

Predict User Mood According to Facebook Postings

Sabaragamu Ralalage Thilini Bhagya Samaraweera

169331R

Faculty of Information Technology

University of Moratuwa

2019

Predict User Mood According to Facebook Postings

Sabaragamu Ralalage Thilini Bhagya Samaraweera

169331R

Dissertation submitted to the Faculty of Information Technology, University of
Moratuwa, Sri Lanka for partial fulfillment of the requirements of Master of Science
in Information Technology

February 2019

Declaration

I hereby declare that this project report entitled “Predict User Mood According to Facebook postings” contains my own work and has not been submitted and will not be submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Name of Student: S.R.T.B.Samaraweera

Signature of Student:

Date:

Supervised by

Name of Supervisor:

Signature of Supervisor:

Date:

Dedication

I would like to dedicate my project “Predict User Mood According to Facebook Postings”,

To my project supervisor Mrs. GTI Karunaratne,

To the Instructors of Faculty of Information Technology at the University of Moratuwa who have supported me to make a success of this project.

Acknowledgments

Firstly, I would like to express my sincere gratitude to my supervisor Ms.GTI Karunaratne for the continuous support of my research, for her patience, motivation, and immense knowledge. Her guidance helped me in all the time of the research and writing of this thesis.

Besides my supervisor, I would like to thank Mr. SC Premarathne for his insightful comments, encouragement and for the support given me to widen my research from various perspectives.

Finally, my sincere thanks also go to Instructors of Faculty of Information Technology at the University of Moratuwa who has supported me to find resources and requirements for the project. Without their precious support, it would not be possible to conduct this research.

Abstract

We live in an era where communication is growing fast in the cyberspace. As a part of this people tend to be online 24 hours of a day and they write postings in social media. The interesting point is people may put a social mask to hide their feelings in the real world, but they reveal it on their post unknowingly. With this context discussion regarding opinion, mining has dominated research in recent years. Because of the ambiguity of human language, it is difficult to extract the sentiment precisely. Using appropriate machine learning approaches this paper explores to extract the polarity of the postings and predict the mood of a user accordingly. It could use to manage, communicate and collaborate with people more effectively and to manage own personal well-being and happiness. This study applies sentiment analysis for analyzing the hidden information present in the text on social media postings. It is an application of Natural Language Processing. In order to perform Sentiment analysis, need to identify the subjective and objective in the text. Because only the subjective text describes the sentimental information. Then the subjective text is preprocessing using various text preprocessing methods to extract the features. Text preprocessing may include stop word removal, stemming, tokenization, conjunction handling, and negation handling. After performing sentiment classification sentiment polarity can be extracted. To achieve this study uses the lexicon sentiment analysis process. Next, for sentiment classification, machine learning approaches can be used. It is an automatic classification technique and classification is performed using text features. The study uses supervised learning techniques. Using predefined emotion classes and sentiment polarity classifier is built accordingly.

Keywords - Natural language processing, Text preprocessing, Sentiment Analysis, Supervised learning.

Table of Contents

Declaration.....	<i>i</i>
Dedication.....	<i>ii</i>
Acknowledgments.....	<i>iii</i>
Abstract.....	<i>iv</i>
Chapter 1.....	1
Introduction.....	1
1.1 Chapter Overview.....	1
1.2 Background and Motivation.....	1
1.3 Aim and Objectives.....	2
1.4 Chapter Summary.....	3
Chapter 2.....	4
Review of others' work.....	4
2.1 Chapter Introduction.....	4
2.1 Text preprocessing techniques.....	4
2.1.1 N-gram.....	4
2.1.2 Tokenization.....	4
2.1.3 Stemming.....	5
2.1.4 Removing stop words.....	5
2.1.5 Negation handling.....	5
2.2 Sentiment classification approaches.....	6
2.2.1 Lexicon based approaches.....	6
2.2.2 Machine learning approaches.....	7
2.2.3 Support Vector Machines Classifier.....	7
Chapter 3.....	9
Technology Adapted.....	9
3.1 Chapter Introduction.....	9
3.2 Python.....	9
3.2 NLTK.....	9
3.3 TextBlob.....	9
3.4 PyCharm 2018 3.5.....	10
3.5 Tkinter toolkit.....	10
3.6 WordNetLemmatizer.....	10
3.7 Chapter Summary.....	10

Chapter 4.....	11
Methodology.....	11
4.1 Chapter Introduction.....	11
4.2 Statement of Research Problem.....	11
4.3 Process.....	11
4.2 Expected outcome/alternative approaches.....	12
4.3 Collecting Data.....	12
4.4 Data Preprocessing.....	12
4.4.1 Removing punctuations and other unnecessary characters.....	13
4.4.2 Removing whitespaces.....	14
4.4.3 Spelling correction.....	14
4.4.4 Tokenization.....	14
4.4.5 Part of Speech Tagging.....	15
4.4.6 Stop word removal.....	15
4.4.7 Lemmatization.....	15
4.4.8 Negation Handling.....	16
4.5 Supervised learning classification approach.....	16
4.5.1 Feature Extraction.....	16
4.5.2 Sentiment Classification.....	17
4.6 Lexicon based approach.....	18
4.7 Chapter Summary.....	18
Chapter 05.....	19
Analysis and Design.....	19
5.1 Chapter Introduction.....	19
5.2 Collecting Data.....	20
5.3 Data Preprocessing.....	20
5.4 Supervised learning classification approach.....	20
5.4.1 Feature Extraction.....	20
5.4.2 Sentiment Classification.....	20
5.5 Lexicon based approach.....	20
5.6 System training and evaluation.....	21
5.7 Graphical User Interface (GUI) design.....	21
5.8 Chapter Summary.....	22
Chapter 6.....	23
Implementation.....	23
6.1 Chapter Introduction.....	23

6.2 Data Preprocessing.....	23
6.2.1 Removing punctuations and other unnecessary characters.....	23
6.2.2 Removing whitespaces.....	23
6.2.3 Spelling correction.....	23
6.2.4 Tokenization.....	24
6.2.5 Stop word removal.....	24
6.2.6 Lemmatization.....	24
6.3 Supervised learning classification approach.....	24
6.4 Lexicon based approach.....	25
6.5 Chapter Summary.....	25
Chapter 07.....	26
Evaluation and Results.....	26
7.1 Chapter Introduction.....	26
7.2 Supervised learning classification approach.....	26
7.2 Confusion Matrix.....	26
7.3 Lexicon based approach.....	27
7.4 Evaluation.....	28
7.5 Chapter Summary.....	29
Chapter 08.....	30
Discussion.....	30
7.1 Chapter Introduction.....	30
7.2 Future works and Limitations.....	30

List of Figures

Figure 4.1: Classification Model.....	11
Figure 4.2: Lexicon Model	11
Figure 4.3: Example for unstructured data	13
Figure 4.4: Punctuations and other characters	13
Figure 4.5: Punctuations and other characters are removed	13
Figure 4.6: Whitespaces are included	14
Figure 4.7: Whitespaces are removed.....	14
Figure 4.8: Spelling mistakes.....	14
Figure 4.9: Spelling correction	14
Figure 4.10: Facebook posting 1.....	14
Figure 4.11: Sentence tokenization.....	15
Figure 4.12: Word tokenization	15
Figure 4.13: Facebook posting 2.....	15
Figure 4.14: POS tagging for each word	15
Figure 4.15: Facebook posting 3.....	16
Figure 4.16: Lemmatized words	16
Figure 4.17: Facebook posting 4.....	16
Figure 4.18: Convert in to ‘not’	16
Figure 5.1: System Diagram	19
Figure 5.2: User GUI	21
Figure 5.3: Data labeling GUI	22
Figure 6.1: White space removing	23
Figure 6.2: Spelling correction	23
Figure 6.3: sentence tokenization	24
Figure 6.4: word tokenization	24
Figure 6.5: stop word removal	24
Figure 6.6: textblob pos tagging	25
Figure 7.1: NB model accuracy	26
Figure 7.4: Measures of confusion matrix	27
Figure 7.5: Confusion matrix for the classification model	27
Figure 7.6: Results of classification model.....	27

Figure 7.7: Confusion matrix for lexicon model28
Figure 7.8: Results of the lexicon model28

List of Tables

Table 4.1: Types of Emotions 12