Alexandria Engineering Journal (2022) **61**, 905–910

HOSTED BY

Alexandria University

Alexandria Engineering Journal

www.elsevier.com/locate/aej
www.sciencedirect.com

An ecommerce recommendation algorithm based on link prediction

Guoguang Liu

School of Economics and Management, Binzhou University, Binzhou 256600, China

Received 21 January 2021; revised 11 April 2021; accepted 26 April 2021
Available online 6 June 2021

KEYWORDS

Recommendation algorithm;
Bipartite graph network (BGN);
Link prediction;
Ecommerce

Abstract In the field of ecommerce, most recommendation algorithms are based on user-item bipartite graph network (BGN). But this kind of recommendation algorithm is severely lacking in accuracy and diversity. In this paper, a novel ecommerce recommendation algorithm is proposed based on BGN link prediction. Firstly, all the user-item data were imported into distance formula to calculate the similarity between the attributes. Then, the BGN was projected into a single-mode network (SMN), making it more efficient to extract potential links from the BGN. On this basis, the potential links were predicted based on similarity. Through experiments on real ecommerce datasets, it was proved that our algorithm has a higher accuracy and coverage than typical recommendation algorithms.

1. Introduction

The boom of ecommerce has brought a huge amount of data on items purchased online. However, much of these data is highly redundant, dragging down the retrieval efficiency of users. This not only suppresses the satisfaction of users, but also brings a huge burden to ecommerce websites. Therefore, ecommerce researchers have devoted much attention to finding the items that interest users efficiently and accurately out of the massive data. Personalized recommender system provides a way to push appropriate items to users according to their purchase records, reviews, and ratings, thereby reducing their access time to such items.

Traditional recommendation algorithms have been widely adopted by ecommerce giants like Amazon and Taobao. However, most of the recommendation algorithms in traditional recommender systems face two problems: (1) the assumption of independence between users ignores the real-world social relationships between users; (2) the algorithms cannot output sufficiently diverse solutions, or overcome the constraint of the cold-start problem, despite achieving relatively high accuracy.

Many real-world problems can be modeled as a complex network [1], in which each node represents an entity of the problem, and each link represents the relationship between two entities. Complex and largescale problems can be described by bipartite graph network (BGN) [2]. It is that the vertex set can be divided into two disjoint subsets, and the two vertices attached to each edge in the graph belong to the two disjoint subsets, and the vertices in the two subsets are not adjacent. Therefore, BGN greatly expands the descrip-

E-mail address: liuguoguang@bzu.edu.cn

Peer review under responsibility of Faculty of Engineering, Alexandria University.

https://doi.org/10.1016/j.aej.2021.04.081
1110-0168

tion ability and application scenarios of complex network model. In some special cases, if we need to study the connection relationship between the same kind of nodes in bipartite graph, the most common method is projection method, which projects one type of node to another in bipartite graph network.

The BGN-based recommendation algorithm models the items and user groups as nodes, regards the item-user relationships as links, and predicts the potential links of the network based on the existing ones. The potential links are the items that may interest users.

Most BGN-based recommendation algorithms recommend popular items to users, but fail to suggest enough unpopular items to users. Thus, the recommendation is severely lacking in diversity. This is because the BGN-based recommendation algorithms only consider the strong relationships in the dataset, without paying attention to the weak relationships or hidden information.

Combined recommendation algorithm generally combines two complementary recommendation algorithms, and improves the recommendation accuracy and coverage of recommendation system, but the calculation efficiency of combined recommendation algorithm is low, and the real-time performance of data push is insufficient. To solve the above problem, the BGN was projected into a single-mode network (SMN). The strong and weak relationships of the BGN were retained in the links of SMN, which considers the number of neighboring nodes, the degree of common neighboring nodes, and the degree of nodes. Then, the enhanced weighted SMN was used to filter the redundant information of the original dataset. The filtering reduces the amount of information in SMN, making link prediction more efficient.

2. Literature review

The most popular recommendation algorithms mainly include collaborative filtering recommendation algorithm [3,4], content-based recommendation algorithm [5,6], combined recommendation algorithm [7,8], and network-based recommendation algorithm [9–11]. The collaborative filtering recommendation algorithms perform poorly on large sparse datasets. To solve the problem, Yu et al. [12] proposed a probabilistic matrix factorization model, which only predicts user ratings of unknown items based on the existing rating data of user-item matrix. Cui et al. [13] proposed a new singular value decomposition (SVD) model, reducing the sparsity of user-item matrix. Jiao et al. [14] added implicit feedback to improve the SVD model, and treated the historical data on user browsing, user ratings, item browsing, and item ratings as new parameters. However, all the above methods assume that users are independent of each other, failing to consider the social relationships between users.

With the rapid development of social networks, the social relationships between users have become an important basis for many ecommerce platforms to push items to users. Besides receiving information, users actively establish relationships with each other. Therefore, it is important for recommendation algorithms to improve the quality of recommender system, using the social relationships between users. Through probability matrix factorization, Belkhadir et al. [15] presented a recommendation algorithm that connects social network and user-item rating matrix via a shared user potential feature space, and experimentally proved that the algorithm is more efficient than popular algorithms in the case of little or no user ratings. Bin et al. [16] integrated the multiple relationships in social networks into collaborative filtering recommendation algorithm, which greatly improves the recommendation accuracy. To overcome data sparsity, Wu et al. [17] put forward a probabilistic factor analysis framework, in which the preferences of each user are fused with those of his/her friends through a social parameter set. Inspired by communication theory, Tian et al. [18] designed an intelligent recommendation algorithm based on mass diffusion in BGN, and demonstrated its high accuracy and diversity of recommendation.

3. Link prediction in BGN

The BGN can be described as $G_b = (U, I, E)$, where U and I are two sets of disjoint nodes; $E \sqsubseteq |U| \times |I|$ is the set of links. There is a link between a $u \in U$ node and an $i \in I$ node, but no link between the nodes in the same node set. In addition, the set of neighboring nodes of nodes in I is normally denoted as $N(I)$.

The BGN can also be represented as a matrix $|U| \times |I|$. Let n and m be the number of nodes in sets U and I, respectively. Then, the matrix G_b can be expressed as an $m \times n$ dimensional adjacency matrix. The matrix element a_{ij} and diagonal matrix $\widehat{A}$ of BGN G_b can be respectively defined as:

$$a_{ij} = \begin{cases} 1, if(u_i, v_j) \in E \\ 0, otherwise \end{cases}$$

$$\widehat{A} = \begin{bmatrix} 0_{m\times n} & A_{n\times m} \\ A^T_{m\times n} & 0_{n\times m} \end{bmatrix}$$

where, $0_{m\times n}$ and $0_{n\times m}$ are all-zero matrices of $n \times n$ and$m \times n$, respectively; $a_{n\times m}$ is a non-zero matrix. This means the adjacency matrix is symmetric. Thus, the BGN G_b can be described by an $A_{n\times m}$ matrix, with each row and column of U set representing a node of set I.

The BGN link prediction aims to find the links that do not exist but will appear in the network. Suppose $G_b = (U, I, E)$ is a BGN at time t. Then, the task of link prediction is to predict a new link in the network at time $t + 1$. The existing BGN link prediction algorithms firstly projected a bipartite graph into a monopartite graph, and then defined the concept of potential edge. The complexity of the prediction algorithms is high. In this paper the pattern covered by the potential link and the weight of the pattern is defined, the credibility of the potential link by the weight of the pattern covered by the potential link is calculated, and it is taken as the score of the actual link on the potential edge, which makes the prediction of the binary network link only in the selected potential edge, and greatly reduces the complexity of the prediction algorithm.

A general practice for BGN analysis is to convert the network into an SMN through projection. The projected network has the typical structure of SMN. To predict the potential link of BGN, the authors firstly projected the BGN into two SMNs. The SMN can be defined as follows:

Let $G_b = (U, I, E)$ be a BGN, and $|U(\mathrm{G})| = m$, $|I(G)| = n$. Then, the two node sets U and I of the BGN can be converted into two SMNs, namely, SMN $G_u = (U, E_u)$ and $G_v = (I, E_v)$:

$$E_u = \{(u_i, u_j) | u_i, u_j \in U, |\exists v_i \in I, v_i \in \Gamma(u_i) \cap \Gamma(u_j)\}$$

$$E_v = \{(v_i, v_j) | v_i, v_j \in I, |\exists u_i \in U, u_i \in \Gamma(v_i) \cap \Gamma(v_j)\}$$

The conversion might sacrifice some topology information of the original network. To prevent the information loss, this paper designs a weighted BGN, and converts it into a weighted SMN through weighted SMN projection. In the weighted SMN, the weight of each link represents the number of common neighboring nodes.

For BGN $G_b = (U, I, E)$, each E_u or E_v link was weighted by function W:

$$E_u = \{(u_i, u_j) | u_i, u_j \in U, |\exists v_i, v_i \in \Gamma(u_i) \cap \Gamma(u_j)\}$$

$$W : (u_i, u_j) \to |\Gamma(u_i) \cap \Gamma(u_j)|$$

$$E_v = \{(v_i, v_j) | v_i, v_j \in I, |\exists u_i, u_i \in \Gamma(v_i) \cap \Gamma(v_j)\}$$

$$W : (v_i, v_j) \to |\Gamma(v_i) \cap \Gamma(v_j)|$$

where, $W(u_i, u_j)$ is the weight between the i-th node and the j-th node. For an SMN, $W(u_i, u_j)$ can be defined as the number of common neighboring nodes for $\Gamma(u_i) \cap \Gamma(u_j)$.

Because it is constructed from real data, each SMN contains a large amount of redundant information, which has a negative impact on the prediction of potential links. Thus, it is necessary to simplify the network and enhance the prediction quality by filtering the redundant links. This paper designs a backbone network extraction algorithm to convert the weighted SMN into an enhanced SMN. The enhanced SMN can be defined as:

From BGN $G_b = (U, I, E)$, two SMNs $G_u = (U, E_u)$ and $G_v = (I, E_v)$ can be obtained. If the weight $W(a, b)$ of the link between a and b is greater than , then the link(a, b) should be retained in the enhanced network EN^t; otherwise, the link should be filtered as a redundant information [17].

The key work of this research is to extract the basic information from the BGN composed of items, users, and ratings. Specifically, a complex BGN was constructed from a large real-world dataset, and a link prediction scheme was designed for the network. Fig. 1 shows the flow diagram of the proposed method, which includes the basic information extraction model. The potential link prediction of the current BGN is another key task.

In the proposed method, firstly, the supplement attribute information of bipartite graph is extracted from network database, and then the network topology of bipartite graph is established, trained and tested. Finally, the link prediction algorithm is constructed by weighted projection of bipartite graph.

Through enhanced BGN projection, the potential links of BGN were extracted, and the potential links were predicted based on similarity. Firstly, the BGN was established by mining the ecommerce big data. Then, the BGN was converted into weighted SMNs. Finally, each SMN was converted into an enhanced network EN^t.

In each SMN, a node pair with internal node pair attributes was detected to predict the potential links [17]. This approach could reduce the number of predicted links, and improve the prediction quality. If two nodes in the BGN G_b interact with each other, there may be potential links between them; if the two nodes have no common neighboring node, it is very unlikely for them to have potential links.

For BGN $G_b = (U, I, E)$, the two projected SMNs can be described as $G_u^t = (U, E_u^t)$ and $G_v^t = (I, E_v^{\alpha})$. If $a \in U$ is the

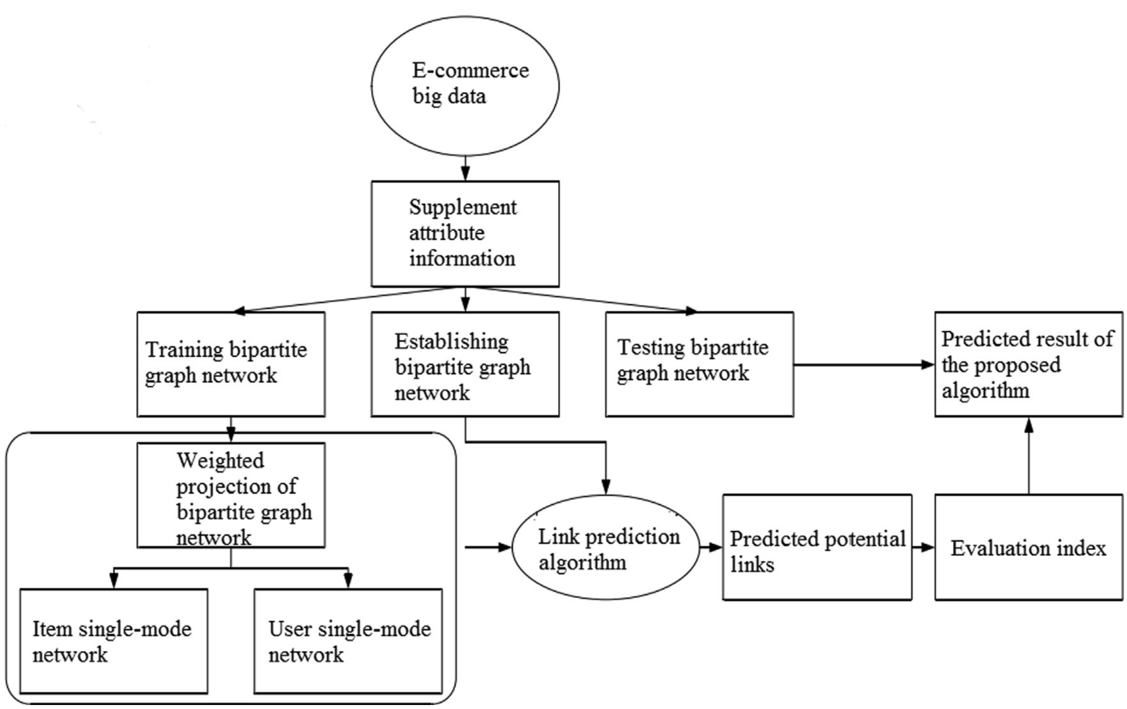

Fig. 1 Flow diagram of the proposed method.

node of SMN $G_u^t = (U, E_u^t)$, then the probability for node a to have a potential link can be described as:

$$PL_a = \{N\{(\Gamma_a)_{train}\}, N = \Gamma(N_1) \cup \Gamma(N_2), \cdots, \cup\Gamma(N_n)$$

where, N is the neighboring nodes of a node Γ_a^α in the trained BGN G_{train}; node a is part of SMN G_u^t;$\Gamma(N_n)$ is the neighboring nodes of node N_n in the BGN; $PL_a = \{p|p \in N \bigwedge p \notin \Gamma_A, (a,p)\}$;$PL = PL_a \cup PL_b, \cdots, \cup PL_m$ is the predicted potential link. Nodes with potential links must satisfy:

$$\Gamma_U(a) \cap \Gamma(p_i) \neq \emptyset and p_i \notin \Gamma(a)$$

where,$a \in U$ and $p_i \in I$ are two nodes; $(a, p_i) \notin E$.

Suppose G_{train} is a trained BGN and G_u^t is a U-projection SMN. The potential link for each $q \in (\Gamma_U(a))_{train} \cap \Gamma(p_i)$ is the pattern of potential link coverage. The greater the number of patterns covered by a potential link, the higher the probability that the potential link will become a real link. Therefore, the number of patterns covered by potential link can measure the probability of potential links. The more the potential link coverage patterns are, the more important each link weight of the potential link coverage in the network are.

Let $G_u = (U, E_u)$ be a weighted SMN projected from BGN $G_b = (U, I, E)$, and $(a, b) \in E_u$ be a link of G_u. Then, the weight $w(a, b)$ of pattern $\{a, b\}$ can be calculated by:

$$w(a,b) = \frac{2}{D_a + D_b} \sum_{\tau \in \Gamma(a) \cap \Gamma(b)} \frac{1}{D_c}$$

where, D_a, D_b, and D_c are the degrees of nodes a, b, and c in BGN G_b, respectively; $\Gamma(a)$ and $\Gamma(b)$ are the neighboring sets of nodes a and b in the BGN. It can be seen that, the smaller the common neighbor degree of nodes a and b, the larger the weight of the pattern.

The probability of each pattern element covered by potential link is equivalent to the total weight of the pattern covered by potential link. Hence, the final total evaluation of potential link (a, p_i) can be calculated by:

$$S(a, p_i) = \sum_{a,b \in \Gamma(a,p_i)} w(a,b)$$

where a, b represents the nodes in network.

The above equation shows that the number of potential links in the high weight pattern is positively correlated with the probability of link prediction. Therefore, the sum of the pattern weights of potential link coverage is the final value of the potential link prediction.

To sum up, the proposed method can be implemented in the following process: A projection SMN containing weak relationships is obtained from the original dataset, the redundant information is filtered out, and the backbone network is extracted. Then, the backbone network is enhanced based on thresholds predefined, according to the high frequency link weights in an initialized single-mode SMN. The proposed method effectively descales the set of potential links, reduces the overall computing time, and maintains nodes with strong relationships.

4. Experiments and results analysis

The performance of the proposed algorithm was verified through multiple experiments. Two indices were selected to measure the performance: recommendation accuracy and recommendation coverage.

4.1. Experimental datasets

Two datasets were used for our experiments: FilmTrust and Epinions [19]. These are two standardized recommender system testing data sets. FilmTrust is a collection of user-item ratings from the website of Film Trust, involving 35,497 ratings, 1,642 users, and 2,071 films. The ratings fall between 0.5 and 4. Epinions is a collection of user-item ratings from Epinions.com. This dataset covers 40,290 users, and 139,728 items. The ratings were given against a 5-point scale. Here, 6,000 users and 12,000 items are randomly selected from the Epinions dataset as the test set. The basic information of the two datasets are shown in Table 1.

The recommendation accuracy was measured by mean absolute deviation (MAE):

$$MAE = \frac{\sum_{i=1}^{W} |c_i - c_p|}{W}$$

where, c_p represents the predicted rating of the item i; c_i is the actual rating of item i; W is the number of ratings in the dataset. The smaller the MAE, the better the recommendation accuracy.

The recommendation coverage, another important metric of the performance of the recommendation algorithm, can be calculated by:

$$R_C = N/|Q|$$

where, N is the number of predicted ratings; $|Q|$ is the total number of ratings in the dataset. The higher the R_C value is, the better the performance of the recommendation algorithm.

The overall performance of the recommendation algorithm can be evaluated by index $F1$:

$$F1 = \frac{2 \times precision \times R_C}{precision + R_C}$$

where, $precision$ can be defined as:

$$precision = 1 - MAE/(c_{max} - c_{min})$$

where, c_{max} and c_{min} are the highest and lowest ratings in the recommendation algorithm, respectively.

In our experiments, the proposed algorithm is compared with different types of recommendation algorithms: collaborative filtering recommendation algorithm (CF) [20], trust-based recommendation algorithm (MT) [21], content-based recommendation algorithm (TCF) [22], and CF-social relationship merging recommendation algorithm (Merge) [23]. In addition,

Table 1 Basic information of FilmTrust and Epinions datasets.

Dataset	FilmTrust	Epinions
Number of users	1,508	6,000
Number of items	2,071	12,000
Number of ratings	35,497	82,461
Rating interval	0.5	1
Rating range	[0.5,4]	[1,5]
Number of comments	1,853	25,905

two BGN-based recommendation algorithms were also chosen as contrastive methods: real-time construction full-weighted bipartite (RTCF) [24], and multi-view bipartite network (MV) [25]. Figs. 2-4 compares the performance of different algorithms on Epinions dataset.

The above results show that the proposed algorithm achieved higher recommendation accuracy and coverage on the large dataset than the other types of recommendation algorithms, and also outshined the other BGN-based algorithms. Comparatively, the BGN-based algorithms were superior to the other types of algorithms. The main reason is that RTCF, MV and the proposed algorithm establish a comprehensive BGN, which retains lots of information of the dataset. In other types of algorithms, however, much information of the dataset is filtered out through preprocessing, for the sake of computing efficiency.

Figs. 5-7 compares the performance of different algorithms on FilmTrust dataset.

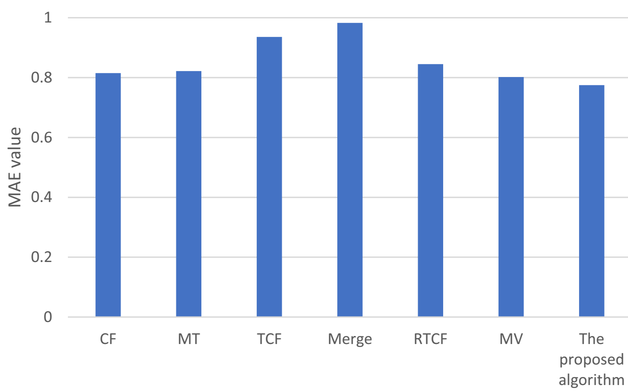

Fig. 2 MAEs of different algorithms on Epinions dataset.

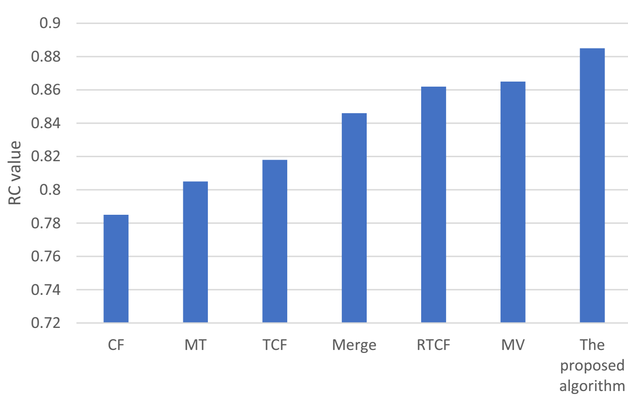

Fig. 3 R_C values of different algorithms Epinions dataset.

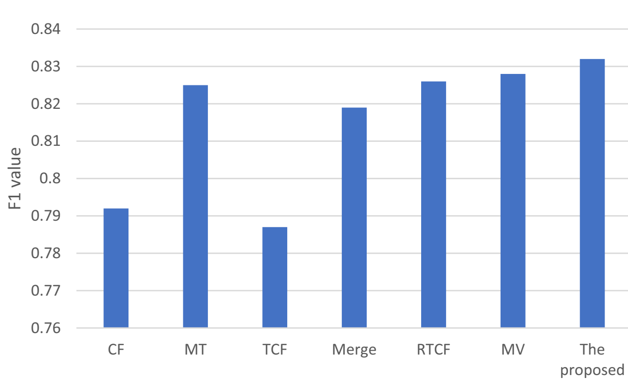

Fig. 4 F1 values of different algorithms on Epinions dataset.

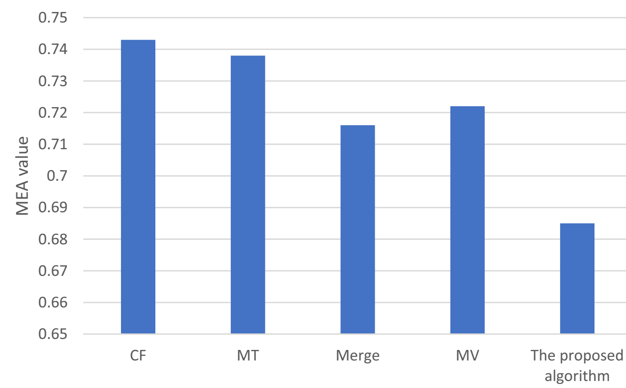

Fig. 5 MAEs of different algorithms on FilmTrust dataset.

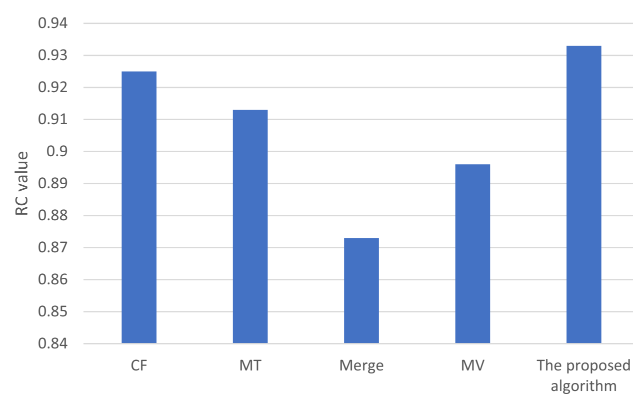

Fig. 6 R_C values of different algorithms on FilmTrust dataset.

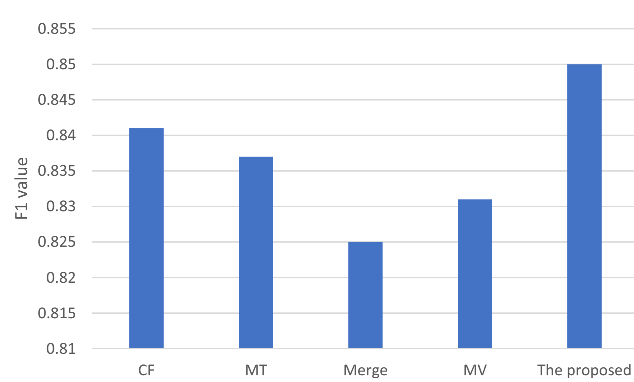

Fig. 7 F1values of different algorithms on FilmTrust dataset.

The above results show that the proposed algorithm outperformed the other types of algorithms in recommendation accuracy and coverage, and also realized better performance than another BGN-based algorithm, that is, MV.

Under the same hardware conditions, the time required by the proposed algorithm reduced about 15% than that of RTCF and MV algorithms.

Comparing the results of the above two sets of experiments, the proposed algorithm is more accurate on small dataset than on large dataset. This is because our algorithm filters redundant information during the extraction of the backbone network projected from the BGN. If the original dataset is large, the filtering operation needs to remove too much redundant information that might affect link prediction accuracy. Of course, our algorithm is still more accurate than other recommendation algorithms on large dataset.

5. Conclusions

When recommender systems deal with big data, most recommendation algorithms only consider the strong relationships in the dataset, failing to recognize the weak relationships or hidden information. This strategy could indeed improve the recommendation accuracy. But the recommendation results will lack diversity. To solve the problem, this paper proposes an enhanced weighted SMN projected from the BGN to extract the backbone network, filter the redundant information of the original dataset, and retain the strong and weak relationships. Through experiments on two real-world datasets, the proposed algorithm was proved superior than other types of recommendation algorithms and other BGN-based recommendation algorithms on both large and small datasets. In the future research, more attention will be paid to improving the accuracy and diversity of recommendations.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] G. Sun, S. Bin, Router-level internet topology evolution model based on multi-subnet composited complex network model, Journal of Internet Technology 18 (6) (2017) 1275–1283.

[2] S. Bin, G. Sun, C.C. Chen, Spread of infectious disease modeling and analysis of different factors on spread of infectious disease based on cellular automata, Int. J. Environ. Res. Public Health 16 (23) (2019) 4683.

[3] Y. Liu, H. Yang, G.X. Sun, S. Bin, Collaborative filtering recommendation algorithm based on multi-relationship social network, Ingénierie des Systèmes d'Information 25 (3) (2020) 359–364.

[4] A. Gholami, Y. Forghani, Improving multi-class Co-Clustering-based collaborative recommendation using item tags, Revue d'Intelligence Artificielle 34 (1) (2020) 59–65.

[5] J. Shu, X. Shen, H. Liu, A content-based recommendation algorithm for learning resources, Multimedia Syst. 24 (2) (2018) 163–173.

[6] B. Cao, X.F. Liu, M. Rahman, Integrated Content and Network-Based Service Clustering and Web APIs Recommendation for Mashup Development, IEEE Trans. Serv. Comput. 13 (1) (2020) 99–113.

[7] G. Sun, S. Bin, M. Jiang, N. Cao, Z. Zheng, H. Zhao, L. Xu, Research on Public Opinion Propagation Model in Social Network Based on Blockchain, CMC-Computers Materials & Continua 60 (3) (2019) 1015–1027.

[8] J. Li, W. Xu, W. Wan, Movie Recommendation Based on Bridging Movie Feature and User Interest, Journal of Computational Science 26 (5) (2018) 128–134.

[9] Y. Wen, J. Hou, Z. Yuan, Heterogeneous Information Network-Based Scientific Workflow Recommendation for Complex Applications, Complexity 4 (2020) 1–16.

[10] H. Zhong, H. Lyu, S. Zhang, Measuring user similarity using check-ins from LBSN: a mobile recommendation approach for e-commerce and security services, Enterprise Information Systems 14 (1) (2019) 1–20.

[11] G. Sun, S. Bin, A new opinion leaders detecting algorithm in multi-relationship online social networks, Multimedia Tools and Applications 77 (4) (2018) 4295–4307.

[12] H. Yu, A Novel Collaborative Recommendation Algorithm Integrating Probabilistic Matrix Factorization and Neighbor Model, Journal of Information and Computational Science 12 (5) (2015) 2011–2019.

[13] Cui, L., Huang, W., Yan, Q. (2017). A novel context-aware recommendation algorithm with two-level SVD in social networks. Future Generation Computer Systems, 86(9), 1495-1470.

[14] J. Jiao, X. Zhang, F. Li, A Novel Learning Rate Function and Its Application on the SVD++ Recommendation Algorithm, IEEE Access 8 (2020) 14112–14122.

[15] I. Belkhadir, E.D. Omar, J. Boumhidi, An intelligent recommender system using social trust path for recommendations in web-based social networks, Procedia Comput. Sci. 148 (2019) 181–190.

[16] S. Bin, G.X. Sun, N. Cao, J.M. Qiu, Z.Y. Zheng, G.H. Yang, L. Xu, Collaborative filtering recommendation algorithm based on multi-relationship social network, CMC-Computers, Materials & Continua 60 (2) (2019) 659–674.

[17] H. Wu, K. Yue, Y. Pei, Collaborative Topic Regression with social trust ensemble for recommendation in social media systems, Knowl.-Based Syst. 97 (4) (2016) 111–122.

[18] G.L. Tian, S. Zhou, G.X. Sun, C.C. Chen, A novel intelligent recommendation algorithm based on mass diffusion, Discrete Dynamics in Nature and Society 11 (2020) 1–9.

[19] M. Ayub, M.A. Ghazanfar, Z. Mehmood, Unifying user similarity and social trust to generate powerful recommendations for smart cities using collaborating filtering-based recommender systems, Soft. Comput. 24 (15) (2020) 11071–11094.

[20] Y. Wu, Y. Zhao, S. Wei, Collaborative filtering recommendation algorithm based on interval-valued fuzzy numbers, Applied Intelligence 50 (3) (2020).

[21] A.D.R. Oliveira, Trust-based recommendation for the social Web, IEEE Lat. Am. Trans. 10 (2) (2012) 1661–1666.

[22] J. Shu, X. Shen, L. Hai, et al, A content-based recommendation algorithm for learning resources, Multimedia Syst. 1 (2017) 1–11.

[23] M. Xin, L. Wu, S. Li, A User Profile Awareness Service Collaborative Recommendation Algorithm Under LBSN Environment, International Journal of Cooperative Information Systems 28 (21) (2019).

[24] S. Bin, G. Sun, Matrix Factorization Recommendation Algorithm Based on Multiple Social Relationships, Mathematical Problems in Engineering 2021 (2021) 1–8.

[25] G. Sun, C.-C. Chen, S. Bin, Study of Cascading Failure in Multisubnet Composite Complex Networks, Symmetry 13 (3) (2021) 523.