Abstract

Using information from opinion sentences in the newspaper is not an easy thing to do because of a lot of cases of anaphora in the sentences in newspaper. This final year project will identify one kind of anaphora called definite noun phrase anaphora by using Support Vector Machine with Feature - Based Radial Basis Function (RBF). This system uses feature set that formed the feature vector as the input to the Support Vector Machine (SVM). This system is divided into several processes, namely preprocessing process, feature extraction process, training process and testing process.

Testing process will be conducted in 4 kind of testing, first is cross validation testing for finding most appropriate model for the system[16]. The second one is parameter testing. In this testing the model found in the cross validation test is used in finding best parameter combination that reach highest accuracy. The result shows that the bigger value of C produce higher accuracy. As for gamma, it is found that its value can make the accuracy stable in particular value depend on the value of C. The next testing result also shows that the distribution of data used in this final year project doesn't affect accuracy. And the last testing shows that system accuracy reaching 100% based on white-box testing.

Keyword: definite noun phrase anaphora, RBF, SVM, feature set