Hybrid Recommendation System

Khagesh Patel Ankush Sachdeva

Mentored by Prof. Amitabh Mukerjee, Dept. of CSE, IIT Kanpur

INTRODUCTION

Recommendation systems are widely used in e-commerce companies like Amazon, Net ix to help users discover items that they might not have found by themselves.

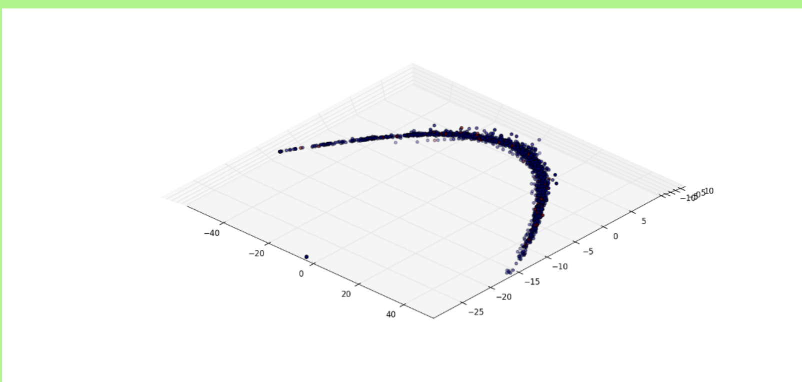
There are a number of techniques that are currently being employed in the industry for this task. We look at some of them and then propose a hybrid model.

The methods we tried to implement are:

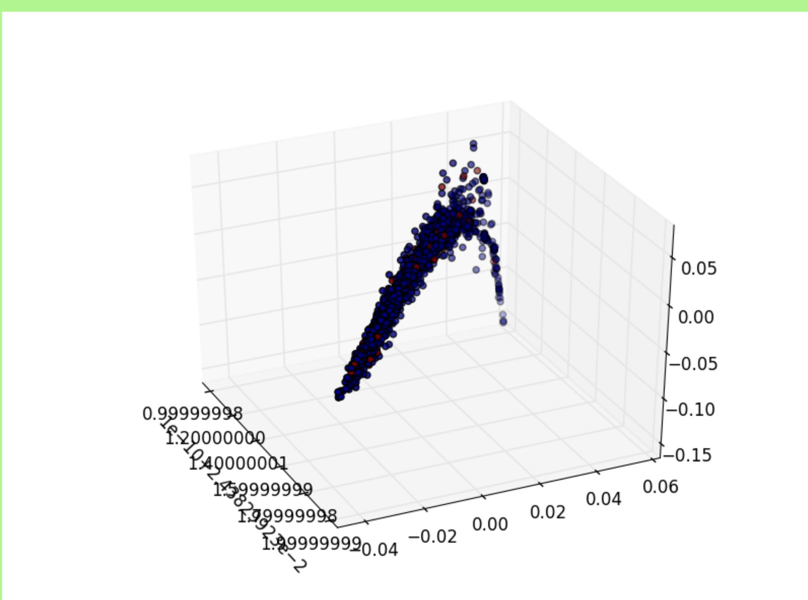
- Slope-one Item-Item collaborative filtering
- K-nearest neighbor user-user collaborative filtering
- K-nearest neighbor item-item collaborative filtering
- SVD
- Incremental SVD
- Incremental SVD with temporal dynamics
- Content based recommendation
- Demographic based recommendation

DATA VISUALIZATION

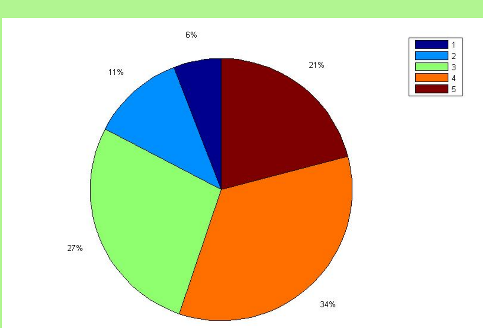
ISOMAP of Movielens Data

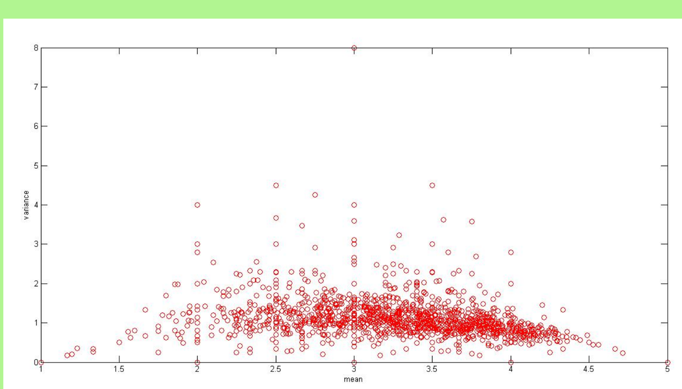


Local Linear Embedding of Movielens Data



General rating behavior





METHODS

- **Slope one item-item collaborative filtering**
  - A regression model of a linear polynomial with slope 1, i.e only one independent variable which is trained. The model, though simple and computationally less intensive gives surprisingly good results.
- **KNN item-item collaborative filtering**
  - Recommending movies based upon the similarity of rated items with k nearest neighbours in the dataset.
  - Similarity criteria - cosine, Euclidean distance, pearson correlation coefficient
- **KNN user-user collaborative filtering**
  - Recommending movies based upon the similarity of users who rated an item with k nearest neighbours in the dataset.
  - Similarity criteria - cosine, Euclidean distance, pearson correlation coefficient
  - Generally, poorer result compared to KNN item-item CF
- **Content based collaborative filtering**
  - Generates a feature for each item based upon the prior knowledge available for that item.
  - For movies - movie genre used to generate the feature vector.
  - Useful for users who have a sparse rating vector.
- **Demographic based collaborative filtering**
  - Generates a feature for each user based upon the prior knowledge available for the user.
  - Age, gender and profession used to generate the feature vector
- **SVD**
  - Projecting each user and item to a lower dimension (15 in our case).
  - Stochastic gradient descent to factorize the rating matrix to user and item feature matrix.
  - Learning rate 0.001, Num of iterations 200
  - $\hat{r}_{ui} = p_u^T q_i + b_u + b_i + \mu$
- Incremental SVD
  - Similar to SVD except for including implicit feedback.
  - Reduced the data dimensionality to 5
  - Learning rate 0.004, Num of iterations 500
  - $\hat{r}_{ui} = q_i^T \left(p_u(t) + |N(u)|^{\frac{-1}{2}} \sum_{j \epsilon N(u))} y_j \right) + b_u + b_i + \mu$
- Incremental SVD with temporal dynamics
  - Similar to Incremental SVD except for time dependent user feature matrix.
  - All rating divided into 25 equally spaced time buckets
  - Learning rate 0.0005, Num of iterations 1100
  - $\hat{r}_{ui} = q_i^T \left(p_u + |N(u)|^{\frac{-1}{2}} \sum_{j \epsilon N(u))} y_j \right) + b_u + b_i + \mu$

RESULTS

RMSE values for all methods on Movilens 100k dataset

Method	RMSE
Slope one (item-item)	1.03136
KNN(user-user)	0.9439889
KNN(item-item)	0.9500658
Content based	1.8461
Demographic based	1.11833
SVD	0.942863
SVD++	0.936
timeSVD++	0.929762
Hybrid	0.915816

CONCLUSION

- Combining KNN, Demographic, content-based and time-SVD++ methods using weighted mean, we achieve the RMSE value 0.91581557 i.e. a 1.5% improvement over the best individual method.
- Even a small improvement in RMS greatly impacts the top 10 suggestions given to the users[5]
- Isomap and locally linear embedding shows that the data has intrinsic lower dimensionality.
- Time-svd++ performed the best individually compared to all other methods.

REFRENCES

[1] Linden, Greg and Smith, Brent and York, Jeremy (2009) Amazon.com recommendations: Item-to- item collaborative filtering

[2] Robert M. Bell, Yehuda Koren, Chris Volinsky (2008) The BellKor 2008 Solution to the Netflix Prize

[3] Francesco Ricci, Lior Rokach, Bracha Shapira, Paul B. Kantor(2010) Recommender Systems Handbook

[4] Yehuda Koren(2010) Collaborative filtering with temporal dynamics

[5] Netflix Community- How useful is lower RMSE http://www.netflixprize.com/community/viewtopic.php?id=828

CONTACT

Ankush Sachdeva - 11120 - sankush@iitk.ac.in

Khagesh Patel - 11362 - kpatel@iitk.ac.in