CS579 Project 2 weekly sample report

Bhanuja Arekatla(A20449753)
Navya Medarametla(A20442648)

Date 10/29/2020

1. Your progress this week

   We will be using the available Yelp Restaurant ratings dataset throughout this project.

   The algorithm that we will be using to generate the recommendation is Bayesian Personalized Ranking, which is a matrix factorization based collaborative filtering algorithm.

Given the dataset and the algorithm the preparation step we'll be doing in the next few code chunks is this:

- The raw **review.csv** contains user's reviews for each restaurant. Here, we will focus on implicit data, and follow the usual procedure of simulating binary implicit feedback data (i.e. whether the user enjoyed the food in a specific restaurant) by retaining only ratings of 4 stars and higher, while dropping lower ratings.
- The raw **restaurant.csv** contains each restaurant's details. We will also eliminate restaurants that had no categories information attached and create a mapping that stores each restaurant's category distribution. In this dataset, each restaurant ii typically has several categories cc associated with it, thus we assign equal probabilities p(c|i)p(c|i) to each genre such that ∑cp(c|i)=1∑cp(c|i)=1 for each restaurant ii. This genre distribution will play a strong role in determining whether our recommendation is well calibrated or not.

2. The problems you encountered and how you solve them
   The main problem we faced was when we are reading the review table is that we got a UTF-8 encoding error for the reviewContent column in the 'review' table. We solved the problem using a lambda function, decoding the columns by ignoring the errors.

3. Your plan for next week

   We are planning to implement collaborative filtering by filling the missing ratings and then get a similarity matrix to rank the items as per the user's work history.
   We will take a look at if the phenomenon of crowding out user's sub-interest occurs with our recommendation, develop a quantitative metric to measure how severe this issue is and implement a post-preprocessing logic that is agnostic of the underlying recommendation algorithm we decided to use to ensure the recommendation becomes more calibrated.

4. Teamwork contribution

   NavyaMedarametla: Read the paper on Calibrated Recommendations https://dl.acm.org/doi/pdf/10.1145/3240323.3240372 and extracted the datafile yelpResData.db from SQLite and exported restaurant.csv and review.csv.

   Bhanuja Arekatla : Found out what are the important tables needed to proceed with further steps to go ahead with calibrated recommendation.