

Name

bee.rs.

Team

Sayer Rippey, Oren Shoham, Peter Fogg.

Problem

There's a lot of beer in the world. You, as someone of discerning taste, require the finest of quality. But you don't have time to try every beer in the world (if only!). How do you know you'll like a beer before you even try it? The answer, of course, is machine learning!

We propose to design a machine learning program to recommend beer to the user, based on beers they like and dislike. That is, given an input set of beers liked and disliked, we will return a set of possibly liked beers as recommendations.

Approach

We will use web scraping and NLP techniques to gather data from a variety of beer review websites,¹ and build a database of beers along with attributes that we will select from this data. Given the input set of beers, we will train a machine learning algorithm to classify other beers as liked or disliked. We then return a set of liked beers to the user, who presumably then tries said beers.

Data

Data will be obtained from the aforementioned beer review websites. A beer will be represented as a vector of attributes, some binary (is it hoppy?) and some multi-valued (the style – IPA, stout, etc.).

Evaluation

Given that the system should evaluate a user's preferences, we will test the system on that user. So if a user likes beers that have been recommended, or if beers which they enjoy are among the recommendations, then the system is performing well. We can also compare the user's recommendations against the set of beers that a user of Beer Advocate or another site with similar preferences would like. Another evaluation would be to take a user from one of

¹beeradvocate.com, thebeercritic.com, ratebeer.com, etc.

the websites, train the algorithm on half of their reviews, test on the other half, and see if it classifies the beers correctly.

Background Reading

- <http://cs229.stanford.edu/proj2012/Wilson-BeerRecommendationEngine.pdf>
- https://en.wikipedia.org/wiki/Recommender_system
- <http://www.cs.washington.edu/education/courses/csep521/07wi/prj/michael.pdf>
- <https://recommender-systems.org>

Tasks

- Scrape web data – Peter, Oren.
- Parse data into vectors consisting of users, beers, beer attributes, and like/dislike – Sayer, Peter.
- Implement ML algorithm – Oren, Sayer.
- Test – everybody.

Schedule

- Obtaining data – Tuesday 4/16.
- Parsing data – Tuesday 4/23.
- ML algorithm – Friday 4/26.
- Testing – Tuesday (Thursday?) 4/30.