MIDTERM REPORT

Qing Wang, Tianqi Wu, Hao Zhu
{qwang55,twu38,haozhu3}@illinois.edu
University of Illinois at Urbana-Champaign
Champaign, IL 61820

November 6, 2019

1 Introduction

As the amount of information available on the web grows rapidly in recent decade, recommender system becomes an inevitable approach to avoid providing users with contents that are not of their interest across a variety of web domains. Recommender systems with implicit feedback have attracted many studies recently as it is easier to collect implicit feedbacks from massive users such as "like" or "dislike" rather than specific ratings. However, as there is no rating information, it is harder to capture user's degree of interest over items and requires more advanced strategies. For review-based web platform such as Yelp, social network of users is easily accessible and can serve as an additional source to better capture user preference. For users with little or no feedback, there is a high chance that they will like the item if most of their friends like it. In this work, we explore the possibility of incorporating social network information with implicit feedbacks to enhance performance of existing recommender systems. We use Bayesian Personalized ranking method for implicit feedback as our baseline model which is only based on user-item interactions. We will design our model that uses both user-item interaction and social network and compare it with baseline model to illustrate the effect of incorporating social network with recommender system.

2 Related Work

2.1 Bayesian Personalized Ranking

Methods like matrix factorization and adaptive knearest-neighbor are commonly used for item recommendation of implicit feedback. However, none of them are optimized for ranking. Bayesian personalized ranking [1] (BPR) is designed to improve the standard learning techniques. It uses maximum posterior estimator derived from a Bayesian analysis as the optimization criterion and stochastic gradient-descent algorithm based on bootstrap sampling as learning models. Unlike traditional methods, the training data consists of both positive and negative pairs and missing values. Also, the training data is for the actual objective of ranking. Whereas, the model relies on the quality of negative sampler and the uniform negative sampler may be inefficient for data sets with large number of items.

2.2 Socially Enabled Preference Learning from Implicit Feedback Data

This work [2] proposes a method to incorporate social network in recommender system with implicit feedback by adding friends' influence term, a weighted sum of user's friends' preferences, in objective function and optimizing it with block Gauss-Seidel process. The advantage is that such model learns user preference and friend importance simultaneously. The model outperforms recommender systems that use only user-item interaction and illustrates that it is beneficial to incorporate social network with recommender systems based on implicit feedback. However, the model assigns equal weight of users' friends across different items while friends usually have different influence on a user for different item categories. A model that adjusts friends' weights based on item category should further improve proposed model.

3 Result

3.1 Dataset

The baseline BPR model is trained and tested on Yelp dataset. There are 141623 user-restaurant interactions, 6858 unique users and 3317 unique restaurants in the dataset. The task is to recommend restaurants for given user id. There is no user rating since we are focusing on implicit feedback. The dataset is split into 80% of training set and 20% of validation set.

3.2 Evaluation Metric

The performance metric used is Recall@K and NDCG@K and the equations are provided:

$$Recall@K(u) = \frac{\sum_{k=1}^{K} rel(k)}{min(K, |I_u|)} \tag{1}$$

$$NDCG@K(u) = \sum_{k=1}^{K} \frac{rel(k)}{log_2(k+1)} \tag{2}$$

where I_u is the set of held-out items for user u and $rel(k)$ is indicator equaling 1 if item at rank k is relevant.

3.3 Results and Discussion

The model is trained for 20 epochs and best validation performance is achieved around epoch 7. The experiemnt is repeated three times and average Recall@50 is 0.330 and NDCG@50 is 0.263. Recall@50 measures the fraction of relevant items retrieved correctly within top-50 ranks and NDCG@50 is discount factor to emphasize the importance of higher ranks. Relative lower NDCG@50 indicates that the model may not be able to provide correct ranking for the recommended restaurants. To summarise, our baseline BPR model does poorly on both training and test set. It may indicate possible underfit of our model. Later on, we will focus on using social network of users to improve the performance of our model.

References

[1] Steffen Rendle et al. Bpr: Bayesian personalized ranking from implicit feedback. *Conference on Uncertainty in Artificial Intelligence, pages 452–461*, 2009.

[2] Tomasz Matuszczyk Stephane Canu Julien Delporte, Alexandros Karatzoglou. Socially enabled preference learning from implicit feedback data. *ECML PKDD 2013*, 2013.