

Pemrosesan teks menjadi suara menggunakan model tacotron2 berdasarkan sataset rekaman podcast dan talk show berbahasa Indonesia = Text-to-speech processing using the tacotron2 model based on a recorded dataset of Indonesian podcast and talk show

Martin Novela, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20525170&lokasi=lokal>

Abstrak

Salah satu faktor keberhasilan suatu model pembelajaran dalam machine learning atau deep learning adalah dataset yang digunakan. Pemilihan dataset yang digunakan untuk pelatihan suatu model sangat penting dalam pemrosesan yang dilakukan. Beberapa model deep learning seperti sistem pemrosesan text-to-speech pada model Tacotron2 sangat bergantung dengan dataset yang dipilih. Hal tersebut dikarenakan dalam pelatihan yang dilakukan untuk pembentukan suatu model pada bahasa tertentu harus menggunakan bahasa tersebut agar kualitas dari suara sintesis yang dihasilkan dapat mendekati pelafalan yang seharusnya untuk bahasa yang ditentukan. Pada penelitian ini dilakukan pelatihan yang bertujuan untuk pembentukan model Tacotron2 pada Bahasa Indonesia, sehingga untuk mendapatkan hasil yang lebih baik, maka digunakan dataset berbahasa Indonesia dari rekaman podcast dan talk show. Pada penelitian ini dilakukan dua skenario untuk mendapatkan hasil yang paling optimal. Skenario pertama adalah melakukan pelatihan pada model Tacotron2 menggunakan dataset berbahasa Indonesia yang terdiri dari banyak pembicara. Skenario kedua adalah melakukan pelatihan pada model Tacotron2 menggunakan dataset berbahasa Indonesia yang terdiri dari satu pembicara. Dalam pemrosesan pelatihan untuk pembentukan model Tacotron2 pada Bahasa Indonesia dilakukan dengan menggunakan pre-trained dan hyperparameter berdasarkan penelitian sebelumnya (Shen, et al., 2018). Hasil dari setiap skenario dievaluasi menggunakan Mean Opinion Score (MOS). Kuesioner disebarikan kepada 25 responden untuk menilai kualitas dari suara yang terbentuk berdasarkan kriteria penilaian pada MOS. Suara yang dibentuk oleh model Tacotron2 pada skenario pertama menghasilkan nilai MOS sebesar 4.01. Sedangkan suara yang dibentuk oleh model Tacotron2 pada skenario kedua menghasilkan nilai MOS sebesar 3.78. Hal tersebut menunjukkan bahwa penggunaan dataset berbahasa Indonesia dari rekaman podcast dan talk show dalam pembentukan model Tacotron2 untuk Bahasa Indonesia menghasilkan kualitas yang cukup baik.

.....One of the success factors of a learning model in machine learning or deep learning is the dataset used. The selection of the dataset used for training a model is very important. Some deep learning models such as the text-to-speech processing system in Tacotron2 are highly dependent on the selected dataset. This is because the training for forming a model in a particular language must use the same language so that the quality of the synthetic sound produced can resemble the proper pronunciation and prosody. In this study, training is aimed to form Tacotron2 model in Bahasa Indonesia. So, to get better results, a dataset in Bahasa Indonesia from recorded podcasts and talk shows are developed. Based on the dataset, two scenarios of training were conducted to obtain the most optimal results. The first scenario is to conduct training on the Tacotron2 model using the dataset consisting of many speakers. The second scenario is to conduct training on the Tacotron2 model the dataset consisting of only one speaker. Both scenarios employed the Tacotron2 model and the hyperparameters used in previous work (Shen, et al., 2018).

The results of each scenario are evaluated using the Mean Opinion Score (MOS). Questionnaires are

distributed to 25 respondents to assess the quality of the sound formed based on the MOS assessment criteria. The sound formed by the Tacotron2 model in the first scenario produces an MOS value of 4,012. While the sound formed by the Tacotron2 model in the second scenario produces an MOS value of 3.78. This shows that the use of Indonesian language dataset from recorded podcasts and talk shows in the formation of the Tacotron2 model for Bahasa Indonesia produces a good quality.