Universiti Teknologi MARA

Sentiment Classification for Malay Newspaper Using Clonal Selection Algorithm

Nur Fitri Nabila Binti Mohamad Nasir

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 FITRI NABILA BINTI MOHAMAD NASIR
2010270306

JANUARY 23, 2013

ABSTRACT

Sentiment classification is technique to analyze the subjective information in the text then mine the opinion. Mostly people are using blog or twitter to collect the sentiment data but not frequently used newspaper because not so many researchers are using newspaper to classify sentiment data as the main source. In this study, sentiment classifier using clonal algorithm selection was developed to categorize sentiment in Malay newspaper (Berita Harian). Another objective was to evaluate the proposed model effectiveness in classifying Malay newspaper's data. In our method, the training of clonal selection algorithm (CSA) is first used to teach algorithm which is intelligent to categorize the sentiment in newspaper's sentences into the polarity (positive, negative and neutral)from the data are collected and the testing was implemented after did the training to test whether a word should be taught correctly or not. Firstly, the data was dividing by ratio 80:20 from 1000 sentences. Therefore, 80% from 1000 sentences will use for training and 20% from 1000 sentences use for testing. Secondly, the data was dividing by ratio 70:30 which are 700 newspaper's sentences as the training data and 300 newspaper's sentences as the testing data. The experimental results show that our method can achieve better performance in clonal selection algorithm sentiment classification and the data collected cannot be used at once in this model because training data is very time-consuming if using all the data. The experiment achieves the best accuracy at 89.0%for ratio 70:30.This model was built with capability to help user in classifying newspaper sentence in easy way.

Keywords: Clonal selection algorithm, newspaper data, sentiment classification

TABLE OF CONTENTS