Cmpe 462 Final Project
Movie Recommendation
Midway Report

Mustafa Onur Eken

April 24, 2016

1 Plan, Progress & Future Work

In this project we aim to attack the problem of item recommendation to customers with latent factor approach. We are going to factorize the given data matrix to actualize the latent factor approach. There are several different methods to achieve such a decomposition. In particular, we are going to implement this factorization via 1) *Stochastic Gradient Descent* [1] and 2) *Singular Value Decomposition* [2]. Afterwards we are going to compare performances of these two approaches. In addition, we are going to expand our latent factor model by taking user and item specific biases into account as described in [1]. The plan is as follows:

1. Generate synthetic training and test data.
2. Assume the basic model (no regularization, no user, item biases), implement SGD algorithm for factorization.
3. Test the algorithm on synthetic data
4. Read the actual *MovieLens* data (N = 100,000 where 80% training, 20% test).
5. Test the algorithm on *MovieLens* data.
6. Assume the basic model (no regularization, no user, item biases), implement SVD algorithm [2] for factorization.
7. Test the algorithm on synthetic data.
8. Test the algorithm on *MovieLens* data.
9. Expand the model by adding regularization terms and biases associated with users, items. Alter the algorithms accordingly.
10. Test on synthetic and *MovieLens* data.
11. Do comparison plots.

Steps [1-5] are currently completed, we are having an issue with the convergence of SGD though. According to our observation when data size grows larger a much more delicate choice of learning rate is needed. Some intuitively guessed values of learning rate, seem not to work well since each of them have been stuck at a different error level due to oscillation. For solving this problem we are going to implement *momentum update* or *adaptive learning rate* [3]. After tackling this problem, we are going to implement an algorithm for computing SVD [2] and follow the remaining steps in order. See the *github* repository if interested in the code further, `https://github.com/oeken/cathar`.

References

[1] Yehuda Koren and Robert Bell and Chris Volinsky. *Matrix Factorization Techniques For Recommender Systems, 2009.*

[2] Sarwar, Badrul and Karypis, George and Konstan, Joseph and Riedl, John. *Application of dimensionality reduction in recommender system-a case study, 2000,* DTIC Document.

[3] Alpaydin, Ethem. Introduction to machine learning, 2014, MIT press.