

Predicting hotel score by review using LSTM and Transformer

Setup

The data might be found [here](#). The data consists of 100k reviews and scores collected from 1500 hotels. Each review includes two texts: positive and negative aspects. The column score may have a real value from 0 to 10. In the presented data, the lowest score is 2.5.

Goal: get the lowest possible MAE for score using only review texts.

LSTM

First, I use a bi-directional LSTM model with 2 layers, and an additional Dropout layer setting 70% of the output to zero to prevent overfitting with L1 Loss from PyTorch. I did not find any gain in performance from separately learning embedding layers for “negative“ and “positive“ columns. Thus, I combine them into one column.

Secondly, I use a bi-directional 2-layer LSTM model with an Attention module and Dropout layer, setting 85% of the neuron output to zero. Again, feeding columns “negative“ and “positive“ separately did not improve the performance but increased the time.

I trained both models for 30 epochs. The result for a regular LSTM is below:

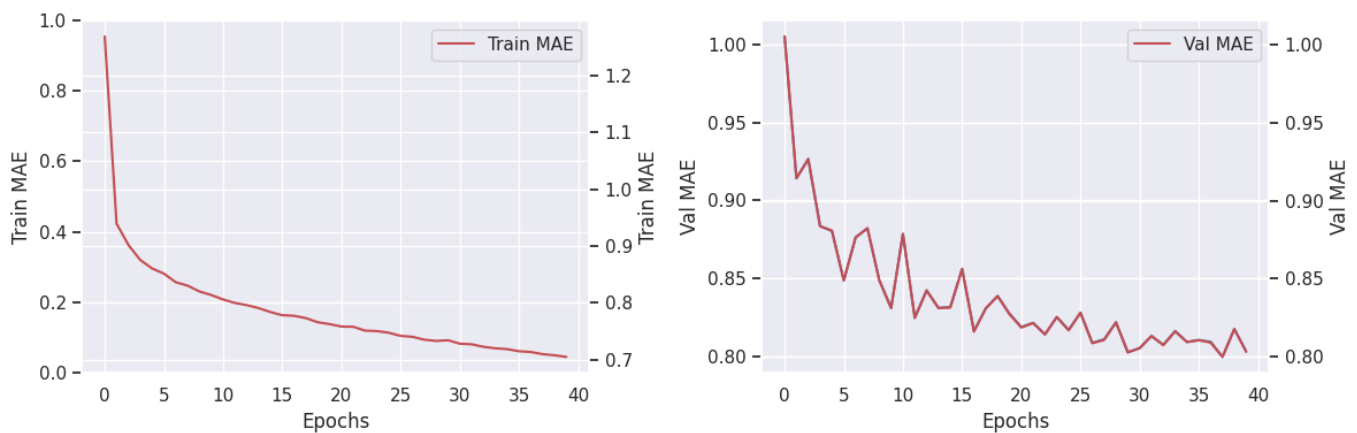


Figure 1: LSTM

Clearly, the model overfits slightly; the MAE for the validation set reaches 0.800, which is 8% error. The model reaches its peak performance after 30 epochs. Next, I present the results for LSTM with an Attention layer:

As a result, the obtained MAE for the validation set is 0.7288, which is 7.29% for the 1 to 10 scale score. Notably, the model with a training set of 5k reviews resulted in 0.8128 MAE. Perhaps increasing the size of the training set would drop MAE further.

It is important to emphasize that I did not add any additional layer at the end and did not modify the LLaMA-3 model to use it as a regressor to predict the score. Thus, the model, contrary to the transformer, predicted the score as a combination of tokens that do not have a regular mathematical order. Since during fine-tuning the performance on the training and validation set is presented as cross-entropy, it is unclear how to translate it into MAE or MSE. For that reason, I do not report the train and validation loss function change.

Conclusion

In this exercise, I used an LSTM, a fine-tuned Transformer, and a fine-tuned LLM to predict the score based on a hotel review. Results are:

- The lowest MAE is reached using a fine-tuned LLM: 7.29% for 1 to 10 scale score
- LSTM + Attention takes fewer epochs to reach the lowest MAE compared to regular LSTM, but tends to overfit rather fast
- Fine-tuned Transformer seems to be the best candidate for the task:
 - 0.7375 MAE - better than LSTM and not that far from LLM
 - fine-tuning is way faster than for LLM