Beer Recommender System

Jason Yingling

University of Central Arkansas

December 11, 2017

Table of Contents

How Do We Choose?

Techniques

Collaborative Filtering: Method finds a subset of users who have similar preferences to the target user and use this subset to offer recommendations.

Content Filtering: Method attributes characteristics to each item. Similar items are computed by summing up the total number of matching characteristics.

What Can it do?

- predict how much you may like a product/service
- create a list of N beset items for you
- create a list of N best users for a product/service
- Explain to you why these items are recommended (Robust system)

Single Value Decomposition

SVD is a matrix factorization technique for producing low-rank approximations. Given *mxn* matrix *A*, the SVD defined as:

$$SVD(A) = USV^T$$

.

SVD error is over all entries, but recommender system data is sparse!

Single Value Decomposition

SVD provides minimum reconstruction error (Sum of Squared Errors):

$$min(U, V, S) \sum_{(i,j) \in A} (A_{ij} - [USV^T]_{ij})^2$$

Where U V are orthogonal matrices and S is a diagonal matrix.

Single Value Decomposition

SVD for recommendation data:

$$R = Q \cdot P^T$$

Where $Q = U$ and $P^T = SV^T$

SVD Algorithm

$$\hat{r} = r_{ui} = q_i^T p_u$$

q_i and p_u can be found such that square error difference between their dot product and known user rating is:

$$minimum(p,q) \sum_{(u,i)\in K} (r_{ui} - q_i^T \cdot p_u)^2$$

Stationary Probability

To reduce error we introduce bias terms:

$$\hat{r}_{ui} = \mu + b_u + b_i + q_i^T p_u$$

If user u is unknown, then the bias b_u and the factors p_u are assumed to be zero. The same applies for item i with b_i and q_i

Average Customers

Thus the final equation to minimize is:

$$\sum_{r_{ui} \in R_{train}} (r_{ui} - \hat{r}_{ui})^2 + \lambda \left(b_i^2 + b_u^2 + ||q_i||^2 + ||p_u||^2 \right)$$

SVD Algorithm

The minimization is performed by stochastic gradient descent:

$$b_u \leftarrow b_u + \gamma(e_{ui} - \lambda b_u)$$

$$b_i \leftarrow b_i + \gamma(e_{ui} - \lambda b_i)$$

$$p_u \leftarrow p_u + \gamma(e_{ui} \cdot q_i - \lambda p_u)$$

$$q_i \leftarrow q_i + \gamma(e_{ui} \cdot p_u - \lambda q_i)$$

where $e_{ui} = r_{ui} - \hat{r}_{ui}$.
The learning rate γ and regularization term λ are used to assist training the model.

D. Goldberg et al., "Using Col- laborative Filtering to Weave an Information Tapestry," Comm. ACM, vol. 35, 1992, pp. 61-70.

. Koren, "Factorization Meets the Neighborhood: A Multifaceted Collaborative Filtering Model," Proc. 14th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, ACM Press, 2008, pp. 426-434

Y.F. Hu, Y. Koren, and C. Volinsky, "Collaborative Filtering for Implicit Feedback Datasets," Proc. IEEE Int'l Conf. Data Mining (ICDM 08), IEEE CS Press, 2008, pp. 263-272.

Y. Koren, "Collaborative Filtering with Temporal Dynamics," Proc. 15th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD 09), ACM Press, 2009, pp. 447-455.