

Prediction of customer star ratings from text reviews using Transformer

1. Motivation and idea

Online user ratings are important to improve service of a company and help users in their decision making before they buy a product. Customer review systems of today ask for two inputs: the text comments and the star rating. We see two potential problems here – i) some customers give ratings which may not be in coherence with their comments; and ii) the star rating as a user input is redundant because they can be influenced by the star ratings given by other customers which are publicly visible on the website and might not reflect the genuine opinion of the user. Deep learning based language models have achieved wide success in text classification. We propose the prediction of user rating (1 to 5) directly from customer reviews (text) using Transformer based neural network architecture [1]. Transformer models are based on self-attention mechanism that we believe to be particularly well suited for language understanding and the problem in hand.

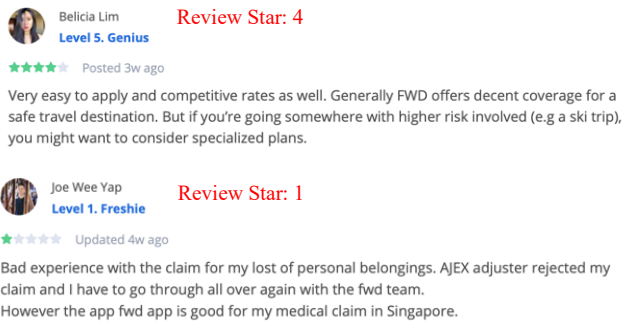


Fig 1. Review (input) and star rating (output)

We have a novel dataset which is real and targeted to customers in Singapore. The dataset is scraped from Seedly (refer to Fig 1), a community platform where customers can post their reviews as per product. Since the data scraped from Seedly is small in size (5,000 text reviews), we shall first train the model using the Yelp dataset (280,000 text reviews, which is public [2]). We shall compare the performance measures of our model with the already existing models for prediction.

2. Methodology

Data Acquisition: We plan to use API to scrap the dataset from [Seedly](#)

Data Exploration: We shall use visualization tools like matplotlib, ggplot and pandas to visualize a sample of the dataset to make sure that the data is consistent or if skewed, to be aware of when we look at the model accuracy.

Model: We shall use Transformers to build the classifier and also compare it with baseline models like BoW and LSTM.

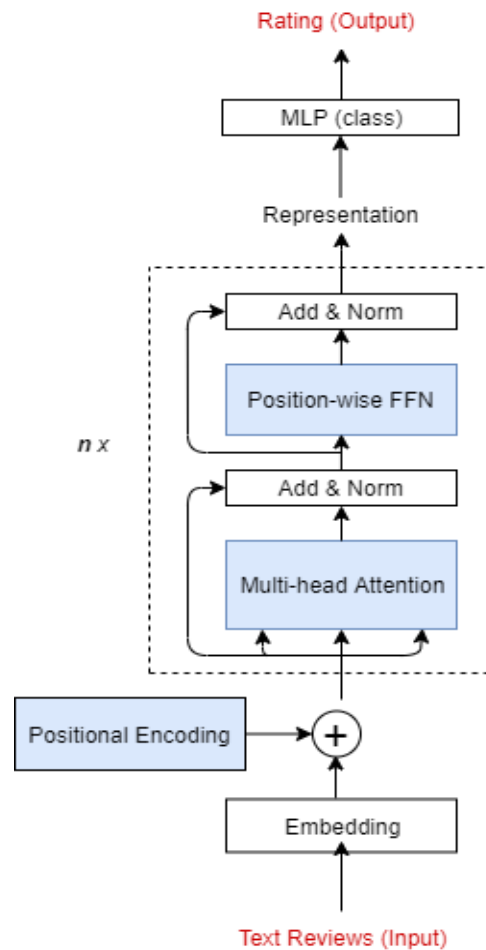


Fig 2. Transformer Model Design

3. Milestones

- ✓ Week 7: Problem statement and plan
- ✓ Week 8: Understanding and scraping data
- Week 9: Data analysis and pre-processing
- Week 10: Build baseline models (BoW, LSTM)
- Week 11: Build transformer-based classifier and compare
- Week 12: Prepare report and supplementary material

4. References

- [1] A. Vaswani, "Attention Is All You Need arXiv:1706.03762v5," no. Nips, 2017.
- [2] M. Tran, "Predicting Product Ratings From