My Beer Recommender

How to find new beers for beer enthusiasts?

Raphael Egan, February 4th, 2021

Why a beer recommender?

Tired of your same-old choices?

Out of your comfort zone?

Potential marketable interests:

- Advertising platform for producers;
- Source of information for determining more area-specific customer preferences;

How to recommend?

A first attempt (draft) using collaborative filtering

Based on reviews from beeradvocate.com (https://data.world/socialmediadata/beeradvocate):

- 1,586,614 reviews of 56,857 beers from 33,355 reviewers, from 2001 to 2011 (172 Mb csv file).
- used it as a starting point for building a collaborative-filtering recommender system.

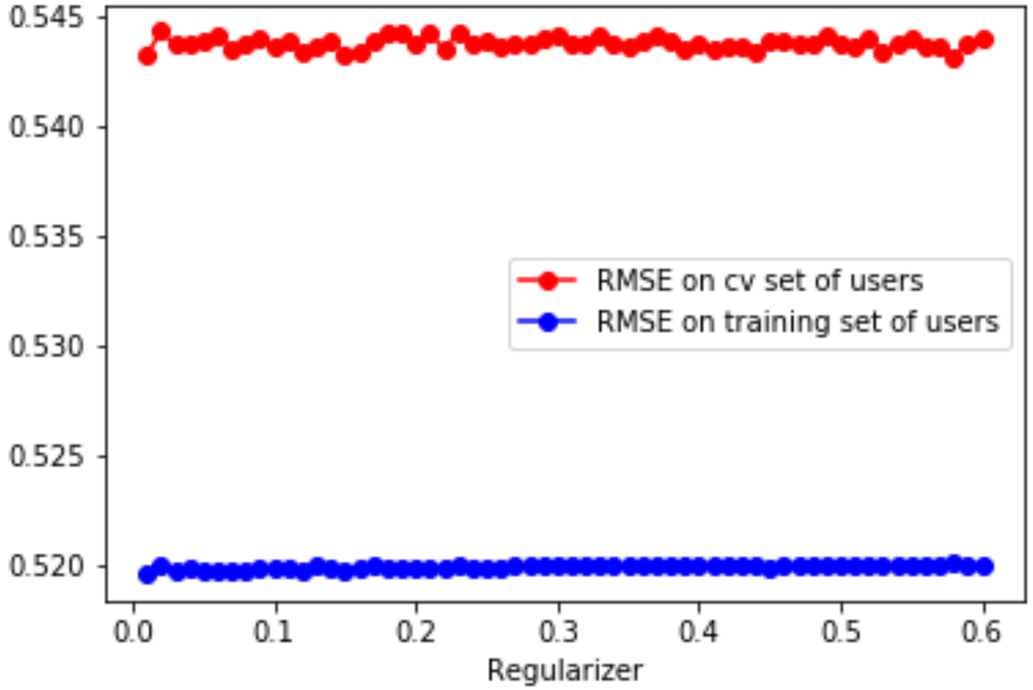
Core concept

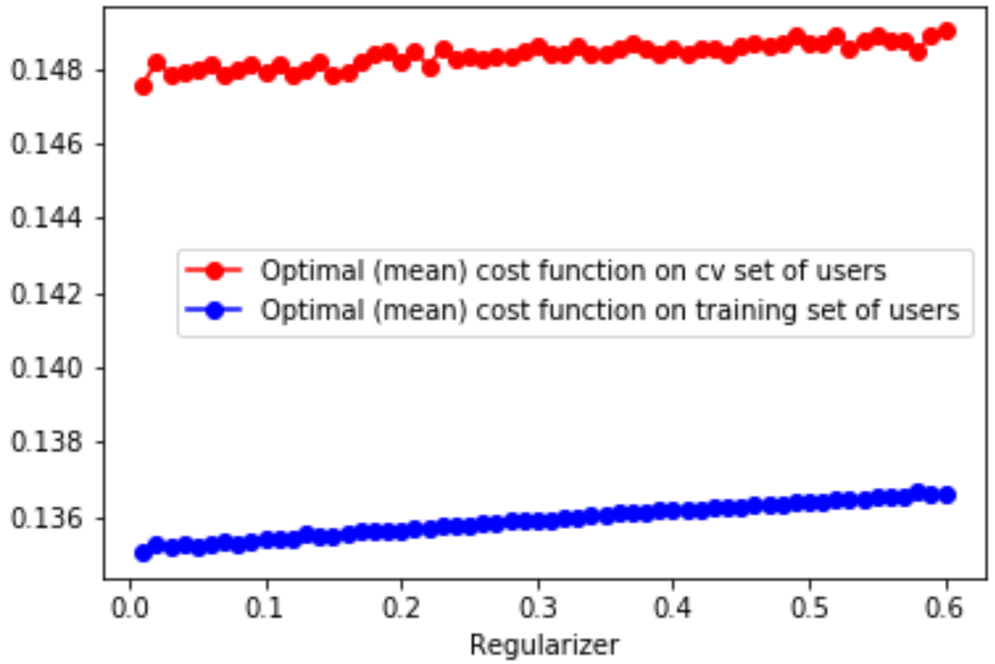
- Use the review data to find n_f core features x_i^b $(i = 1, \ldots, n_f)$ "defining" every beer b considered.
- Correspondingly, every user u is characterized by $1 + n_f$ "preferences factors" θ_i^u $(i = 0, \ldots, n_f)$.
- The overall rating for beer b by user u is estimated via $\hat{r}^{u,b} = \theta_0^u + \sum_{i=1}^{n_f} \theta_i^u x_i^b$.
- Learn x_i^b (and $\theta_i^{u^{\mathrm{t}}}$ as well), as

$$\underset{x_i^b, \theta_j^{u^{\mathrm{t}}}}{\operatorname{argmin}} \frac{1}{2} \left(\sum_{b, u^{\mathrm{t}} : r^{u^{\mathrm{t}}, b} \exists} \left(\hat{r}^{u^{\mathrm{t}}, b} - r^{u^{\mathrm{t}}, b} \right)^2 + \lambda \left(\sum_{i,b} \left(x_i^b \right)^2 + \sum_{j, u^{\mathrm{t}}} \left(\theta_j^{u^{\mathrm{t}}} \right)^2 \right) \right)$$

Training, and generalization

- Trained on the 2,000 most reviewed beers with the 1,000 most active corresponding reviewer profiles (reduced the data set for the sake of execution time, mostly);
- Fixed the hyper-parameter $n_f = 5$ (educated guess from the data file);
- Swept $\lambda \in [0.01, 0.02, \ldots, 0.6]$ seeking the best regularizer for generalization to a cross validation set of 200 other reviewers...





A first user (myself)

(Picked $\lambda = 0.15$)

```
I am considering a collaborative filtering strategy, trained on the
2000 most reviewed beers in the data base and the 1000 most active
reviewer profiles considering these beers. I give 5 features per
beer and I use a regularization parameter of 0.15
--------------------------------------------------------------------------------
|                                                                              |
|                Hello, I am your beer assistant and recommender system        |
|                     Let's find new beers for you to check out!               |
|                                                                              |
--------------------------------------------------------------------------------
I would need to know you and your tastes better before I can make
any recommendation: please take some time to think about beers
you already know and let me know how much you like/dislike them,
on a scale from 1 to 5. In order to give sensible recommendations,
I need you to share your feelings about 5 beers, at least
[Since I am currenlty in training, I may not know some beer you
mention and I may ask you to find other ones if possible.
Thank you for your understanding ;-)]

Tell me the name of a beer you know (or type QQQ if you are done): chimay
I found these possible matches in my data_base, which one did you have in mind?
Enter 0 for Chimay Grande Réserve (Blue)
Enter 1 for Chimay Première (Red)
Enter 2 for Chimay Tripel (White)
[if none of the above matches your thought, enter any other number]

Which one did you mean? 0

How much did you like it (on a scale from 1 to 5): 4.2

Tell me the name of a beer you know (or type QQQ if you are done):
```

MATCHES:

A first user (myself)

(Picked $\lambda = 0.15$)

1.0

4.1

3.0

4.1

4.3

4.8

4.2

My beer recommender suggestions:

> 5.0 ?

> 5.0

> 5.0 ?

4.9 ?

Further work and improvements

- Better handle sparsity of rating data to train on larger data sets (distributed computing?)
- Find similar reviewer profiles to improve recommendations (K-means clustering?)
- Exploit and couple with other data sources;
- Suggestions?