Universiti Teknologi MARA

Sentiment Analysis Using Clonal Selection Algorithm for Twitter's Data

Fatimah bt. Mamat

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 2012

.....

FATIMAH BINTI MAMAT

2010226342

ABSTRACT

Twitter's is a microblogging social networking website that has a large and rapidly growing user base. Thus, the website provides a rich bank of data in the form of "tweets", which are short status update from Twitter's user that must be written in 140 characters or less. As an increasingly popular platform for conveying opinions and thoughts, it seems natural to mine Twitter for potentially interesting trends regarding prominent topics in the news or popular culture. The sentiment analysis using clonal selection algorithm for twitter's data system was developed to achieve the main objective which is to classify the twitter's messages according three sentiments which are positive, negative and neutral. Clonal selection algorithm was used in this project because there are no researcher are focus on that technique for classify twitter's data. This project can be used for marketing area because of the data was about review on I-phone. Nevertheless, it's only accepts English standard word. In order to achieve the main objective, five phases of methodology was been implemented which are preliminary study, data preparation, model development, model evaluation & prototype development and last but not list is documentation. The evaluation conducted in this project has shown by accuracy is testing process. It used to check whether the data have been classifier correctly or incorrectly. Two experiments were carried out with different amount of data. At the first experiment, 200 data was used and the accuracy was 60 percent, while decrease data into 125 during experiment two, the accuracy was 56 percent only.

Keywords – clonal selection algorithm, text mining, sentiment analysis, twitter.

TABLE OF CONTENTS