SEQUENCE MODELING FOR TWEET CLASSIFICATION: DETECTING PERSONAL HEALTH MENTIONS USING LSTM AND BI-LSTM NETWORKS

COURSE : CSC4093-NEURAL NETWORKS AND DEEP LEARNING
NAME : T.S.W.PATHIRANA
REG. NUMBER : S19836

1. Introduction

With the rapid growth of social media, platforms like Twitter serve as major sources for sharing real-time information, including personal health experiences. Automatically classifying tweets as or not as personal health mentions (PHM) can greatly support public health surveillance, disease tracking, and research. In this project, I implemented and compared two deep learning architectures "LSTM" and "Bi-LSTM" for binary tweet classification. The analysis focuses on model performance, computational efficiency, and overfitting behavior, with a particular focus on practical aspects of model implementation and evaluation using Keras.

2. Data and Methodology

2.1. Dataset

- **Training Data** : phm_train.csv (9991 labeled tweets)
- **Test Data** : phm_test.csv (3331 labeled tweets)
- **Columns** : tweet_id , label, tweet
- **Methods** : Binary classification [PHM(1) vs. non-PHM(0)]
- **Preprocessing** : Removed HTML tags and anything non-alphabetic, shifted all characters to lower case, and removed stopwords.
- **Special Implementation** : The validation set was created using the validation split=0.2 argument in Keras during model training, which automatically allocates 20% of the training data for validation. This approach ensures efficient tracking of model generalization without manual data splitting but assumes that the data is shifted to avoid bias.

2.2. Model Architectures

LSTM: A recurrent neural network aimed to deal with the vanishing gradient problem present in traditional RNNs.

- Embedding layer – 32 dimensions
- Single LSTM layer – 64 units
- Dense output layer with sigmoid activation for binary classification

Model: "sequential_16"

Layer (type)	Output Shape	Param #
embedding_16 (Embedding)	(None, 10, 32)	405,120
lstm_16 (LSTM)	(None, 64)	24,832
dense_16 (Dense)	(None, 1)	65

Total params: 430,017 (1.64 MB)
Trainable params: 430,017 (1.64 MB)
Non-trainable params: 0 (0.00 B)

Bi-LSTM: An extension of LSTM that processes input sequences in both forward and backward directions, potentially capturing more context.

- Embedding layer – 32 dimensions
- Bidirectional LSTM layer – 64 units per direction
- Dense output layer with sigmoid activation

Model: "sequential_17"

Layer (type)	Output Shape	Param #
embedding_17 (Embedding)	(None, 10, 32)	405,120
bidirectional_8 (Bidirectional)	(None, 128)	49,664
dense_17 (Dense)	(None, 1)	129

Total params: 454,913 (1.74 MB)
Trainable params: 454,913 (1.74 MB)
Non-trainable params: 0 (0.00 B)

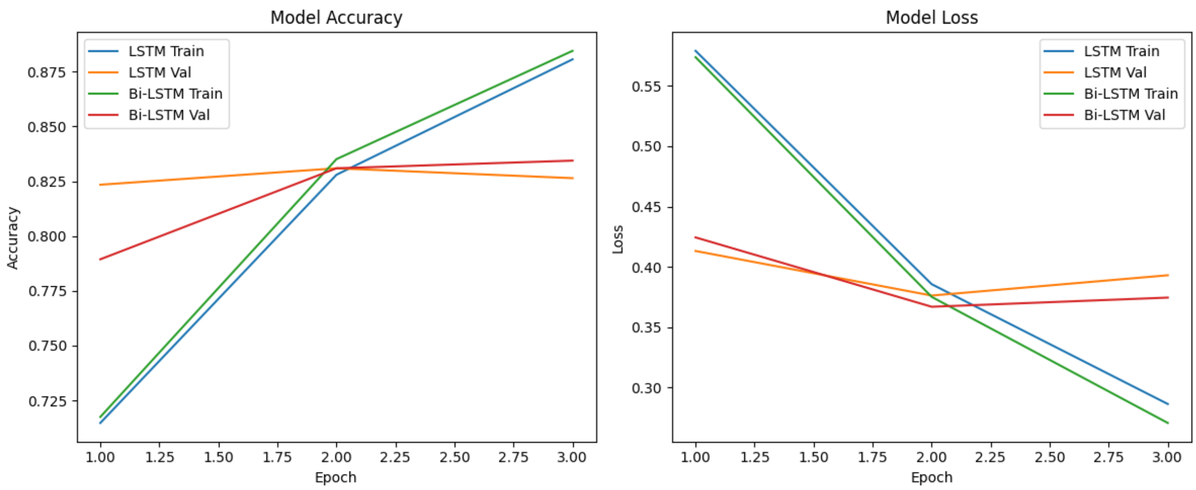
3. Performance Comparison & Discussion

3.1. Performance Comparison Table

Model	Correct Predictions	Wrong Predictions	Test Accuracy	Test Loss
LSTM	2685	646	80.61%	0.4316
Bi-LSTM	2704	627	81.18%	0.4310

3.2. Performance Plots

Training and validation accuracy (left) and loss (right) across epochs for both LSTM and Bi-LSTM models.



3.3. Discussion on Model Performance

- The Bi-LSTM model achieved a slightly higher test accuracy (81.18%), correctly predicting 2704 tweets, while LSTM model achieved (80.61%) with 2685 correct predictions.
- Test loss for Bi-LSTM (0.4310) is marginally lower than that of LSTM (0.4316), indicating slightly better generalization.
- Bi-LSTM made fewer wrong predictions (627) than LSTM (646)

Training and Validation:

LSTM Training:

- Training accuracy increased from 0.6752 to 0.8807 over 3 epochs.
- Validation accuracy peaked at 0.8309 in epoch 2 and then decreased slightly.
- Training loss decreased from 0.6360 to 0.2886.
- Validation loss reached its lowest level in epoch 2 (0.3763) and then increased, indicating slight overfitting.

Bi-LSTM Training:

- Training accuracy increased from 0.6863 to 0.8952 over 3 epochs.
- Validation accuracy gradually improved to 0.8344 in epoch 3.
- Reached 0.8344 in epoch 3.
- The epochs decreased from 0.6268 to 0.2604.
- The validation loss dropped to 0.3670 in epoch 2, and then increased slightly to 0.3746 in epoch 3.

3.4. Interpretation:

Both models generalize well to the test set, with Bi-LSTM achieving a slight advantage (+0.57% accuracy) and a slightly lower loss. This improvement is consistent with its higher validation accuracy during training.

4. Conclusion

Both LSTM and Bi-LSTM models learned effectively to classify tweets as personal health mentions, achieving over 80% of test accuracy. Bi-LSTM provided a slight improvement in accuracy (0.57%) at the cost of significantly higher computational resources. The validation loss and accuracy trends show that both models start to overfit after the second epoch, suggesting that further regularization or early stopping may be beneficial. For this task, the LSTM model offers a strong balance between accuracy and efficiency, and Bi-LSTM may be preferred if the highest possible accuracy is desired, and resources allow.