

Team Champagne project progress

Sentiment Analysis of Google Reviews

ZongYun Li(netID: zyl2), Chi Han(netID: chihan3)

November 14, 2022

1) Which tasks have been completed?

We have already collect a dataset from google's comments by crawler. It contains 10.2k comments about American restaurants with the rating labels.

2) Which tasks are pending?

Our next plan is to conduct analysis on the sentiment aspects in the collected data. Specifically, the following tasks are pending:

- Data filtering: we need to filter out data where the rating scores are most likely mislabeled by users.
- Model training: We plan to build a sentiment analysis tool built upon BERT and RoBERTa.
- Evaluation and result analysis: Finally, we plan to quantitatively evaluate the model's performance on the dataset we collected.

3) Are you facing any challenges?

If requests are too often, Google will send you a request to check. So I use Google's colab to crwal the data, and I find that the maximum number of data that can be collect one time is probably 1460.