SI650 Project Proposal

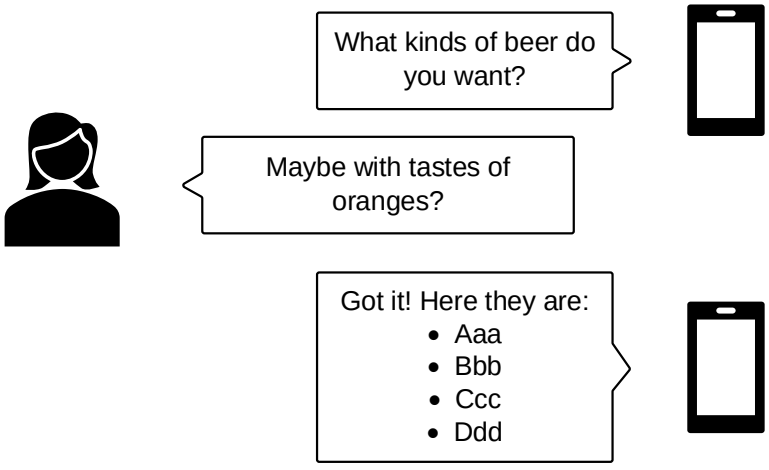
Yu Xia
xiayuu@umich.edu

1. Motivation

As a beer lover, drinking difference kinds of beer has been one of my favorite relaxing activities. However, every time when I would like to try a new flavor of beer, I get stuck on the all those different names of beer and have no clue which one might be the one I want. Since I am not a beer expert and the names of the beer usually do not tell me how they taste like, I have to take a lot of time searching them one by one to learn a bit more knowledge before making a decision. What if there is a system that can recommend beer to me after I tell it what flavor do I want? A beer recommender system based on user reviews might be something that I am thinking of.

2. System Overview

In this project, I would like develop individually a content-based beer recommender system. The recommender system will be built on a data collection of user reviews to different kinds of beer. The system takes a user query on the beer as input and then returns a list of beer names matching the user's taste.



Beer Recommendation

3. **Dataset**

I derive the dataset "BeerAdvocate" from Dr. Julian McAuley's ICDM 2012 paper, Learning attitudes and attributes from multi-aspect reviews. The link to the dataset is https://cseweb.ucsd.edu/~jmcauley/datasets.html#multi_aspect.

According to description on the website, the dataset includes beer reviews from sensory aspects such as taste, look, feel, and smell. A sample of the data is shown below:

{'beer/name': 'Sausa Weizen', 'beer/beerId': '47986', 'beer/brewerId': '10325', 'beer/ABV': '5.00', 'beer/style': 'Hefeweizen', 'review/appearance': '2.5', 'review/aroma': '2', 'review/palate': '1.5', 'review/taste': '1.5', 'review/overall': '1.5', 'review/time': '1234817823', 'review/profileName': 'stcules', 'review/text': 'A lot of foam. But a lot.\tIn the smell some banana, and then lactic and tart. Not a good start.\tQuite dark orange in color, with a lively carbonation (now visible, under the foam).\tAgain tending to lactic sourness.\tSame for the taste. With some yeast and banana.'}

From all the fields, I will use only two of them, "beer/name" and "review/text" to build my recommender system.

4. **Implementation Detail**

To be more specific, I plan to use TF-IDF, cosine similarity, BM25 and BERT to implement different kinds of weighting functions. I will then use them to get scores for the retrieval tasks and compare the performances between them.

5. **Evaluation Metric**

I will use Normalized Cumulative Discounted Gain of the top k results to measure the performance. The ground-truth relevance of the recommendation results is obtained through data annotation. The annotating beer recommendation results might be challenging as it is not a common domain for most people. Fortunately, the "BeerAdvocate" datasets come with two additional data annotated by beer experts. With such data, it would be much easier to obtain the ground-truth label to measure the performances of the models.

6. **Tentative Timeline**

1) data preprocessing – week1
2) weighting functions implementation – week 2-3
3) retrieval module and basic interface – week 4-5
4) retrieval result and comparison analysis – week 6
5) conclusion and project report writing – week 7