Yu Lu Liu

Final Project MAIS 202 Deliverable 3

Final Results

In my preliminary results, the model achieved a F1 value of 0.77 on the test set, despite scoring very well on the training set and moderately well on the validation set. So, as proposed in the conclusion of Deliverable 2, I would first attempt to address the issue of <u>overfitting</u> by varying two hyperparameters:

1. <u>Embedding dimension</u>
   I explored several values: 96, 48, and 12. Overall, the variation of the embedding space did not affect the results significantly, within only around 0.02%. Indeed, an embedding dimension of 12 achieved a F1 value of 0.79 on the test set, so I preserved this value for the final model.

2. <u>Dropout rate</u>
   In my initial model, the dropout rate was 0.2. I tried increasing it to 0.3 and 0.4, but just like with the embedding dimension, this hyperparameter variation did not affect the performance of the model and even weakened it at a dropout rate of 0.4. The rate of 0.2 is thus kept unchanged.

The final solution for overfitting is <u>early stopping</u>. I observed that, while training the model, the validation accuracy starts decreasing towards the 5th iteration. I thus set the final epoch number to 5, instead of 25 epochs from the initial setup.

Second, I would like to increase the <u>accuracy</u>. After exploring several kernels on Kaggle working on the same dataset, I realized that LSTM's performance rarely exceeded 0.85 accuracy, but that Bidirectional LSTM, which visit the input sequences in both directions, generally performed better. I thus decided to use a <u>Bidirectional LSTM</u> for the final model. Furthermore, I doubled the <u>size of dataset</u>. Upon discovering that the two versions of News Headlines Dataset for Sarcasm Detection from R. Misra and A. Prahal (2019) were not overlapping, I merged them to create a dataset double the size of the initial one.

Following these modifications, the final model achieved:

	Preliminary Results	Final Results	Improvement
Precision	83.2%	88.0%	+ 4.8%
Recall	72.4%	88.0%	+ 15.6%
Accuracy	81.4%	89.1%	+ 7.7%
F1	76.6%	88.6%	+ 12.0%

Final Demonstration Proposal

For the final demonstration, I decided to implement a webapp. As proposed in Deliverable 1, the webapp would allow the user to input a news headline (or just a short sequence of words) and would display whether the input was a sarcastic statement or not.

Since I am a novice at web development, I plan to use the Flask implementation example provided in Workshop 2 as a skeleton.