Implicit Feature Extraction from Customer Reviews Using Supervised Learning

Atheesan Sornalingam

179306F

M.Sc. in Computer Science

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

March 2020

Implicit Feature Extraction from Customer Reviews Using Supervised Learning

Atheesan Sornalingam

179306F

This dissertation submitted in partial fulfillment of the requirements for the Degree of M.Sc. in Computer Science specializing in Data Science, Engineering, and Analytics

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

March 2020

DECLARATION

I declare that this is my own work and this PG Diploma Project Report does not incorporate without acknowledgment 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 acknowledgment 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.

…….. ……...

Atheesan Sornalingam Date

I certify that the declaration above by the candidate is true to the best of my knowledge and that this project report is acceptable for evaluation for the CS5999 PG Diploma Project.

…….. ……...

Dr. Uthayasanker Thayasivam Date

ABSTRACT

Every online product selling applications are having review systems for their customers to review the products that they have purchased. Customers' reviews about the product will be either negative or positive and some reviews will give the meaning explicitly and some reviews will have implicit meaning. Nowadays most of the people do purchasing through online as a result there are thousands of reviews for a single product. On the other hand, these reviews will be useful for other customers to decide whether to purchase the product or not by going through the reviews. Mining implicit features from the customer reviews is a fundamental requirement for extracting customers' opinions and summarizing. This research focuses on extracting implicit features from reviews for opinion mining using a word embedding model. It removes noisy words and learn the model parameters automatically and extract the implicit features from customer reviews. Most of the existing researches have focused on implicit feature extraction from Chinese web reviews and only few attempts are made to extract implicit features from English web reviews. Implicit feature extraction was done through supervised, semi-supervised and unsupervised learning approaches. This research focuses on supervised aspect extraction using deep learning. This research proposes a novel and yet simple CNN model employing two types of pre-trained embeddings for aspect extraction: general-purpose embeddings and domain-specific embeddings associated with a Word Embedding based Correlation (WEC) model by integrating advantages of both the translation model and word embedding to extract implicit features. WEC model can score their correlation score for each word in review and feature. Then the CNN is used to identify the feature where the input for CNN is similarity matrix generated using the correlation scores. CNN gives the matching score of the review feature pair as the output and the review’s corresponding feature will be identified from the feature set.

Acknowledgements

I would like to express profound gratitude to my advisor, Dr. Uthayasanker Thayasivam, for his invaluable support by providing relevant knowledge, materials, advice, supervision and useful suggestions throughout this research work. His expertise and continuous guidance enabled me to complete my work successfully.

I would like to thank all my colleagues for their help in finding relevant research material, sharing knowledge and experience and for their encouragement.

I am as ever, especially indebted to my parents for their love and support throughout my life. I also wish to thank my loving friends, who supported me throughout my work. Finally, I wish to express my gratitude to all my colleagues at my work place, for the support given to me to manage my MSc research work.

Table of Contents

LIST OF FIGURES

Page

LIST OF ALGORITHMS

Page

LIST OF TABLES

Page