

Abstract

As an individual, people have so many questions in their entire life. Sometimes they could solve it on their own, but sometimes they need other people to find the answers to that questions. Because of that, many Community Question-Answering Forum are formed to help solve each other's questions. The widespread use of the internet nowadays formed a lot of Online Community Question-Answering. The Online Community Question-Answering not only covers the academic issues, but also about how to living in other country like Qatar Living Forum. However, most of the times the answer that were given did not match the question. It takes some times to choose the right answer that correspond with the answer. For that reason, a rank-based system which can rank the correlation between the question and the answer is needed.

This research is based on SemEval 2016: Semantic Evaluation Exercise Task 3 Subtask A about the Question-Comment Similarity. The steps that being taken on this study are pre-processing dataset based on SemEval 2016 question-answering similarity, feature extraction to help the classification process using Lexical Similarity Feature, Semantic Similarity Feature, Non-Textual Feature Group, and Heuristic. The classifier used in this study is Logistic Regression and Support Vector Machine (SVM). The value of the classification results are used to rank the most suitable answers to the questions. The measurement of the system performance is done by using Mean Average Precision (MAP) calculation.

The results of this study shows the influence of the similarity between the questions and answers. Lexical Similarity Feature, especially Cosine Similarity Sub-Feature and LCS shows that the higher the value of the feature, the closer the answer to the question. On the other hand, the distribution of Semantic Similarity value using Wu Palmer Algorithm in each class is more evenly distributed, so that it is quite difficult to distinguish the characteristics of each class. Non Textual Feature Group is helping to classify the answers and improves 4% of the accuracy. Compared with SemEval 2016 Task 3 participants, this study placed the 8th rank from 13 participants.

Keyword: Community Question Answering, Question Answering System, similarity measure, classifier, CQA answer rank, MAP.