Universiti Teknologi MARA

Sentiment Analysis for Malay Newspaper (SAMNews) Using Negative Selection Algorithm

Nur Amalina Binti Redzuan

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

January 2013

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.

.....

NUR AMALINA BINTI REDZUAN

2010453932

JANUARY 23, 2013

ABSTRACT

Newspapers express sentiments during reporting on recent events every day. It also can be known as a new domain in textual type for sentiment analysis that deals with many suggestions. The newspaper is documented in long sentences but the contents that express the sentiment is compressed and clearly understand by human. However, for the machine learning text representation it cause some problems because of the noisy text. As the solution, this project is conducted on the purpose to determine the polarity of the sentiment in the newspapers sentences. This project is implemented based on five phases in methodology part which consists of background study, data collection and preparation, prototype design, prototype development and evaluation and documentation. Sentiment Analysis for Malay Newspaper (SAMNews) is constructed using the negative selection algorithm which is able to classify the sentiment in newspaper's sentences into the polarity (positive, negative or neutral) automatically repose on detectors word. The sentiment analysis in this project utilized 1000 newspaper's sentences for the training and classification phase and testing data to evaluate the average of accurateness. The evaluation is made on three experiments which in Experiment I used 700 newspaper's sentences as the training data and 300 newspaper's sentences as the testing data. The accuracy of this experiment is about 59.99%. In Experiment II, 800 newspaper's sentences and 200 newspaper's sentences are used as the training data and testing data. The accuracy of this experiment is increased about 58.58%. While in Experiment III used 900 newspaper's sentences as the training data and 100 newspaper's sentences as the testing data and the accuracy is improved to 65.81%. In future, a comparative study on Artificial Immune System and other techniques or algorithms can be carried out to enhance the performance of the classification model.

Keywords – negative selection algorithm (NSA), text mining, sentiment analysis, newspaper.

TABLE OF CONTENTS

CHAPTER THREE: METHODOLOGY

CHAPTER FOUR: RESULT AND ANALYSIS