

Abstract

In a question-and-answer forum, the identification of question similarity is used to determine how similar two questions are. This procedure makes sure that user-submitted questions are compared to the questions in a database for matches to improve system performance on the online Q&A platform. Currently, question similarity is mostly done in foreign languages. The purpose of this research is to identify question similarities and evaluate the effectiveness of the methods used in Indonesian language questions. The data used is a public dataset with labeled pairs of questions as 0 and 1 where label 0 for different pairs of questions and label 1 for the same pairs of questions. The method used is Recurrent Neural Network (RNN) with the Manhattan Distance approach to calculate the similarity distance between two questions. The question pairs are taken as two inputs with a reference label to identify the similarity distance between the two question inputs. We evaluated the model using three different optimizers namely RMSprop, Adam, and Adagrad. The best results were obtained using the Adam optimizer with 80:20 ratio split-data and overall accuracy is 76%, precision is 74%, recall is 98.8%, and F1-score is 85.1%.

Keywords: *Manhattan Distance, semantic measurement, similarity questions, RNN.*
