

Progress report – Group LGD

1. What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members. What is the task?

Team members: Zhihao Xu (zhihaox3), Qikai Yang (qikaiy2), Hang Qi (hangq2).

Captain: Zhihao Xu

2. Which tasks have been completed?

First, we read the ACL 2011 paper "Learning Word Vectors for Sentiment Analysis" from Stanford University, and learned about the basic structure and features of the IMDb movie reviews dataset. The basic structure and characteristics of the IMDb movie reviews dataset. Our goal is to try to input movie reviews into the model and have it predict whether it is a positive or negative movie review. We have completed the pre-processing of the dataset. The dataset contains the tokenized data for our models. In addition, we have completed a Convolutional Neural Network model from scratch, and achieved a prediction accuracy of 78% for the test set. 78% accuracy on the test set.

3. Which tasks are pending?

We are also trying to save sequence information to complete a Recurrent Neural Network model. We are also trying to integrate the GloVe features with the preprocessing stage. Finally, if things go well, we may try to complete the generating fake reviews model to automatically generate fake movie reviews with the model.

4. Are you facing any challenges?

A major difficulty is the availability of computational resources. Cloud computing services such as Colab have time and speed constraints. This makes it take longer to complete the project because we need to change a variety of hyperparameters and run them multiple times.