Summarization and Sentiment Analysis of Reviews

Aayush Indrapratap Singh
North Carolina State University
asingh48@ncsu.edu

Bhavya Omprakash Agrawal
North Carolina State University
bagrawa@ncsu.edu

Diksha Paliwal
North Carolina State University
dpaliwa@ncsu.edu

1 Project Idea

With the internet coverage reaching out to remote locations, the visits on online retailers have increased exponentially. As a general tendency, people look at suggestions and feedback from the crowd to decide on whether to purchase commodities online. This project will include a technique to provide users with concise and accurate reviews of the product.

We intend to utilize data, gathered by web scraping, to create a model that creates a summary of the reviews. The model will perform abstractive summarization to retain the sense of statements. The same operations will be performed on the original reviews. The summary will be then fed to a sentiment analyzer to generate polarity of the review.

1.1 Dataset

Live Amazon Reviews will be scraped and utilised to develop classification models.

1.2 Software Requirements

We intend to develop a LSTM Seq2Seq (Encoder-Decoder) Model to Summarize the reviews and thereafter find the polarity. So, our software requirements are:

- Python Environment
- Libraries: NumPy, Pandas, Spacy, SciKit-Learn (ML Algorithms), BeautifulSoup (Web Scraping)
- Framework: TensorFlow (Deep Neural Networks)

1.3 Papers for reference

- C. Rain. Sentiment analysis in amazon reviews using probabilistic machine learning. Swarthmore College, 2013.
- Y. Xu, X. Wu, and Q. Wang. Sentiment analysis of yelps ratings based on text reviews, 2015.

1.4 Work Division(Tentative)

- Diksha: Web Scrapping and Data Analysis
- Bhavya: Sentiment Analysis
- Aayush: Summarization model

2 Midterm Milestone

We intend to complete the data pre-processing stage for our model, including collection, cleaning and processing. Additionally, we look forward to develop word embeddings of the dataset, and thereafter develop the Encoder-Decoder Summarizing Model.