping (Semi-supervised) 12
2.3.2.2 Double propagation 12
2.3.2.3 Dependency parsing 13
2.3.2.4 Lexicon-based 14
2.3.2.5 Word alignment or graph-based 14
2.3.2.6 Statistical based 15
2.3.3 Supervised Methods 15
2.3.3.1 Hidden Markov Model (HMM) based 15
2.3.3.2 Conditional random field (CRF) 16
2.3.4 Deep Learning 17
2.3.4.1 Transformers 18
2.3.5 Implicit Aspects 21
2.3.5.1 Unsupervised Methods 21
2.3.5.2 Semi-Supervised Methods 21
2.3.5.3 Supervised Methods 22
2.3.6 Summary 23
2.4 Data Pre-Processing 23
2.5 Feature Selection 24
CHAPTER 3 : RESEARCH METHODOLOGY 27

3.1 Customer Reviews Data Collection 27
3.2 Product Aspects 28
3.3 Data Preprocessing 32
3.4 Data Annotation 33
3.5 Exploratory Data Analysis 34
3.6 Classification Algorithms 36
3.6.1 Convolution Neural Network (CNN) 36
3.6.2 BERT 37
3.6.3 RoBERTa 37
3.6.4 Ensemble Methods 38
3.7 Evaluation Metrics 38
CHAPTER 4 : SYSTEM EVALUATION 39
4.1 Inter-Annotator Agreement 39
4.2 Evaluation Results 40
4.3 Error Analysis 42
4.3.1 Fit 42
4.3.2 Size 42
4.3.3 Material 43
4.3.4 Comfortability 43
4.3.5 Quality 44
4.3.6 Price 44
4.3.7 Color 45
4.3.8 General 45
CHAPTER 5 : CONCLUSION 46

5.1 Future Improvements 47

References 49

Appendix 57

List of Figures

Figure 1: Importance of customer reviews....2
Figure 2 : BERT input representation. 19
Figure 3: BERT fine-tuning 19
Figure 4: Apparel product aspects....29
Figure 5: Data Preprocessing 32
Figure 6: Sample of Annotated dataset 33
Figure 7: Aspect Distribution....35
Figure 8: Number of words distribution 36

List of Tables

Table 1: Aspect Based Sentiment Analysis 7
Table 2 : Customer reviews 28
Table 3 : Explicit and Implicit aspects 31
Table 4 : Summary of dataset 34
Table 5 : Inter-Annotator Agreement 39
Table 6 : : Evaluation Results 41
Table 7 : Error Analysis - Fit 42
Table 8 : Error Analysis - Size 42
Table 9 : Error Analysis - Material 43
Table 10 : Error Analysis – Comfortability 43
Table 11 : Error Analysis - Quality 44
Table 12 : Error Analysis - Price 44
Table 13 : Error Analysis - Color 45
Table 14 : Error Analysis - General 45