Word2Vec Model Usage

Run the following commands from root prior to using this model

pip install tensorflow

pip install keras

pip install gin-config

pip install tensorflow-datasets

pip install tensorflow-addons

pip install pandas

pip install gensim

pip install sentencepiece

Note: For tensorflow, the version required is 2.2.

1. Preparation of Dataset
   - Navigate to data/final3.csv in Word2Vec folder.
   - Any new command should be inserted as %Input text,%Intent Number(Comma separated)
   - Once the dataset is modified, navigate to /train folder inside Word2Vec folder and delete all the folders present inside.
   - Open terminal from root of Word2Vec folder and run $python data.py.
   - This will generate %Number_of_Intents folders inside the train directory.
2. Training the model
   - From the root run $python text_classification.py
   - The model is saved as "model.tflite" in the root and also generates 2 more files : vocab.txt and labels.txt, Which are required for inference and testing.
   - To change the max sequence length of the input text, open text_classification.py and change the seq_len parameter in the line:
     "model_spec = AverageWordVecModelSpec(seq_len=10,)".
     Further you can alter the embedding dimension and dropout_rate paramters by adding in the same line as follows:
     "model_spec = AverageWordVecModelSpec(seq_len=10,wordvec_dim=16,dropout_rate =0.2)".
     Shown are the values used now.
3. Testing the model
   - Navigate to data/test.csv and edit/add/remove any test cases.
   - Run $python testing.py from root.
   - Testing Accuracy will be printed on screen.

The above testing script uses three files during inference and hence these should be imported wherever the model is used for testing.

1. model.tflite
2. vocab.txt
3. labels.txt