Yelp Review Rating Prediction

Daniel {Palamarchuk, Udekwe, Sabanov}, Amy Hilla
Final Project Group 11

Team Members

- Daniel Palamarchuk d4n1elp@vt.edu
- Daniel Udekwe daudekwe@vt.edu
- Daniel Sabanov sdaniel19@vt.edu
- Amy Hilla ahilla@vt.edu

Outline

- Problem description
- Approach
- Results
- Lesson learned
- Future work

Problem Description

- Over 244.4 million reviews posted for 5 million businesses as of 2021 (via Yelp-Press)
- Many "garbage" reviews with no useful content
- Many reviews where the content of the review does not match the number of stars

Proposed Solution:

We propose to use sentiment analysis to predict the number of stars based on the content of the review, allowing for an automated way to remove "garbage" reviews, improving user experience.

Approach

- Bidirectional Encoder Representations from Transformers model (BERT)
  - Similar to GPT and ELMo, has unique method of capturing context (bidirectional)
- Output continuous values between 1 and 5 stars
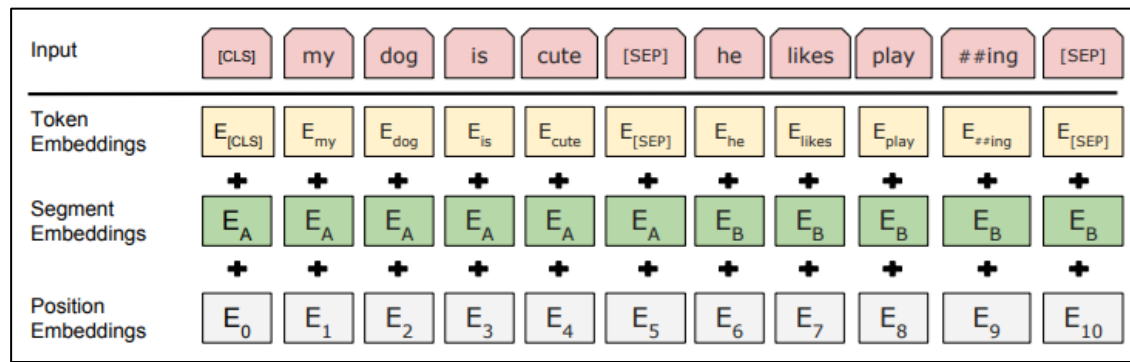
BERT input representation, Devlin et al. 2018

Approach

Data Composition:

- Reviews with emotional language
- Star count

Stars	% of Dataset
5	46.23%
4	20.78%
3	9.90%
2	7.79%
1	15.31%

Recommended Reviews

Search reviews

Yelp Sort   Date   Rating   Elites

English 859

★★★★★

First off, thank you mr. Palomino for creating all day happy hour.

I love the thought of being able to relax and grab a drink at a fancy place at anytime I wish - no matter what time of day!
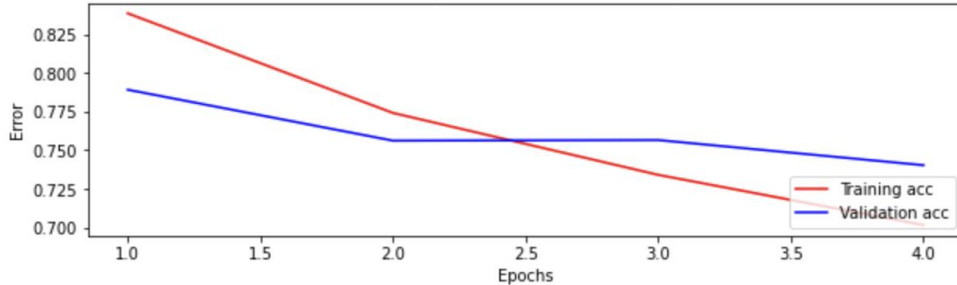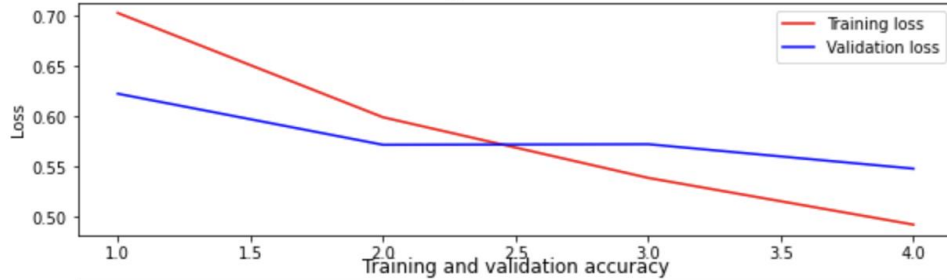
Since I have had many opportunities to pop in, I have pretty much tried everything on your HH menu.
My top 2 faves so far...
Basil gimlet and the crab dip!
It doesn't disappoint!

Was this review ...?

Useful   Funny   Cool

Results

Training and validation loss

Training and validation accuracy

- We trained the model for 4 epochs
- RMSE of 0.738 for the test data
- RMSE of 0.74 for the training data on the last epoch

{'loss': 0.5445756912231445, 'root_mean_squared_error': 0.7379537224769592}

Lessons Learned

```
...  Training model with https://tfhub.dev/tensorflow/small_bert/bert_en_uncased_L-4_H-512_A-8/1
     Epoch 1/4
       256/34952 [................................] - ETA: 2:20:16 - loss: 7.8833 - root_mean_squared_error: 2.8077
```

Original training run

- Computational limitations
  - We had to reduce the size of the dataset
    - Ultimately ended up using 10% of total dataset, and with all the ratings having an equal count
  - Using Colab vs. running locally
  - Caused us to max out on how much improvement could be made in terms of processing speed
- Difficulty with TensorFlow dependencies
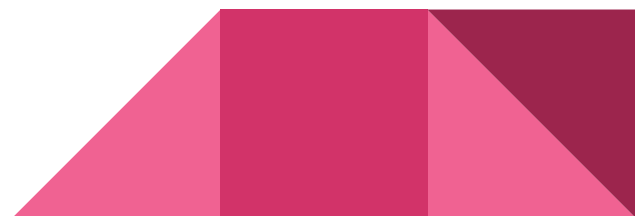
Future Work & Improvements

- Going to try to round to the nearest star and make a confusion matrix
- Look at how much review length affects accuracy
  - Longer reviews are likely to be more emotional and could thus have higher polarity
- Try to get access to a machine with greater processing power

Future Work

- Add more attributes to the model
  - Review Ranking by other users
  - Previous user reviews

Sources

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," CoRR, vol. abs/1810.04805, 2018. [Online]. Available: http://arxiv.org/abs/1810.04805

J. Morris, "Does model size matter? a comparison of bert and distilbert," 2022. [Online]. Available: https://wandb.ai/jack-morris/david-vs-goliath/reports/Does-Model-Size-Matter-A-Comparison-of-BERT-and-DistilBERT%E2%80%93VmlldzoxMDUxNz

Turc, Iulia, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Well-read students learn better: On the importance of pre-training compact models." arXiv preprint arXiv:1908.08962 (2019).

"Yelp Open Dataset," *Yelp*. [Online]. Available: https://www.yelp.com/dataset.

"Classify text with Bert," TensorFlow, 16-Feb-2023. [Online]. Available: https://www.tensorflow.org/text/tutorials/classify_text_with_bert.