Final Assignment: Analysis

Ethan Luster

12/12/2021

Introduction:

The Amazon fine foods database, which can be found at https://www.kaggle.com/snap/amazon-fine-food-reviews?select=Reviews.csv, contains over 500,000 reviews of food pulled from Amazon.com. These reviews include ratings in the form of a five-star scale and the text of the review themselves. This project will seek to build an effective text classification model that can predict whether a review is positive or negative based on the text of the review.

To convert the five-star scale into a form that can be used for classification, it was decided to treat four- and five-star reviews as positive, and one-to-three-star reviews as negative. Once the data was modified to match this, the text of the reviews was processed.

The following models represent a series of attempts to build a model with the highest possible validation efficiency.

Model 1:

The first model uses a simple RNN layer. It achieves a final validation accuracy of 72%. The model is also highly sporadic, with validation accuracy going up and down significantly epoch to epoch. The model also finishes with a validation loss of 74%. Using other types of may produce a better result.

Model 2:

This model uses a GRU layer with the same sample size. It achieves a higher validation accuracy of approximately 81%. This is measurably better and achieves a higher level of stability from epoch to epoch. The validation loss does reach 88%, which is measurably worse than even simple RNN.

Model 3:

This model uses a LSTM layer with the same training sample size. While the accuracy starts around 80%, it falls to 75% by the final model. This did not occur in most previous attempts using RNN, but that’s not what happened this run. The validation loss is also 86%.

The inconsistency from run to run is likely caused by the small sample size. Luckily, the dataset contains a significantly larger sample size that we can use. For models 4 and 5, the training data will increase from 500 reviews to 5000 reviews.

Model 4:

This model uses a GRU layer with 5000 training samples. This model achieves a validation accuracy of 84%, with the final four epochs staying at this amount. The validation loss is 42%, which is significantly better than the GRU model with a lower sample size.

Model 5:

This model uses an LSTM layer with 5000 (not 25000). The model achieves a final validation accuracy of 82.5%. This is lower than the GRU layer, and has a higher level of variability. The validation loss is 55%.

In most runs using an LSTM layer with the same sample size, the model performed slightly better and had more stability than a GRU layer. Even with a ten-fold increase in training samples, variability remaining significant.

The following model use an LSTM layer since my previous experimentation suggested in would work better. The models also use 25000 training samples in an effort to increase validation accuracy and decrease variability.

Model 6:

This model is the same as model 5, except for the increased sample size. It achieves a validation accuracy of 87%, which is higher than both the previous LSTM model and the GRU layer. The validation loss is 34%.

Model 7:

This model adds a dropout layer after the LSTM layer. The model finished with a validation accuracy of 84%. The dropout layer appears to introduce more variability, with the next to last epoch having a validation accuracy of over 87%. The validation loss is 36%.

Model 8:

The next model includes the dropout layer and an increase in the output dimension of the embedding layer. The model's accuracy rises and then falls, with a final validation accuracy of 85.5% The validation loss is 36%.

Model 9:

The final model uses the same layers as model 7, except for a dropout rate of 0.2, rather than 0.5. This model finishes with a validation accuracy of 86% percent. However, this model is still highly variable. The validation loss is 37.5%.

Summary of Results:

Model:	Validation Accuracy	Validation Loss
1	72%	74%
2	81%	88%
3	75%	86%
4	84%	42%
5	82.50%	55%
6	87%	34%
7	84%	36%
8	85.50%	36%
9	86%	37.50%

Final analysis:

Of all the models, the LSTM model without dropout (Model 6) worked best. Not only did it achieve the highest validation accuracy, but it also had the lowest loss. The dropout layers did not appear to work well as a tool to improve the performance of models with LSTM layers. Though the decrease in validation accuracy and increase in loss what no more than a few percentage points, the variability from epoch to epoch was significant.

As the dataset of 500,000 reviews is significantly more than the 25,000 reviews used in the final models, there is significant room for improvement of the model from this point. The greatest single factor in improving model performance is the sample size available to train the model on. The second biggest factor is the layer used after the embedding layer, with the LSTM layer achieving higher consistency in most runs. With a model using all the reviews, it is likely this variability would all but disappear.