Universiti Teknologi MARA

Sentiment Analysis Using Negative Selection Algorithm for Twitter's Messages

Nazirah Binti Che Alhadi

Thesis submitted in fulfillment of the requirements for
Bachelor of Computer Science (Hons)
Faculty of Computer & Mathematical Sciences

July 2012

DECLARATION

I certify that this thesis and the research to which it refers are the product of my own work and that any ideas or quotation from the work of other people, published or otherwise are fully acknowledged in accordance with the standard referring practices of the discipline

JULY 15, 2012

NAZIRAH BINTI CHE ALHADI

2010608328

ABSTRACT

Micro-blogs as a new textual domain offer a unique proposition for sentiment analysis. Their short document length suggests any sentiment they contain is compact and explicit. It can pose difficulties for standard machine learning document representations because of the short length coupled with their noisy nature. The aim of this project is to classify Twitter's messages into sentiment categories based on the important keywords. This project methodology consists of five phases which are preliminary study, data collection and preparation, model development, model evaluation and documentation. This project is designed using negative selection algorithm to automatically classify the Twitter's messages into its sentiment's category based on important keyword recognition. In order to develop this model classification and prototype, 480 Twitter's messages were used as training data and 120 Twitter's messages for testing data to determine the accuracy of the classification model. The accuracy of this model is about 60 percent. Second experiment was carried out by reducing the data to 240 for training data and 60 data for testing. The accuracy for second experiment is improved to 63.33 percent.

Keyword: Sentiment Analysis, text mining, negative selection algorithm (NSA), Twitter.

TABLE OF CONTENTS

CONTENT		PAGE
DECLARATION		i
ACKNOWLEDGEMENT		ii
APPROVAL		iii
ABSTRACT		iv
LIST OF FIGURES		ix
LIST OF TABLES		x
CHAPTER 1 : INTRODUCTION		1
1.1	Background Study	1
1.2	Problem Statement	2
1.3	Project Objectives	3
1.4	Project Scope	3
1.5	Significant of Study	3
1.6	Overview of Research Methodology	4
1.7	Summary	5
CHAPTER 2 : LITERATURE REVIEW		6
2.1	Knowledge Discovery in Database (KDD)	6
	2.1.1 Data Mining	7

2.1.2 Text Mining 7

2.2 Text Classification 9

2.3 Text Preprocessing 10

2.4 Sentiment Anaysis 13

2.5 Twitter's Messages 13

2.6 Classification Techniques for Text Mining 15

2.6.1 Support Vector Machine (SVM) 15

2.6.2 Naïve Bayes Classifier 16

2.6.3 Artificial Immune System (AIS) 17

2.6.3.1 Artificial Immune Network (AINET) 18

2.6.3.2 Clonal Selection Algorithm (CLONALG) 19

2.6.3.3 Negative Selection Algorithm (NSA) 20

2.6.4 NSA for Microblog Text Classification 20

2.7 Summary 23

CHAPTER 3 : RESEARCH METHODOLOGY 24

3.1 Introduction 24

3.2 Project Development Methodology 24

3.2.1 Preliminary Study 26

3.2.1.1 Knowledge Acquisition 26

3.2.1.2 Literature survey on sentiment analysis method 27

3.2.2 Data Collection and Preparation 27

3.2.2.1 Data Preparation 29