HAL
open science

Online Shopping Recommendation with Bayesian Probabilistic Matrix Factorization

Jinming Wu, Zhong Liu, Guangquan Cheng, Qi Wang, Jincai Huang

To cite this version:

Jinming Wu, Zhong Liu, Guangquan Cheng, Qi Wang, Jincai Huang. Online Shopping Recommendation with Bayesian Probabilistic Matrix Factorization. 2nd International Conference on Intelligence Science (ICIS), Oct 2017, Shanghai, China. pp.445-451, 10.1007/978-3-319-68121-4_48 . hal-01820909

HAL Id: hal-01820909

https://inria.hal.science/hal-01820909v1

Submitted on 22 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Online Shopping Recommendation with Bayesian Probabilistic Matrix Factorization

Jinming Wu, Zhong Liu, Guangquan Cheng*, Qi Wang, Jincai Huang

Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology
cgq299@nudt.edu.cn

Abstract. Recommendation system plays a crucial role in demand prediction, arousing attention from industry, business, government and academia. Widely employed in recommendation system, matrix factorization can well capture the potential relationships between users, items and latent variables. In this paper, we focus on a specific recommendation task on the large scale opinion-sharing online dataset called Epinions. We carried out recommendation experiments with the Bayesian probabilistic matrix factorization algorithm and the final results showed the superior performance in comparison to six representative recommendation algorithms. Meanwhile, the Bayesian probabilistic matrix factorization was investigated in depth and the potential advantage was explained from the model flexibility in parameters' adjustment. The findings would guide further research on applications of Bayesian probabilistic matrix factorization and inspire more researchers to contribute in this domain.

Keywords: Recommendation system, Bayesian Probabilistic Matrix Factorization, Monte Carlo Markov Chain

1 Introduction

The past several decades have witnessed the great necessity in designing novel recommendation systems, keeping pace with the fast development of information technologies. The increasing popularity of recommendation systems has influenced human life in many aspects. The potential value of recommendation is being excavated in fruitful applications, such as online social voting recommendation on social networks like Facebook, research paper recommendation for accelerating the speed of crucial scientific findings[1], multimedia recommendation for satisfying users' interest[2], citation recommendation for sorting relevant papers[3]. Interestingly, these commercial or academic applications have generalized well and at the same time brought us convenience in a personalized way. However, the demand has already confronted difficulties and challenges from both data quality and volume, requiring more novel algorithms or frameworks to deal with. A specific instance in recommendation research lies in that the ratings on items by users tend to be sparse and imbalanced, leading to the cold start phenomenon in the domain. On the other hand, the boom of internet allows the interactive behaviors like rating the goods bought before or feeding back the preferences of

items such as foods, movies or music. In China, some famous corporations including Alibaba, Xiecheng Company have perceived the importance of feed-backing comments from customers and there is no doubt that accurate predictions on customers' preferences no matter in the future or later would bring much profits. It is obvious that the potential relationships between the users we take interests in may contribute to the recommendation in some sense and such social information involved in the model forms the representative family of recommendation algorithms as collaborative filtering.

To address these issues, we firstly make a brief survey regarding the probabilistic matrix factorization (PMF) algorithm which is proved to be effective in sparse and imbalanced recommendation dataset with the social information well exploited [4]. Distinguished from the traditional methodologies which mainly capture the information from the user-item ratings, we combined the social information with the rating circumstances for the performance promotion in our study and a superior adaptive Bayesian based model was employed in shopping recommendation.

The remainders of paper are arranged as follows. Section2 elaborates the related developments of probabilistic matrix factorization and Bayesian based recommendation system. The experiments on the collected dataset using Bayesian probabilistic matrix factorization are described in section3 with some detailed analysis. Section4 summarizes our work and light the direction in our future research.

2 Related Work and Employed Methodology

2.1 Review on Probabilistic Matrix Factorization

Probabilistic matrix factorization (PMF) is regarded as one of the outstanding algorithms in several recommendation competitions like Netflix Prize, compared to several collaborative filtering methodologies. The essence of the PMF is to make use of the low rank approximation and assume the potential distribution of dataset, overcoming the shortness of matrix factorization model. After both of user and item feature matrix obtained by low-rank approximating the rating matrix, we utilize the feature matrix to complete the potential missing ratings. Such model maintains well prediction accuracies in practice.

Due to the advantages of PMF, it has successfully attracted researchers' attention as well as interest. D. Dueck et al. proposed a probabilistic sparse matrix factorization algorithm, which well explained the uncertainty of noise in different extents[5]. Based on the matrix factorization ideology, R. Salakhutdinov et al. made assumption that user-feature matrix, item-feature matrix and user-item matrix were Gaussian in prior distribution, and then by modelling the latent variables, the PMF were derived in first time[6]. Multiple rating indexes were integrated into a user-item weighted matrix, assuming the latent matrix obeyed the Gaussian distribution and density probabilistic distribution conditionally correlated with user-feature and item-feature matrixes[7]. Another adaptation on the PMF was SoRec recommendation model[8], where user-feature vectors from both PMF model and confidence based matrix factorization model were combined to enhance the performance. Apart from the theoretical research on the PMF,

some applications have also driven the interest of adaptation on PMF. Kernel probabilistic matrix factorization [7][9].

2.2 A Brief introduction to the Bayesian Probabilistic Matrix Factorization

In this section, we would focus our attention on the mechanism of Bayesian probabilistic matrix factorization (BPMF) algorithm proposed by R Salakhutdinov et al. [10] and the essence of the recommendation algorithm would be elaborated.

The PMF is the basic model for the BPMF. Generally speaking, the PMF is a linear probabilistic model for solving recommendation tasks with some plausible information containing N users, M items and N*M rating matrix as R. Missing elements in R are ratings which the user never response to the item. The element r_{ij} in R records the preference score of the user i towards the item j.

We apply the low rank approximation to the rating matrix R and then obtain two matrix in low dimensions as

$$\mathrm{R} \cong \mathrm{U}^T * V(1)$$

Where the D*N matrix U and D*M matrix V are respectively user-feature matrix and item-feature matrix in low rank representations, describing the latent representations of users and items respectively.

The probabilistic graphical model for PMF can be translated in the Fig1.

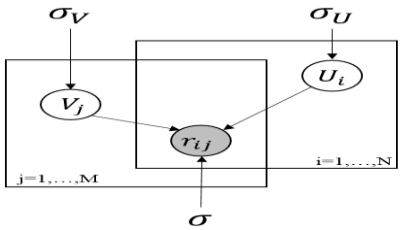

Fig. 1. The probabilistic graphical model of PMF

Assuming that the difference between the actual rating r_{ij} and the predicted rating $r_{ij}^{\sim}$ obeys the Gaussian distribution with mean 0 and variance σ, we derive

$$\mathrm{r}_{\mathrm{ij}} - U_i^T V_j \sim Gaussian(0, \sigma^2)(2)$$

By the constant shifting, the distribution of rating r_{ij} can be represented as

$$\mathrm{r}_{\mathrm{ij}} \sim Gaussian(U_i^T V_j, \sigma^2)(3)$$

Meanwhile, the assumption on the prior distribution on both latent user and latent item is the Gaussian distribution as follows:

$$\mathrm{p}(\mathrm{U}|\sigma_{\mathrm{u}}^2) = \Pi_{i=1}^{N} N(U_i|0, \sigma_U^2 I)(4)$$

$$\mathrm{p}(\mathrm{V}|\sigma_{\mathrm{V}}^2) = \Pi_{j=1}^{M} N(V_j|0, \sigma_V^2 I)(5)$$

Based on the Bayesian inference, we can acquire the loss function by minimizing the log of the posterior distribution accompanied with quadratic regularization terms.

$$\mathrm{E} = \frac{1}{2}\Sigma_{i=1}^{N}\Sigma_{j=1}^{M} I_{ij}\left(r_{ij} - U_i^T V_j\right)^2 + \frac{\lambda_U}{2}\Sigma_{i=1}^{N}\left|\left|U_i\right|\right|_{Fro}^2 + \frac{\lambda_V}{2}\Sigma_{j=1}^{M}\left|\left|V_j\right|\right|_{Fro}^2 \quad (6)$$

Where the equations follow $\lambda_{\mathrm{U}} = \frac{\sigma^2}{\sigma_{\mathrm{U}}^2}$ and $\lambda_{\mathrm{V}} = \frac{\sigma^2}{\sigma_V^2}$.

During the past years in the research, the adaptations on the feature dimension D and prior distribution's assignment make great improvements on the capability of PMF's generalization.

For the BPMF, the U and V are assumed to follow the Gaussian distributions instead of constants.

$$\mathrm{p}(\mathrm{U}|\mu_{\mathrm{U}}, \mathrm{Z}_U) = \Pi_{i=1}^{N} N(U_i|\mu_U, \mathrm{Z}_U^{-1}) \quad (7)$$

$$\mathrm{p}(\mathrm{V}|\mu_{\mathrm{V}}, Z_V) = \Pi_{j=1}^{M} N(V_j|\mu_V, \mathrm{Z}_V^{-1}) \quad (8)$$

Furthermore, the Wishart distribution W_0 is introduced as the priors for hyperparameters $\Theta_{\mathrm{U}} = \{\mu_U, \mathrm{Z}_U\}$ and $\Theta_{\mathrm{V}} = \{\mu_V, \mathrm{Z}_V\}$.

$$\mathrm{p}(\Theta_{\mathrm{U}}|\mu_0, v_0, W_0) = p(\mu_U|\mathrm{Z}_U)p(\mathrm{Z}_U) \quad (9)$$

$$\mathrm{p}(\Theta_{\mathrm{V}}|\mu_0, v_0, W_0) = p(\mu_V|\mathrm{Z}_V)p(\mathrm{Z}_V) \quad (10)$$

Where we set $\mathrm{v}_0 = D$, $\mu_0 = 0$ and $\mathrm{W}_0 = \mathrm{I}_{\mathrm{D}}$.

In the prediction process of BPMF, the rating can be obtained by marginalizing over model parameters and hyper parameters as follows.

$$\mathrm{p}\left(\mathrm{r}_{\mathrm{ij}}^*\middle|R, \mu_0, v_0, W_0\right) = \int\int p\left(r_{ij}^*\middle|U_i, V_j\right)p(U, V|R, \mu_U, \mathrm{Z}_U, \mu_V, \mathrm{Z}_V)p(\mu_U, \mathrm{Z}_U, \mu_V, \mathrm{Z}_V|\mu_0, v_0, W_0)d\{\mu_U, \mathrm{Z}_U, \mu_V, \mathrm{Z}_V\} \quad (11)$$

In our research, the Monte Carlo approximation is utilized for the computation.

$$\mathrm{p}\left(\mathrm{r}_{\mathrm{ij}}^*\middle|R, \mu_0, v_0, W_0\right) \approx \frac{1}{K}\Sigma_{k=1}^{K} p(r_{ij}^*|U_i^k, V_j^k) \quad (12)$$

Where the samples $\{U_i^k, V_j^k\}$ are generated from a Markov chain. Detailed information on MCMC's application to the BPMF refers to [10].

3 Experiments and Analysis

3.1 Evaluation Metrics and Selected Dataset

In our research, the root means square error (RMSE) is chosen as the evaluation metric. More specifically, for the rating r_{ij} which measures the user i's preference towards the item j and the predicted rating $r_{ij}^{\sim}$ derived by the recommendation system, the RMSE is computed as

$$\mathrm{RMSE} = \left(\Sigma_{\mathrm{i,j}\in\mathrm{T}}\left(r_{ij} - r_{ij}^{\sim}\right)^2\right)^{\frac{1}{2}} / |T| (13)$$

Due to the severe penalty on inaccurate predictions using square error in RMSE, the evaluation metric is regarded as a strict but reasonable criteria.

The dataset in our experiments is from the electronic business website as Epions (www.Epinions.com). The ratings reflecting the preference of users towards items are collected, processed and prepared in user-item matrix. Another crucial information matrix describes the friend relationship between users is also provided in Table1.

Table 1. Related information about Epinions

Total Users	Total Items	Density of Ratings	Density of User Relationships	Averaged Ratings on Items	Averaged Friends
49289	139738	664824 (0.0097%)	487183 (0.0201%)	4.76	9.88

In comparison to the commonly used dataset like Movielens and IMBD, the Epnions behaves rather sparse in social relationships.

3.2 Results and Analysis

To verify the effectiveness of the BPMF, we selected some frequently used popular algorithms which were believed to perform well in several recommendation tasks as baselines. Specifically, they were SVD[11], NMF[12], BaselineOnly[13], SlopeOne[14], KNNBaseline[13] and CoClustering[15]. In the experiments, both of the parameters λ_U and λ_V were set as 0.002 and the turns of iteration was 200. Some other parameters were selected as the default. Aiming at determining the optimal feature dimension D, we ranged the parameter from 5 to 20 at the step width 5 and the 5-fold cross validation was performed in final selection. The results were arrange in the Table2.

Table 2. The RMSE in different D value

D	5	10	15	20
BPMF	1.2232	1.1678	0.9978	1.4471

From the table, we can notice that the value of RMSE decreased steadily in the small interval while such metric fluctuated severely when D was over 15. So 15 was chosen as the optimal value for the feature dimension.

A persuasive metric is the k-fold evaluation to average the testing results, in which the k-1 folds are prepared for model fitting with the last fold for testing and such manipulation is performed k times in one turn by ranging all of k-1 folds.

Furthermore, we carried out experiments using the above-mentioned seven algorithms in 10 independent turns and provided the results in average. The fold partition ranged from 1 to 5 with the step width 1. The Table3 recorded the averaged RMSE in ten turns using various algorithms in different folds and it revealed the superior performance of BPMF with the lowest RMSE.

Table 3. Averaged k-fold RMSE in different algorithms

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
NMF	1.1168	1.1186	1.1173	1.1153	1.1151
SVD	1.0914	1.0941	1.0969	1.0926	1.0935
BaselineOnly	1.1084	1.1086	1.108	1.1094	1.1091
SlopeOne	1.1062	1.1061	1.1067	1.1082	1.1057
KNNBaseline	1.0963	1.0942	1.0967	1.0936	1.0933
CoClustering	1.1155	1.1169	1.1163	1.1142	1.1128
BPMF	0.9959	0.9955	0.9951	0.9952	0.9954

To further validate the outstanding performance of BPMF in statistics, the paired t-test was introduced to test the results of 10 independent turns with seven algorithms in the case when the 5-fold RMSE was chosen as the metric. Specifically, the 10 results in each turn obtained by BPMF were compared with those derived by NFM, SVD, BaselineOnly, SlopeOne, KNNBaseline and Coclustering respectively with the help of SPSS. The testing results indicated the BPMF outperformed others significantly. And the compared results were mapped into three element tuple as win\tie\loss. Then the counts of win\tie\loss in 10 turns with some baseline algorithm were summarized in the Table4 as follows.

Table 4. Win\Tie\Loss Statistical Results in Comparison Experiments

	NMF	SVD	BaselineOnly	SlopeOne	KNNBaseline	CoClustering
BPMF	9\1\0	8\1\1	10\0\0	9\0\1	8\2\0	10\0\0

The potential reason for obtaining satisfying results using BPMF was that the Bayesian framework was more flexible in searching optimal hyper parameters and the model maintained high complexity to fit the dataset of high volume.

4 Conclusions

In the end, the Bayesian probabilistic matrix factorization was employed in online item recommendation backgrounded in the opinion-sharing platform. The BPMF showed its superiority in recommendation task, deriving the lowest RMSE in comparison to six representative algorithms. The conclusions were supported by the paired t-test in ten independent experiments and the persuasive results motivated us to research in depth in other applications. In the future, we would design more variant matrix factorization algorithms incorporating the other social information to promote the performance in recommendations.

References

1. Beel, J., Gipp, B., Langer, S., Breitinger, C.: Research-paper recommender systems: a literature survey. International Journal on Digital Libraries 17, 1-34 (2016)
2. Lee, W.P., Tseng, G.Y.: Incorporating contextual information and collaborative filtering methods for multimedia recommendation in a mobile environment. Multimedia Tools & Applications 1-21 (2015)
3. Liu, H., Kong, X., Bai, X., Wang, W., Bekele, T.M., Xia, F.: Context-Based Collaborative Filtering for Citation Recommendation. IEEE Access 3, 1695-1703 (2015)
4. Fuchs, M., Zanker, M.: Multi-criteria Ratings for Recommender Systems: An Empirical Analysis in the Tourism Domain. Lecture Notes in Business Information Processing 123, 100-111 (2012)
5. Dueck, D., Frey, B.J.: Probabilistic Sparse Matrix Factorization. University of Toronto Technical Report Psi (2004)
6. Salakhutdinov, R., Mnih, A.: Probabilistic Matrix Factorization. In: International Conference on Neural Information Processing Systems, pp. 1257-1264. (2007)
7. Yang, X., Guo, Y., Liu, Y., Steck, H.: A survey of collaborative filtering based social recommender systems. Computer Communications 41, 1-10 (2014)
8. Ma, H., Yang, H., Lyu, M.R., King, I.: SoRec:social recommendation using probabilistic matrix factorization. In: ACM Conference on Information and Knowledge Management, CIKM 2008, Napa Valley, California, Usa, October, pp. 931-940. (2008)
9. Zhou, T., Shan, H., Banerjee, A., Sapiro, G.: Kernelized Probabilistic Matrix Factorization: Exploiting Graphs and Side Information. SDM (2012)
10.Salakhutdinov, R., Mnih, A.: Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. In: International Conference on Machine Learning, pp. 880-887. (2008)
11.Kim, J.K., Cho, Y.H.: Using Web Usage Mining and SVD to Improve E-commerce Recommendation Quality (2003)
12.Luo, X., Zhou, M., Xia, Y., Zhu, Q.: An Efficient Non-Negative Matrix-Factorization-Based Approach to Collaborative Filtering for Recommender Systems. IEEE Transactions on Industrial Informatics 10, 1273-1284 (2014)
13.Koren, Y.: Factor in the neighbors: Scalable and accurate collaborative filtering. Acm Transactions on Knowledge Discovery from Data 4, 1-24 (2010)
14.Sun, L., Jingjiao, L.I.: SlopeOne Collaborative Filtering Recommendation Algorithm Based

on Dynamic k-Nearest-Neighborhood. Journal of Frontiers of Computer Science & Technology 5, 857-864 (2011)
15.George, T., Merugu, S.: A scalable collaborative filtering framework based on co-clustering. In: IEEE International Conference on Data Mining, pp. 4 pp., (2005)