Survey Paper on Response Aware Probabilistic Matrix Factorization

Siddharth N. Dhage[1], Chetan J. Awati[2], Sonam Kharade[3]

[1]Post Graduate Student, Department Of Technology, Shivaji University, Kolhapur, Maharashtra, India
[2]Assistant Professor, Department Of Technology, Shivaji University, Kolhapur, Maharashtra, India
[3]Assistant Professor, Department Of Technology, Shivaji University, Kolhapur, Maharashtra, India

---***---

Abstract - *Nowadays, Internet has become crucial for our lives and its importance is growing day by day. This popularity causes World Wide Web to have huge amount of information. Now days, it is more difficult for the users to find the information they are looking for due to the increasing size and complexity of many web sites. A web site can be personalized or pages that are related to the user's interest may be selected to help users find the information they are looking for more easily. As the web sites continue to grow, recommender systems have become valuable resources for users who try to find a smart way to get what they like from the large volume of data available to them. Basically, recommender systems use the opinions of users of a system to help individuals identify the information or products most likely to be of interest to them or relevant to their needs. Recommender systems are used in collaborative filtering.*

Key Words: **Opinions, recommender system, collaborative filtering.**

1. INTRODUCTION

We use e-commerce websites for different purposes like reading books, listening songs, watching movies, shopping, etc. Some of the e-commerce websites are Amazon, Netflix, Flipkart, etc. These e-commerce websites have huge amount of data, therefore user get confused for which book to read, which song to listen or which movie to watch because searching required item from large data requires more time and increase work. So, to provide users with appropriate items, recommender systems are used.

Collaborative Filtering techniques can usually be classified into memory based CF methods and model based CF methods. Previously proposed CF methods mainly focus on manipulating the explicitly observed rating scores to understand user's likings for future prediction. An explicit rating score clearly indicates a user's liking on a selected item as well as items hidden properties. The ratings that a user gives to different items shows information of user interest. The rating values that an item received from different users also carry information on intrinsic properties of the item. The rating information indeed can present users preferences on different items.

1.1 Assumptions

For a recommender system to work, we need to make a few assumptions based on which we can predict the preferences of a user. Without making assumptions, it is impossible to make any recommendation. We describe two assumptions that are made in recommender systems

- Preference Consistency through Time: The first assumption we made in recommender systems is that most users' preferences are more or less consistent throughout a short period of time. In other words, if a user enjoys a certain type of music today, we assume that she will enjoy the same genre in the near future. This assumption is needed so that the collected past information can be utilized to produce recommendations for the future.

- Preference Consistency across Users: Another assumption made in most recommender systems is that users who shared preferences in the past tend to share similar preferences in the future. For example, Items liked by a user tend to be enjoyed by the users similar to the user and the user-item-rating matrix is not a full rank matrix. The user-item-rating matrix is a way of organizing the collected rating data in the form of a sparsely filled matrix. The i^{th} row and j^{th} column of the matrix contains the rating assigned by specific user to specific item.

2. RELATED WORK

Collaborative filtering, unlike content-based filtering which use content analyzer to construct items features and match them with a user's profile to make recommendations, rely solely on the past ratings assigned by users to items. In collaborative filtering, we try to extract intrinsic insights such as users who buy diaper often buy beer at the same time, be it explicit such as the ones in neighborhood based methods or implicit ones as is in model-based methods. Recommender systems are gaining popularity in providing personalized online services. With the increase in the use of e-commerce sites, it has become very easy for the users to find the items of their interest without wasting a lot of time. Websites like Amazon and EBay examples for Recommender systems which provide recommendations to the users based on their search history and purchase history. Recommender systems provide recommendations of almost all the items ranging from books to movies to music. Facebook and Twitter are also recommender sites which provide recommendations for friends. NetFlix.com is very famous as a movie recommender website.

These methods are deployed in real-world recommender systems. Author Greg Linden, Brent Smith, and Jeremy York have explained example of Amazon.com. Here recommender algorithms are used to personalize online store for each user. Author G.Adomavicius and A.Tuzhilin classified Collaborative filtering methods into two main categories: memory based methods and model-based methods [2].

2.1 Memory Based Methods

Memory based methods are very popular and widely used in commercial websites. Memory based methods make rating predictions based on entire collection of previously rated items by the users. These systems use statistical techniques to find the set of users called neighbours. It combines the preferences of neighbour users using different algorithms and gives a prediction or top recommendations for the active user. This system is called used based collaborative filtering. Memory-based methods can be subdivided into user-based methods and item-based methods [3].

2.2 Model Based Methods

Model-based approaches provide a systematic way to train a predefined compact model in the training phase that explains observed ratings, which is then used to make predictions. Usually, model-based collaborative filtering methods can achieve better performance. There are both rating-oriented methods and ranking-oriented methods. There are various models like aspect models, the latent factor models, the Bayesian hierarchical model, restricted Boltzmann machines, SVD++, Probabilistic Matrix Factorization (PMF), multi-domain collaborative filtering, graphical models, pair-wise tensor factorization and matrix factorization with social regularization, etc. For real-world recommender system dealing with large scale data, low-rank matrix approximation method performs well [4].

3. RESPONSE AWARE COLLABORATIVE FILTERING

In online rating systems, user's ratings have twofold information. Rating value indicates a user's likeliness on a specific item and also an item's hidden properties. The ratings that user give to different items shows information of user's interest. The rating values that an item received from different users also carry information on intrinsic properties of the item. The ratings also disclose user's response patterns whether the items are rated or not. This data can be used to improve the model performance. However, previously proposed methods usually assume that all the users would rate all the inspected items, or more generally, randomly select inspected items to rate. These methods fit the user's ratings directly and ignore the key factor, user's response patterns. The ignorance will degrade the model performance. We explore last ignored response information to further boost recommender system's quality. Practically, the assumption of all examination or arbitrarily rate is false in real-world ranking schemes. Users will not rate all the selected items or arbitrarily select the visited items for rating.

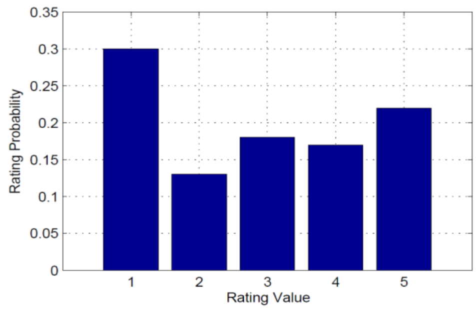

Fig.1 (a) User Selected Ratings

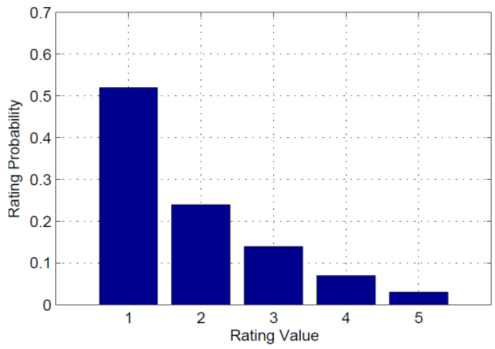

Fig.1 (b) Randomly Selected Ratings

Fig. 1 shows the distribution of rating scores collected from a real-world system, the Yahoo! Music's Launch Cast Radio service [5]. Fig. 1a shows the distribution of rating scores on those items that users choose to rate, while Fig. 1b shows the distribution of rating scores for the songs which are randomly selected from the whole music pool and asked for rating by the same group of users. From figure we can say that, user give higher ratings to the selected items whereas ratings given to randomly selected items is less.

RAPMF consist of two model, data model and response model. Here unobserved ratings are considered. Data model generates user ratings on items using inner product of two low rank featured matrices generated by using PMF and In response model response patterns are assumed based on whether the ratings are observed or not. In response model, deterministic Bernoulli parameter is given to the observed ratings and step function is assumed on the unobserved ratings [1].

4. EXPERIMENTS AND RESULTS

In a recommender system, there are three types of relations regarding an item to a user i.e. un-inspected, inspected-unrated, and inspected-rated. Old methods mainly focus on inspected rated data and do not consider response pattern. Various experimental protocols used are traditional protocol and realistic protocol and adversarial protocol. In traditional protocol, the exercise set and the trial set are arbitrarily selected from visited-rated items together with the users who have rated items and the matching rating scores. In realistic protocol, the training set is randomly selected from observed-rated items, but the trial set is arbitrarily selected from un-observed items. In adversarial protocol, the exercise set is arbitrarily selected from non-visited-rated items; nonetheless the test set is arbitrarily selected from visited-unrated items.
In the experiment, we use root mean square error (RMSE), a popular metric in collaborative filtering [6] to evaluate the performance of different models.

Experiments are performed on both a synthetic dataset and two real-world datasets. The synthetic dataset provides both benchmark information on user-item ratings and user item response patterns. Hence, we can use it to evaluate all the compared models under all three evaluation protocols. The real-world Yahoo! dataset is collected from Yahoo! Music's Launch Cast Radio service and is particularly prepared for evaluating the realistic protocol in real-world recommender systems [5]. For other benchmark real-world datasets, we select MovieLens to evaluate the performance on traditional protocol.

4.1 Model Performance

The average performance for all datasets is given in Table 1. For the synthetic dataset, 10 independent trials are run and test them on three protocols. For the Yahoo! dataset, we run 10 independent trials on the traditional protocol and 10 different initialization on the realistic protocol. We do not test it under the adversarial protocol because we do not have the inspected-unrated information. For the MovieLens dataset, we only test the traditional protocol on the default five-fold separating data.
The improvement shows the improvement of RAPMF-r over other models in percentage (%). Sensitivity analysis is carried out under the realistic setting in single test [7].

For the synthetic dataset, the best performance attains when σ=0.04, while it is 0.008 for the Yahoo! dataset, and 0.02 for the MovieLens dataset.

5. CONCLUSIONS

Collaborative Filtering techniques are gaining more importance now days, because there is lot of information on the internet but user want specific information which can be provided by using collaborative filtering techniques. Existing systems uses only visited items hence gives biased recommendations therefore, we introduced a new method which considers non visited items while giving recommendations. Recommendations given using response aware probabilistic matrix factorization method are more accurate.

REFERENCES

[1] IEEE Transaction on "Boosting Response Aware Model-Based Collaborative Filtering".

[2] Z. Huang, H. Chen, and D. Zeng, "Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering," ACM Trans. Inf. Syst., vol. 22, no. 1, pp. 116–142, 2004.

[3] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," Adv. Artif. Intell., vol. 2009, 2009, pp. 1–19.

[4] C. Cheng, H. Yang, I. King, and M. R. Lyu, "Fused matrix factorization with geographical and social influence in location-based social networks," in Proc. 6th AAAI Conf. Artif. Intell., Toronto, ON, Canada, 2012, pp. 17–23.

[5] B. M. Marlin, R. S. Zemel, S. T. Roweis, and M. Slaney, "Collaborative filtering and the missing at random assumption," in Proc. 23rd Conf. Uncertainty Artif. Intell., 2007, pp. 267–275.

[6] R. Salakhutdinov and A. Mnih, "Probabilistic matrix factorization," in Proc. Adv. Neural Inf. Process. Syst., 2007, pp. 1257–1264.

[7] G. Ling, H. Yang, I. King, and M. R. Lyu, "Response aware modelbased collaborative filtering," in Proc. 28th Conf. Uncertainty Artif.Intell., 2012, pp. 501–510.