

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 thought 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 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA-2018). Our system was ranked among the top ten systems (8th) 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

Figure 1.1	An emotion intensity example.	2
Figure 2.1	Plutchik's wheel of emotions	6
Figure 3.1	Machine Learning Hierarchy	13
Figure 3.2	An isolated neuron	13
Figure 3.3	A three layer Feed-forward Neural Network	14
Figure 3.4	Convolution operation on a matrix	15
Figure 3.5	Unrolled RNN Structure	15
Figure 3.6	Common RNN module types	16
Figure 3.7	Different Settings of Transfer Learning	17
Figure 3.8	Transfer learning with Fine tuning	18
Figure 3.9	Transfer learning with Fixed Model	19
Figure 3.10	Cross validation iterations	21
Figure 4.1	Approach used in this study for emotion analysis	22
Figure 4.2	Overview of the implicit emotion classification architecture	24
Figure 4.3	FNN used as as final implicit emotion classifier	25
Figure 4.4	High-level LSTM-Conv Network Architecture	26
Figure 4.5	Overview of emotion intensity prediction architecture	28
Figure 4.6	Recurrent-Convolutional Neural Network	29
Figure 4.7	Emotion Intensity Regression - EITL Module	30
Figure 6.1	High-level Architecture of Emotion Visualization/Summarization Architecture	41
Figure 6.2	Emotion visualization and summarization platform	42
Figure 6.3	Individual emotion intensity visualization with the visualization platform	42

LIST OF TABLES

Table 1.1	Example Tweets classified according to emotion	2
Table 2.1	Results for systems evaluated on SemEval-2018 Task 1	7
Table 2.2	Comparison of system results for emotion annotations in [1]	8
Table 2.3	Emotion Intensity models at WASSA-2017	9
Table 2.4	Emotion Intensity regression models at SemEval-2018 Task 1	11
Table 3.1	Transfer Learning Approaches	16
Table 4.1	Embedding Models used in Experiments	26
Table 5.1	Emotion words used when collecting Tweets	32
Table 5.2	Distribution of the Implicit Emotion Dataset	32
Table 5.3	Network Parameters for LSTM-CNN	32
Table 5.4	Network parameters for FNN	33
Table 5.5	Evaluation of LSTM-CNN for different word embeddings	34
Table 5.6	Results of FNN for best feature combinations	34
Table 5.7	The number of tweets in the SemEval-2018 Affect in Tweets Dataset	36
Table 5.8	Model and training hyper-parameters for ECCU and EIPU models	37
Table 5.9	Parameters used for training EITL model	37
Table 5.10	Performance scores for ECCU compared with the benchmark systems	38
Table 5.11	Performance scores of emotion intensity prediction model	38
Table 5.12	Examples for word level importance heat-map visualizations	39

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

Declaration of the Candidate and Supervisor	i
Acknowledgement	ii
Abstract	iii
List of Figures	iv
List of Tables	v
List of Abbreviations	vi
Table of Contents	vii
1 Introduction	1
1.1 Problem	3
1.2 Motivation	3
1.3 Research Objectives	3
1.4 Contributions	4
2 Literature Survey	5
2.1 Emotion Classification	5
2.2 Emotion Intensity Prediction	8
3 Background	12
3.1 Artificial Neural Networks	12
3.2 Neural Network Layer Types	13
3.2.1 Dense Layer	13
3.2.2 Convolutional Layer	14
3.2.3 Recurrent Layers	14
3.3 Transfer Learning	15
3.3.1 Transfer Learning with Fine-tuning	17
3.3.2 Transfer Learning with Fixed Model	18
3.4 Word Embedding	19
3.4.1 Word2vec	19
3.4.2 Glove	19
3.4.3 FastText	20

3.5	Cross Validation	20
4	Methodology	22
4.1	Implicit Emotion Classification	22
4.1.1	Tweet Pre-processor	23
4.1.2	Models	24
4.1.3	Feature Extraction	25
4.2	Emotion Intensity Prediction	27
4.2.1	Tweet Pre-processor	27
4.2.2	Models	27
5	Evaluation	31
5.1	Implicit Emotion Classification	31
5.1.1	Dataset	31
5.1.2	Experimental Setup	32
5.1.3	Results	33
5.1.4	Discussion	35
5.2	Emotion Intensity Prediction	35
5.2.1	Dataset	35
5.2.2	Experimental Setup	36
5.2.3	Results	36
5.2.4	Discussion	38
6	Twitter Emotion Analysis Platform	41
7	Conclusion	43
7.1	Future Work	44
	References	45