AFFECT LEVEL OPINION MINING OF TWITTER STREAMS

Wijesuriya Arachchige Yasas Pramuditha Senarath

188005H

Thesis/Dissertation submitted in partial fulfillment of the requirements for the degree Master of Science in Computer Science and Engineering

Department of Computer Science & Engineering

University of Moratuwa
Sri Lanka

July 2019

DECLARATION

I declare that this is my own work and this dissertation 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 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).

Signature: Date:

The above candidate has carried out research for the Masters thesis/Dissertation under my supervision.

Signature of the Supervisor: Date:

ACKNOWLEDGEMENTS

I would first like to express my gratitude to my supervisor Dr. Uthayasanker Thayasivam and other advisers from DataSEARCH center who always shared their wisdom and experience to make this research study success. They constantly steered me towards right direction whenever they though I needed it. This research was supported by DataSEARCH center at University of Moratuwa, Sri Lanka.

I am sincerely grateful for Dr. Surangika Ranathunga and Dr. C.R.J. Amalraj for reviewing this research study at progress reviews and providing their valuable suggestions to improve the project outcomes.

I would also like to thank Mr. Kumaran Kugathasan for helping me to coordinate research activities. Further, I thank both academic and non-academic staff of Department of Computer Science and Engineering for giving their fullest support towards this project.

Last but not least, I would like to express my gratitude to all the members of DataSEARCH center of Department of Computer Science and Engineering for their constant monitoring and invaluable suggestions to make this study a success.

ABSTRACT

Affect Level Opinion Mining of Twitter Streams

Twitter is a social media platform which is used by millions of users to express their opinions freely. However, it is almost impossible to analyze the opinion manually due to the sheer number of Tweets generated per day. Therefore, automated analysis of emotions in Tweets, which is also known as affect level opinion mining in the literature is crucial. Emotion analysis in this study is performed at two levels: Emotion Category Classification and Emotion Intensity Prediction.

One key challenge in identifying emotion categories is the presence of implicit emotions. This study introduces a model that enables reuse of the same deep neural network architecture with different word embeddings for the extraction of different features related to implicit emotion classification. We presented this model at 9^{th} Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA-2018). Our system was ranked among the top ten systems (8^{th}) amidst constrained corpus usage. Our implicit emotion classifier outperformed the baseline system by more than 8%, achieving a 68.1% macro F1-Score.

We solved the emotion intensity task with transfer learning techniques. Among the models used in transferring features were a sentiment classifier, emotion classifier, emoji classifier and emotion intensity predictor. Our transfer learning based intensity predictor outperformed existing best in two out of four emotions. We were able to achieve an average Pearson score of 79.81%. Additionally, we propose a technique to visualize the importance of each word in a tweet to get a better understanding of the model.

Finally, we developed a web-platform that utilizes our emotion analysis models to summarize and view the opinion of a group of tweets.

Keywords: Emotion Classification; Emotion Intensity Prediction; Sentiment Analysis; Opinion Mining;

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABBR	Abbreviation
API	Application Programming Interface
ML	Machine Learning
FNN	Feed-foreword Neural Network
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long-Short Term Memory Network
SVM	Support Vector Machine
SVR	Support Vector Regression

TABLE OF CONTENTS