International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 04 | Apr 2019 www.irjet.net p-ISSN: 2395-0072

Boosting Response Aware Model-Based Collaborative Filtering

Sabirahmad Attar[1], AnujaAwati[2], Prof. C.S Arage[3], Prof. M.P. Gaikwad[4]

[1,2]*Student: CSE, Sanjay GhodawatInstitutes*
[3,4]*Asst.prof CSE Department, Sanjay Ghodawat Institutes*

---***---

Abstract—*Recommender systems are promising for providing personalized favourite services. Collaborative filtering (CF) technologies, making prediction of users' preference based on users' previous behaviours, have become one of the most successful techniques to build modern recommender systems. Several difficult problems occur in antecedently planned CF strategies: First is most CF methods ignore users' response patterns and should yield biased parameter estimation and suboptimal performance; Second is some CF methods adopt heuristic weight settings, which lacks a systematical implementation; Third is the multinomial mixture models may weaken the computational ability of matrix factorization for generating the data matrix, thus increasing the computational cost of training. To resolve these issues, we incorporate users' response models into the probabilistic matrix factorization (PMF), a popular matrix factorization CF model, to establish the Response Aware Probabilistic Matrix Factorization (RAPMF) framework. More specifically, we make the assumption on the user response as a Bernoulli distribution which is parameterized by the rating scores for the observed ratings while as a step function for the unobserved ratings. Moreover, we have a tendency to speed up the algorithmic program by a mini-batch implementation and a crafting programming policy. Finally, we design different experimental protocols and conduct systematically empirical evaluation on both synthetic and real-world datasets to demonstrate the merits of the proposed RAPMF and its mini-batch implementation.*

Keywords — Cold start technique, recommended system, mini-batch learning, collaborative filtering, RAPMF.

1 INTRODUCTION

RECENTLY, online shopping and entertainment services are growing explosively. Popular service providers, e.g., Amazon, Netflix, iTunes Match, Yahoo! Music, etc., have contributed to building up platforms for consumers to buy new products or rate them. As a coin has two sides, these platforms can provide users attractive services to improve their lifestyle, they also introduce inundated choice which increases users' information overload. Matching consumers' taste and presenting the most appropriate products to them is a key to enhance users' satisfaction and loyalty in using these online services. Hence, recommender systems, providing personalized favourite recommendations, have been prevalently adopted in these services to boost the sales of retailers and trigger the growth of business. Due to the prominence of the commercial value and technical challenges, recommender techniques have attracted the interests of researchers from academia and practitioners from industry [2], [4], [6]. Collaborative filtering (CF) technologies, aiming to automatically predict consumers' preferences by analyzing their previous behaviours, e.g., the transaction history or product ratings, become mainstream techniques for recommender systems. These techniques can usually be classified into memory-based CF methods and model based CF methods, see [2], [6] and the references therein. Overall, previously proposed CF methods mainly focus on manipulating the explicitly observed rating scores to understand users' preferences for future prediction. An explicit rating score clearly indicates a user's preference on a particular item as well as an item's inherent features. The scores that a user assigns to different items convey information on what the user likes and what the user dislikes. 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. However, valuable implicit information of users'response patterns, i.e., some items are rated while others not, is usually less explored in existing CF methods. Several pieces of research publications have been conducted to exploit users' response patterns. For example, the original problem is formulated as the one-class collaborative filtering task, where a heuristic weight in the range of 0 to 1 is introduced to calibrate the loss on those unseen ratings [8], [9] or the user information is embedded to optimize the weight on the unseen ratings via users' similarity. The multinomial mixture model is combined with conditional probability tables with Bernoulli distribution to model the non-random response. This work is also extended to specify the probability that a rating is missing in a logistic form which depends on both the values of the underlying ratings and the identity of the items. The previous work, however, may suffer from some practical limitations: 1) the heuristic weight setting methods may lack a systematic way to model users' response patterns; 2) the multinomial mixture models may weaken the computational ability of generating data matrix and increase the computational cost of training the model. To overcome the above limitations, in this paper, we propose a Response Aware Probability Matrix Factorization (RAPMF) framework by expanding the Bernoulli response patterns to probability matrix factorization for users' ratings in [13]. Different from previously proposed methods, we present a succinct assumption on response patterns and further investigate the properties and effectiveness of the proposed RAPMF. We highlight the key contributions of this article as follows: • First, our proposed RAPMF framework consists of a data model and a response model. The data model generates users' ratings on items via the inner product of two low-rank feature matrices captured by probabilistic matrix factorization (PMF), a popular model-based matrix factorization method. Meanwhile, the response model is assumed following a Bernoulli distribution. That is, the response patterns are assumed based on whether the ratings are observed or not, where a deterministic Bernoulli parameter is given to the observed ratings while a step function is assumed on the unobserved ratings. The treatment of the response for unobserved ratings allows us to marginalize the underlying response and the data model on the unobserved ratings. This assumption is more precise and easy to marginalize the missing responses. This is slightly different from the setup in [13], where the responses may depend on the latent features. • Second, we seek the optimal solution of RAPMF by gradient descent, which consumes the time complexity of O(N × M), where N and M are the number of users and items, respectively. It is too expensive for real-world recommender systems, which contain over millions or even billions of users and items. To resolve the computational issue, we realize a mini-batch implementation for RAPMF and reduce its training cost to O(B2) for each mini-block with B users and B items. The mini batch is executed in parallel via multiple threads with a crafting scheduling policy. In an extreme case, when B = 1, the algorithm is equivalent to a parallel implementation of stochastic gradient ascent. Hence, the mini-batch implementation will reduce the training consumption of RAPMF largely while maintaining the same test cost. • Third, we design different experimental protocols to reveal different distributions on the training data and the test data. We conduct model evaluation on both synthetic and real-world datasets under different protocols to compare the model performance. Our experimental results demonstrate that our proposed RAPMF contains several merits. The rest of the paper is organized as follows. In Section 2, we present the preliminaries on the basic model setup and a motivating example, review several existing work, and the motivation of the work. In Section 3, we develop the proposed RAPMF model on how to incorporate response models into PMF and elaborate its properties. In Section 4, we present the mini-batch learning implementation for RAPMF. In Section 5, we conduct comparison on the models and present detailed explanation. Finally, we conclude the paper in Section 6.

2. PRELIMINARIES AND RELATED WORK

Response patterns

	i1	i2	i3	i4	i5
u1	0	1	0	0	0
u2	0	1	0	1	0
u3	1	0	1	1	1
u4	0	0	0	0	0
u5	1	1	0	1	1

In the following, we will first present the basic setup and the objective of this paper with a motivatingexample. After that, we will review three main topics related to our work.

Ratings

	i1	i2	i3	i4	i5
u1	0	5	0	0	0
u2	0	4	0	5	0
u3	4	0	4	4	5
u4	0	0	0	0	0
u5	5	4	0	5	5

2.1 Setup

Example Let D = {1,2,...,D} be the set of rating scores (grades) in the range 1 to D. For example, in the Yahoo!Music's LaunchCast dataset, D is 5 and therefore the rating values range from 1 (indicating no interest) to 5 (implying a strong interest). Collecting all data of N users and M items from a recommender system can form an N ×M matrix X, where a row of the matrix indicates a user's ratings on the items and a column of the matrix represents the ratings on a specific item. Usually, the observed matrix X is highly sparse. For example, in the Yahoo!Music's LaunchCast dataset, only about 2% of the ratings are observed. Formally, we denote Ω as the set of the indexes of the observations in X and likewise ¯ Ω for the unobserved data. Hence, we separate X into two sets, XΩ and X¯ Ω, for the observed ratings and unobserved ratings, respectively, where Xij = a ∈D, if (i,j) ∈ Ω 0, if (i,j) ∈ ¯ Ω . (1) Correspondingly, we can then construct the fully observed response matrix R as Rij = 1, if (i,j) ∈ Ω 0 if (i,j) ∈ ¯ Ω (2) Hence, R = RΩ ∪R¯ Ω and RΩ ∩R¯ Ω = ∅. In most of previously proposed CF methods, users response patterns are ignored, which is equivalent to assuming the missing of users' ratings on items occurs randomly. That is, all users would rate all the inspected items, or more generally they will randomly select the inspected items to rate. It should be noted that in real-world recommender systems, this assumption may be violated. To verify this phenomenon, we show in Fig. 1 for the distributions of rating scores collected from a real-world system, the Yahoo!Music'sLaunchCast Radio service. The distribution of rating scores on those items that users choose to rate, while 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. Obviously, these two distributions are dramatically different. For those songs that the users have rated, more items are rated on high scores than those randomly selected from the music pool. This is a compelling evidence showing that the assumption that all the users would rate all the inspected items or randomly select items to rate is unlikely to be true. The investigation of the Yahoo!Music LaunchCast data indicates that users are more likely to rate items they do love or hate than those neutral to them. Table 1 again gives us a vivid example of skewed ratings of five users on five items and their corresponding response patterns. This extreme case (ratings skewed to either 4 or 5) clearly shows that without considering users' response patterns, user-based approaches [3] and item-based approaches [4] are more likely to predict rating values in the range of 4 to 5. The extreme example implies that the response patterns have to be taken into account to enhance model performance. Hence, in this paper, we aim to boost the model performance by exploiting both partially-observed rating matrix $X\Omega$ and fully-observed response matrix R.

2.2 Missing Data Theory

In the literature, missing data theory has established a systematic framework to explore missing response patterns. In the following, we review this theory and elaborate how it can be utilized in collaborative filtering because ignoring the missing responses will yield biased parameter estimation. Following missing data theory, we can model the available data in Eq. (1) and Eq. (2) as a two-step procedure. First, a data model $P(X|\theta)$ parameterized by θ generates the full data matrix X. Then, a response model $P(R|X,\mu)$ determines which element sin X are observed. Hence, we can take a parametric joint probability on the partially observed data matrix $X\Omega$ and the fully observed response matrix R, conditioned on the model parameters, θ and μ as follows: $P(R,X\Omega|\mu,\theta) = P(R|X\Omega,\mu,\theta)P(X\Omega|\mu,\theta)$ (3) $= P(R|X\Omega,\mu)P(X\Omega|\theta)$ In Eq. (3), the response parameter, μ, is not related to users' rating and therefore we discard it in calculating the probability $P(X\Omega|\theta)$. The probability, $P(R|X\Omega,\mu)$ is also referred to as the missing data model. In the following, we use response model and missing data model interchangeably. According to the missing data theory, there are three kinds of missing data assumptions:

• **Missing Completely At Random (MCAR):** This is the strongest independence assumption. Whether there is a response is fully determined by a parameter, which is irrelevant to users' ratings and the model's latent variables. One typical example where MCAR holds is that given an inspected item, whether it will be observed or not is a Bernoulli trail with probability μ. That is, $P(R|X,\mu) = P(R|X\Omega,\mu) = P(R|\mu)$. (4)

• **Missing At Random (MAR):** The probability of observing a particular response can only depend on the observed elements of the data vector. In other words, the probability of response is not related to missing data values or the model's latent variables. The assumption can be formulated as follows: $P(R|X,\mu) = P(R|X\Omega,\mu)$. (5)

• **Not Missing At Random (NMAR):** The response patterns depend on either the unobserved data vectors, the unobserved values of latent variables, or the observed data. This assumption requires an explicit response model to learn unbiased model parameters. The response model has to be incorporated into the data model for estimating missing ratings

2.3 Collaborative Filtering Techniques

Collaborative filtering (CF) approaches are effective recommendation techniques to filter out irrelevant information only based on users' previous behaviours and to provide items/products that users may be interested [2], [3],[6]. Due to effective performance, they have been successfully deployed in various real-world recommender systems [4], [6]. Based on different assumptions, CF approaches are usually classified into two main categories: memory-based methods and model based methods [2], [3], [6]. Memory-based methods are very popular and applied widely in commercial websites [4], [7]. These methods make predictions based on users' previous ratings to compute similarity between users or items. They can further be classified into user-based methods and item-based methods with the facts that neighbour users share similar tasks and users tend to assign similar ratings to similar items, respectively [3], [7]. The success of memory-based methods relies on accurately computing the paired similarity between users and items from previously observed ratings. However, for those unobserved ratings, the information is discarded. The response patterns are usually ignored in these methods. Some other methods, e.g., nearest neighbour regression [6], may be able to correctly identify relevant neighbours for a user or an item in the presence of nonrandom missing data using common similarity measures like Pearson correlation. If data are not missing at random, these models will yield the predicted results bias. Clearly, as referred to the data in Table 1, user based approaches [3] and item-based approaches [4] are more likely to predict rating values in the range of 4 to 5. Model-based approaches, instead of manipulating the ratings directly, train a predefined compact model based on partially-observed user-item rating data to recover the whole matrix. Various models lie in this category, including the aspect models the latent factor model [4] the Bayesian hierarchical model restricted Boltzmann machines SVD++ multi-domain collaborative filtering pair-wise tensor factorization and matrix factorization with social regularization etc. Among model-based approaches, low-rank matrix approximation methods have demonstrated their efficiency and good performance for real-world recommender systems in dealing with large scale data [5], [10]. Currently, there are two main streams of work trying to include the response patterns in the CF methods. One line of work is to explore the response patterns into the one-class collaborative filtering task [8], [9]. SVD++ with implicit feedback [5] follows similar framework, but embedded users' rating and un-rating behaviours by a latent unknown matrix. In these methods, when the ratings are unobserved, a heuristic weight in the range of 0 to 1 is introduced to calibrate the loss [8], [9] while the ratings are set to 0. The weight on the unseen ratings is also optimized by calculating users' similarity from the embedded users' profile information. However, these methods do not directly explore users' missing response patterns and integrate them with the ratings. The other line of work models the response patterns through missing data theory. In the multinomial mixture model is combined with conditional probability tables with Bernoulli distribution to model the non-random response. This work is also extended to specify the probability that a rating is missing in a logistic form which depends on both the value of the underlying rating and the identity of the item. These methods model users' ratings matrix via the multinomial mixture model and discard the effectiveness and interpretability of the matrix factorization approaches [5], [10]. The PMF for users' data generation model has also incorporated the Bernoulli response patterns [13]. However, the assumption on the missing response patterns can further be simplified. The insufficiency of previous work motivates our exploration of the missing response patterns and matrix factorization model in this paper.

2.4 Online Learning

Online learning is a family of efficient and scalable machine learning algorithms [11]. Different from traditional batch-trained learning algorithms which require that all training data are available prior to the learning task, online learning promptly update the predictive model when a new instance appears [12]. It can avoid the cost of retraining effort largely when a new instance appears. Hence, it is more appropriate for recommender systems to capture users' preference as in real-world systems, ratings are obtained sequentially. Nowadays, online learning has been extended and explored in collaborative filtering. The implementation of these algorithms can be categorized as Perception like algorithms and stochastic gradient descent approaches [1], [8]. In the work casts online collaborative filtering as an online ranking algorithm. However, it requires to know users' all preferences and it is undesirable for real-world scenario. The implementation of collaborative filtering in stochastic gradient descent is efficient and allows to scale for building on the-fly recommender systems. In online collaborative filtering is conducted on the Probabilistic Latent Semantic Analysis (PLSA) model for the personalized news recommendation. In [1], online collaborative filtering is performed on the low-rank approximation matrix factorization with and without feature models. In online collaborative filtering is also conducted on low rank approximation matrix factorization models to attain good rating fitting and ranking orders. The online collaborative filtering framework is also extended in to deem each user a search task and reformulate the problem as a multi-task learning problem. The performance is boosted by including users' similarity information as the relationship of tasks. The previously proposed work is also possible and promising for parallelization. The success of previous work motivates us to explore the implementation of stochastic gradient descent in RAPMF. However, since our RAPMF requires the information of both observed and unobserved ratings, to allow balance, we consider the mini-batch method which requires a small bunch of data to update the models, but enjoys the additional advantages of parallelization speedups.

CONCLUSION

In this paper, we established a Response Aware Probabilistic Matrix Factorization (RAPMF) framework to unify users' response behaviours and a popular model-based collaborative filtering technique, Probabilistic Matrix Factorization. More specifically, we model the probability of whether a user will rate an item by engaging the Bernoulli distribution, where the parameters are determined by the rating scores, and assume the corresponding probability of unobserved ratings as a step function. More significantly, we speed up the algorithm by mini-batch implementation and conduct a crafting scheduling policy on automatically selecting free-blocks to further accelerate the original batch-trained RAPMF algorithm. Empirically, we verify the performance of RAPMF under carefully designed experimental protocols and show that RAPMF performs best when it tries to fulfill the ultimate goal of real-world recommender systems, i.e., recommending items to those users who do not see the items before, but may be interested in them. The empirical evaluation demonstrates the potential of our RAPMF model in real-world recommender system deployment. There are several interesting directions worthy of future consideration. The first direction is to incorporate other side information to boost the model performance. Another promising avenue is to investigate how to model the response when the response patterns are hidden. For example, in location-based social network applications such as Foursquare and Google Latitude, the check-in frequency can be obtained explicitly, but the response regarding whether a user will check-in a place or not is unknown. The third direction is to design a smart way in efficiently tuning the hyper-parameters or to design a learning scheme in obtaining the model parameters automatically.

REFRENCES:

[1] J. Abernethy, K. Canini, J. Langford, and A. Simma. Online collaborative filtering. Technical report, University of California at Berkeley, 2007.

[2] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Trans. Knowl. Data Eng., 17(6):734–749, 2005.

[3] J. S. Breese, D. Heckerman, and C. M. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In UAI, pages 43–52, 1998.

[4] G. Linden, B. Smith, and J. York. Amazon.com recommendations: Item-to-item collaborative filtering. IEEE Internet Computing, 7:76– 80, January 2003.

[5] Y. Koren. Factor in the neighbors: Scalable and accurate collaborative filtering. ACM Trans. Knowl. Discov. Data, 4(1):1:1–1:24, Jan. 2010.

[6] X. Su and T. M. Khoshgoftaar. A survey of collaborative filtering techniques. Adv. Artificial Intellegence, 2009, 2009.

[7] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. Grouplens: An open architecture for collaborative filtering of netnews. In CSCW, 1994.

[8] R. Pan and M. Scholz. Mind the gaps: weighting the unknown in large-scale one-class collaborative filtering. In KDD, pages 667– 676, 2009.

[9] R. Pan and M. Scholz. Mind the gaps: weighting the unknown in large-scale one-class collaborative filtering. In KDD, pages 667– 676, 2009.

[10] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. In NIPS, 2007.

[11] H. Yang, M. R. Lyu, and I. King. Efficient online learning for multi-task feature selection. ACM Transactions on Knowledge Discovery from Data, 2013.

[12] P. Zhao, S. C. H. Hoi, and R. Jin. Duol: A double updating approach for online learning. In NIPS, pages 2259–2267, 2009.

[13] G. Ling, H. Yang, I. King, and M. R. Lyu. Response aware modelbased collaborative filtering. In UAI, pages 501–510, 2012.