http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (14): 3025-3029, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.3025.3029

Improved Graph Model Algorithm for E-Commerce Evaluation and Recommendation

Zhi-yuan Li
School of Management, Harbin University of Science and Technology, Harbin, 150000, China

Abstract: As information technology rapidly develops, personalized recommendation has now been widely used in E-commerce. Therefore, to effectively utilize the large amount of item information for evaluation and recommendation has become the research focus. In this study, a comprehensive evaluation and personalized recommendation approach is presented based on Graph Model Algorithm. To demonstrate the comprehensiveness and flexibility of this model, three methods of recommendations are compared and the principle of projection matrix is applied in this study, maximizing the initial information. The empirical results show that the proposed model in this study is practical, accurate and efficient, giving more recommendation information. Moreover, it can represent different combinations of items, users and the transaction information and it has the potential to accommodate many graph search techniques for evaluations and recommendations.

Key words: Evaluation and recommendation, graph model algorithm, information retention, maximization, E-commerce

INTRODUCTION

With the continuous advancement of information technology and the rapid growth of network data, the overwhelmed product information online presents great challenges to E-commerce. E-consumers are greatly influenced by the product evaluation and recommendation made by the previous users while E-businesses often find it difficult to promote products appropriate to specific users (Li *et al*., 2013). Therefore, how the large amount of item information can be effectively utilized to support better decision-making has become the focus of the researchers who often face two challenges. One is how to represent the diverse information about the users and items. The other is how to build a model that is flexible enough to incorporate recommendation approaches. Evaluation and recommendation, both as the information processing, are consistent in an effective approach to reflect the available information more comprehensively and effectively. Since evaluation is to seek similarities and recommendation is to keep differences, their integration can reserve as much information as possible and reflect the link between items and users.

An improved Graph Model Algorithm (GMA) proposed in this study can fully represent the user-item information. It contains nodes (users and items) and links (transactions and similarities) that capture various types of e-commerce information. The methodology is thoroughly demonstrated in evaluation and recommendation respectively and a systematic empirical study is conducted to test the validity of the proposed GMA.

LITERATURE REVIEW

Generally, evaluation and recommendation approaches develop into four categories, namely, Collaborative Filtering Approach (CFA), Content-based Approach (CBA), GMA and hybrid approach (Zhang *et al*., 2013). The rating matrix of CFA is usually very sparse, greatly reducing the recommendation accuracy (Liu *et al*., 2011). CBA has difficulty in extracting multimedia information features and it only recommends users to buy similar items (Chen and Wang, 2013). For the sustainable development of E-commerce, these problems need to be solved urgently.

The evaluation and recommendation based on GMA use the user-item relations of graph structure to more effectively solve the problems above. GMA which makes use of the selection relations between users and items to dynamically allocate the items' resources, is relatively new. Aggarwal *et al*. (1999) introduced a directed graph of users in recommender systems, where the directed edges corresponded to the notion of predictability. Based on this graph, the personalized recommendations could be generated via a few reasonably short (strongly predictable) directed paths joining multiple users. Mirza *et al*. (2003) also presented a novel framework to study recommendation algorithms in terms of the 'jumps' that they made to connect people to artifacts. This approach emphasized the reachability via an algorithm within the implicit graph structure underlying a recommender dataset and allowed them to consider questions relating algorithmic parameters to properties of the datasets. They illustrated the approach with a common jump called the 'hammock' using movie recommender datasets. Flexible recommendation approaches were also to be explored using such a graph model. Li *et al*. (2010) constructed a unified GAM for sentence-based opinion retrieval and compared with the existing approaches and the experimental results showed that their graph-based model achieved significant improvement. Zhang *et al*. (2010) built a personalized recommendation via integrated diffusion on user-item-tag tripartite graphs.

In summary, to address the needs of the comprehensive evaluation and to support the flexible recommendation approaches, an E-commerce evaluation and recommendation model based on GMA has an advantage on solving the sparsity problem with both high accuracy and little computing time.

METHODOLOGY

Problem description and similarity

Problem description: The online evaluation and recommendation systems are simplified as the following model. M users evaluate N items in α-point scale (usually α = 5). The rating vector given by each user is $v^m_{\alpha N\times 1}$, where m∈{1, 2, ..., M}. When User m doesn't evaluate Item K:

$$\begin{aligned} &V^m_{1\times 1} = 0,\\ &1\in\{ak-a+1, ak-a+2, ..., ak\}\\ &k\in\{1, 2, ..., N\}\end{aligned} \tag{1}$$

when User mevaluates Item K:

$$\begin{aligned} &V^m_{1\times 1} = 1, && 1 = ak-z+1\\ &V^m_{1\times 1} = 0, && 1\in\{ak-a+1, ak-a+2, ..., ak\}\setminus\{ak-z+1\}\\ & && k\in\{1, 2, ..., N\},\quad z\in\{1, 2, ..., a\}\end{aligned} \tag{2}$$

The rating matrix $A^m_{\alpha N\times\alpha N}$ of User M is as follows:

$$A^m_{\alpha N\times\alpha N} = \frac{v^m_{\alpha N\times 1}(v^m_{\alpha N\times 1})^T - \text{diag}(v^m_{\alpha N\times 1})}{n^m - 1} \tag{3}$$

where, n^m represents the total number of items that User m evaluates and $\text{diag}(v^m_{\alpha N\times 1})$ represents the diagonal matrix with the diagonal as $v^m_{\alpha N\times 1}$.

Then sum all the users' rating matrix to get the system rating matrix $A_{\alpha N\times\alpha N}$.

$$A_{\alpha N\times\alpha N} = \sum_{m=1}^{M} A^m_{\alpha N\times\alpha N} = \sum_{m=1}^{M} \frac{v^m_{\alpha N\times 1}(v^m_{\alpha N\times 1})^T - \text{diag}(v^m_{\alpha N\times 1})}{n^m - 1} \tag{4}$$

Matrix $A^m_{\alpha N\times\alpha N}$ is the information processing of rating system transformed into mathematical expression with the matrix which views the evaluation issue based on GMA. In GMA, an item is regarded as a point and the evaluation is regarded as a side to constitute the form of a graph. Since $A^m_{\alpha N\times\alpha N}$ is actually the matrix of the rating graph of User m and the matrix and graph express the same information, the scores and the quantity of the items evaluated can be obtained. The rating graph is the characterization of scores by the users in the system, without losing any information related to evaluation during the process.

Sum up $A^m_{\alpha N\times\alpha N}$ of all the users and then the rating matrix $A_{\alpha N\times\alpha N}$ of the entire system will be obtained. The matrix depicts the link between the items in the system, containing the information of both the users and the evaluated items. The link, namely, the side of the graph, is generated through the evaluation of each user. This approach has two significant properties. On one hand, $A^m_{\alpha N\times\alpha N}$ is a symmetric matrix, with the sum of the rows being 1 or 0. When the sum is 1, it means User m evaluates the item corresponding to the row. When the sum is 0, it means the corresponding properties are not supported. The sum of the columns means the same due to the symmetry of the matrix. On the other hand, the sum of rows of $A_{\alpha N\times\alpha N}$ is the sum of rows of all $A^m_{\alpha N\times\alpha N}$, whose value represents the number of the users in favor of the score of an item.

Methods of similarity: Following is a brief introduction to the recommendation algorithm based on the Item-based Similarity (IBS) and User-based Similarity (UBS) (Liu *et al*., 2010; Pan *et al*., 2010).

Method based on IBS: The recommendation method based on IBS can be defined as follows:

$$\text{sim}(\alpha,\beta) = \frac{\sum_{i\in U_\alpha\cap U_\beta}(v_{i,\alpha}-\bar{v}_\alpha)\cdot(v_{i,\beta}-\bar{v}_\beta)}{\sqrt{\sum_{i\in U_\alpha\cap U_\beta}(v_{i,\alpha}-\bar{v}_\alpha)^2}\sqrt{\sum_{i\in U_\alpha\cap U_\beta}(v_{i,\beta}-\bar{v}_\beta)^2}} \tag{5}$$

where, U_α and U_β respectively represent the user set who evaluate Item α and β, $v_{i,\alpha}$ and $v_{i,\beta}$, respectively represent the scores of Item α and β evaluated by User I, $v_{i,\alpha}$ and $\bar{v}_\alpha$ and $\bar{v}_\beta$, respectively represent the average score of Item α and β. Thus, the similarity sim(α, β) based on IBS can be defined. Ω_i is defined as the set of all the items evaluated by User i. Item γ is the item to be evaluated and k items have the highest similarity to Item γ. Define Φ as the set of k items and $\Theta\in\Omega_i$. Then $p_{i,\gamma}$ is the score of Item γ evaluated by User i defined as follows:

$$p_{i,\gamma} = \bar{v}_\gamma + \frac{\sum_{\alpha\in\Theta} sim(\gamma,\alpha)(v_{i,\alpha} - \bar{v}_\alpha)}{\sum_{\alpha\in\Theta} sim(\gamma,\alpha)} \quad (6)$$

Method based on UBS: The recommendation method based on UBS can be defined as follows:

$$sim(i,j) = \frac{\sum_{\alpha\in U_i\cap U_j}(v_{i,\alpha} - \bar{v}_i)\cdot(v_{j,\alpha} - \bar{v}_j)}{\sqrt{\sum_{\alpha\in U_i\cap U_j}(v_{i,\alpha} - \bar{v}_i)^2}\sqrt{\sum_{\alpha\in U_i\cap U_j}(v_{j,\alpha} - \bar{v}_j)^2}} \quad (7)$$

where, U_i and U_j, respectively represent the set of all the items evaluated by User i and User j, $\bar{v}_i$ and $\bar{v}_j$, respectively represent the average score of all the items evaluated by Users i and User j. Thus the similarity sim(i, j) based on UBS can be defined. Ω_α is defined as the set of all the evaluators on Item α. For User i, who does not evaluate the item, his score on Item α will be calculated based on UBS. Define Θ as the set of k users with the highest similarity to User i and $\Theta\in\Omega_\alpha$. Then $p_{i,\alpha}$ is the score of Item α by User i defined as follows:

$$p_{i,\alpha} = \bar{v}_i + \frac{\sum_{j\in\Theta} sim(i,j)(v_{j,\alpha} - \bar{v}_j)}{\sum_{j\in\Theta} sim(i,j)} \quad (8)$$

Graph model algorithm

Evaluation procedure of GMA: The most information of matrix $A_{\alpha N\times\alpha N}$ can be represented, reaching the maximum value through the column vector $X_{\alpha N\times 1}$ as follows:

$$\max \quad \frac{(A_{\alpha N\times\alpha N}X_{\alpha N\times 1}, X_{\alpha N\times 1})}{(X_{\alpha N\times 1}, X_{\alpha N\times 1})} \quad (9)$$

The maximization problem is equivalent to obtaining the maximum value of matrix $A_{\alpha N\times\alpha N}$ on the direction of vector $X_{\alpha N\times 1}$. The larger the value is, the more information is reflected. Therefore, the solution of the above equation leads to the solution of the largest eigenvalue of matrix $A_{\alpha N\times\alpha N}$ and the corresponding eigenvector (Bellman, 1970). Obviously, $A_{\alpha N\times\alpha N}$ is a non-negative irreducible matrix. According to Perron-Frobinus Theorem (Luenberger, 1979), if the sum of columns of $X_{\alpha N\times 1}$ is limited to 1, there exists the only solution for the above optimization problem.

The i-th element of $X_{\alpha N\times 1}$ is defined as x_i in $X_{\alpha N\times 1}$ obtained previously. The vector of each evaluation score of an item will be as follows:

$$\hat{x}_i = \frac{x_i}{\sum_{j=1}^{\alpha} x_{(\lceil \frac{i}{\alpha} \rceil - 1)\times\alpha + j}} \quad (10)$$

where, $\lceil \frac{i}{\alpha} \rceil$ means the ceiling of $\frac{i}{\alpha}$

Then the score g_n of Item n is as follows:

$$g_n = \sum_{i=1}^{\alpha} \hat{x}_{\alpha(n-1)+i} \times (\alpha + 1 - i) \quad (11)$$

Thus, this score, as the rating of an item on the website, can be a comprehensive rating reflecting the evaluation of the item by all the users.

Recommendation procedure of GMA: In order to infer the personalized recommendation method based on a user, it is necessary to delete the inconsistent evaluation information from the initial $A_{\alpha N\times\alpha N}$ of User m, obtaining $A^m_{\alpha N\times\alpha N}$. Let v^m_i as the i-th element of rating vector $v^m_{\alpha N\times 1}$ by User m.

If User m doesn't evaluate Item j, then:

$$A^m_{i,:} = A_{i,:}, j\in\{1,2,\cdots,N\},\ i\in\{(j-1)\alpha+1,(j-1)\alpha+2,\cdots,\alpha j\} \quad (12)$$

If User m evaluates Item j, then:

$$A^m_{i,:} = A_{i,:}, i = \alpha j + 1 - z,\ z\in\{1,2,\cdots,\alpha\} \quad (13)$$

$$A^m_{i,:} = 0,\ z\in\{1,2,\cdots,\alpha\},\ i\in\{(j-1)\alpha+1,(j-1)\alpha+2,\cdots,\alpha j\}\setminus\{\alpha j+1-z\} \quad (14)$$

Thus the defined $A_{\alpha N\times\alpha N}$ doesn't contain the inconsistent evaluation score by User m but retain the consistent information of those items that User m never evaluates. Similar to the evaluation algorithm in this study, replace $A_{\alpha N\times\alpha N}$ with $A^m_{\alpha N\times\alpha N}$, obtaining the score g^m_n of any item by User m. The evaluation and recommendation approaches both retain the most initial information through maximization.

As can be seen through the comparison of methodologies, the tradition UBS or IBS is not established on a comprehensive link perspective like GMA. UBS only focuses on all the evaluation information of one item and ignores the link between the item and other items established through the evaluation process. IBS ignores the link between users. Both UBS and IBS fail to represent all the links in the evaluation system. Theoretically and mathematically, the information processing method based on GMA retains more raw information compared with the existing evaluation and recommendation methods.

EMPIRICAL STUDY

Sample selection and data sources: In this study, the database of a movie recommendation website MovieLens (http://movielens.org/login) is applied. MovieLens, part of the Department of Computer Science and Engineering at

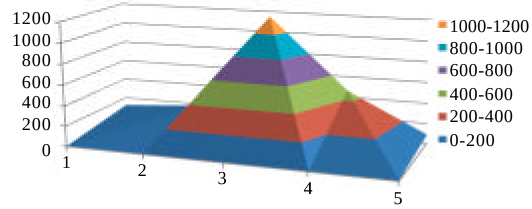

Fig. 1: Distribution of the evaluation scores

the University of Minnesota has thousands of users who have provided millions of ratings. MovieLens is an experimental platform to give effective movie recommendations. The movies are rated from 1 to 5 stars, with 1 as the worst and 5 as the best.

Program and calculation: Since maximization is equivalent to obtain the largest eigenvalue and the corresponding eigenvector. The proposed approach is fast and efficient and proper for solving the online evaluation and recommendation.

Evaluation generation: All the evaluation data of 2000 movies are calculated with the evaluation method proposed in this study and the algorithm above. The distribution of evaluation scores are shown as below.

It can be found from the figure that the number of 3-star is the largest of all, that of 4-star is the second. This figure shows the distribution of evaluation scores obtained with the evaluation method in this study.

Recommendation generation: In order to verify the validity and accuracy of the recommendation algorithm, the initial datasets are divided randomly, 80% as the sample set and 20% as the test set. The proposed recommendation algorithm is applied in the sample set, obtaining the corresponding scores of the movies. Compare the calculated score with the recommendation algorithm in the sample set and the actual scores in the test set. The accuracy of the model can generally be reflected by Mean Absolute Error (MAE) defined as follows:

$$\mathrm{MAE} = \frac{\sum_{n=1}^{N} | v_{i,\alpha}^{n} - \mathrm{round}(\hat{v}_{i,\alpha}^{n}) |}{N} \quad (15)$$

where, N is the total scores in the test set $v_{i,\alpha}^{n}$ represents Datum n, i.e., the score on Movie α by User i, $\hat{v}_{i,\alpha}^{n}$ represents the score on Movie α by User i based on the recommendation algorithm and round(*) indicates rounding.

Table 1: Comparisons of the three recommendation methods

Method	MAE	D (%)
GMA	0.77	38.10
IBS	0.86	30.90
UBS	0.84	31.50

Meanwhile, the recommendation accuracy D is defined as:

$$D = \frac{\aleph(| v_{i,\alpha}^{n} - \mathrm{round}(\hat{v}_{i,\alpha}^{n}) | = 0)}{N} \quad (16)$$

Where:

$$\aleph(| v_{i,\alpha}^{n} - \mathrm{round}(\hat{v}_{i,\alpha}^{n}) | = 0)$$

represents the number of the accurate recommendations. After the calculation, Table 1 shows the value of MAE and the value of D of the three evaluation methods, namely, GMA, IBS and UBS.

By comparing the data in Table 1, we can see that GMA has the highest accuracy of all the three methods and the lowest MAE. The higher the accuracy is and the lower the MAE is, the better performance of the model has. The accuracy of GMA is 38.1 and 7.2% higher than that of IBS, whose is only 0.06% lower than that of UBS. The MAE of GMA is 0.77, 0.09 lower than that of UBS, whose is only 0.02 lower than that of IBS. To sum up, although these three methods are useful to some extend in terms of rating, GMA has a large and obvious advantage over the other two recommendation methods.

CONCLUSIONS

Evaluation and recommendation approaches have been greatly applied to deal with overload information and to help utilize useful information online. In the principle of GMA, the evaluation information given by the users is transformed into a graph which retains the most initial evaluation information, laying the foundation for further comprehensive evaluations and personalized recommendations. In this study, we described a generic graph model for E-commerce evaluation and recommendation. To demonstrate the comprehensiveness and flexibility of this model, we experimented with three methods of recommendations and the principle of projection matrix is applied in this study, maximizing the retention of the information. The results showed that GMA achieved better performance than IBS and UBS. The proposed model in this study is practical, accurate and efficient, giving more recommendation information. Moreover, it can represent different combinations of items, users and the transaction information and it has the potential to accommodate many graph search techniques for evaluations and recommendations.

Future work may include (1) using sentiment lexicon-based method for opinion identification and evaluation; and (2) further mining the customer information in terms of gender, occupation, hobbies, etc., so that the graph can be multi-dimensional and the corresponding matrix can become the matroid.

ACKNOWLEDGEMENT

This study is sponsored by Special Missions in Humanities and Social Sciences Research funded by Department of Education, Heilongjiang, China (1252xs333).

REFERENCES

Aggarwal, C.C., J.L. Wolf, K.L. Wu and P.S. Yu, 1999. Horting hatches an egg: A new graph-theoretic approach to collaborative filtering. Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining, August 15-18, 1999, San Diego, CA., USA., pp: 201-212.

Chen, L. and F. Wang, 2013. Preference-based clustering reviews for augmenting e-commerce recommendation. Knowl. Based Syst., 50: 44-59.

Li, B., L. Zhou, S. Feng and K.F. Wong, 2010. A unified graph model for sentence-based opinion retrieval. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, July 11-16, 2010, Association for Computational Linguistics, Stroudsburg, PA, USA, pp: 1367-1375.

Li, Y.M., C.T. Wu and C.Y. Lai, 2013. A social recommender mechanism for e-commerce: Combining similarity, trust and relationship. Decis. Support Syst., 55: 740-752.

Liu, J.G., T. Zhou, B.H. Wang, Y.C. Zhang and Q. Guo, 2010. Degree correlation of bipartite network on personalized recommendation. Int. J. Modern Phys. C, 21: 137-147.

Liu, Z.X., N. Zhang and J.M. Li, 2011. One personal recommendation algorithm based on collaborative filtering and network structure. Complex Syst. Compl. Sci., 8: 45-50.

Luenberger, D.G., 1979. Introduction to Dynamic Systems. Wiley, USA.

Mirza, B.J., B.J. Keller and N. Ramakrishnan, 2003. Studying recommendation algorithms by graph analysis. J. Intell. Inform. Syst., 20: 131-160.

Pan, X., G.S. Deng and J.G. Liu, 2010. Information filtering via improved similarity definition. Chin. Phys. Lett., Vol. 27. 10.1088/0256-307X/27/6/068903

Zhang, L., L.S. Bai and T. Zhou, 2013. Crossing recommendation based on multi-b2c behavior. J. Univ. Electron. Sci. Technol. China, 42: 154-161.

Zhang, Z.K., T. Zhou and Y.C. Zhang, 2010. Personalized recommendation via integrated diffusion on user-item-tag tripartite graphs. Phys. A: Stat. Mech. Appl., 389: 179-186.

Bellman, R., 1970. Introduction to Matrix Analysis. McGraw Hill, New York.