Acta Technica 62 (2017), No. 7A, 497–506

A SVM recommendation model based on similarity evaluation and collaborative filtering of multi angle knowledge units

Liming Xu [1]

Abstract. The traditional user based collaborative filtering recommendation algorithm only considers the user's score when calculating the similarity between users, but ignores the differences between different projects. Aiming at the unsatisfactory performance of traditional methods in data sparseness, a collaborative filtering recommendation algorithm based on item label information is proposed for each user to select neighbors. Firstly, based on user rating matrix to determine the initial neighbor, calculate the target users for each target neighbor; when the minimum number of nearest neighbor scoring the goal of the project or not, consider adding the development from the tag information according to the nearest neighbor; for the target item rating prediction. The experimental results show that the algorithm improves the accuracy of similarity calculation, effectively alleviates the sparsity of user rating data, and improves the accuracy of prediction.

Key words. Recommendation system, Collaborative filtering, Similarity Measurement, User evaluation, Knowledge unit.

1. Introduction

In the recent decade, E-commerce has developed substantially and competition among merchants gradually requires merchants to actively master more accurate user demand and preference, thus pertinently providing users with services. Hence, a recommendation filtering algorithm with high accuracy and high performance becomes especially important [1]. Collaborative filtering recommendation can fully utilize relation between information and has high execution efficiency. It can get better recommendation result and thus become the hotspot of current research [2].

Domestic and foreign scholars and experts have conducted a lot of deep researches

[1] Beijing Wuzi University, Beijing, China

into collaborative filtering recommendation and up to now there are many collaborative filtering recommendation algorithms [3–6]. Each collaborative filtering recommendation algorithm has different working thoughts and has its own advantage and obvious defect in actual application: such as sparse data, cold starting, poor extensibility [7–9]. To solve these deficiencies, some scholars have proposed adopting association rule for data mining, Bayesian network, neural network, support vector machine and other technologies [10–12] to improve recommendation precision of recommendation system and obtain favorable recommendation result. However, users' interests are subject to comprehensive action and influence of many factors. In addition, the similarity value calculation of current collaborative filtering recommendation algorithm is not scientific, lacks rationality, ignores interest information of users. The recommendation precision is to be further improved [13].

2. Classical collaborative filtering recommendation algorithm

2.1. Classical collaborative filtering recommendation algorithm

The working steps of classical collaborative filtering recommendation algorithm are as follows:

(1) Establish evaluation matrix of item evaluation by users $R = \{r_{ij}\}_{m\times n}$, where m and n respectively indicate user number and item number, and r_{ij} is evaluation of user i for item j, which can be used to describe user judgment and preference, specifically as shown in Table 1.

Table 1. User-item evaluation matrix

	Item$_1$	Item$_2$	......	Item$_n$
User$_1$	r_{11}	r_{12}	......	r_{1n}
User$_2$	r_{21}	r_{21}	......	r_{21}
......	...	...	......	...
User$_m$	r_{m1}	r_{m1}	......	r_{m1}

(2) Obtain user similarity value according to evaluation value and sort similarity values to select k neighbors.

(3) Estimate users and item evaluation values according to k neighbors. Suppose "the nearest neighbor" of i is S_i and evaluation value of i for item x is P_{ix}, then:

$$P_{ix} = \overline{r_i} + \frac{\sum_{j\in S_i} sim(i,j) \times (r_{jx} - \overline{r_j})}{\sum_{j\in S_i} |sim(i,j)|}\,. \tag{1}$$

Where, $sim(i,j)$ indicates similarity between user i and user j; $\overline{r_i}$ and $\overline{r_j}$ indicate average values of all evaluated items in i and j.

2.2. Traditional similarity calculation method

Traditional similarity calculation methods mainly include: cosine similarity and correlation similarity

(1) Cosine similarity: user evaluation is a vector. When the user does not evaluate specific items, it can be deemed that the evaluation value is 0. The calculation formula of similarity $sim(i,j)$ between user i and user j is:

$$sim(i,j) = \frac{\sum\limits_{x=1}^{n} (r_{ix} \times r_{jx})}{\sqrt{\sum\limits_{x=1}^{n} r_{ix}^2}\sqrt{\sum\limits_{x=1}^{n} r_{jx}^2}}. \tag{2}$$

Where: r_{ix} and r_{jx} are respectively evaluation values of user i and user j for item x.

2) Pearson similarity: Pearson similarity only considers item set commonly evaluated by two users, deducting average value of all evaluated items. Common evaluation of two users i and j is I_{ij} $(I_{ij} = I_i \cap I_j)$. The calculation formula of Pearson similarity is:

$$sim(i,j) = \frac{\sum\limits_{x \in I_{ij}} (r_{ix} - \overline{r_i})(r_{jx} - \overline{r_j})}{\sqrt{\sum\limits_{x \in I_{ij}} (r_{ix} - \overline{r_i})^2}\sqrt{\sum\limits_{x \in I_{ij}} (r_{jx} - \overline{r_j})^2}}. \tag{3}$$

Where $\overline{r_i}$ and $\overline{r_j}$ are respectively average evaluation values of user i and user j for all items.

3. Collaborative filtering recommendation algorithm

Classical algorithm has defects such as non-scientific calculation of similarity value, lack of rationality, negligence of users' interest information and others, leading to large recommendation error and unreliable recommendation result [14]. To improve recommendation precision and exploit interest information in user evaluation, a new similarity calculation method is proposed in the Thesis.

3.1. User evaluation similarity

User evaluation similarity can be used to describe nonlinear change trend of two users for evaluation of the same item. Hence, nonlinear function is introduced to describe user evaluation similarity. Then similarity calculation formula of two users for evaluation of the same item is:

$$sim_1(i,j,x) = 2\left(1 - \frac{1}{1 + \exp\left(-\left|r_{ix} - r_{jx}\right|\right)}\right). \tag{4}$$

3.2. Interest tendency similarity

Each user has its own evaluation habit. For one specific item, some users give high scores and some users give low scores. Therefore, average evaluation of user describes the user's interest in certain objective. The calculation formula of interest tendency similarity of users i and j for the same item is:

$$sim_2\left(i,j,x\right)=\frac{1}{1+\exp\left(-\left(r_{ix}-\overline{r_i}\right)\left(r_{jx}-\overline{r_j}\right)\right)}. \tag{5}$$

3.3. Confidence degree of user evaluation similarity

When two users give similar scores for certain item, it not necessarily shows that the two users are similar, for similarity has one confidence degree. Hence, Jaccard function is used to measure confidence degree. Specific calculation formula is:

$$sim_3\left(i,j\right)=\frac{\left|I_i\cap I_j\right|}{\left|I_i\cup I_j\right|}. \tag{6}$$

Where I_i indicates item set evaluated by user i. On the whole, final calculation formula of similarity is:

$$sim_{score}\left(i,j\right)=\left(\frac{1}{\left|I_{ij}\right|}\sum\nolimits_{x\in I_{ij}} sim_1\left(i,j,x\right)\times sim_2\left(i,j,x\right)\right)\times sim_3\left(i,j\right). \tag{7}$$

3.4. User attribute similarity

Classical collaborative filtering recommendation algorithm only realizes recommendation through information related to existing users and is unable to accurately evaluate information of new users, leading to extremely high probability of cold starting. When there are a few user evaluation items, recommendation is made through user attribute similarity. With increasing of user evaluation item, recommendation is made through user evaluation. Hence, sigmoid function is introduced to combine user attribute recommendation and user evaluation recommendation and realize smooth transition of them. Suppose feature vector of user i is $Attr_i$=(a_{i1}, a_{i2}, ..., a_{in}), n is number of user attribute. If attribute m of user i and user j is the same, $sim_{Attr}(i,j,m)=1$, otherwise $sim_{Attr}(i,j,m)=0$. Therefore, similarity calculation formula for attributes of user i and user j is:

$$sim_{Attr}\left(i,j\right)=\sum\nolimits_{m\in Attr} w_m\cdot sim_{Attr}\left(i,j,m\right). \tag{8}$$

Where, w_i is weight of attribute i.

3.5. Fusion of user attribute similarity and user evaluation similarity

The calculation formula for fusion of user attribute similarity and user evaluation similarity is:

$$sim\,(i,j) = \alpha \cdot sim_{Attr}\,(i,j) + \beta \cdot sim_{score}\,(i,j)\ . \tag{9}$$

$$\alpha{=}2 \times \left(1 - \frac{1}{1+\exp\left(-\left|I_i\right|\right)}\right)\ . \tag{10}$$

$$\beta + \alpha = 1\,. \tag{11}$$

3.6. Working steps of algorithm in the Thesis

Step1: collect user attribute dimension and corresponding data, establish attribute matrix.

Step2: collect user attribute evaluation data and corresponding value, calculate evaluation value and establish similarity evaluation matrix.

Step3: analyze two matrixes and obtain user similarity matrix through combining them;

Step4: obtain K neighbors of user i through similarity matrix, estimate value of unevaluated item x through Equation (12) and obtain corresponding recommendation scheme according to results.

$$\widehat{r_{ix}} = \frac{\sum_{k\in N_K} sim\,(i,x)\cdot r_{kx}}{\sum_{k\in N_K} sim\,(i,x)}\ . \tag{12}$$

Where, N_K is K neighbors which are the most similar to user; r_{ki} is evaluation of user k for x.

4. Simulation experiment

4.1. Data set

In the computer with Intel(R) Core i5-3337U 3.0GHz CPU, 4GB RAM, Windows XP operating system, Visual C++ programming is adopted for simulation testing. Data come from public data set *MovieLens* and its description is specifically shown in Literature [11].

4.2. Comparison algorithm and evaluation standard

To make experimental results of algorithm in the Thesis more persuasive, collaborative filtering recommendation algorithm in Literature [15] and Literature [16] is selected for comparative experiment and mean absolute error (MAE) is selected as evaluation criterion of algorithm performance. Its definition is as follows:

$$MAE = \frac{\sum_{i=1}^{N} |p_i - q_i|}{N}. \tag{13}$$

Where N is size of test set, p_i is predicted evaluation value of recommended algorithm and q_i is actual evaluation value of user.

4.3. Results and analysis

Comparison of recommendation precision: When the nearest neighbor is 35, recommendation algorithm is adopted to solve the problem and specific results are shown in Fig. 1. It can be clearly seen from Fig. 1 that MAE value of collaborative filtering recommendation algorithm in the Thesis is lower than comparison algorithm. It effectively improves recommendation precision and obtains ideal recommendation results.

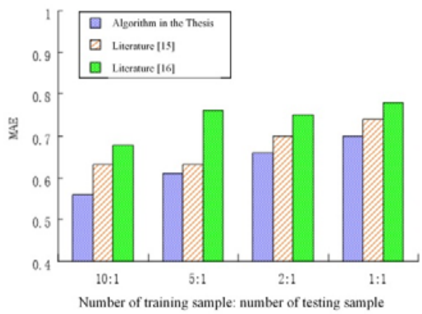

Fig. 1. Comparison of recommendation precision of different algorithms

Result analysis under the condition of cold starting: To simulate cold starting conditions, 10 users are selected and their evaluation information is deleted. Results are shown in Fig. 2. Through detailed analysis of Fig. 2, it can be known that the collaborative filtering recommendation algorithm combining user evaluation and attribute similarity in the Thesis can solve the current difficulty that recommendation algorithm cannot be implemented under cold starting conditions, improve recommendation precision and obtain superior recommendation results.

Performance comparison under different sparseness: recommendation error under different sparseness is shown in Fig. 3. There is approximately linear change relationship between data sparseness and MAE. However, under equal conditions, compared with comparison results of Literature [15] and Literature [16], it can be found that MAE of recommendation result for algorithm in the Thesis is smaller. Hence, recommendation precision for algorithm in the Thesis is superior to algorithm of Literature [15] and Literature [16] under equal conditions.

Universality testing: To verify the universality of collaborative filtering recommendation algorithm combining user evaluation and attribute similarity in the

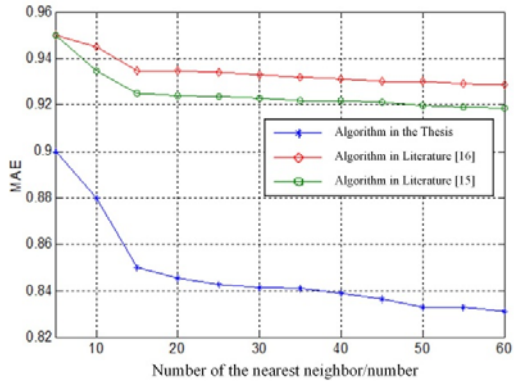

Fig. 2. Comparison of algorithm performance under cold starting condition

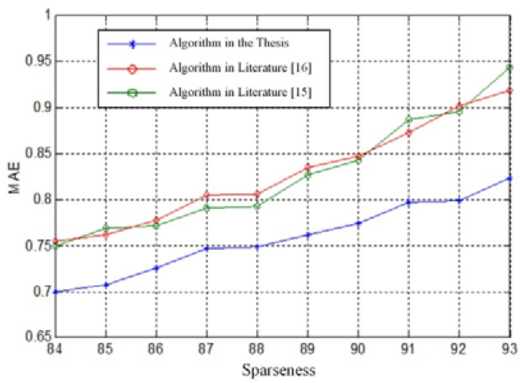

Fig. 3. Performance comparison of three algorithms under different sparseness

Thesis, Book-Crossing data set is selected for simulation testing and there is information of 287558 users and their 1491807 evaluation data for 231797 E-books. We adopt evaluation system for evaluation modeling with data in the interval between [0,10], with 1 for the highest evaluation and 0 for the lowest evaluation. Experimental results of different algorithms are shown in Fig. 4. It can be known from Fig. 4 that compared with other collaborative filtering recommendation algorithms, MAE of collaborative filtering recommendation algorithm in the Thesis is the smallest and the recommendation precision is higher, proving superiority and good universality of algorithm in the Thesis again.

5. Conclusion

In view of the traditional collaborative filtering algorithm in the similarity calculation process, the calculation, virtual high distortion, not the data sparsity in the proposed discrete and close to the collaborative filtering algorithm based on the user's interest degree. The core point of this algorithm is more user rating items,

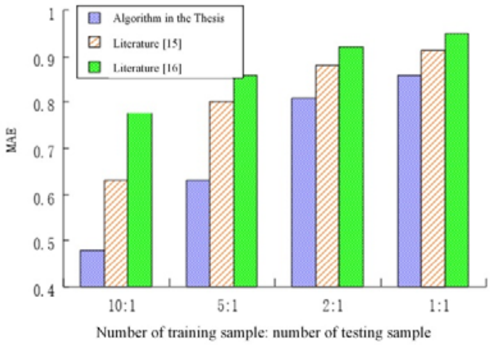

Fig. 4. Performance comparison with classical algorithm

common rating items and more, with the same score more items, the similarity is more. From the two aspects of user rating, full information and user interest preferences, the algorithm integrates the discrete quantity and user interest closeness to measure the similarity between users. The experimental results show that the proposed algorithm can effectively improve the recommendation quality of the information recommendation system, and also maintain good performance in the case of extremely sparse data.

Acknowledgement

Beijing Wuzi University 2017 Youth Foundation.

References

[1] W. S. Pan, S. Z. Chen, Z. Y. Feng: *Investigating the Collaborative Intention and Semantic Structure among Co-occurring Tags using Graph Theory.* 2012 International Enterprise Distributed Object Computing Conference, IEEE, Beijing, (2012), . 190–195.

[2] J. W. Chan, Y. Y. Zhang, and K. E. Uhrich: *Amphiphilic Macromolecule Self-Assembled Monolayers Suppress Smooth Muscle Cell Proliferation*, Bioconjugate Chemistry, *26* (2015), No. 7, 1359–1369.

[3] Y. Y. Zhang, E. Mintzer, and K. E. Uhrich: *Synthesis and Characterization of PEGylated Bolaamphiphiles with Enhanced Retention in Liposomes*, Journal of Colloid and Interface Science, *482* (2016), 19–26.

[4] J. J. Faig, A. Moretti, L. B. Joseph, Y. Y. Zhang, M. J. Nova, K. Smith, and K. E. Uhrich: *Biodegradable Kojic Acid-Based Polymers: Controlled Delivery of Bioactives for Melanogenesis Inhibition*, Biomacromolecules, *18* (2017), No. 2, 363–373.

[5] Z. Lv, A. Halawani, S. Feng, H. Li, & S. U. Réhman: *Multimodal hand and foot gesture interaction for handheld devices.* ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), *11* (2014), No. 1s, 10.

[6] Y. Z. Chen, F. J. Tang, Y. Bao, Y. Tang, G. D. Chen: *A Fe-C coated long period fiber grating sensor for corrosion induced mass loss measurement.* Optics letters, *41* (2016), 2306–2309.

[7] Y. Du, Y. Z. Chen, Y. Y. Zhuang, C. Zhu, F. J. Tang, J. Huang: *Probing Nanostrain via a Mechanically Designed Optical Fiber Interferometer.* IEEE Photonics Technology Letters, *29* (2017), 1348–1351.

[8] W. S. Pan, S. Z. Chen, Z. Y. Feng: *Automatic Clustering of Social Tag using Community Detection.* Applied Mathematics & Information Sciences, *7* (2013), No. 2, 675–681.

[9] Y. Y. Zhang, Q. Li, W. J. Welsh, V. P. Moghe, and K. E. Uhrich: *Micellar and Structural Stability of Nanoscale Amphiphilic Polymers: Implications for Anti-atherosclerotic Bioactivity*, Biomaterials, *84* (2016), 230–240.

[10] J. W. Chan, Y. Y. Zhang, and K. E. Uhrich: *Amphiphilic Macromolecule Self-Assembled Monolayers Suppress Smooth Muscle Cell Proliferation*, Bioconjugate Chemistry, *26* (2015), No. 7, 1359–1369.

[11] D. S. Abdelhamid, Y. Y. Zhang, D. R. Lewis, P. V. Moghe, W. J. Welsh, and K. E. Uhrich: *Tartaric Acid-based Amphiphilic Macromolecules with Ether Linkages Exhibit Enhanced Repression of Oxidized Low Density Lipoprotein Uptake*, Biomaterials, *53* (2015), 32–39.

[12] Y. Y. Zhang, A. Algburi, N. Wang, V. Kholodovych, D. O. Oh, M. Chikindas, and K. E. Uhrich: *Self-assembled Cationic Amphiphiles as Antimicrobial Peptides Mimics: Role of Hydrophobicity, Linkage Type, and Assembly State*, Nanomedicine: Nanotechnology, Biology and Medicine, *13* (2017), No. 2, 343–352.

[13] R. Majumder: *Fast collaborative filtering through approximations: US*, US 20070124698 A1[P], (2007).

[14] R. Majumder: *Fast collaborative filtering through sketch function based approximations: US*, US 7624095 B2[P] (2009).

[15] J. Chen, T. Takiguchi, Y. Ariki: *A robust SVM classification framework using PSM for multi-class recognition*[J]. Eurasip Journal on Image & Video Processing, (2015), No. 1, 1–12.

Received May 7, 2017