

CSC 180 Intelligent Systems (Fall 2021)

Final Project Proposal

Paper Title: Instant Feedback on your Amazon Review

Project Type: B

Team Members:

- Kai Kit Lok (Brian), 301108938
- Kevin Estrada, 216179224
- Cody Wuco, 301090621

Project Description/Abstract: This project seeks to give instant feedback on a consumer's review. Allowing them to know whether their review will be of value to others or not.

Motivation: There is an over abundance of reviews on products, and it would be nice to know how useful your review would be when you post it, so you can use the feedback to write better reviews without having to babysit your posts over extended periods of time.

Background: A single Amazon product can have many thousands of reviews, and there are many people who pride themselves on their reviews. The feedback those people would be looking for tends to come slowly when a product sells slowly or your review is buried in many other reviews. This could also be used to find good reviews before they are buried, and push them closer to the top of the reviews list.

Data sets:

<http://snap.stanford.edu/data/web-Amazon.html>

This dataset consists of reviews from [amazon](#). The data span a period of 18 years, including ~35 million reviews up to March 2013. Reviews include product and user information, ratings, and a plaintext review. **Note:** this dataset contains potential duplicates, due to products whose reviews Amazon merges. A file has been added below (possible_dupes.txt.gz) to help identify products that are potentially duplicates of each other.

System/Algorithmic design: describe what models/algorithms you plan to use in your project.

- We will create two models for this project.
 - Convolutional using TF IDF
 - Fully connected using TF IDF
 - Convolutional using TF IDF with word embedding
 - Fully connected using TF IDF with word embedding
- The Fully connected model without word embedding will be lifted from project 1.
- Logistic regression

Evaluation plan: We will test the model against reviews and see if the model will be capable of predicting the amount Helpfulness votes each review would get.

- We will split the data into train and test part
- We will use an RMSE and lift chart.

Task Division:

- Brian - Data preprocessing, Fully connected model with word embedding
- Cody - Code review, CNN without word embedding
- Kevin - Gathering results and reviewing the report, CNN with word embedding

We normally review each other's work and give advice and help. Therefore, every member's work is not limited to the task they were assigned, and the work gets divided more evenly around times available, which allows everyone to learn the material and does not force anyone to wait around while the other team members.

Our schedules aren't very easy to predict, pair programming and take over each other's work has been working very well for us. This is due to the projects being small enough for everyone to understand. We divide the work into much smaller pieces on the fly, and we can easily break away from the project when an issue arises.