Research Article
A Novel Intelligent Recommendation Algorithm Based on Mass Diffusion

Guanglai Tian,1 Shuang Zhou,1 Gengxin Sun,1 and Chih-Cheng Chen2,3

1School of Data Science and Software Engineering, Qingdao University, Qingdao 266071, China
2Department of Industrial Engineering and Management, Chaoyang University of Technology, Taichung 413310, Taiwan
3Information and Engineering College, Jimei University, Xiamen, Fujian 361021, China

Correspondence should be addressed to Gengxin Sun; sungengxin@qdu.edu.cn and Chih-Cheng Chen; 3343033397@qq.com

Received 1 June 2020; Revised 17 October 2020; Accepted 31 October 2020; Published 16 November 2020

Academic Editor: Chi-Hua Chen

Copyright © 2020 Guanglai Tian et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Social recommendation algorithm is a common tool for recommending interesting or potentially useful items to users amidst the sea of online information. The users usually have various relationships, each of which has its unique impact on the recommendation results. It is unlikely to make accurate recommendations solely based on one relationship. Based on user-item bipartite graph, this paper establishes a multisubnet composited complex network (MSCCN) of multiple user relationships and then extends the mass diffusion (MD) algorithm into a novel intelligent recommendation algorithm. Two public online datasets, namely, Epinions and FilmTrust, were selected to verify the effect of the proposed algorithm. The results show that the proposed intelligent recommendation algorithm with two types of relationships made much more accurate recommendations than that with a single relationship and the traditional MD algorithm.

1. Introduction

The era of information technology is defined by the explosion of all sorts of information. The wealth of information not only provides human with more choices but also causes serious information overload [1–5]; amidst the sea of information, it is difficult for users to find the information that they really need. As a result, how to provide users with useful information from massive data has become a research hotspot. To solve information overload, many recommendation systems have been developed to predict user preference based on their historical data and recommend interesting or potentially useful information to them [6–8].

The common recommendation methods can be divided into collaborative filtering (CF) recommendation [9], content-based recommendation [10], spectrum analysis recommendation [11, 12], and hybrid recommendation [13, 14]. The recommendation methods based on user or item similarity are also quite popular, namely, mass diffusion (MD) algorithm and heat spreading (HeatS) algorithm. The MD is a classic algorithm capable of making accurate recommendations of items to users. By the principle of energy distribution, the algorithm spreads energy from the initial items to different items and recommend the items with relatively high energy at a high probability.

After predicting user preference, Bin et al. [15] evaluated the popularity of each item and used it to modify the MD algorithm. Based on mass diffusion and heat spreading in physical kinetics, Hu et al. [16] summarized the mechanism of energy distribution and realized a recommendation mechanism based on the tripartite of user, item, and topic. Li et al. [17] managed to improve recommendation quality through singular value decomposition (SVD) of the recommendation results.

The above studies mainly recommend items to users based on the item scores rated by users. However, none of them have considered the influence of multiple social relationships on user preference. In fact, the preference of a user is affected by his/her trusted friends. Therefore, this paper constructs a complex network with multiple social relationships. The network incorporates multiple user
relationships into the recommendation system, which effectively improves the recommendation accuracy.

2. Multisubnet Composited Complex Network (MSCCN) Model

There are various systems that can be modeled in our daily life, such as the urban system, ecosystem, transport system, and recommendation system. In a complex system, each object can be treated as a node, and each relationship between objects can be represented as an edge, turning the system into a complex network [18–20].

The MSCCN model [20–23] can describe multiple relationships among heterogeneous individuals in complex systems. For example, users and items can be considered nodes, while the user relationships and user-item relationships can be regarded as edges in the recommendation system.

By integrating multiple subnets and relationships among complex systems, the MSCCN model can be expressed as a four-tuple $G = (V, E, R, F)$, where $V = \{v_1, v_2, \ldots, v_m\}$ is the set of nodes, $V = \{v_1, v_2, \ldots, v_m\} \subseteq V \times V$ is the set of edges among the nodes, $R = R_1 \times \cdots \times R_i \times \cdots \times R_n = \{(r_1, 1, r_1, \ldots, r_n) \mid r_i \in R_i, 1 \leq i \leq n\}$ is the set of relationship types, and $R = \{r_1, r_2, \ldots, r_m\}$ as a four-tuple $G = (V, E, R, F)$.

The structure of a typical MSCCN model is presented in Figure 1, where $R = R_1 \times R_2$, $R_1 = \{r_1\}$, and $R_2 = \{r_2\}$. As shown in Figure 1, the edges $v_1, v_2, v_3, v_4$ only have $R_1$ relationship, the edges $v_2, v_3, v_5, v_6$ only have $R_2$ relationship, and the edge $v_3, v_4$ has both $R_1$ and $R_2$ relationships.

3. MD Algorithm Based on MSCCN Model

3.1. MD Algorithm. As proposed by Li et al. [17], the MD algorithm can be implemented as follows: First, a unit of energy is placed on the items selected by the target user. Then, the energy spreads from the items to the users who select them. At this time, the energy $h_i$ obtained by user $u_i$ can be calculated by

$$h_i = \sum_{\beta=1}^{\gamma} a_{\beta\beta} f_\beta,$$

where $\gamma$ is the degree of the item $o_{\beta\beta}$; $\gamma$ is the number of items; $f_\beta$ is the energy of the item $o_{\beta\beta}$; $a_{\beta\beta}$ is the selection variable of items. If item $o_{\beta\beta}$ is selected by user $u_i$, then $a_{\beta\beta} = 1$.

Next, the energy that is spread to users is distributed to the items selected by users. After the distribution, the energy $f'_a$ obtained by item $o_a$ can be expressed as

$$f'_a = \sum_{i=1}^{\delta} \frac{a_{ia} h_i}{d_i},$$

where $d_i$ is the degree of user $u_i$; $\delta$ is the number of users.

Let $\vec{f}$ and $\vec{f}'$ be the initial and final energy of items, respectively. Then, the entire diffusion process can be defined as

$$\vec{f}' = W^M \vec{f},$$

where $W^M$ is the state transition matrix. Each element $w_{ab}^M$ of the matrix satisfies

$$w_{ab}^M = \sum_{i=1}^{\gamma} \frac{a_{ia} a_{ib}}{d_i} \delta.$$  

After diffusion, the unselected items are arranged in descending order by the resource quantity, forming a recommendation list. This recommendation list is generated solely based on user-item relationships, without considering the social relationships of the target user.

The workflow of the MD algorithm is illustrated in Figure 2. It can be inferred that $\vec{f} = (1, 0, 1, 0)$ is the initial energy of items and $\vec{f}' = (19/24, 5/24, 5/8, 3/8)$ is the final energy of items after being processed by MD algorithm.

3.2. MD Algorithm Based on MSCCN Network (SMD). By the MD algorithm, the recommendation list is generated solely based on user-item relationships, failing to consider the social relationships of the target user. To overcome the defect, this paper introduces the social relationship into the recommendation system by setting up a multirelationship network, allowing the initial energy of item nodes to spread along the edges in the network.

Let $r_1$ and $r_2$ be user-item relationship and user relationship in the multirelationship network, respectively. It is assumed that when weights of the two types of relationship are both 1, then the proportionality coefficient $s_{f_1}$ of $r_1$ and that $s_{f_2}$ of $r_2$ satisfy $s_{f_1} + s_{f_2} = 1$. Assuming that $s_{f_1} = p$, then $s_{f_2} = 1 - p$, $p \in (0, 1)$. After introducing one type of social relationship, the SMD algorithm can be implemented as follows:

(i) Step 1. One unit of energy is distributed to the items selected by the target user, serving as the initial energy of items.
(ii) Step 2. The initial energy of items spread to the users who have selected them. The energy $h_i$ obtained by user $u_i$ can be described as

$$h_i = \sum_{\beta=1}^{n} a_{i\beta}^1 f_{\beta}$$

where $a_{i\beta}^1$ and $a_{i\beta}^2$ are selection variables of items; $k_{ij}^r$ is the degree of item $o_{ij}$ for relationship $r_1$. If $a_{i\beta}^1 = 1$, user $u_i$ and item $o_{ij}$ have $r_1$ relationship. If $a_{i\beta}^2 = 1$, user $u_i$ and item $o_{ij}$ have $r_2$ relationship.

(iii) Step 3. The energy spreads to users and items along edges in the composite network. Then, the energy $h_j$ obtained by user $u_j$ and that $f_a^c$ obtained by item $o_a$ can be, respectively, described as

$$h_j = \sum_{i=1}^{m} \left( \frac{a_{1ij}^1 s_j f_{ij}^1 + a_{2ij}^2 s_j f_{ij}^2}{k_{ij}^r} \right)$$

$$f_a^c = \sum_{i=1}^{m} \left( \frac{a_{1ia}^1 s_i f_{ia}^1 + a_{2ia}^2 s_i f_{ia}^2}{k_{ij}^r} \right)$$

(iv) Step 4. The energy allocated to the user via $r_2$ relationship further spreads to the items. Then, the energy $f_a^m$ obtained by the item $o_a$ can be depicted as

$$f_a^m = \sum_{i=1}^{m} \frac{a_{1ia}^1 h_i}{k_{ij}^r}$$

Thus, the total energy $g_a$ obtained by item $o_a$ can be defined as

$$g_a = f_a^c + f_a^m.$$
Figure 3: The diffusion process of SDM1 algorithm.

Figure 4: The diffusion process of SDM2 algorithm.
If $q = 0$ or $q = 1$, that is, $sf_2 = 0$ or $sf_3 = 0$, the SMD2 algorithm degenerates into the SMD1 algorithm.

4. Results and Discussion

4.1. Experimental Data. Two public online datasets, Epinions and FilmTrust, were selected to evaluate the performance of the proposed algorithm. The Epinions dataset contains 40,272 users, 139,738 items, 487,182 user relationships, and more than 664,000 item scores rated by users. The FilmTrust dataset provides 1,050 users, 2,071 items, >1,800 user relationships, and 35,497 item scores rated by users. The original data were preprocessed by removing the relationships of the users that have not selected any items.

4.2. Evaluation Indices. Fivefold cross validation was employed to test the performance of the SMD algorithm. Specifically, the preprocessed data were randomly split into five subsets. For each experiment, a random subset was chosen as the test set, and the remaining subsets were chosen as training sets. Five experiments were conducted to ensure that the SMD algorithm is tested on each subset. The results of the five experiments were averaged to obtain the final result.

Then, the mean ranking score (RS) [15] was introduced to evaluate the ranking accuracy of the SMD algorithm, and the Hamming distance [24] was adopted to measure the diversity of recommendation results.

Suppose the target user $u_t$ has chosen the item $o_j$ in the test set. Then, the ranking $r_{ij}$ of item $o_j$ in the recommendation list can be calculated. For the mean RS of all the items chosen by the user from the test set, the higher the value of mean RS, the better the accuracy of the recommendation algorithm. The RS of user $u_t$ can be calculated by

$$RS_t = \frac{1}{|E_t^p|} \sum_{(ia) \in E_t^p} RS_{ia},$$

where $E_t^p$ is the number of items preferred by user $u_t$ in the test set; $ia$ is the item $a$ preferred by user $i$ in the test set.

The Hamming distance can be defined as

$$H_{ut}(L) = 1 - \frac{Q_{ut}(L)}{L},$$

where $u$ and $t$ are two users; $Q_{ut}(L)$ is the number of overlapping items in the recommendation lists of the two users; $L$ is the length of the recommendation list. If $H_{ut}(L) = 1$, the two recommendation lists have no overlapping items; if $H_{ut}(L) = 0$, the two recommendation lists are identical.

4.3. Results Analysis. The numerical simulation was performed to determine the values of $p$ and $q$. Figure 5 presents the simulation results of SMD1 on the two datasets with $q = 0$ and $p$ changing between various values.

As shown in Figure 5, the optimal RS value was reached at $p = 0.4$ on Epinions dataset, indicating that SMD1 has the highest accuracy at $p = 0.4$. Similarly, it can be inferred that SMD1 has the highest accuracy at $p = 0.9$ on FilmTrust dataset.

Let $O_u$ and $O_v$ be the sets of items chosen by users $u$ and $v$, respectively. The greater the number of overlapping items in the two sets is, the more likely that the two users have the same interests, and the greater they influence each other. The user similarity can be defined as

$$f_{uv} = \frac{|O_u \cap O_v|}{|O_u \cup O_v|}$$  \hspace{1cm} (13)

If $f_{uv} > 0.2$, the two users have similar interests.

Figure 6 displays the simulation results of SMD2 with both $p$ and $q$ changing between various values.

As shown in Figure 6, the RS value reached the minimum at $p = 0.4$ and $q = 0.2$ on Epinions dataset. This means, on Epinions dataset, the SMD2 has the highest accuracy at $p = 0.4$ and $q = 0.2$. Similarly, the SMD2 has the highest accuracy at $p = 0.9$ and $q = 0.9$ on FilmTrust dataset. Since $r_3$ relationship is denser than $r_2$ relationship, the recommendation accuracy increases as $q$ approaches 1 on FilmTrust dataset.

For comparison, traditional recommendation algorithms, namely, HeatS algorithm and Hybrid algorithm [25], were also applied on Epinions dataset and FilmTrust dataset. The RS values of the Hybrid algorithm were minimized on the two datasets, when the regularization parameter $\lambda$ was 0.67 and 0.5, respectively. Table 1 compares the results of the contrastive algorithms at the optimal values of $p$ and $q$.

As shown in Table 1, SMD2 achieved slightly higher recommendation accuracy than SMD1. The accuracy of the two algorithms was marked higher than that of the traditional algorithms on the two datasets. The results indicate that the recommendation system becomes more accurate, thanks to the introduction of multiple social relationships.

The diversity of the recommendation list is also an important indicator of recommendation quality. Figure 7 shows the diversity trends of recommendation list on Epinions dataset and FilmTrust dataset, with the list length changing between various values.

As shown in Figure 7, the length of the recommendation list is positively correlated with the similarity between recommendation lists for different users. On the two datasets, SMD2 and SMD1 had lower recommendation diversity than the MD algorithm. The relatively low diversity is the result of the following: After user relationship(s) is introduced, the recommendation list for a user is affected by his/her historical scores and also the historical scores of the relevant users. In other words, the recommendation list of a user will bear a high resemblance to that of users, who has social relationships with the user. The HeatS algorithm achieved higher recommendation diversity than the MD algorithm because it attracts the historical scores of the user to the less popular items.

In the real world, some users only select very few items. For implicitly, the users who have purchased fewer than 30
items were defined as small-degree users, and the other users as generous-degree users. The small-degree users generate relatively few data to be used for recommendation. Therefore, the recommendation to such users is usually of low accuracy. Then, the users with $k_i \leq 30$ ($k_i$ is the degree of user $u_i$) on the two selected datasets were treated as a small-degree user group. Figure 8 shows the recommendation results of SMD for this user group on the two datasets, with $q = 0$ and $p$ changing between various values.

As shown in Figure 8, SMD1 algorithm achieved the minimum RS at $p = 0.2$ and $p = 0.6$, respectively, on Epinions and FilmTrust datasets. The $p$ value decreased from the scenario in Figure 5, where SMD1 achieved the highest accuracy at $p = 0.4$ and $p = 0.9$, respectively. The $p$ value determines the number of resources that need to be transmitted through the social network to reach the item during the diffusion process. The smaller the $p$ value, the greater the resource demand.

| Dataset      | MD   | HeatS  | Hybrid     | SMD1      | SMD2      |
|--------------|------|--------|------------|-----------|-----------|
| Epinions     | 0.18216 | 0.20215 | 0.17063 | 0.16203 | 0.16128 |
| FilmTrust    | 0.04032 | 0.04232 | 0.04015 | 0.03894 | 0.03891 |
To further explore the relationship between parameters $p$ and $k_i$, the RS values were grouped by user degree. Figure 9 shows the $p$ values corresponding to the optimal RS values on the two datasets.

As shown in Figure 9, the relationships between $p$ and $k_i$ exhibited a consistent trend on the two datasets: the smaller the degree of the target user, the smaller the $p$ value for the RS to reach the optimal value. A user with a small degree chooses only a few items and provides limited information for the recommendation algorithm. Therefore, the recommendation accuracy for small-degree users needs to be improved based on social network information.

To demonstrate the superiority of SMD in the recommendation for small-degree users, the small-degree users satisfying $k_i \leq 30$ were grouped by $k_i$, as $p$ and $q$ reached the optimal values on the two datasets. Figure 10 compares the optimal RS values under SMD and MD for small-degree users.

As shown in Figure 10, the highest RS values were achieved by MD algorithm on Epinions dataset, followed in
turn by SMD1 and SMD2. That is, for small-degree users, SMD outperforms the MD in recommendation accuracy; the advantage increases with the reduction of user degree. The same was observed on FilmTrust dataset. To sum up, SMD is more effective than MD in the recommendation for small-degree users.

5. Conclusions

This paper introduces social information to the MSCCN model to create a complex network of multiple relationships. On this basis, the MD algorithm was improved into the SMD. Experimental results show that the SMD made a highly accurate recommendation to users because it considers the potentially useful items of different users in the social network. It is also learned that SMD2 outperformed SMD1 and MD, indicating that the integration of two types of relationships is better than incorporating only one type of relationship or considering no social relationship. In other words, the recommendation accuracy can be greatly improved by introducing multiple social relationships between users. In addition, the experiments also demonstrated that...
the SMD is highly effective in making accurate recommendations to small-degree users. The future research will discover the implicit social relationships between users and evaluate their impacts on recommendation results.

Data Availability

The basic data used in this paper were downloaded from two public datasets online: the Epinions http://www.trustlet.org/opinions.html and the FilmTrust https://www.librec.net/datasets.html.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was funded by Shandong Provincial Natural Science Foundation, China, Grant no. ZR2017MG011.

References

[1] D. Bawden, C. Holtham, and N. Courtney, “Perspectives on information overload,” Aslib Proceedings, vol. 51, no. 8, pp. 249–255, 2013.

[2] Y. Liu, H. Peng, and J. Wang, “Verifiable diversity ranking search over encrypted outsourced data,” Computers. Materials & Continua, vol. 55, no. 1, p. 37, 2018.

[3] S. Fekri-Ershad, “Gender classification in human face images for smart phone applications based on local texture information and evaluated kullback-leibler divergence,” Traitement Du Signal, vol. 36, no. 6, pp. 507–514, 2019.

[4] R. Meng, S. G. Rice, J. Wang, and X. Sun, “A fusion steganographic algorithm based on faster R-CNN,” Computers. Materials & Continua, vol. 55, no. 1, pp. 1–16, 2018.

[5] B. K. Lee and W. N. Lee, “The effect of information overload on consumer choice quality in an on-line environment,” Psychology & Marketing, vol. 21, no. 3, pp. 159–183, 2010.

[6] L. Lü, M. Medo, C. H. Yeung, Y.-C. Zhang, Z.-K. Zhang, and T. Zhou, “Recommender systems,” Physics Reports, vol. 519, no. 1, pp. 1–49, 2012.

[7] Y. Li, ”Design and implementation of intelligent travel recommendation system based on internet of things,” Ingénierie Des Systèmes D’Information, vol. 23, no. 5, pp. 159–173, 2018.

[8] N. Hasanzadeh and Y. Forghani, “Improving the accuracy of M-distance based nearest neighbor recommendation system by using ratings variance,” Ingénierie Des Systèmes D’Information, vol. 24, no. 2, pp. 131–137, 2019.

[9] J. B. Schafer, D. Frankowski, J. Herlocker et al., “Collaborative filtering recommender system,” in The Adaptive Web, Lecture Notes in Computer Science, vol. 4321, pp. 291–324, Springer-Verlag, Berlin, Germany, 2007.

[10] M. J. Pazzani and D. Billsus, “Content-based recommendation systems,” in The Adaptive Web, Lecture Notes in Computer Science, vol. 4321, pp. 325–341, Springer-Verlag, Berlin, Germany, 2007.

[11] S. Maslov and Y.-C. Zhang, “Extracting hidden information from knowledge networks,” Physical Review Letters, vol. 87, no. 24, pp. 248701–248705, 2001.

[12] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins, “Eigentaste: a constant time collaborative filtering algorithm,” Information Retrieval, vol. 4, no. 2, pp. 133–151, 2001.

[13] M. J. Pazzani, “A framework for collaborative, content-based and demographic filtering,” Artificial Intelligence Review, vol. 13, no. 5/6, pp. 393–408, 1999.

[14] K. Yoshii, M. Goto, K. Komatani, T. Ogata, and H. G. Okuno, “An efficient hybrid music recommender system using an incrementally trainable probabilistic generative model,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, no. 2, pp. 435–447, 2008.

[15] S. Bin, G. Sun, N. Cao et al., “Collaborative filtering recommendation algorithm based on multi-relationship social network,” Computers, Materials & Continua, vol. 60, no. 2, pp. 659–674, 2019.

[16] J. M. Hu and X. Lin, “Design and implementation of recommendation algorithm based on user-socialized resource-vocabulary three-part graph,” Information Studies: Theory & Application, vol. 39, no. 3, pp. 130–134, 2016.

[17] W. Li, J. Qi, Z. Yu, and D. Li, “A social recommendation method based on trust propagation and singular value decomposition,” Journal of Intelligent & Fuzzy Systems, vol. 32, no. 1, pp. 807–816, 2017.

[18] B. Beirami and M. Mokhtarzade, “Spatial-spectral random patches network for classification of hyperspectral images,” Traitement Du Signal, vol. 36, no. 5, pp. 399–406, 2019.

[19] W. Meng, C. Mao, J. Zhang, J. Wen, and D. Wu, “A fast recognition algorithm of online social network images based on deep learning,” Traitement Du Signal, vol. 36, no. 6, pp. 575–580, 2019.

[20] G. Sun and S. Bin, “A new opinion leaders detecting algorithm in multi-relationship online social networks,” Multimedia Tools and Applications, vol. 77, no. 4, pp. 4295–4307, 2018.

[21] G. Sun and S. Bin, “Router-level internet topology evolution model based on multi-subnet composited complex network model,” Journal of Internet Technology, vol. 18, no. 6, pp. 1275–1283, 2017.

[22] S. Bin and G. Sun, “Optimal energy resources allocation method of wireless sensor networks for intelligent railway systems,” Sensors, vol. 20, no. 2, p. 482, 2020.

[23] G. Sun, S. Bin, M. Jiang et al., “Research on public opinion propagation model in social network based on blockchain,” Computers, Materials & Continua, vol. 60, no. 3, pp. 1015–1027, 2019.

[24] A. Bellogín, I. P. Castells, and I. Cantador, “Statistical biases in information retrieval metrics for recommender systems,” Information Retrieval Journal, vol. 20, no. 6, pp. 606–634, 2017.

[25] T. Zhou, Z. Kuscsik, J.-G. Liu, M. Medo, J. R. Wakeling, and Y.-C. Zhang, “Solving the apparent diversity-accuracy dilemma of recommender systems,” Proceedings of the National Academy of Sciences, vol. 107, no. 10, pp. 4511–4515, 2010.