

Intent Classification for Bank Customer Queries

Yumi Heo
230003122
MSc Data Science
yumi.heo@city.ac.uk

Intermediate results and implementation details

When we first created the baseline of LSTM, the model was not learning at all. When we looked for the cause, we found that the maximum length given for padding was based on the longest text length in the training data, which was around 300. After adjusting the padding value to the median value, the model was able to increase accuracy through training.

When comparing model sizes while training each LSTM model and saving the model each time it was trained, the size of the LSTM with GloVe model was always the smallest. As stated in the report, this may be because the embedding size is relatively light because it does not completely match the vocabulary dictionary, but additional research is needed.

Although DistilBERT can show higher performance by accepting and learning from longer sequences of data than LSTM, it is worth noting that it still shows high performance on preprocessed data of shorter length.

The model test was made possible to run using Google Colab's CPU without using cuda. Clearly, the test time for DistilBERT is several minutes longer than that for the LSTM model.