Team Rotten Tomatoes

Greta Sharoyan
Georgia Institute of Technology
gsharoya@gatech.edu

Peng Chun
Georgia Institute of Technology
pchun9@gatech.edu

Amari Farnaz
Georgia Institute of Technology
farnaz_amiri@gatech.edu

Robert Steward
Georgia Institute of Technology
rsteward7@gatech.edu

1. Project Summary

There is increasing need for personalized recommendations in various online domains, such as e-commerce and entertainment. Traditional recommendation systems often rely on basic features such as item popularity and user preferences, which may not be sufficient to provide accurate and relevant recommendations. BERT, as a state-of-the-art language model, has shown promising results in capturing the semantic and contextual meaning of text data. Thus, incorporating BERT embeddings in the recommendation system could potentially enhance its performance and accuracy.

2. What we will do

We will develop a content based movie recommender system using Bidirectional Encoder Representations from Transformers (Bert) embeddings. A key advantage of BERT is that it has been trained a very large corpus of text. Additionally, BERT has the advantage that it captures context of words achieved by its use of a bi-directional transformation vector. The data set of interest will contain movie titles along with pertinent information such as popularity plot description. A user may then enter a set of one or more key words such as "comedy, rated PG" from which our Deep Learning model will then output several suggested movie recommendations ranked by a metric of similarity.

3. Resources

Research into recommender systems in general and movie recommendations in particular is not new as many companies are highly monetarily incentivized to give its users pertinent options. Yet at the same time, these models are constantly evolving as market pressures force company strategies to evolve in an ever more competitive environment for market share. "BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformers" by Pengjie Ren et al. is one such current research paper that provides user recommendations in a contextual sequential setting. Furthermore, introduction tutorials can also be found on the web both here "https://towardsdatascience.com/creating-a-hybrid-content-collaborative-movie-recommender-using-deep-learning-cc8b431618af" and here "https://medium.com/analytics-vidhya/recommendation-system-using-bert-embeddings-1d8de5fc3c56".

4. Datasets

While at this stage in the project, we have yet to begin implementation and formally settle on a project data source, several possible resources are available. For example Kaggle has dataset available here "https://medium.com/analytics-vidhya/recommendation-system-using-bert-embeddings-1d8de5fc3c56" which depicts the top trending YouTube videos of 2021. Another viable data set for us may be found here "https://grouplens.org/datasets/movielens/" which contains over 25M different ratings for modern movies. Several other applicable data sets also are available at "NetFlix Prize", "IMDB", "Flixter". We will explore all these options to see which of these data set option are the most amenable for us to achieve are stated purpose above within the project constraints of the course.