Automatic Recommendation Method of Network Data Based on Big Data Technology

Xiaorui Huang 1*, Shuai Guo 2, Yunchu Zhang 3, Peidong Zhang 4, Shiyuan Xu 5

1Chongqing University of Posts and Telecommunications, Chongqing, Chongqing, 400065, China
2Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, 210023, China
3Chongqing University of Posts and Telecommunications, Chongqing, Chongqing, 400065, China
4University of Science and Technology Beijing, Beijing, Beijing, 100083, China
5North China University of Technology, Beijing, Beijing, 100083, China
*Corresponding author’s e-mail: xiaoruihuang99@gmail.com

Abstract. With the advent of the Internet era in modern society, the surge in network data has greatly increased the difficulty for users to obtain demand information. In order to provide customers with information that may be of interest to them, this paper proposes a belief collaborative recommendation optimization algorithm, which introduces DP synthesis rules and Smets synthesis rules to improve traditional DS synthesis rules and establishes network data recommendation models. This method realizes automatic recommendation of network data according to the self-property and historical behavior data of network customers. The experimental results show that the automatic recommendation method for network data proposed in this paper has higher recommendation accuracy.

1. Introduction

The shopping recommendation search engine that exists on the online shopping platform can provide users with products that may be of interest. However, due to the existence of a large amount of information on the Internet, the search recommendation results usually contain some redundant information that exceeds users’ expectations. Personalized recommendation came into being under the background of this demand, and it provides people with a new mode of obtaining information. The recommendation method is an information filtering system that predicts a user's "rating" or "preference" for an item [1-2]. Recommended items include movies, music, news, books, academic papers, search queries, and other products. In this paper, a confidence-based collaborative recommendation algorithm based on evidence theory is introduced. A soft scoring method is used to measure the uncertainty of the user's evaluation of the item using a soft scoring mechanism. The K-nearest algorithm is used to calculate the user's neighbor neighborhood. Based on the product evaluation, this paper uses the improved evidence combination rules to build a credible collaborative recommendation model, and finally recommends the fused high evaluation items to the target users. Finally, the data Epinions are used to verify the effectiveness of the improved model [3-4].
2. Combination rule

The combination rules are defined as follows

**Definition 1**: Let \( \theta = \{ \theta_1, \theta_2, \ldots, \theta_n \} \) represent a complete set of all possible values of \( X \), and all elements in \( \theta \) are mutually exclusive, and \( \theta \) is called the identification frame of \( X \). Evidence theory is based on the power set \( 2^\theta = \{ \alpha : \alpha \subseteq \theta \} \).

**Definition 2**: Let \( \theta \) be the identification frame of \( X \), then the mapping \( M \) from the set \( 2^\theta \) to \([0,1]\) is the basic belief assignment (BBA) function on \( 2^\theta \), if

\[
\begin{align*}
\sum_{\alpha \subseteq \theta} M(\alpha) &= 1 \\
M(\emptyset) &= 0
\end{align*}
\]

In the formula, \( M(\alpha) \) is the reliability distribution of the event \( \alpha \), and represents the degree of trust in \( \alpha \). The basic reliability assigned to the empty set \( \emptyset \) represents the inconsistency and incompleteness of the identification framework.

**Definition 3**: D-S evidence combination rule. Let \( M_1 \) and \( M_2 \) be the BBAs corresponding to the two evidences \( E_1 \) and \( E_2 \) in the identification frame \( S \), and the focal element is \( \alpha_i \subseteq \theta \), then the composition rule is [5-6]:

\[
[M_1 \oplus M_2] (\alpha) = M(\alpha) = \begin{cases} 
\sum_{\alpha_i \cap \alpha_j = \alpha} M_1(\alpha_i)M_2(\alpha_j) / (1-N), & \alpha \neq \emptyset \\
0, & \alpha = \emptyset 
\end{cases}
\]

In formula (2),

\[
N = \sum_{\alpha_i \cap \alpha_j = \emptyset} M_1(\alpha_i)M_2(\alpha_j)
\]

**Definition 4**: Dubois and Prade combination rule. Let \( M_1 \) and \( M_2 \) be the BBAs corresponding to the two evidences \( E_1 \) and \( E_2 \) in the identification frame \( S \), and the focal elements are \( \alpha_i \) and \( \beta_j \), respectively, then the combination rule is:

\[
M(\alpha) = \begin{cases} 
\sum_{\alpha_i \cap \beta_j = \alpha} M_1(\alpha_i)M_2(\beta_j), & \alpha \neq \emptyset \\
\sum_{\alpha_i \cap \beta_j = \emptyset, \alpha_i \cup \beta_j = \alpha} M_1(\alpha_i)M_2(\beta_j), & \alpha_i \cap \beta_j = \emptyset, \alpha_i \cup \beta_j = \alpha \\
0, & \alpha = \emptyset 
\end{cases}
\]

**Definition 5**: Smets composition rule. Let \( M_1 \) and \( M_2 \) be the BBAs corresponding to the two evidences \( E_1 \) and \( E_2 \) in the identification frame \( S \), and the focal elements are \( \alpha_1 \) and \( \alpha_2 \) respectively, then the composition rule is:

\[
M(\alpha) = \begin{cases} 
\sum_{\alpha_i \cap \alpha_2 = \alpha} M_1(\alpha_1)M_2(\alpha_2), & \alpha \neq \emptyset \\
\sum_{\alpha_i \cap \alpha_2 = \emptyset} M_1(\alpha_1)M_2(\alpha_2), & \alpha = \emptyset 
\end{cases}
\]

3. Confidence collaborative recommendation algorithm based on improved evidence combination rules

In recent years, many studies have shown that when using D-S composition rules for evidence reasoning, conclusions that do not conform to common sense will appear, which directly affects the accuracy and reliability of inference decisions. In order to better integrate the scores of different users on items, improve the synthesis rules, introduce DP synthesis rules and Smets synthesis rules, and
establish a confidence collaborative recommendation algorithm that improves the evidence combination rules. The steps are as follows:

1) Soft scoring production mechanism: according to partial probability models and power set approaches, the traditional "hard scoring" data is converted into "soft scoring".

2) Community mining: use label propagation algorithm (LPA) to classify users through the user relationship network, calculate the similarity of users in each community, and use the improved KNN algorithm to screen neighboring users.

3) Information fusion: use different evidence combination rules to fuse the items scores of neighboring users, and recommend the items with high scores after fusion to the users to be recommended.

4. Case analysis
This article selects the Epinions dataset, which consists of two parts: hard score and user relationship. In the hard scoring section, the scoring value ranges from 1 to 5 with a step size of 1. Each user has evaluated at least 20 items. In the user relationship network, each user knows at least 10 other users. The test data set has a total of 52,378 users, evaluated 175,824 items, and generated a total of 763,346 score records. These 52,378 users have 625,541 layers of trust.

4.1. Recommendation process under D-S synthesis rules
The Label Propagation Algorithm (LPA) was used to divide 52378 users into 4 groups. Some users in different groups are shown in Table 1.

| Group 1 | Group 2 | Group 3 | Group 4 |
|---------|---------|---------|---------|
| 1       | 59      | 164     | 381     |
| 2       | 106     | 172     | 399     |
| 3       | 148     | 177     | 405     |
| 4       | 254     | 210     | 411     |
| 5       | 406     | 336     | 414     |
| 6       | 492     | 389     | 482     |
| 7       | 673     | 525     | 596     |
| 12      | 785     | 744     | 599     |
| 17      | 869     | 757     | 612     |
| 23      | 878     | 802     | 654     |
| ...     | ...     | ...     | ...     |

The D-S evidence theory was used to fuse the scores of each item to obtain the item recommendation table in Table 2.

| Number | User ID | Recommended Item Rating |
|--------|---------|-------------------------|
| 1      | 257     | 849 819 | 2418 3654 |
| 2      | 864     | 456 509 | 991 1569 |
| 3      | 1025    | 432 591 | 796 1996 |
| 4      | 3246    | 134 596 | 641 1414 |
| 5      | 4604    | 455 995 | 2356 3498 |
| 6      | 4908    | 374 459 | 1844 2652 |
| 7      | 6726    | 496 449 | 1541 2943 |
| 8      | 7108    | 659 954 | 2652 3496 |
| 9      | 7587    | 48 478 949 1966 3195 |
| 10     | 8962    | 949 569 1496 1949 4962 |
The accuracy of the proposed model is evaluated by calculating the values of the mean absolute error (DS-MAE) and confusion matrix (DS-Recall). The DS-MAE value represents the probability of error. A small value indicates that the evaluation result is good and the accuracy is high. In the experimental results, a higher recovery rate is expected. The larger the DS-Recall value, the higher the accuracy. First divide the data into 10 groups, and then calculate the accuracy of the results. The average absolute error (DS-MAE) of the 10 test groups is 0.7654, and the average of the DS-Recall is 0.5283.

4.2. Comparison of results of improved recommendation algorithms under different synthesis rules

The confidence collaborative recommendation model of evidence theory is improved. DP synthesis rules and Smets synthesis rules are introduced to establish a confidence collaborative recommendation model that improves evidence combination rules. Based on the model improvement, the accuracy of the recommendation results under different synthesis rules is calculated and compared with the results of the D-S combination rule. Using the Smets synthesis rule, the average value of DS-MAE is 0.5792, and the average value of DS-Recall is 0.5475. After using the Smets synthesis rule, DS-MAE decreases and DS-Recall increases, which indicates that the overall accuracy will also increase.

The average value of the DS-MAE of the DP synthesis rule is 0.4396, and the average value of the DS-Recall is 0.5846. After using the DP synthesis rule, the DS-MAE is significantly reduced and the DS-Recall is increased. This shows that the recommended accuracy of the DP synthesis rule has been a substantial improvement.

Figure 1 shows the DS-MAE value and DS-Recall value obtained after using different synthesis rules. Through comparison, it can be found that the overall accuracy rate is improved after using the DP synthesis rule.
5. Conclusion
The uncertainty information processing method of social networks is a hot and difficult point in the research of recommendation systems. In the framework of evidence theory, this paper introduces a soft scoring system, combines a group recognition algorithm and a K-nearest neighbor algorithm, and proposes an ECR algorithm to recommend items and test the sensitivity of the data. Considering the limitations of traditional D-S combination rules, this paper introduces new synthesis rules and recommendation algorithms ECR-Sm and ECR-DP to improve the accuracy of recommendation. Based on the research of these algorithms, the dataset Epinions is tested and the test accuracy is evaluated. The results show that the accuracy of the improved synthetic rule recommendation model is improved.

References
[1] Al-Hawawreh, M., Hartog, D.F., Sitnikova, E. (2019) Targeted Ransomware: A New Cyber Threat to Edge System of Brownfield Industrial Internet of Things. IEEE Internet of Things Journal, 6: 7137–7151.
[2] Bergasa-Suso, J., Sanders, D.A., Tewkesbury, G.E. (2005) Intelligent browser-based systems to assist Internet users. IEEE Transactions on Education., 48: 580–585.
[3] Shafer, G.A. (1994) Mathematical Theory of Evidence. Princeton University Publishing, Princeton, New Jersey.
[4] Wickramarathne L.T., Premaratne, K., Kubat, M., et al. (2011) CoFiDS: A Belief-Theoretic Approach for Automated Collaborative Filtering. IEEE Transactions on Knowledge and Data Engineering., 23: 175-189.
[5] Chen, T., Que, P. (2006) Target recognition based on modified combination rule. Journal of Systems Engineering and Electronics., 17: 279–283.
[6] Liu, Y., Pal, R.N., Marathe, R.A, et al. (2018) Weighted Fuzzy Dempster–Shafer Framework for Multimodal Information Integration. IEEE Transactions on Fuzzy Systems., 26: 338–352.