

Abstract

The growth of the internet makes user to be able to find information easily. Community Question Answering (CQA) is a facility that allow user to do question answering freely and easily. Most discussion forum frees user to write question or answer, therefore the answer is certainly varies, there are answers that provide a solution and does not provide a solution. In this research, we will classify questions that are available in Qatar Living discussion forum. The purpose of identification is to assist whenever we want to specify answer that belong to good, bad, and potential class. Answer classification is performed with supervised learning.

Classification process performed on data which represented by feature such as Similarity feature (semantic similarity and cosine similarity), Topic model, Textual feature (author), and Non Textual feature (special word, heuristic/link, head word, emoticon, and question mark). Generally, this research is done with three step: first preprocessing data, then feature extraction, and the last is answer classification process. Preprocessing that is consist of tokenization, stopword removal, and lemmatization. Differences this research between previous research, JAIST is selection feature. JAIST using Word matching feature group, Special-component feature group, Non textual feature group, Topic model, Word vector, and Translation based feature.

Based on the evaluation of this research, the author found that classification have been implemented has an accuracy 72,36 % and Macro F1 54,10 %. If compared with the results of SemEval 2015, this research was ranked 3 out of 12 participant with Macro F1 score as a baseline for the ranking.

Keyword: community question answering, supervised learning, semantic similiarity, topic model, qatar living.