

INTISARI

KLASIFIKASI PASANGAN PERTANYAAN MENGGUNAKAN *LONG SHORT-TERM MEMORY RECURRENT NEURAL NETWORK*

Oleh:

Dsk Feby Inka Putri

13/347478/PA/25256

Dalam sebuah situs tanya jawab, permasalahan yang umumnya muncul adalah ketika salah satu atau beberapa pengguna lainnya mengajukan pertanyaan yang sama dalam diskusi yang berbeda. Hal ini akan menjadi sebuah permasalahan dalam efisiensi sebuah situs tanya jawab serta penggunaan ruang penyimpanan yang besar untuk informasi yang sama. Pertanyaan-pertanyaan dengan maksud yang sama akan tersebar di berbagai forum diskusi, padahal jika digabungkan akan lebih baik dan memaksimalkan jawaban oleh pengguna lain.

Solusi awal dari permasalahan tersebut adalah dengan membangun sebuah sistem yang melakukan deteksi otomatis terhadap pasangan pertanyaan yang duplikat. Dimana pada penelitian ini, dilakukan pemodelan pada bahasa alami untuk menemukan kalimat yang duplikat satu sama lainnya dengan menggunakan *Long Short-Term Memory Recurrent Neural Network*. Data yang digunakan dalam penelitian ini adalah dataset yang dirilis oleh situs tanya jawab Quora, dimana data terdiri dari pertanyaan-pertanyaan yang telah dipasangkan satu sama lain untuk mengetahui apakah kedua pertanyaan tersebut duplikat atau tidak.

Dari hasil penelitian diperoleh arsitektur terbaik dalam pembelajaran pasangan pertanyaan Quora, dengan nilai akurasi yang dicapai sebesar 76,01%.

Kata kunci: klasifikasi pertanyaan, LSTM, RNN.

ABSTRACT

SENTENCE-PAIR CLASSIFICATION USING LONG SHORT-TERM MEMORY RECURRENT NEURAL NETWORK

By:

Dsk Feby Inka Putri

13/347478/PA/25256

In a question and answer site, a common problem arises when one or more other users ask the same question in different discussions.. This will be an issue in the efficiency of a question and answer site and the use of large storage space for the same information. Questions with the same intentions will spread across different discussion forums, which would be better if they are combined in a one canonical page and it will maximize answers by other users.

The initial solution to the problem is to build a system that performs automatic detection of duplicate pair of questions. Where in this study, modeling on natural language is done to find sentences that duplicate one another by using Long Short-Term Memory Recurrent Neural Network. The data used in this study is the dataset released by the Quora question and answer site, where the data consists of questions that have been paired with each other to see whether the two questions are duplicate or not.

From the research results obtained the best architecture in learning Quora questions pair, with the value of accuracy achieved by 76,01%.

Keyword: sentence classification, LSTM, RNN