LETTER

Integrated Collaborative Filtering for Implicit Feedback Incorporating Covisitation

Hongmei LI†, Xingchun DIAO†, Jianjun CAO††a), Yuling SHANG†, *Nonmembers*, *and* Yuntian FENG†, *Member*

SUMMARY Collaborative filtering with only implicit feedbacks has become a quite common scenario (e.g. purchase history, click-through log, and page visitation). This kind of feedback data only has a small portion of positive instances reflecting the user's interaction. Such characteristics pose great challenges to dealing with implicit recommendation problems. In this letter, we take full advantage of matrix factorization and relative preference to make the recommendation model more scalable and flexible. In addition, we propose to take into consideration the concept of covisitation which captures the underlying relationships between items or users. To this end, we propose the algorithm Integrated Collaborative Filtering for Implicit Feedback incorporating Covisitation (ICFIF-C) to integrate matrix factorization and collaborative ranking incorporating the covisitation of users and items simultaneously to model recommendation with implicit feedback. The experimental results show that the proposed model outperforms state-of-the-art algorithms on three standard datasets.
key words: *collaborative filtering, implicit feedback, matrix factorization, covisitation, relative preference*

1. Introduction

Traditionally, collaborative filtering (CF) is one of the most successful recommendation technologies, which aims at predicting the ratings on items by considering users' explicit ratings on items, and make recommendations by ranking items in the order of ratings [1], [2]. However, in many practical recommendation systems, most feedbacks are not explicit but implicit, such as clicks, collections, visitations and so on. They are more convenient and effortless for people to express themselves than explicit ratings. Compared with explicit ratings, implicit feedbacks are more pervasive and easy to collect, but this kind of data only contains a small portion of positive examples, that is we only know users' likes other than dislikes, which is imbalanced and presents a new challenge to recommendation [3], [4], [7]. Hence, how to tackle the massive, high-dimensional and only positive data is an important issue in collaborative filtering.

Many popular algorithms for traditional CF, such as matrix factorization (MF), have attracted many attentions for its effectiveness and efficiency and variant MF tailored for implicit feedback recommenders have been proposed. MF aims at predicting users' ratings for items and performing personalized ranking indirectly. For instance, Pan, et al. [4] proposed the frame based on weighted low rank matrix factorization by treating all the missing data as negative examples. Hu, et al. [5] propose a SVD model by treating data as indication of positive and negative preference associated with varying confidence levels. Notice that the above methods all conform to the assumption that the unobserved data are negative examples, which inevitably brings noises into the very model and degrades learning efficiency due to the consideration of both observed and unobserved data. In order to address the issues, [6] proposed to weight the missing data based on item popularity, and enhanced the effectiveness to some extent. Shi, et al. [9] proposed Collaborative Less-is-More Filtering mode by directly optimizing the ranking measurement, that is mean reciprocal rank.

Apart from the above methods designed for the item prediction of indirect personalized ranking, the method based on learning to rank aims at directly optimizing for ranking. For example, Rendle, et al. [7] introduced an optimization criterion for personalized ranking based on maximum posterior estimator derived from a Bayesian analysis of the problem. Then, Rendle, et al. [8] introduced an improving pairwise ranking learning method for item recommendation from implicit feedback.

Owing to the favorable performance of both matrix factorization methods and learning to rank methods, we propose a novel model utilizing their advantages to tailor recommendation with only implicit feedbacks. First, inspired by [3], we propose to introduce relative preference degree to represent and rank users' feedback behavior. This is due to the fact that a user selects an item other than the others just because he/she has a relatively higher preference degree towards to the item than others. Then, we translate the recommendation problem into probability optimization considering relative preference degree represented by matrix factorization.

In addition, in order to enhance recommendation performance, we also try to incorporate the information underlying in user-item interactions implying the neighboring relationships in the latent semantic space among items or users. Here, we introduce the concept of item covisitation proposed by [10], where the covisitation is defined as an event in which two items are clicked by the same user within a certain time interval. The covisitation of items reflects the neighbor relationship to some degree. Inspired by that, we also propose the concept of user covisitation, which implies that the users visiting the same item during a certain time span may have similar preference. After that, we try to in-

Manuscript received January 29, 2017.
Manuscript publicized April 17, 2017.
†The authors are with PLA University of Science and Technology, Nanjing, China.
††The author is with Nanjing Telecommunication Technology Institute, Nanjing, China.
a) E-mail: xinxizhiliang@163.com
DOI: 10.1587/transinf.2017EDL8020

corporate covisitation information into CF recommendation model appropriately through the graph Laplacian based on manifold learning [11]. To the best of our knowledge, there is no similar work applying it to the personalized recommendation with implicit feedback. To this end, we propose the algorithm Integrated Collaborative Filtering for Implicit Feedback incorporating Covisitation (ICFIF-C) to integrate matrix factorization and collaborative ranking incorporating item and user covisitation information simultaneously to model recommendation with implicit feedback.

2. Problem Definition and Notations

Assume i users give implicit feedbacks on j items, and the feedbacks are available in preference matrix $\mathbf{P} \in \{0,1\}^{n\times m}$. $\mathbf{P}_{ij}$, denotes the implicit feedback of user i for item j. $\mathbf{P}_{ij} = 1$ denotes the observed data representing a user has preference on some certain item, and $\mathbf{P}_{ij} = 0$ for the otherwise situation. $\mathbf{U} \in \mathbf{R}^{n\times k}$ and $\mathbf{V} \in \mathbf{R}^{m\times k}$ respectively represents latent feature matrices of n users and m items with k dimensional latent feature vectors. Our goal is to predict the preference of users for items, and recommend a ranking list.

3. Integrated Collaborative Filtering Model Based on Matrix Factorization and Relative Preference

3.1 Relative Preference Degree

Definition 1 Preference Degree (PD)

Inspired by matrix factorization, we assume that the preference of user i on item j is determined by a small number of unobserved factors. Preference Degree PD_{ij} is conducted by nonlinearly combining user-specific factor vector $\mathbf{U}_i$ with item-specific factor vector $\mathbf{U}_j$ using sigmoid function, and it is formulated as

$$PD_{ij} = \frac{\exp(\mathbf{U}_i\mathbf{V}_j^T)}{\exp(\mathbf{U}_i\mathbf{V}_j^T) + 1} \tag{1}$$

Where the rating PD_{ij} is mapped to the interval (0, 1). However, the preference degree is somehow absolute, and may be insufficient to exactly represent users' preference to some specific item. Intuitionally, a user selects an item just because he/she has a relatively higher preference towards the item than others. Specifically, if preference degree of user i on item j is far higher than others, user i would have a higher probability to choose item j. So we give a definition of relative preference degree.

Definition 2 Relative Preference Degree (RPD)

Relative Preference Degree RPD_{ij} denotes relative preference of user i towards the item j. It characterizes the relative level of the overall preference degree on all the items and is formulated as

$$RPD_{ij} = \frac{PD_{it}}{\sum_{t=1}^{m} PD_{it}} \tag{2}$$

Where m denotes all the items in the datasets, and $\sum_{t=1}^{m} PD_{it}$ means overall preference degree on all the items.

3.2 Probabilistic Model for Collaborative Filtering

Given the relative preference degree, we then define the probability distribution over all users formulated as

$$p(\mathbf{R}|\mathbf{U},\mathbf{V}) = \prod_{(i,j)\in S} f(RPD_{ij}|\mathbf{U},\mathbf{V}) \tag{3}$$

Where $f(x) = 1/(1+\exp(-x))$ is the sigmoid function. S denotes the whole set of implicit feedbacks. We place zero-mean spherical Gaussian priors on user and item factor vectors as follows

$$p(\mathbf{U}|\sigma_U^2) = \prod_{i=1}^{n} N(\mathbf{U}_i|0,\sigma_U^2\mathbf{I}) \tag{4}$$

$$p(\mathbf{V}|\sigma_V^2) = \prod_{j=1}^{m} N(\mathbf{V}_j|0,\sigma_V^2\mathbf{I}) \tag{5}$$

Then, based on Bayesian theory, the log-posterior distribution over the user and item factors is given distribution over all users formulated as

$$\begin{aligned} &\ln p(\mathbf{U},\mathbf{V}|\mathbf{R}) \propto \ln p(\mathbf{R}|\mathbf{U},\mathbf{V}) \ln p(\mathbf{U}) \ln p(\mathbf{V}) \\ &= -\sum_{(i,j)\in S} \ln(1+\exp(-RPD_{ij})) \\ &- \frac{1}{2\sigma_U^2}\|\mathbf{U}\|_F^2 - \frac{1}{2\sigma_V^2}\|\mathbf{V}\|_F^2 + C_0 \end{aligned} \tag{6}$$

Where $\|.\|_F^2$ denotes Frobenius norm, C_0 is a constant that does not depend on the parameters. Learning in this model is performed by maximizing the log-posterior over the user and item factors, and is equivalent to minimizing the objective function formulated as

$$\begin{aligned} &\min_{\mathbf{U},\mathbf{V}} E = \sum_{(i,j)\in S} \ln(1+\exp(-RPD_{ij})) \\ &+ \lambda_{\mathbf{U}}\|\mathbf{U}\|_F^2 + \lambda_{\mathbf{V}}\|\mathbf{V}\|_F^2 \end{aligned} \tag{7}$$

Where λ_U, λ_V denotes the regularization parameter to avoid overfitting. Compared with traditional CF method, our method maximizes the probability of user's relative preference other than absolute rating. This kind of treatment, on one hand, fits the initiation of users' implicit feedback behavior, on the other, avoids introducing noise when assuming negative examples.

4. Recommendation Model Incorporating Covisitation

4.1 Covisitation

Definition 3 Item covisitation

Item covisitationis an event in which two items are clicked by the same user within a certain time interval. It captures the simple intuition: "Users who viewed this item also viewed the following items" [10]. For instance, in news recommendation domain, the news browsed by a user in a time period is likely to present the same or similar topic.

The scores of covisitation can be easily obtained from users' implicit feedbacks with timestamp. Concretely, each element of the covisitation matrix $\mathbf{C}_V$ represents the time discounted number of covisitation instances, which can be measured by the count of how often they were covisited by the same user in a short time slot. The higher value of $\mathbf{C}_V$ means the more similarity or correlation between items.

The covisitation matrix is symmetric since we do not care about the order. Inspired by Item covisitation, we also propose to give the definition of User covisitation.

Definition 4 User covisitation

User covisitation captures the simple intuition: "users visiting the same item during the same period of time may have significant similarity with each other". For instance, in point-of-interest recommendation, users tend to have similar preference visiting the same position in an approximate time period. User covisitation matrix $\mathbf{C}_U$ can be calculated in the similar way as $\mathbf{C}_V$.

We can see that covisitation matrix could capture the underlying relationship among items and users respectively. It breaks limit of independency of users and items in traditional methods.

4.2 Recommendation Model with Graph Regularization Incorporating Covisitation

To incorporate the covisitation relationship information into the recommender system, we consider to integrate graph Laplacian [11] into the optimization subject function formulated as Eq. (7) by adding an extra regularization term, so as to penalize the differences between the item-specific and user-specific latent feature vectors when there exists covisitation between them. Graph Laplacian [10] is capable of capturing the underlying relationships between the similar objects based on manifold learning. Specifically, Item graph Laplacian is defined as

$$\mathbf{L}_{CV} = diag(\mathbf{C}_V \mathbf{1}) - \mathbf{C}_V \tag{8}$$

Where $\mathbf{1}$ denotes a vector with all one elements. $diag(.)$ is a diagonal matrix. Then item graph Laplacian can be integrated into subject Eq. (7) by adding regularization term $tr(\mathbf{V}^T\mathbf{L}_{CV}\mathbf{V})$ into the minimization problem. Analogically, user graph Laplacian can also be integrated into subject function by adding regularization term $tr(\mathbf{U}^T\mathbf{L}_{CU}\mathbf{U})$. The minimization problem is formulated as

$$\begin{aligned}\min_{\mathbf{U},\mathbf{V}} E &= \sum \ln(1+\exp(-RPD_{ij}))^{O_{ij}} \\ &+ \lambda_{\mathbf{U}}\|\mathbf{U}\|_F^2 + \lambda_{\mathbf{V}}\|\mathbf{V}\|_F^2 \\ &+ \eta_{\mathbf{U}} tr(\mathbf{U}^T\mathbf{L}_{CU}\mathbf{U}) + \eta_{\mathbf{V}} tr(\mathbf{V}^T\mathbf{L}_{CV}\mathbf{V})\end{aligned} \tag{9}$$

Where $\eta_{\mathbf{U}}$ and $\eta_{\mathbf{V}}$ denote regularization parameters. It can be clearly observed that the covisitation information is naturally embedded into regularization for the item-specific latent factors and incorporated into the naive model. We call our model as ICFIF-C. Then, we perform gradient descent using learning rate α in $\mathbf{U}$ and $\mathbf{V}$ to minimize objective function given by Eq. (9).

5. Experimental Results and Analysis

In this section, we evaluate the effectiveness of the proposed methods on two standard datasets: Movielens100K† and Foursquare [12], which both contain time stamps of the ratings. For Movielens100K, we relabel ratings 4 and 5 as 1 and remove other ratings below 4. For Foursquare, we draw a subsample that every user has at least 10 checks and each check has at least 20 users. We evaluate the performance of different methods on the test set using two metrics: Mean Average Precision (MAP) [4] and Area Under Curve (AUC) [7]. The higher value of MAP and AUC means the better performance.

For MAP, we randomly divide data into training and test datasets with a 80/20 splitting ratio [4]. For AUC, we use leave one out scheme to construct train and test set [7]. We compare our method ICFIF, ICFIF-C with two baselines: WMFM (Weighted Matrix Factorization Model) [4], and IFRM (Recommendation Model for Implicit Feedback) [3].

In the setup of the experiments, the parameters $\mathbf{U}$ and $\mathbf{V}$ are initialized by randomly sampling from norm distribution. Our ICFIF-C has five parameters:$\{\lambda_{\mathbf{U}}, \lambda_{\mathbf{V}}, \eta_{\mathbf{U}}, \eta_{\mathbf{V}}, \alpha\}$. In the process of selecting parameters, we adjust one parameter while fix the others. The ranges of parameters are: $\lambda_{\mathbf{U}} \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$, $\lambda_{\mathbf{V}} \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$, $\eta_{\mathbf{U}} \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$, $\eta_{\mathbf{V}} \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$, $\alpha \in \{10^{-3}, 5*10^{-3}, 10^{-2}, 5*10^{-2}, 10^{-1}\}$. The resulting parameter values for Movielens100K under MAP are: $\lambda_{\mathbf{U}} = \lambda_{\mathbf{V}} = 10^{-2}$, $\eta_{\mathbf{U}} = 10^{-4}$, $\eta_{\mathbf{V}} = 10^{-3}$, $\alpha = 5*10^{-3}$, and $\lambda_{\mathbf{U}} = \lambda_{\mathbf{V}} = 10^{-2}$, $\eta_{\mathbf{U}} = 10^{-3}$, $\eta_{\mathbf{V}} = 10^{-2}$, $\alpha = 5*10^{-3}$ under AUC. The resulting parameter values for Foursquare under MAP are: $\lambda_{\mathbf{U}} = \lambda_{\mathbf{V}} = 10^{-2}$, $\eta_{\mathbf{U}} = 10^{-3}$, $\eta_{\mathbf{V}} = 10^{-4}$, $\alpha = 10^{-3}$, and $\lambda_{\mathbf{U}} = \lambda_{\mathbf{V}} = 10^{-2}$, $\eta_{\mathbf{U}} = 10^{-2}$, $\eta_{\mathbf{V}} = 10^{-3}$, $\alpha = 10^{-3}$ under AUC. We evaluate the performance of different methods on different datasets along with changed number of latent features k ranging from 10 to 50. After selecting the optimal parameters, we repeat each compared experiment for 10 iterations and compute the average value of results. The results on Movielens100K dataset is shown in Table 1 and Table 2, and the results on Foursquare dataset is shown in Table 3 and Table 4.

From the experimental results, we can obtain the following conclusions: 1) Generally, our method ICFIF-C performs better than others with a stable and slightly increasing performance along with feature numbers. The results demonstrate the effectiveness of model incorporating covisitation information. 2) On both of the datasets, the recommendation accuracy of ICFIF is better than the other two traditional methods. This demonstrates the priority of our integrated model in consideration of users' relevant prefer-

†http://grouplens.org/datasets/movielens

Table 1 The recommendation accuracy under MAP (%) on Movielens100K dataset.

Features / Algorithm	10	20	30	40	50
WMFM	12.4085	12.7356	12.7356	13.0462	13.0462
IFRM	**13.0462**	12.8066	12.8066	12.8066	13.1987
ICFIF	12.5901	**12.8876**	**13.0514**	13.1855	13.2707
ICFIF-C	12.6894	12.6894	12.6894	**13.2601**	**13.3483**

Table 2 The recommendation accuracy under AUC on Movielens100K.

Features / Algorithm	10	20	30	40	50
WMFM	0.8551	0.8606	0.8644	0.2356	0.8644
IFRM	0.8597	0.8645	0.8677	0.8697	0.8701
ICFIF	0.8619	0.8659	0.8695	0.8718	0.8720
ICFIF-C	**0.8628**	**0.8675**	**0.8718**	**0.8737**	**0.8746**

Table 3 The recommendation accuracy under MAP (%) on Foursquare dataset.

Features / Algorithm	10	20	30	40	50
WMFM	8.3685	8.3685	8.8123	8.9302	9.0451
IFRM	8.4536	8.7099	8.9053	9.0566	9.0949
ICFIF	8.5000	8.7523	8.9387	9.1045	9.1434
ICFIF-C	**8.6513**	**8.8711**	**8.9909**	**9.1581**	**9.2113**

Table 4 The recommendation accuracy under AUC on Foursquare.

Features / Algorithm	10	20	30	40	50
WMFM	0.7362	0.7408	0.7424	0.7447	0.7456
IFRM	0.7401	0.7444	0.7458	0.7469	0.7473
ICFIF	0.7448	0.7475	0.7480	0.7489	0.7492
ICFIF-C	**0.7488**	**0.7521**	**0.7534**	**0.7547**	**0.7556**

ence by matrix factorization. 3) On both of the datasets, ICFIF-C performs better than other models, which demonstrates the significance of incorporating item and user covisitation information. This phenomenon reveals that items visited in some short period by the same user present certain similarity which can be captured by item covisitation. Besides, users visiting the same items in an approximate time span present similar preference, and it can be rightly captured by user covisitation.

6. Conclusion

In this letter, we propose an integrated collaborative filtering for implicit feedback incorporating covisitation, named ICFIF-C. The proposed method can harness the advantages of matrix factorization and relevant preference to fully fit the scenario with only implicit feedbacks with great scalability and flexibility. Furthermore, ICFIF-C can incorporate covisitation and manifold structures to enhance the model. Through the learning process, ICFIF-C can finally find the robust latent factor vectors of users and items and perform recommendation. The experiments on the datasets prove the effectiveness of our proposed method.

Acknowledgments

This work was supported by the Natural Science Foundation of China under Grant 61371196. The authors would like to thank the anonymous reviewers for their constructive comments.

References

[1] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," Computer, vol.42, no.8, pp.30–37, 2009.

[2] K. Verstrepen and B. Goethals, "Unifying nearest neighbors collaborative filtering," Proc. 8th ACM RecSys Conference on Recommender Systems, pp.177–184, ACM, 2014.

[3] J. Yin, Z.S. Wang, Q. Li, and W.J. Su, "Personalized recommendation based on large-scale implicit feedback," Journal of Software, vol.25, no.9, pp.1953–1966, 2016 (in Chinese).

[4] R. Pan, Y. Zhou, B. Cao, N.N. Liu, R. Lukose, M. Scholz, and Q. Yang, "One-class collaborative filtering," Proc. 8th IEEE International Conference on Data Mining, Pisa, Italy, pp.502–511, 2008.

[5] Y. Hu, Y. Koren, and C. Volinsky, "Collaborative filtering for implicit feedback datasets," Proc. 8th IEEE ICDM International Conference on Data Mining, pp.263–272, IEEE, 2008.

[6] X. He, H. Zhang, M.-Y. Kan, and T.-S. Chua, "Fast matrix factorization for online recommendation with implicit feedback," Proc. 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.549–558, ACM, 2016.

[7] S. Rendle, C. Freudenthaler, Z. Gantner, and L.S. Thieme, "BPR: Bayesian personalized ranking from implicit feedback," Proc. 25th Conf. Uncertainty in Artificial Intelligence, pp.452–461, 2009.

[8] S. Rendle and C. Freudenthaler, "Improving pairwise learning for item recommendation from implicit feedback," Proc. 7th ACM International Conference on Web Search, Data Mining, pp.273–282. ACM, 2014.

[9] Y. Shi, A. Karatzoglou, L. Baltrunas, M. Larson, N. Oliver, and A. Hanjalic. "CLiMF: Collaborative less-is-more filtering," Proc. 23rd ACM Conf. Artificial Intelligence, pp.3077–3081, 2013.

[10] A.S. Das, M. Datar, A. Garg, and S. Rajaram, "Google news personalization: Scalable online collaborative filtering," Proc. 16th International Conference on World Wide Web, WWW '07, pp.271–280, 2007.

[11] D.M. Cvetkovic, M. Doo, and H Sachs, Spectra of graphs: Theory and application, Wiley, New York, 1998.

[12] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N.M. Thalmann, "Time-aware point-of-interest recommendation," Proc. 36th International ACM Conference on Research and Development in Information Retrieval, pp.363–372, ACM, 2013.