GimmeBeer Personalized Beer Recommender

TEAM 35

Qiong Wu qwu304@gatech.edu | Yan-lin Chen ychen3017@gatech.edu | Yang Yu yyu391@gatech.edu | Yinuo Jiang yjiang375@gatech.edu | Zixin Wang zwang802@gatech.edu | Zhijian Guo zguo321@gatech.edu

MOTIVATION / INTRO

- Over 300,000 different beer brands across 150 types of beer globally generating $280 billion market
- However, there are very few recommendation systems in market providing innovative visualization to make recommendation process transparent and interactive
- Through our application, beer lovers shall get customized recommendations across different beer types

INNOVATION

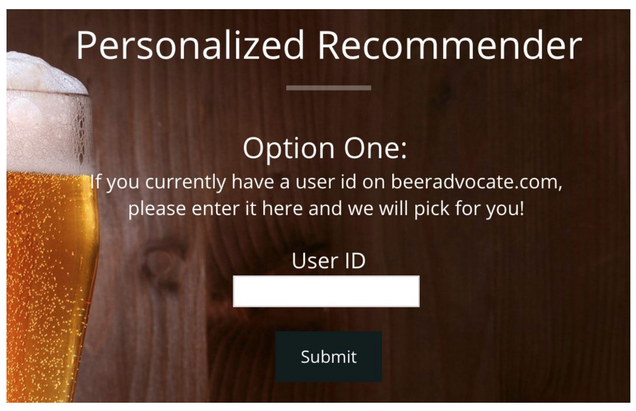
- Combination of multiple algorithms: content-based filtering, collaborative filtering, attribute weights & text mining. It empowers not just finding similar beers but also recommending beers favored by other users with similar tastes
- Visualization of result with circle size indicating likelihood that the user will enjoy that beer and opacity indicating popularity of beer.
- Key words extracted from reviews of the recommended beers displaying in form of word cloud which tells the characteristics of the beers

DATA

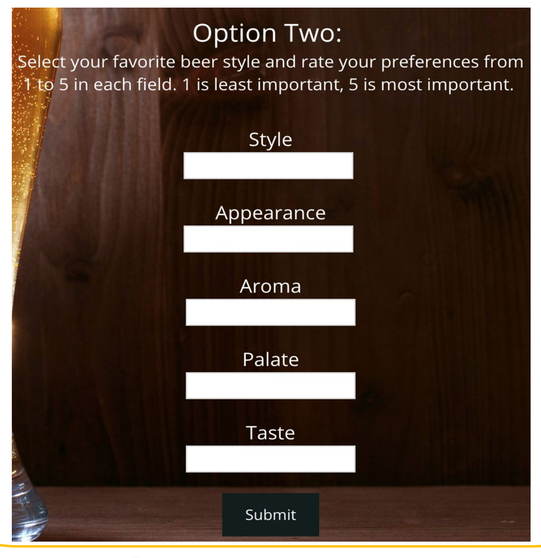
- Our dataset contains beer reviews collected from the BeerAdvocate (www.beeradvocate.com)
- The reviews are consolidated into a csv file and we downloaded from the data.world website (https://data.world/socialmediadata/beeradvocate).
- Size of the raw data file is approximately 1.2 GB, and it contains 12 columns and 1,586,615 rows.
- Our algorithms use beer name, beer style, 4 attribute ratings and 1 overall rating as well as a brief text review.

Input

Option 1 — Current BeerAdvocate users can get recommendations simply by entering beeradvocate user id.

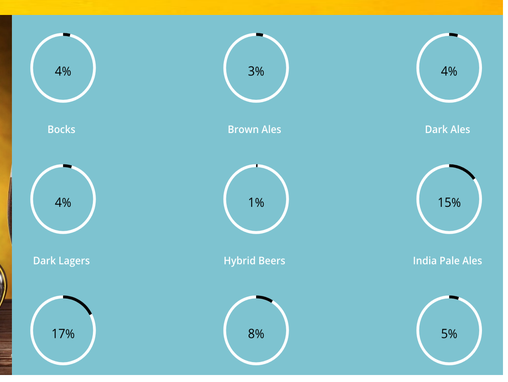


Option 2 — New users can select a preferred beer style and rate importance on appearance, aroma, palate and taste to get recommendations.



APPROACHES

- Three algorithms and one visualization tool utilized in the recommendation engine
- A dynamic website with interactive visualization and animated circles built by d3.js and jQuery



Collaborative Filtering (CF)

An algorithm that analyzes user similarity to recommend beers that similar users like

Content-Based Filtering (CBF)

An algorithm that analyzes similar beers based on user reviews for recommendations

Attribute Weights (AW)

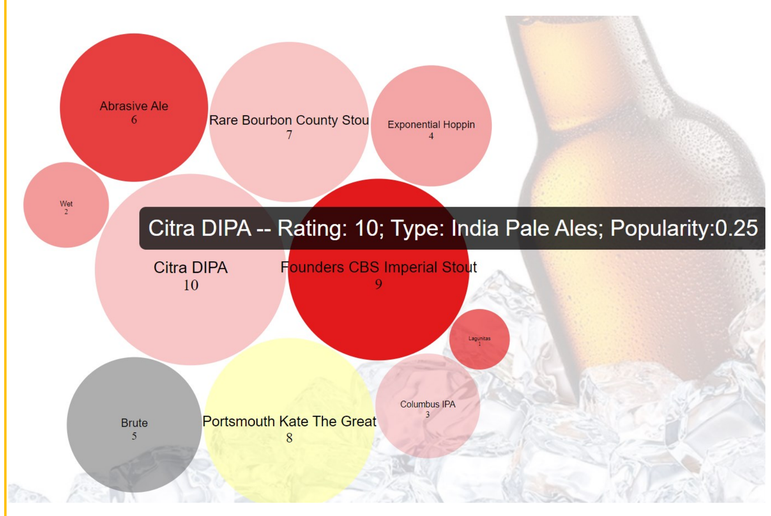
An algorithm that recommends beers based on preference weights of the new user

Text Mining (TM)

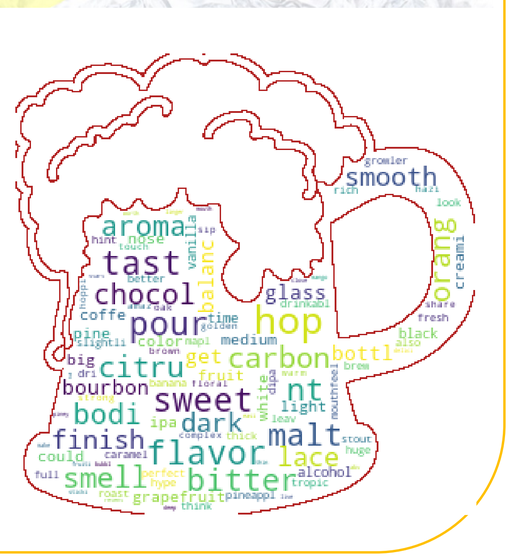
Text based analysis that summarizes key words for the beers recommended

Output

Bubble Chart — Beers are represented by circles. Colors represent different beer styles. Circle size shows the ranking of the recommendation. Opacity tells number of reviews for that beer showing the popularity of the beer.



Word Cloud — A dynamic word cloud that grabs the keywords of reviews from other users who share similar preferences.



EXPERIMENTS & RESULTS

- We evaluate our approaches mainly using two indicators: meaningful visualization of recommendation and run-time of the algorithm.
- For CBF approach, we applied clustering, and spent significant time on testing various clustering algorithms. We found the best clusters (Figure 1) by using K-Means with dimension reduction.
- Running time of recommender engine affects user experience. Running time of CF depends on how fast correlations being computed. We explored using truncated SVD to reduce factors for faster calculation, but training time or model preloading time was still a bottleneck. All similarities were calculated, but it still took about a long time to load data. We then explored storing user-user relations in smaller files and running calculations only when required, this significantly reduced run time to less than 30s (Figure 2).
- CBF also has long runtime. Various options were explored as well, but we were unable to achieve the same reduction in run time, therefore it is dropped from our final system.
- For the TM approach, We have gone through 1.5milion reviews stemming & removing stop words. The most frequently used words were mostly generic adjectives such as ‘good’, ’nice’ etc. We have removed these words to give a more informative word cloud as shown in the word cloud figures.

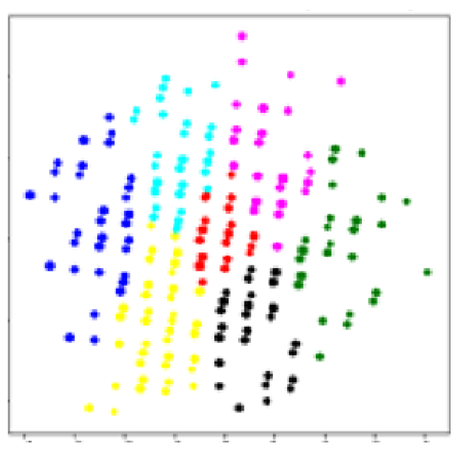


Figure 1: Clustering

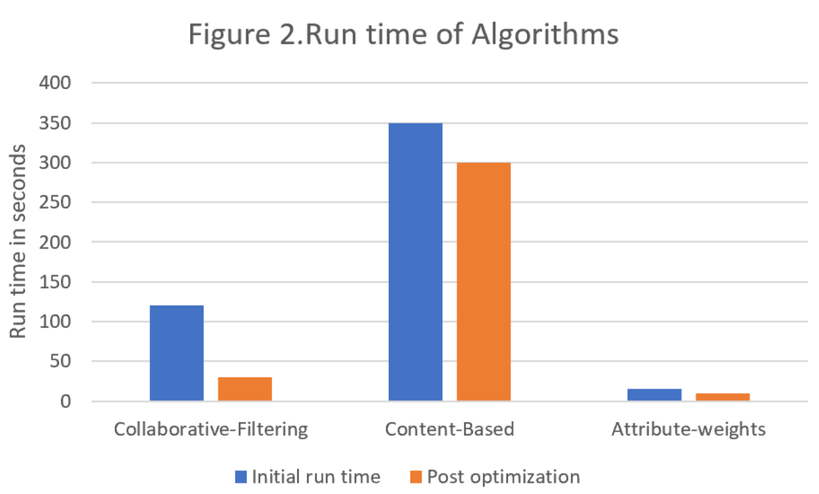


Figure: word-cloud 1

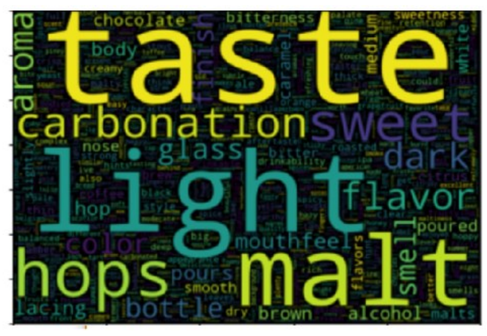


Figure: word-cloud 2