Irony detection

Xieyidan Abuliezi

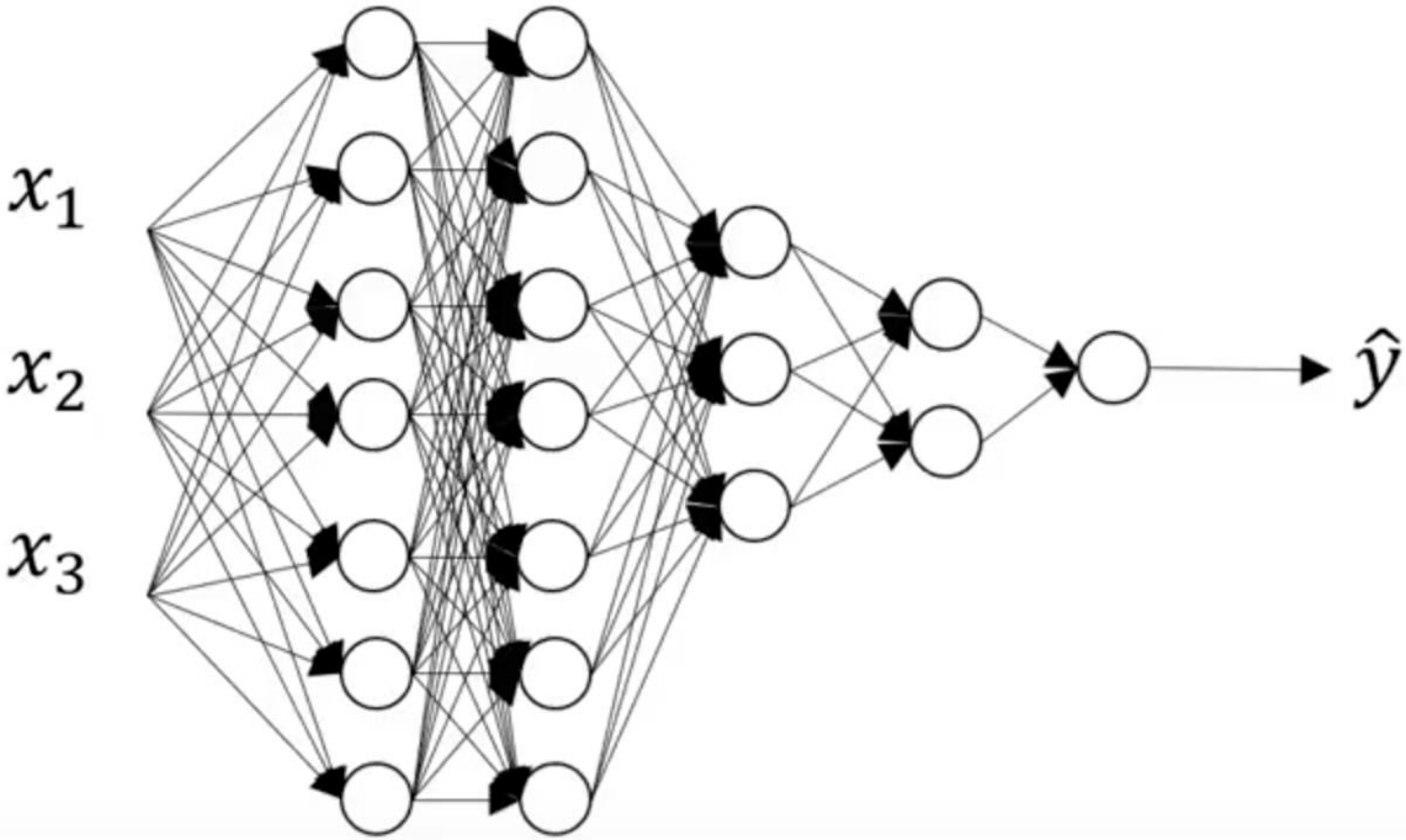
Corpus

- - Given: One larger data set (tweets, SemEval-2018 Task 3), one smaller data set (reddit comments, Kaggle) for irony detection.
- - Task: Predict whether comments are ironic (e.g. LSTM+logistic regression). How can one domain (large data set) help prediction on another domain (smaller dataset)?
- hier (Task A: ironic vs non-ironic):
- https://competitions.codalab.org/competitions/17468
- und hier:https://www.kaggle.com/rtatman/ironic-corpus
- - Effect of Dropout/SpatialDropout1D: Make model more robust by removing words from the input during training.
- - Effect of pre-trained word embeddings.

Auxiliary Library

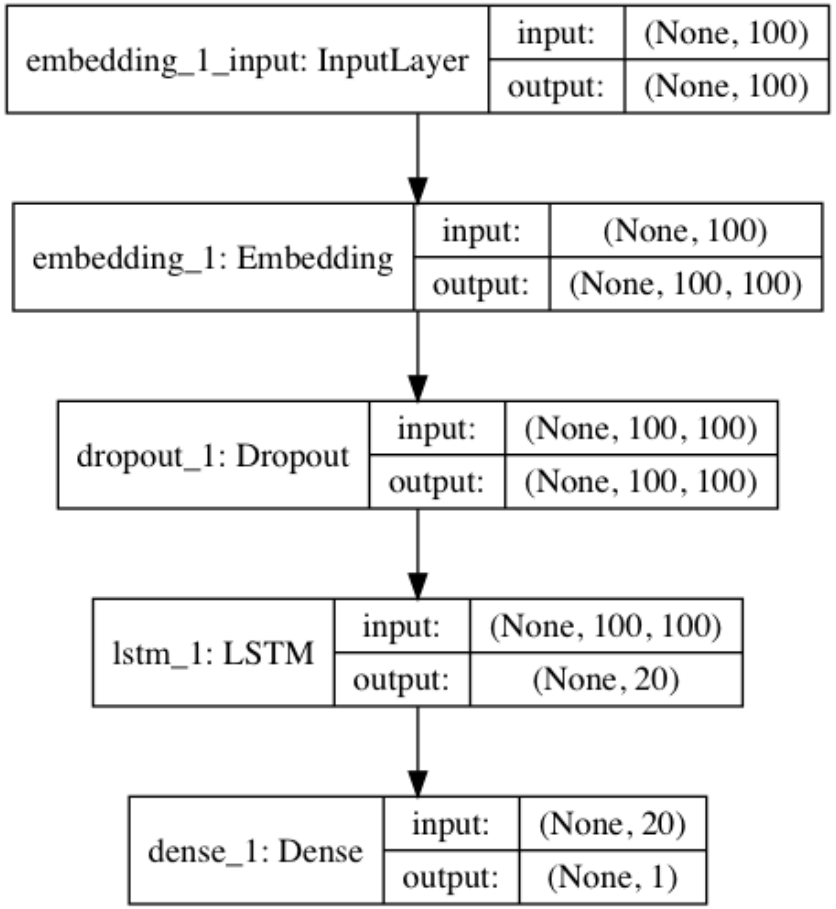
- GloVe: Twitter (2B tweets, 27B tokens, 1.2M vocab, uncased, 25d, 50d, 100d, & 200d vectors, 1.42 GB download): glove.twitter.27B.zip
- external word embedding weights.
- TwitterTokenizer from nltk
- Train_test_split from Scikit Learn

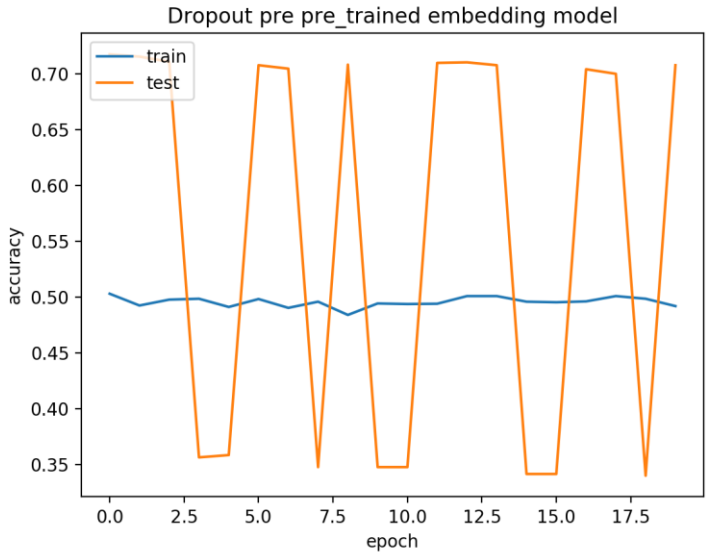
Dropout



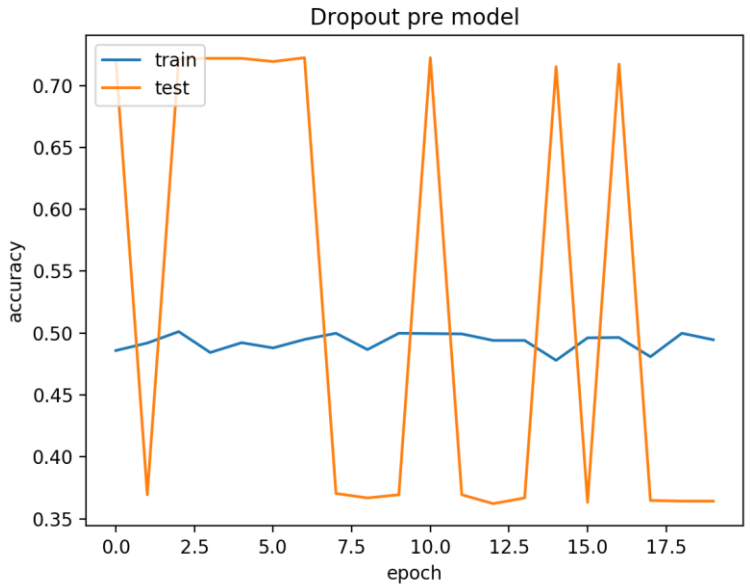
Dropout on input layers

Dropout





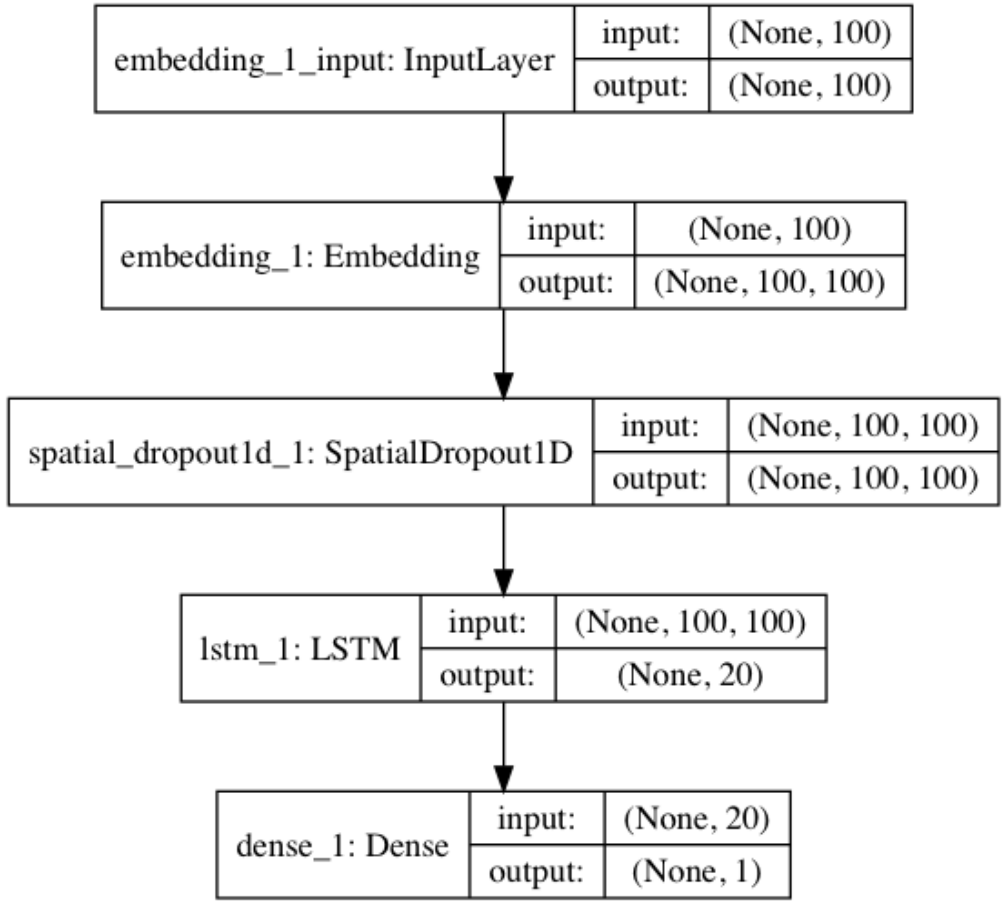
Dropout(0.2)

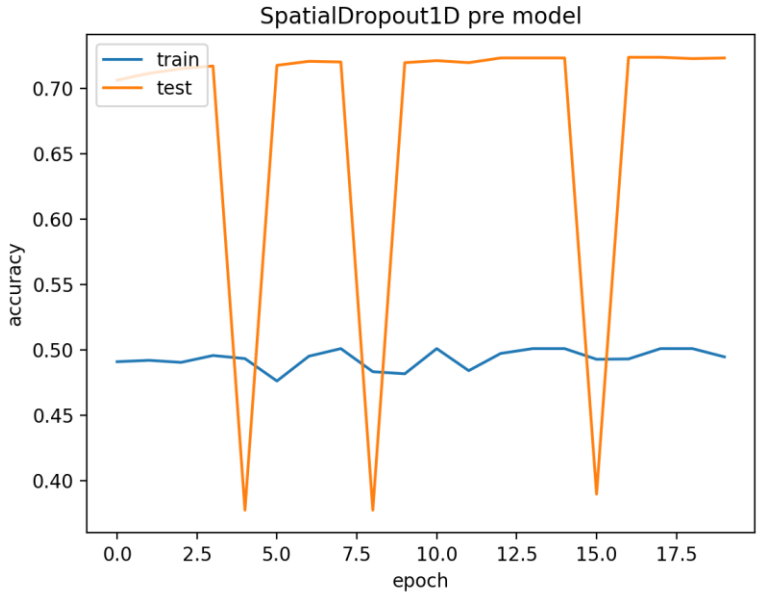


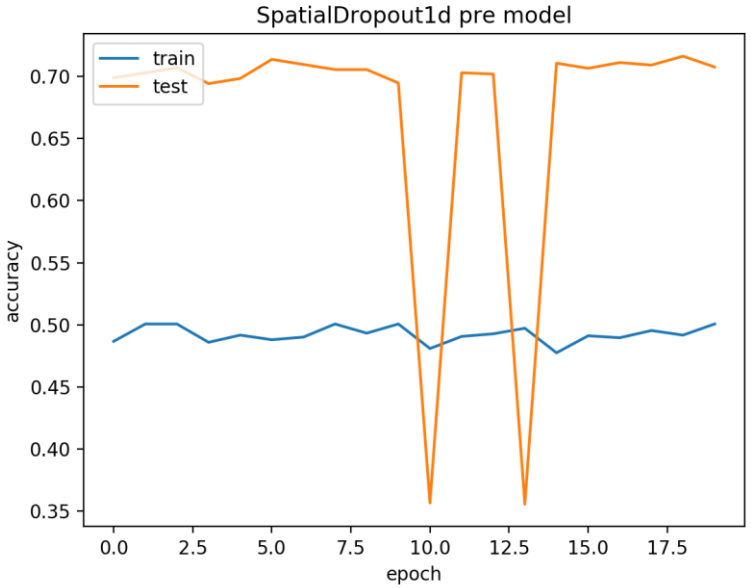
Dropout(0.5)

Dropout on input layers

SpatialDropout1D Mode

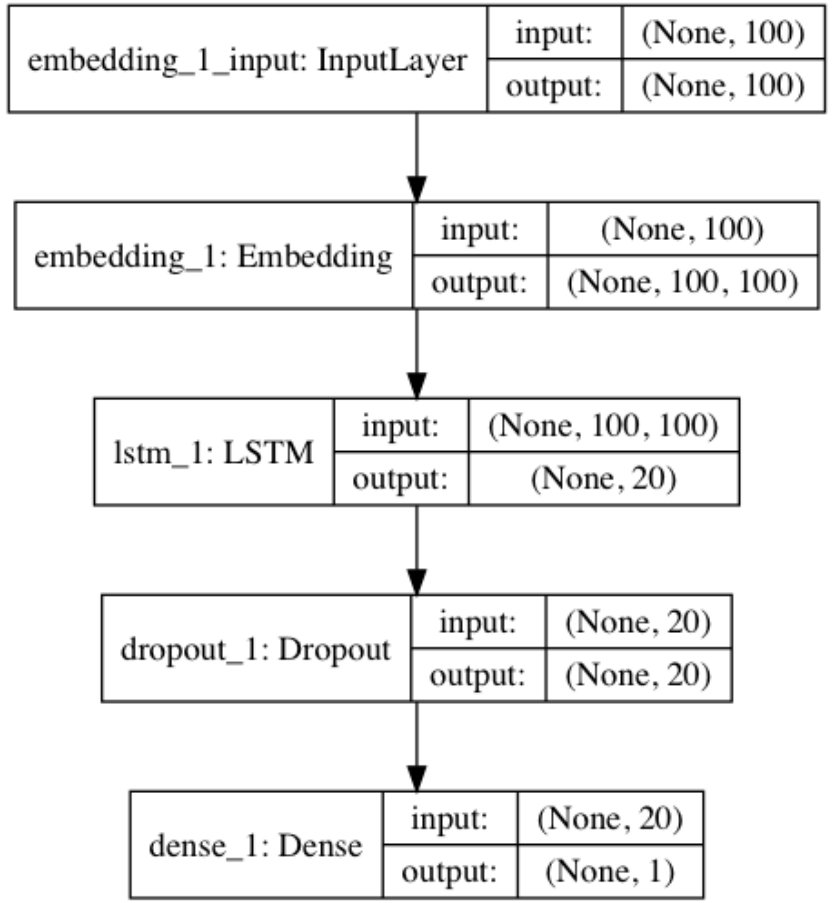


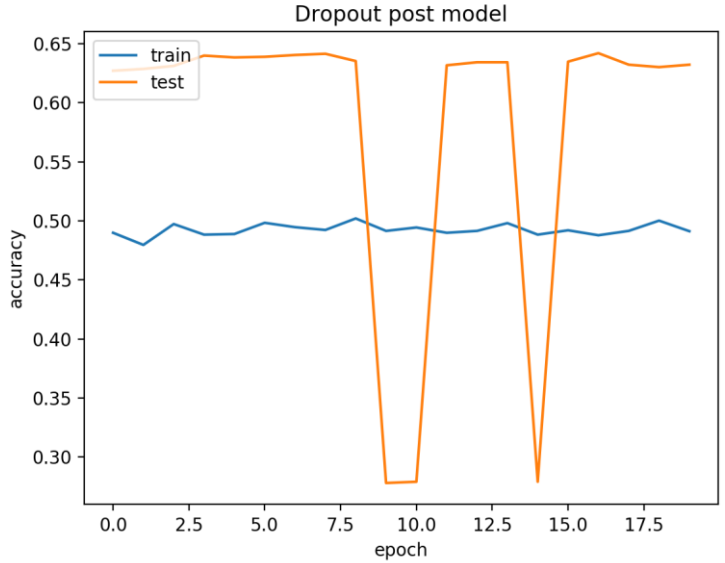


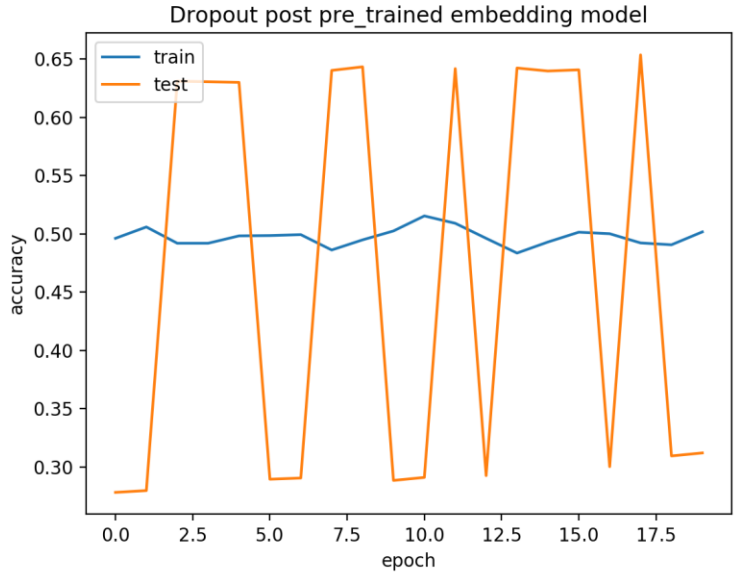


Dropout on Hidden Laye

Dropout Model

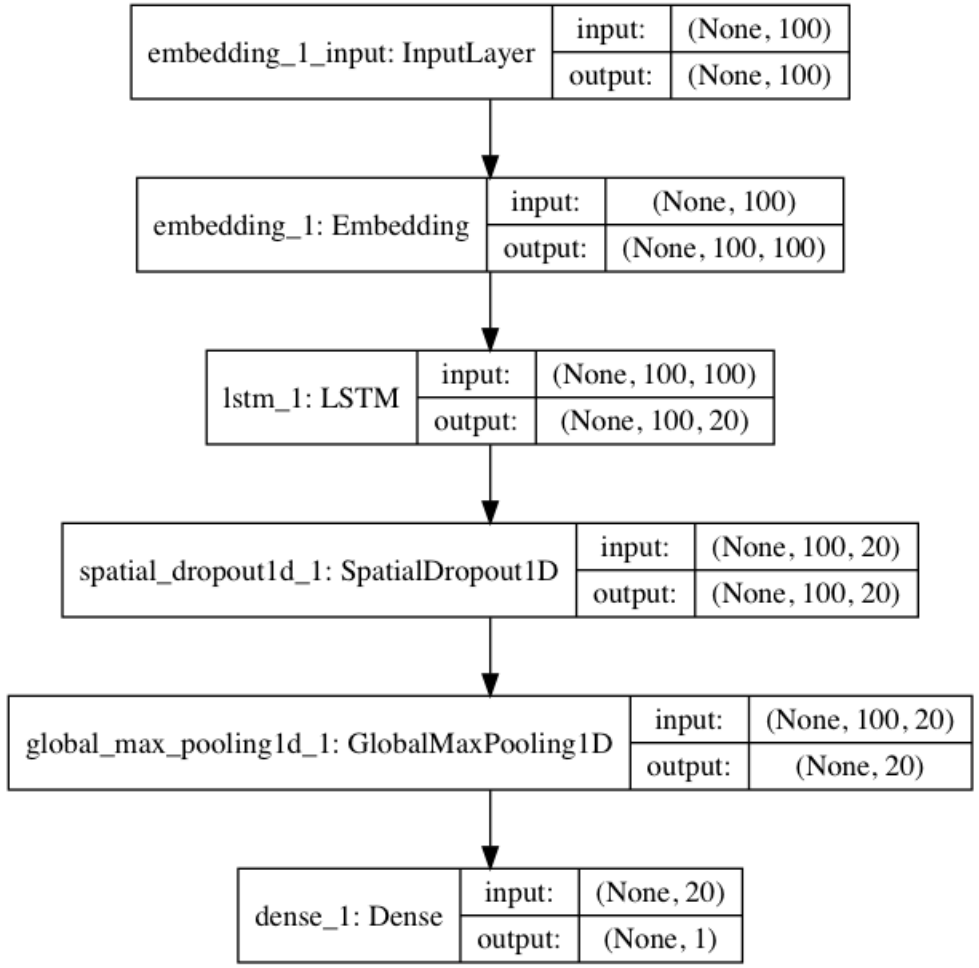


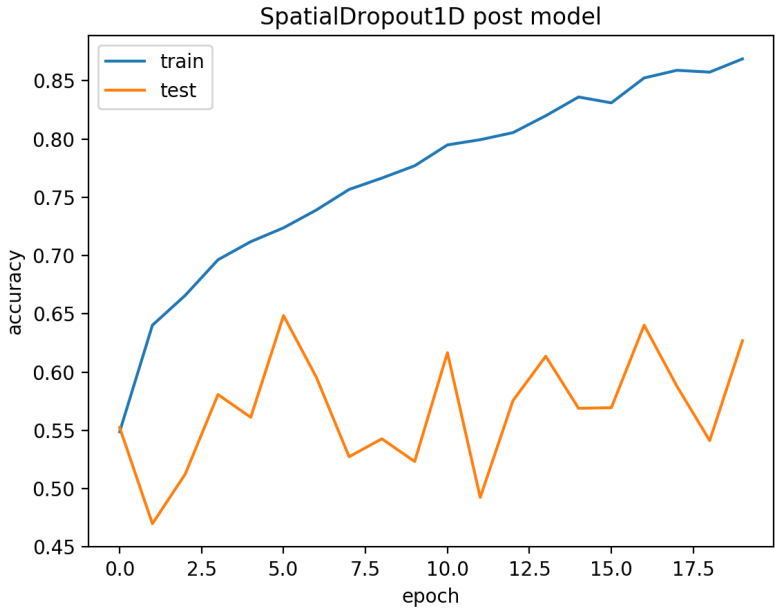


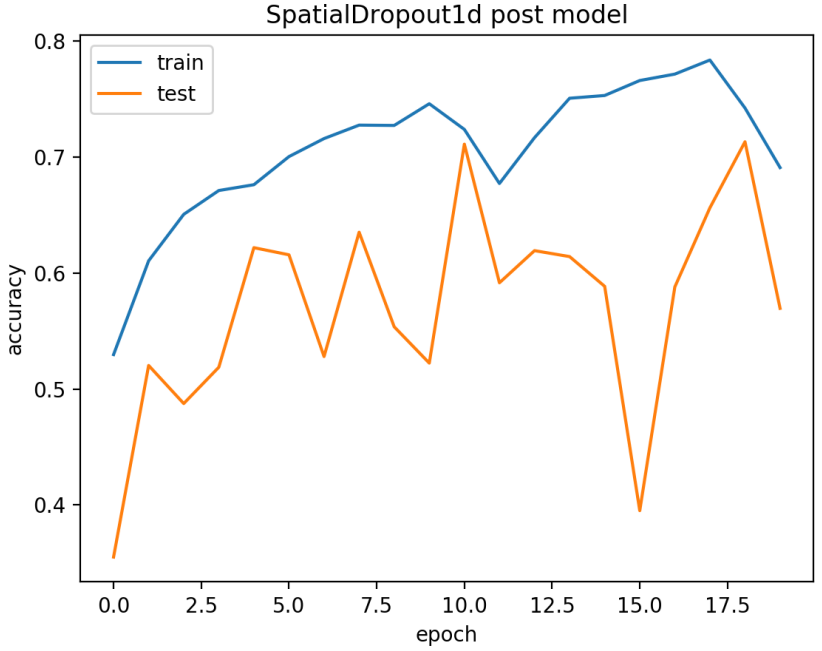


Dropout on hidden layers

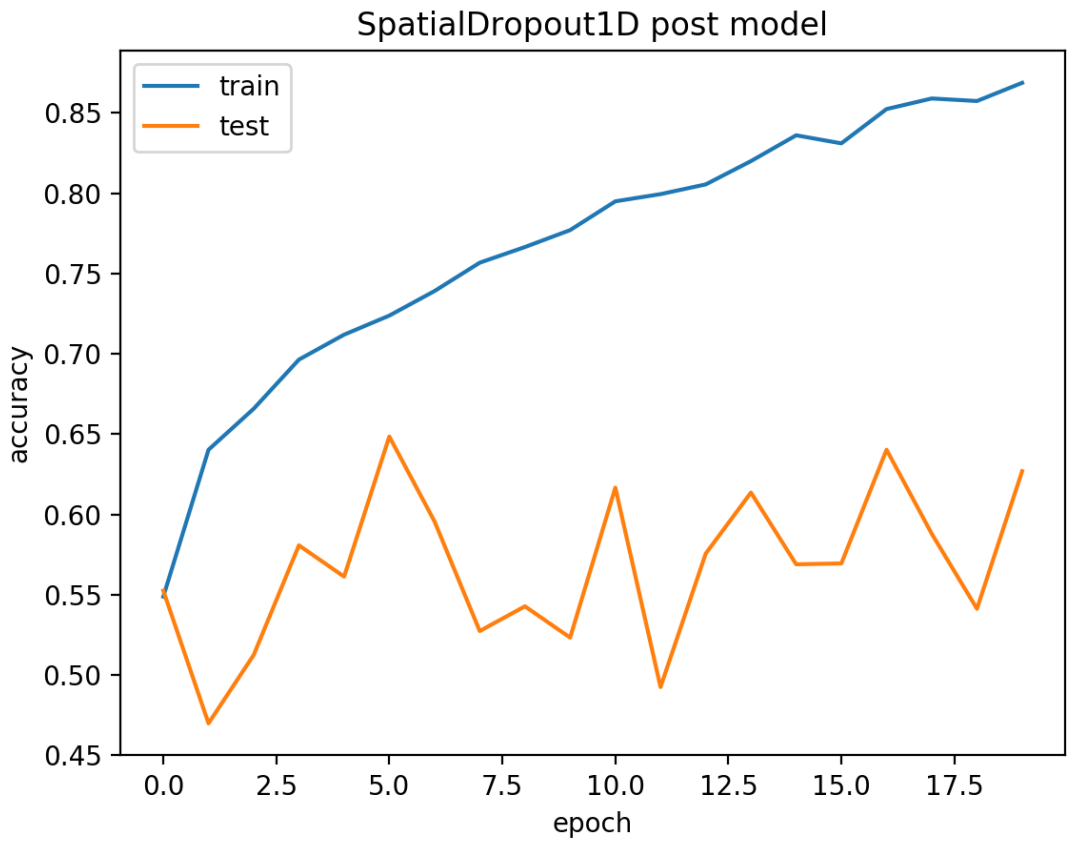
SpatialDropout1D Mode

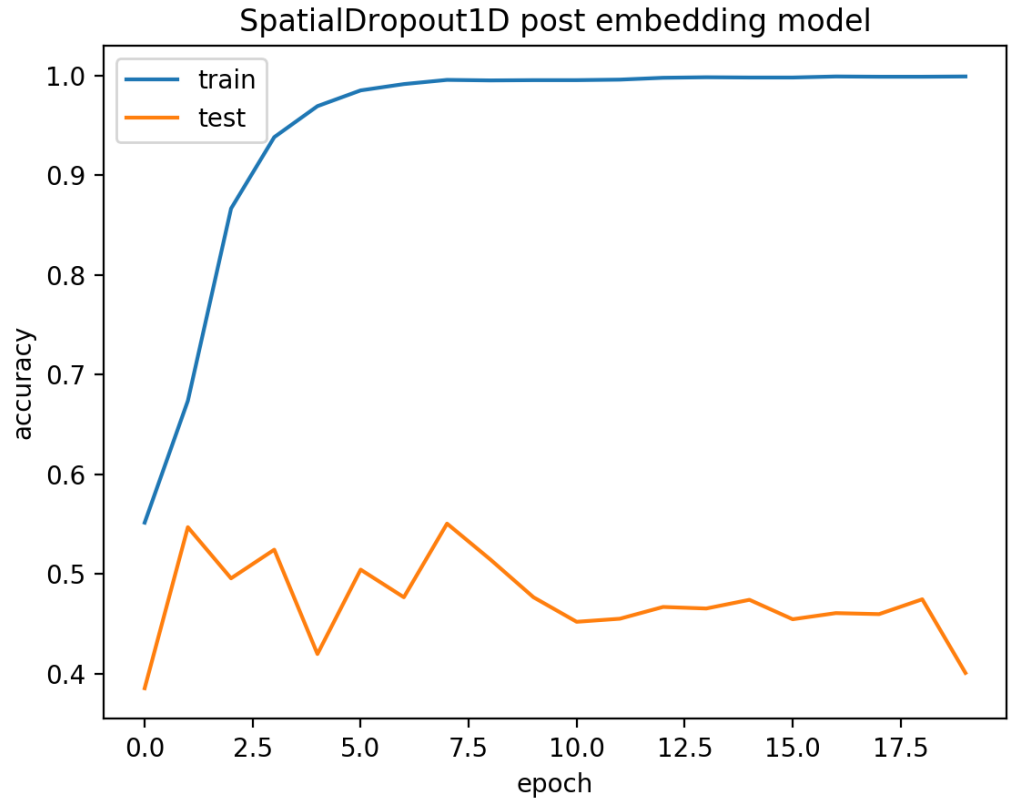




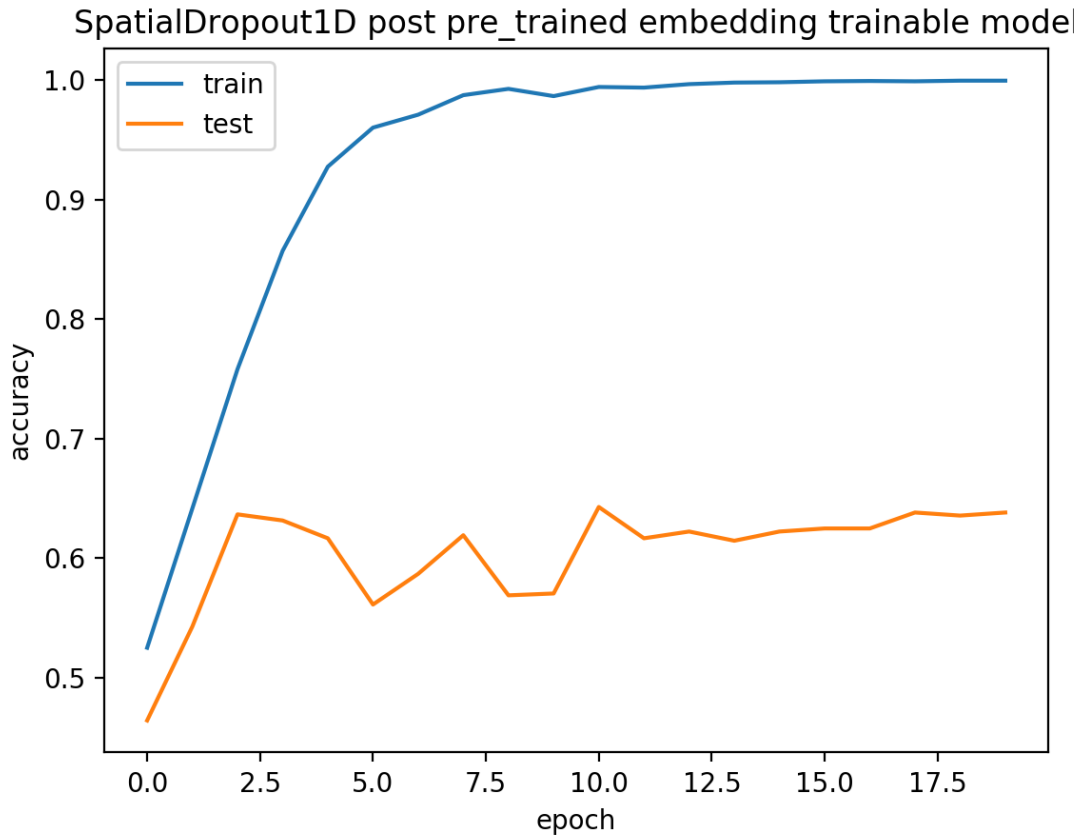


pre_trained embedding/ embedding





Best Result



Dropout(0.5)

```
    if embedding_matrix is not None:
        model.add(Embedding(vocab_size, embedding_size,
                            weights=[embedding_matrix],
                            input_length=MAX_LEN,
                            trainable=False))
    else:
        model.add(Embedding(vocab_size, embedding_size))

    if regularization=='pre':
        if DropoutType=='Dropout':
            model.add(Dropout(0.5))
            model.add(LSTM(hidden_size))
        elif DropoutType=='SpatialDropout1D':
            model.add(SpatialDropout1D(0.5))
            model.add(LSTM(hidden_size))


    elif regularization=='post':

        if DropoutType=='SpatialDropout1D':
            model.add(LSTM(hidden_size,return_sequences=True))
            model.add(SpatialDropout1D(0.5))
            model.add(GlobalMaxPool1D())

        elif DropoutType=='Dropout':
            model.add(LSTM(hidden_size))
            model.add(Dropout(0.5))

    elif regularization=='None':
        model.add(LSTM(hidden_size))


    model.add(Dense(1, activation='sigmoid'))


    model.compile(loss='binary_crossentropy',
                  optimizer='adam',
                  metrics=['accuracy'])
```