Harnessing heterogeneous social networks for better recommendations: A grey relational analysis approach

Lijuan Weng[a], Qishan Zhang[a,*], Zhibin Lin[b], Ling Wu[c]

[a] School of Economics and Management, Fuzhou University, Fuzhou, China

[b] Durham University Business School, Mill Hill Lane, Durham DH1 3LB, United Kingdom.

[c] School of Mathematics and Computer Science, Fuzhou University, Fuzhou, China

Abstract

Most of the extant studies in social recommender system are based on explicit social relationships, while the potential of implicit relationships in the heterogeneous social networks remains largely unexplored. This study proposes a new approach to designing a recommender system by employing grey relational analysis on the heterogeneous social networks. It starts with the establishment of heterogeneous social networks through the user-item bipartite graph, user social network graph and user-attribute bipartite graph; and then uses grey relational analysis to identify implicit social relationships, which are then incorporated into the matrix factorization model. Five experiments were conducted to test the performance of our approach against four state-of-the-art baseline methods. The results show that compared with the baseline methods, our approach can effectively alleviate the sparsity problem, because the heterogeneous

* Corresponding author.

E-mail address: M180710006@fzu.edu.cn (L. Weng), zhangqs136@163.com (Q. Zhang), zhibin.lin@durham.ac.uk (Z. Lin), wuling1985@fzu.edu.cn (L. Wu).

social network provides richer information. In addition, the grey relational analysis method has the advantage of low requirements for data size and efficiently relieves the cold start problem. Furthermore, our approach saves processing time, thus increases recommendation efficiency. Overall, the proposed approach can effectively improve the accuracy of rating prediction in social recommendations and provide accurate and efficient recommendation service for users.

Keywords: Recommender system, Heterogeneous social network, Grey relational analysis, Recommendation algorithm, Implicit social relationships, User profile.

1. Introduction

The proliferation of social media has profoundly changed people's behavior patterns and social interactions. Social network services enable people to communicate without the restrictions of time, space, distance, and cost, creating an excellent development environment for a social recommender system. A reliable and accurate forecasting model is of great importance (Altan & Karasu, 2019; Karasu, Altan, & Bekiros, 2020), because it contributes to the accuracy of rating prediction in recommendations and helps users to make effective decisions, particularly when they are faced with a massive amount of information. Recently, recommender systems have received growing research attention (Alabdulrahman & Viktor, 2021; Jakomin, Bosnić, & Curk, 2020; Walek & Fojtik, 2020). In the pursuit of a high-quality forecasting model, scholars have proposed various frameworks that incorporate data mined from social networks into the recommender systems (Camacho & Alves-Souza, 2018; Deng, Huang, & Xu, 2014; Jiang et al., 2015). By incorporating the user's social relation information, a social recommender system can accurately recommend items to users and effectively improve the recommendation performance.

Most of the extant studies in the social recommender systems are based on explicit social relationships, which do not always exist (Tang et al., 2017). Scholars have recently started to incorporate implicit relationships into the recommendation process to improve the recommendation performance, for example, Al-Sabaawi, Karacan, and Yenice (2020); Ahmadian et al. (2020); Li, Liu, Ren, and Chang (2020). The afore-mentioned studies are all based on homogeneous social networks. The nodes and links in such networks are of the same type. However, there are various different relationships on social networks. As shown in Fig. 1, there are relationships of trust, following, and friendship, and these relationships only exist in a certain aspect. For example, user u_1 follows user u_3 because they have the same preference for movies, but their tastes for food may be different. Therefore, algorithms based on the equality of user nodes and similar relationships may produce inaccurate recommendations.

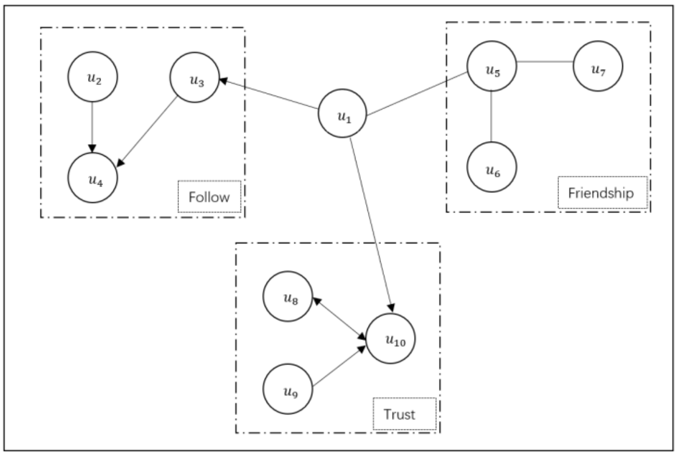

Fig. 1. Social network relationships

To solve the above problems, several studies have tried to use heterogeneous social networks to develop the recommendation model (Wang, Song, Li, Zhang, & Han, 2015; Wang, Xia, Tang, Wu, & Zhuang, 2017). A heterogeneous social network refers to a social network where the type of entity is greater than one, or the type of edge is greater than one. The multiple types of entities and edges in a heterogeneous social network provide researchers with richer information, which can be effectively used to alleviate the sparsity problem and boost the performance of recommendations. Moreover, the information in the heterogeneous social network can be used to calculate the implicit relationships between users (Yu, Gao, Li, Yin, & Liu, 2018). However, calculating the implicit relationships based on the heterogeneous social network for recommendation remains largely unexplored in the literature. Furthermore, using the traditional cosine method, Pearson correlation coefficient method, and Jaccard similarity method to calculate user similarity has many problems such as potential errors, complicated calculation, and the rating prediction obtained is often not accurate enough to achieve the expected effect in the top-N recommendations.

A recommender system can be regarded as a grey system. The term "grey" refers to the uncertainty in data or the noise in the data, suggesting that the grey system's relational analysis requires limited information to obtain the results (Deng, 1989). The essence of grey relational analysis is to compare the similarities between a reference with various alternatives and such comparisons enable decisions to be made with limited information (Liu & Lin, 2010). It is based on the grey system theory, which was first proposed by Deng in 1982 (Deng, 1989). Grey system theory has been proved to be particularly useful in dealing with data uncertainties (Liu & Lin, 2010; Škrinjarić, 2020). The main advantages of grey relational analysis are that its methodology is non-parametric, it is relatively simple in comparison with other major modeling alternatives and it is less subjective, compared to methodologies that require the input of weights from the decision-makers (Liu & Lin, 2010). Despite the various advantages, they have not been adequately exploited for the research on recommender systems (Tao & Dang, 2018).

This study therefore proposes a new Social Recommendation algorithm based on Grey Relational Analysis (SRGRA), with consideration of the heterogeneity of traditional social networks to improve the rating prediction accuracy and alleviate the problems of cold start and data sparsity. This is a hybrid model that overcomes the shortcomings of a single model, and improves the prediction accuracy (Altan & Karasu, 2020; Altan, Karasu, & Bekiros, 2019). What differentiates our method from the existing methods is that we design the recommendation method based on heterogeneous social networks, which is established through the user-item bipartite graph, user social network graph, and user-attribute bipartite graph. In addition, we adopt a grey relational analysis to identify implicit social relationships based on user profiles.

This study makes several contributions to the recommender system literature. First, this paper is among the first studies that incorporate the grey relational analysis method into the recommender system. This new approach has the capacity to mine data on the implicit social relationships, thus it can provide accurate recommendations for users who are not directly connected in the heterogeneous social networks. Second, the implicit social relationships of the target users are identified by integrating user profiles, which can effectively alleviate the problems of cold start and data sparsity, thus improving the accuracy of rating prediction. Third, the implicit social relationships are integrated into the matrix factorization model, which can recommend items that closely suit the target users' preferences. The experimental results confirm that our proposed method reduces the rating prediction error, especially for cold-start users.

The remainder of this paper is organized as follows. Section 2 provides a review of the related studies. Section 3 introduces the proposed method in detail. Section 4 presents the experimental results. Section 5 concludes and discusses the implication of this research.

2. Related works

Social media are prevalent in our everyday life. Users love to share their experiences with "friends" online, forming rich social relationships (Ahmadian et al., 2020). These data can be harnessed in the recommender system to produce effective recommendations that match users' needs and preferences. Recent recommender systems have increasingly used data mined from social media (Wu et al., 2016). Integrating social network information into the traditional recommender system helps to predict the user's rating on a certain item (Mao, Lu, Zhang, & Zhang, 2017). Several approaches have been proposed for advancing recommendation accuracy. For example, Qian, Zhao, Tang, and Zhang (2016) propose a framework based on users' global rating reputation and local rating similarity. Guo, Luo, Dong, and Yang (2018) adopt a private graph-link analysis for applying differential privacy more accurately in the recommendation. Lai, Lee, and Huang (2019) propose a model that is based on the data of product popularity, user interactions, and trust relationships. Li, Xiong, Wang, Chen, and Xiong (2019) investigate the impact of users on their neighbors by analyzing the topology of social networks, and propose an indirect interactive recommendation method. Zhang et al. (2019) propose a session-based recommendation algorithm through social context analysis and collaborative filtering according to the target user's behavior and the perceived intention and identity. Zhang, Li, Wang, and Yang (2020) use social network and user rating matrix to construct a new matrix, and then propose a social recommendation algorithm based on stochastic gradient matrix factorization.

The above social recommendation algorithms can improve the performance of a social recommender system to a certain extent, but they are mainly based on explicit social relationships, which do not always exist in social networks, and even if explicit social relations do exist, the data are always sparse and noisy (Tang, Hu, & Liu, 2013). To solve these problems, researchers have started to consider implicit social relationships. For example, Ahmadian et al. (2020) propose a link prediction model based on Dempster-Shafer theory to calculate the implicit relationship between users in social recommender systems. Ngaffo et al. (2020) develop a recommendation algorithm model that is based on the inference of implicit trust relationships while integrating a time feature into the system. Li et al. (2020) adopt a "trust propagation and aggregation strategy" to identify the indirect trust between users in social networks. Reafee et al. (2016) employ link prediction techniques to extract the implicit relation in social networks and propose a method based on both explicit and implicit friendships. The above scholars infer implicit social relationships from explicit social relationships, while others identify implicit social relationships through a user-item rating matrix. Ma (2013) suggests that the top-k similar users of each user can be identified by calculating the Pearson correlation coefficient between each user. Taheri et al. (2017) suggest the use of Hellinger distance to extract the implicit social relationship in the user-item bipartite graph.

The effect of social recommendation algorithms can deteriorate rapidly with the gradual sparseness of social networks, however, the rich information in heterogeneous social networks can be explored and harnessed to effectively alleviate the sparsity problem (Wang, Song, Li, Zhang, & Han, 2015). In heterogeneous social networks, users are linked to each other through different types of nodes and linked relationships, and the multiple types of entities and edges in heterogeneous social networks can be harnessed for improving recommendation performance. For example, Pham, Li, Cong, and Zhang (2016) model the interaction between different types of nodes and links, and transform the recommendation problem into a node proximity calculation problem on a heterogeneous graph. Yu, Gao, Li, Yin, and Liu (2018) model the whole system as a heterogeneous information network, and calculate the similarity of users through meta-path based embedding representation learning to identify implicit friends.

User profiles can be used to construct heterogeneous social networks and the information can be used to calculate the implicit relationship between users. Recommendation algorithm based on user profiles integrates users' gender, age, occupation, hobbies, and other information into the detection of user similarity. It does not need domain knowledge and historical data, and does not rely on the attributes of items, so it can effectively alleviate the cold start problem (Gogna & Majumdar, 2015). Zhao et al. (2016) suggest applying user profiles extracted from social media and product information as inputs into the recommendation modeling. Bertani, Bianchi, and Costa (2020) propose an algorithm that learns user-profiles and combines novelty and item popularity to generate personalized recommendations. Mazhari, Fakhrahmad, and Sadeghbeygi (2015) measure the similarity between two users based on their user profile, thereby recommending friends for the target user.

Grey relational analysis has the unique capability of deriving useful information from limited, incomplete, or noisy data (Liu & Lin, 2010). Given its various advantages, grey relational analysis has been developed and successfully applied in many fields from social to natural sciences (Kayacan, Ulutas, & Kaynak, 2010). Sun, Guan, Yi, and Zhou (2018) apply the grey relational analysis to the Hesitant Fuzzy Sets for pattern recognition, which has been shown to outperform the conventional fuzzy measures. Škrinjarić (2020) uses grey relational analysis as a tool for portfolio selection, and reorganizes dynamic portfolios based on the results of grey relational analysis. Wu and Zhang (2015) propose a social network link community detection algorithm based on k-means and grey relational analysis, which can accurately identify overlapping communities in social networks. However, the application of grey relational analysis in recommender systems is rather scarce, with the exception of Tao and Dang (2018), who construct a collaborative filtering recommendation algorithm based on grey relational clustering, and apply it to the collaborative filtering recommender system, solving the high sparsity and high dimension data problems. The above works show that grey relational analysis has a strong potential in advancing recommender systems research. Therefore, this paper uses the grey relational analysis to calculate the similarity between users in the heterogeneous social networks, so as to recommend the items that closely match target users' preferences.

3. Proposed approach

3.1. Preliminaries

The basic definitions and concepts used in this paper are as follows.

Definition 1 Heterogeneous Social Network (HSN): Given a directed graph $G = (V, E, T)$, where V represents the entity set, E represents the edge set. Each entity v and edge e are tied with a mapping function $\phi(v): V \to T_V$ and $\phi(e): E \to T_E$, where T_V and T_E represent the possible types of entity and edge in the graph, respectively. When the number of entity types $|T_V| > 1$ or the number of edge types $|T_E| > 1$, such a social network is called a heterogeneous social network. As shown in Fig. 2, this heterogeneous social network is established by the user-item bipartite graph, user social network graph and user-occupation bipartite graph, which contain three types of entities and three types of edges.

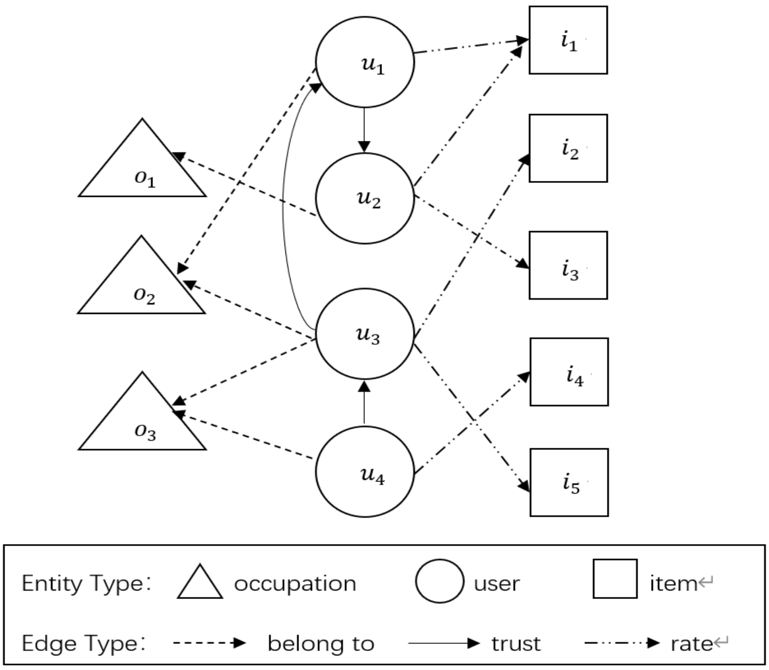

Fig. 2. An example of HSN

Definition 2 Implicit social relationship: An implicit social relationship refers to the relationship between two users who have similar preferences but have no direct connection in a social network. As shown in Fig. 2, user u_1 and user u_4 do not have a connection edge, they do not know each other. However, if both their professions are lawyer, their demand for legal books may be the same, that is, their preferences are similar, so there is an implicit social relationship between them.

Definition 3 User profile: A user profile is also called user statistical data or user attributes. It refers to the characteristics of the user, including gender, age, education level, occupation, postcode and others. A user profile is generally defined in the following format: $P_{u_i} = \{P_{u_i}(1), P_{u_i}(2), P_{u_i}(3), \dots, P_{u_i}(n)\}$, where $P_{u_i}(n)$ represents the nth attribute value of user u_i, $n \geq 3$. Usually the information that a new user is required to provide at registration is about a user's profile. As shown in Fig. 2, the occupation is an attribute of the user.

Definition 4 Grey relational degree: Let the system behavior sequence $X_f = \{x_f(1), x_f(2), \ldots, x_f(n)\}$, X_i denotes the sequence of characteristic behavior system, where $i, f = \{1,2,\ldots,m\}, m \geq 3, i \neq f$. For $\xi \in (0,1)$, the grey relational coefficient is given by:

$$\gamma\left(x_i(k), x_f(k)\right) = \frac{\min_f \min_k |x_i(k)-x_f(k)| + \xi \max_f \max_k |x_i(k)-x_f(k)|}{|x_i(k)-x_f(k)| + \xi \max_f \max_k |x_i(k)-x_f(k)|} \qquad (1)$$

where $k = \{1,2,\ldots,n\}, n \geq 3$, ξ represents the distinguishing coefficient, so the grey relational degree of X_i and X_f can be obtained as follows:

$$\gamma(X_i, X_f) = \frac{1}{n}\sum_{k=1}^{n} \gamma\left(x_i(k), x_f(k)\right) \qquad (2)$$

3.2. Algorithms

In this section, we first establish the heterogeneous social network through the user-item bipartite graph, user social network graph and user-attribute bipartite graph. Secondly, based on the user profile, we identify the implicit social relationships through grey relational analysis. Then, we incorporate these implicit social relationships into the matrix factorization model for generating recommendations. Finally, we analyze the computational complexity of the proposed method. The main symbols and notations utilized in this paper are shown in Table 1.

Table 1. Notations

Symbol	Description
G_{ML}	a heterogeneous social network of dataset 100K MovieLens
G_r	a bipartite graph of user-movie rating
G_s	a user social network graph
G_o	a bipartite graph of user-occupation
$\gamma(X_i, X_f)$	grey association degree between user i and user f
$sim(u_i, u_f)$	similarity between user i and user f
s	the threshold of similarity
$\{u_1, u_2, \cdots, u_m\}$	the set of users
m	the number of users
$\{v_1, v_2, \cdots, v_n\}$	the set of items
n	the number of items
R	the rating matrix
R_{ij}	the rating given by user i on item j
$\hat{R}_{ij}$	the predicted rating
U_i	latent feature vectors of user i
V_j	latent feature vectors of item j

3.2.1. Construction of HSN

We use the 100K MovieLens dataset (Harper & Konstan, 2015) for the experiments. Fig. 3 is an illustration of how to build the 100K MovieLens heterogeneous social network G_{ML}, with occupation as an example attribute of the user profile. Rating information and user social relationships in the 100K MovieLens dataset are represented in the form of user-movie rating bipartite graph G_r and the user social network G_s, respectively. G_o denotes the user-occupation bipartite graph. By constructing the heterogeneous social network, we can discover the rich information shared by the three networks, thereby identifying the implicit social relationships of the target user.

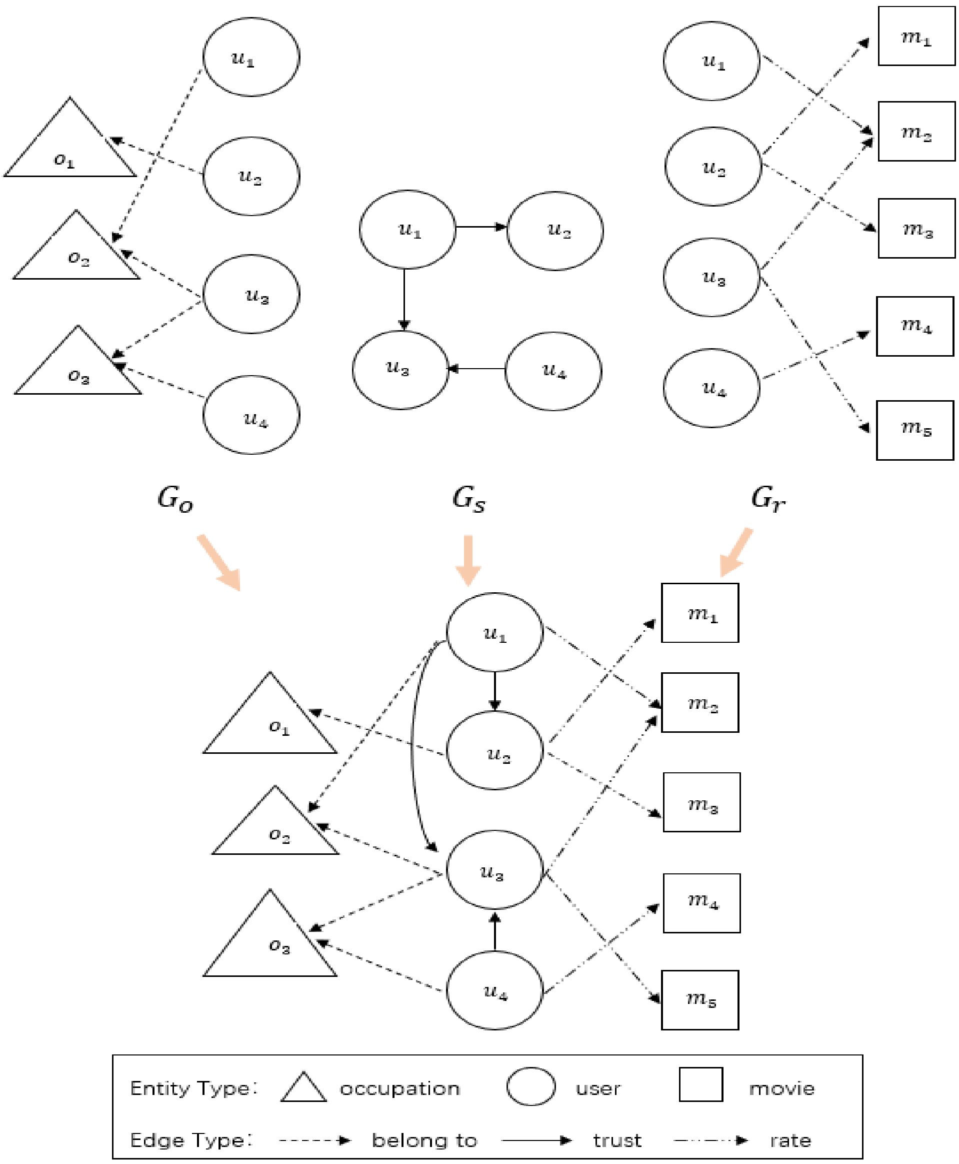

Fig. 3. The 100K MovieLens heterogeneous social network G_{ML}

3.2.2. Identification of implicit social relationships

We first quantify the user profile in the heterogeneous social network, then calculate the similarity between users by using grey relational analysis, and identify the implicit social relationships of the target user.

Quantification of the user profile. We extract four user attributes from the 100K MovieLens dataset. Since most of these four attributes are not numerical, they need to be quantified and converted into more reasonable numerical results for similarity calculation. As shown in Table 2, (1) Age is classified into 7 categories: 1-7 years old category is quantified as 1, 8-16 years old category is quantified as 2, 17-29 years old category is quantified as 3, and so on. (2) Gender: the male is quantified as 0, and the female is quantified as 1. (3) Occupation: According to the international standard occupational classification system, the occupational information is simply divided into 20 categories for quantification, with a value ranging from 1-20. (4) Zipcode: Due to the typical structural relationship between post codes or zip codes, for example, the first digit of the U.S. zip code represents a certain group of states in the United States, and the second and third digits together represent a region (or a city), and the fourth and fifth digits represent a more specific area, so the top-3 digits can be used as the quantitative result. Fig. 4 shows the comparison before and after quantification of the attributes of 5 users in the 100K MovieLens dataset.

Table 2. List and Quantification of Demographic Attributes

Attributes	Values	Quantization
Age	7,10,11,13,14…	1-7,8-16,17-29,30-39,40-49,50-59,60+
Gender	Male, Female	Male=0, Female=1
Occupation	technician, other, writer…	technician=1, other=2, writer=3…
Zip Code	1002,2215,10016…	100,221,100…

User_ID	Age	Gender	Occupation	Zip Code
1	24	M	technician	85711
2	53	F	other	94043
3	23	M	writer	32067
4	24	M	technician	43537
5	33	F	other	15213

Quantification

User_ID	Age	Gender	Occupation	Zip Code
1	3	0	1	857
2	6	1	2	940
3	3	0	3	320
4	3	0	1	435
5	4	1	2	152

Fig. 4. Comparison before and after quantification of user attributes

Similarity calculation based on the grey relational analysis. The grey relational degree calculated by the grey relational analysis is used to determine the similarity between users. The grey relational similarity consists of the average value of the discrete relational coefficient, so it can improve the distinguishability of the similarity between users to a certain extent. From the quantified user profile, the similarity of users in the heterogeneous social network can be calculated. The basic process is shown in Fig. 5. Firstly, the target user u_i is selected and his or her quantified user attributes X_i are taken as the sequence of characteristic behavior system. Then, according to Eq. (1), the grey relational coefficient of the target user sequence and other user sequences are calculated. Finally, the grey relational degree of the target user sequence and other user sequences are calculated using Eq. (2). The similarity between users $sim(u_i, u_f)$ can be obtained by the grey relational degree $\gamma(X_i, X_f)$. Set the similarity threshold s, $s \in (0,1)$, when $sim(u_i, u_f) \geq s$, it is considered that there is an implicit social relationship between user u_i and user u_f, that is, the two users have similar preferences.

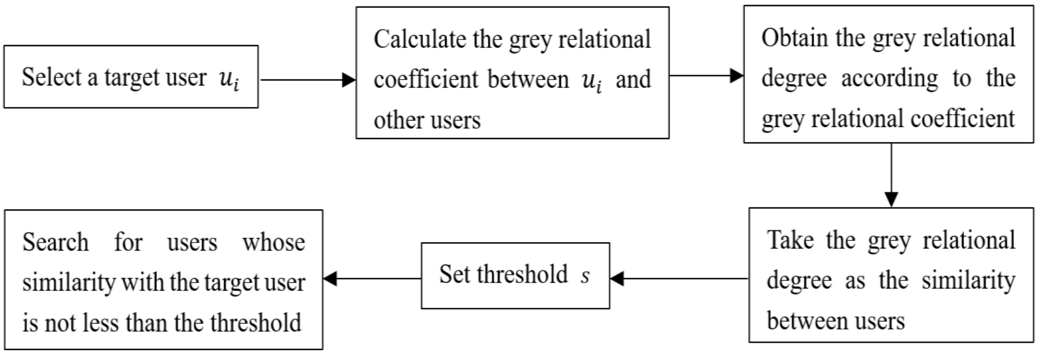

Fig. 5. Similarity calculation process based on grey relational analysis

After identifying the implicit social relationships for each user, we construct a social relation matrix, i.e., the implicit social relation matrix, IC. For any two users u_i and u_f, if there is an implicit social relationship between them, the value of IC_{if} in the matrix IC is the similarity between users $sim(u_i, u_f)$, otherwise it is 0. The construction of the implicit social relation matrix IC is as follows:

$$IC = \begin{bmatrix} IC_{11} & IC_{12} & \cdots & IC_{1n} \\ IC_{21} & IC_{22} & \cdots & IC_{2n} \\ \vdots & \vdots & & \vdots \\ IC_{n1} & IC_{n2} & \cdots & IC_{nn} \end{bmatrix}, \quad \text{where} \quad IC_{if} = \begin{cases} sim(u_i, u_f), \text{if } u_i \text{ has an implicit social relationship with } u_f \\ 0, \text{if } u_i \text{ does not have an implicit social relationship with } u_f \end{cases} \quad (3)$$

3.2.3. Matrix Factorization Model incorporating implicit social relationships

The Matrix Factorization (MF) method (Koren, Bell, & Volinsky, 2009) decomposes the rating matrix into two low-rank latent feature matrices. Given a rating matrix R, R_{ij} in the matrix R indicates that the user u_i rates the item v_j, $R \in \mathbb{R}^{m \times n}$, where m denotes the number of users and n denotes the number of items. As shown in Fig. 6, MF approximately decomposes the rating matrix into the product of the low-rank latent feature matrix of users and items, U and V are used to represent the low-rank latent feature matrix of users and items respectively, $U \in \mathbb{R}^{k \times m}$, $V \in \mathbb{R}^{k \times n}$, the dimension of the latent feature vector $k << min\ (m, n)$, then $R \approx U^T V$. When R_{ij} is missing, MF can predict the rating by the inner product of the user's latent feature vectors U_i and the item's latent feature vectors V_j, that is, $\hat{R}_{ij} = U_i^T V_j$, where $\hat{R}_{ij}$ denotes the predicted rating.

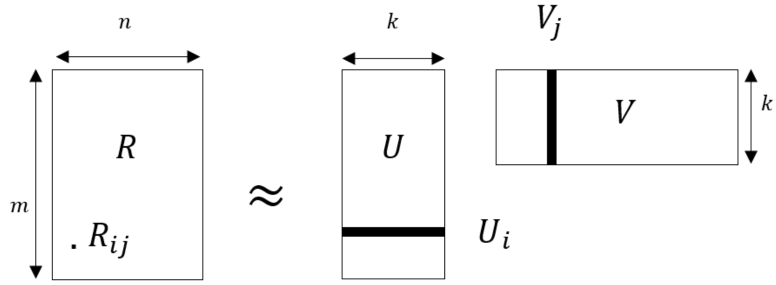

Fig. 6. Matrix Factorization (MF)

We incorporate the implicit social relationships of the target user into the MF as a social regularization term to constrain the objective function of MF, which is defined as follows:

$$\mathcal{L} = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}\left(R_{ij} - U_i^T V_j\right)^2 + \frac{\lambda_U}{2}\sum_{i=1}^{m}\|U_i\|_F^2 + \frac{\lambda_V}{2}\sum_{j=1}^{n}\left\|V_j\right\|_F^2 +$$
$$\frac{\lambda_S}{2}\sum_{i=1}^{m}\sum_{j \in F_i^+} sim\left(u_i, u_f\right)\left\|U_i - U_f\right\|_F^2 \qquad (4)$$

where I_{ij} is a binary function indicating whether user u_i has rated item v_j. If the user u_i has rated the item v_j, its value is equal to 1, otherwise, it is 0. The second and third members of the objective function are regularization terms, which are used to avoid overfitting, λ_U and λ_V control the degree of regularization. The last member of the objective function is used to minimize the preference of user u_i and his or her implicit friends. λ_S controls the degree of the social constraint, F_i^+ represents users who have an implicit social relationship with user u_i, and $sim(u_i, u_f)$ is used to differentiate all users who have an implicit social relationship with user u_i. We use stochastic gradient descent (SGD) to optimize the objective function and find the local minimum value. In each iteration, all the observed ratings are estimated by the latent feature vector, and the corresponding vectors are updated as follows:

$$\frac{\partial \mathcal{L}}{\partial U_i} = -\sum_j \left(R_{ij} - U_i^T V_j\right) V_j + \lambda_U U_i + \lambda_S \sum_{f \in F_i^+} sim\left(u_i, u_f\right)\left(U_i - U_f\right) \qquad (5)$$

$$\frac{\partial \mathcal{L}}{\partial V_j} = -\sum_i \left(R_{ij} - U_i^T V_j\right) U_i + \lambda_V V_j \qquad (6)$$

$$U_i \leftarrow U_i - \gamma \frac{\partial \mathcal{L}}{\partial U_i} \qquad (7)$$

$$V_j \leftarrow V_j - \gamma \frac{\partial \mathcal{L}}{\partial V_j} \qquad (8)$$

where γ denotes the learning rate.

The whole training algorithm is shown in Algorithm 1.

Algorithm 1: Learning algorithm for SRGRA

Input: The heterogeneous social network G_{ML}, user profile P_{u_i}, the max iteration step of SRGRA I_{max}, the threshold value s, the rating matrix R, the number of latent factor k, the learning rate γ, regularization parameters λ_U、λ_V、λ_S

Output: Latent user matrix U and latent item matrix V

Create the heterogeneous social network G_{ML}
for i = 1 to m **do**
  Quantify P_{u_i}
end for
for i = 1 to m **do**
  for f = 1 to m **do**
  Calculate $\gamma(X_i, X_f)$ according to Eq. (2)
  $sim(u_i, u_f) = \gamma(X_i, X_f)$
  end for
end for
if $sim(u_i, u_f) \geq s$ **then**
  Initialize U and V randomly
  while not convergence **do**
    for (u_i, v_j, R_{ij}) in R **do**
      Make prediction $\hat{R}_{ij} = U_i^T V_j$
      Error $e = R_{ij} - \hat{R}_{ij}$
      Calculate $\frac{\partial \mathcal{L}}{\partial U_i}$ according to Eq. (5)
      Calculate $\frac{\partial \mathcal{L}}{\partial V_j}$ according to Eq. (6)
      Update $U_i \leftarrow U_i - \gamma \frac{\partial \mathcal{L}}{\partial U_i}$
      Update $V_j \leftarrow V_j - \gamma \frac{\partial \mathcal{L}}{\partial V_j}$
    end for
  end while
end if

3.2.4. *Complexity analysis*

The main cost in learning SRGRA model is the calculation of the loss function $\mathcal{L}$ and its gradients against the feature vectors of users and items. The computational complexity to calculate the loss function is O ($k|R| + k|IC|$), where k is the dimensionality of feature space, $|R|$ and $|IC|$ refer to the number of nonzero entries of R and IC, respectively. Due to the sparsity of R and IC, the values of $|R|$ and

$|IC|$ are much smaller than the matrix cardinality. The computational complexities of calculating the gradients $\frac{\partial \mathcal{L}}{\partial U_i}$ and $\frac{\partial \mathcal{L}}{\partial V_j}$ are O $(k|R|)$ and O $(k|R|)$, respectively. Therefore, the total computational complexity in one iteration is O $(k|R| + k|IC|)$, which indicates that the overall computational complexity of our method is linear with the number of ratings and implicit social relationships. This complexity analysis shows that our proposed method is scalable for large-scale datasets.

4. Experiments and evaluations

4.1. Datasets and metrics

We selected the 100K MovieLens dataset to verify the effectiveness of our proposed method. 100K MovieLens is a movie sharing and rating website, which accepts user reviews of movies and provides a corresponding movie recommendation list. The dataset contains 943 users, 1682 movies, 100,000 ratings and the user profile, which includes information on age, gender, occupation, postcode, and others. Users rate a movie from 1~5, and each user rates at least 20 movies. The sparsity of the rating data is 93.7%.

We used the following metrics to verify the effectiveness of our proposed method: mean absolute error (MAE), root mean squared error (RMSE), Precision, Recall and F1-measure. The MAE and RMSE can be calculated by:

$$MAE = \frac{1}{N}\sum\nolimits_{(i,j)\in I}\left|\hat{R}_{ij} - R_{ij}\right| \qquad (9)$$

$$RMSE = \sqrt{\frac{1}{N}\sum\nolimits_{(i,j)\in I}\left(\hat{R}_{ij} - R_{ij}\right)^2} \qquad (10)$$

where N refers to the number of predicted ratings, I refers to the data of all test sets, $\hat{R}_{ij}$ denotes the predicted rating and R_{ij} denotes the true rating. The smaller the value of the MAE and RMSE, the better the prediction effect.

Precision, Recall and F1-measure can be calculated using the following equations:

$$\text{Precision} = \frac{\sum_{u \in U_{test}} |L_{(u)} \cap T_{(u)}|}{\sum_{u \in U_{test}} |L_{(u)}|} \quad (11)$$

$$\text{Recall} = \frac{\sum_{u \in U_{test}} |L_{(u)} \cap T_{(u)}|}{\sum_{u \in U_{test}} |T_{(u)}|} \quad (12)$$

$$\text{F1-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (13)$$

where $L_{(u)}$ denotes the set of recommended items to user u, which is defined based on the predictions of actual ratings. $T_{(u)}$ denotes the set of favorite items of user u, which is defined in terms of the ground truth R_{ij} in the test set. U_{test} is the set of users in the test set.

4.2. Algorithms compared in the experiment and parameter settings

4.2.1. Algorithms compared in the experiment

The following state-of-the-art recommendation algorithms were used as the baselines for comparison with SRGRA.

MF (Koren et al., 2009): This method uses the user-item rating matrix for the recommendation, which is the basic matrix factorization method.

$\mathrm{SR}^{\mathrm{u+-}}_{\mathrm{i+-}}$ (Ma, 2013): This method adopts the Pearson Correlation Coefficient approach to calculate all the implicit social information, including similar and dissimilar users as well as similar and dissimilar items, and then plug in the implicit social information into the matrix factorization framework.

Hell-TrustSVD (Taheri et al., 2017): This method extracts the implicit trust relationship from users' ratings to items by describing Hellinger distance between users, and incorporates the inferred trust scores into a matrix factorization method.

ITRA (Li et al., 2020): This method obtains the target user's trusted neighbors through a trust expansion algorithm, and then more available trust information hidden behind recommender systems is mined to generate recommendation results.

4.2.2. Parameter settings

For all the baseline methods, according to the parameter settings suggested in the previous works and the results of multiple parameter tuning, the optimal parameters were selected as the final results for comparison. For MF, $SR^{\mathrm{u}+-}_{\mathrm{i}+-}$, Hell-TrustSVD and SRGRA, we set the regularization coefficient λ_U=λ_V=0.1 and all the learning rate γ=0.001. For ITRA, we set the trust parameter α=0.5.

4.3. Experimental results and discussion

Five experiments were conducted for the test of the impact of similarity threshold s, global test, different rating sparsity test, the Top-N recommendation test and efficiency evaluation test, respectively. We used the 5-fold cross-validation method as an experimental method. The rating data were randomly divided into 5 parts. We chose 4 of them as the training set each time, namely 80% of the rating data, and used the remaining 20% of the rating data as the test set. The final evaluation data were obtained by taking the average of the five evaluation results.

4.3.1. *Impact of similarity threshold s*

Given that the similarity threshold plays an important role in determining how many implicit social relationships should be included in SRGRA, we first examined how the similarity threshold s influenced the accuracy of our model and determined the optimal value of s. We adjusted the value of s from 0.1 to 0.95 and fixed other parameters at the optimal values, and then evaluated the performance of SRGRA when the feature vector dimension is 10. The experimental results are shown in Fig. 7, where the horizontal axis represents the similarity threshold s, and the vertical axis represents MAE and RMSE. The results show that:

First, as the similarity threshold s gradually increases, the prediction accuracy of our SRGRA model became better. It reached the best result when the similarity threshold s was equal to 0.85. This could be because users with higher implicit correlation with the target users were included in the recommendation process, while those with low correlation and noise were removed.

Second, with further experiments, the MAE and RMSE values increased dramatically, but the prediction accuracy of the proposed method became worse. This could be because a larger s leads to a greatly reduced number of users who have implicit social relationships with the target user, which leads to the failure in utilizing the implicit social relationship information, thus reducing the accuracy of rating prediction.

As a result, the similarity threshold s should be set to 0.85 such that the accuracy of rating prediction could be at its best for the next experiments.

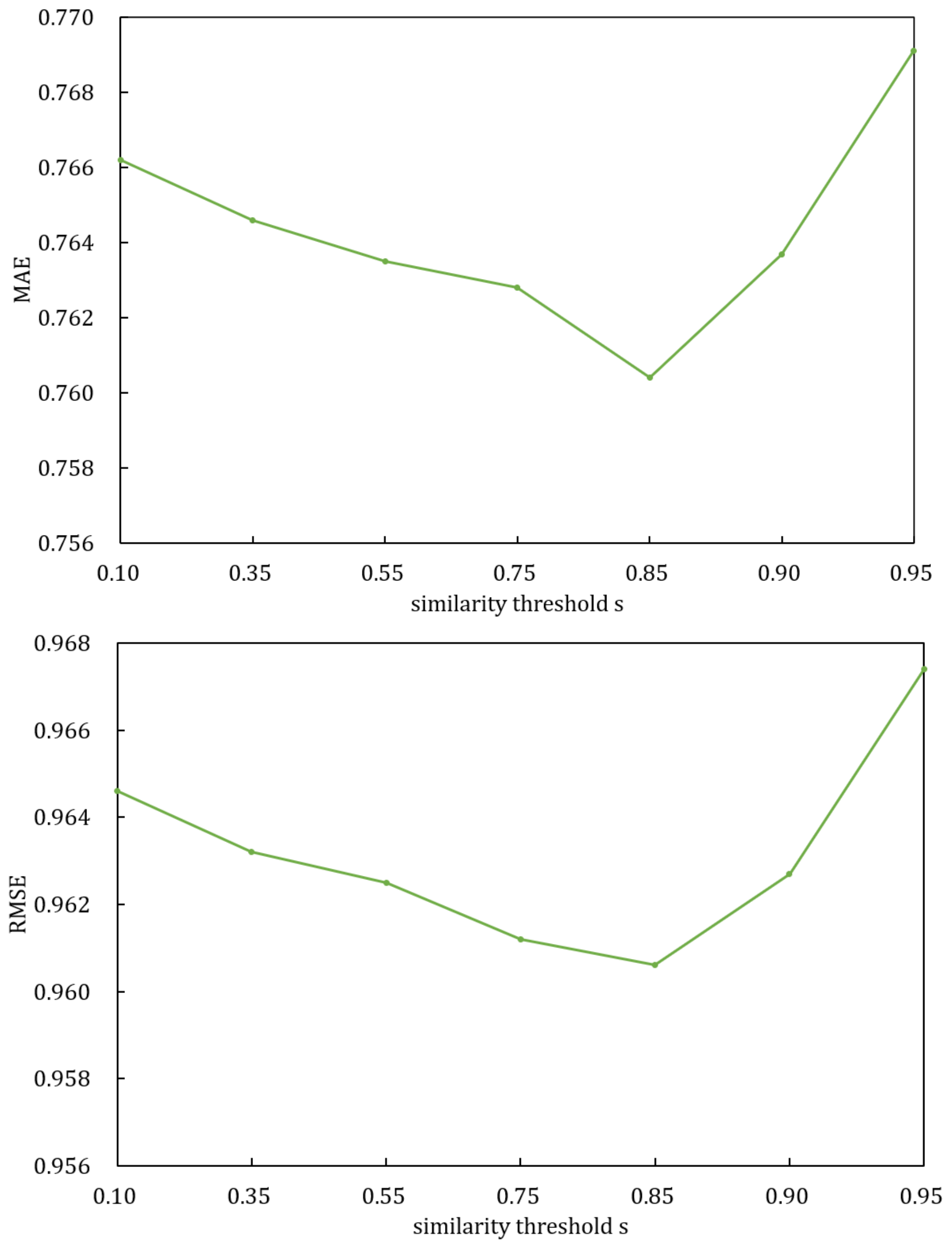

Fig. 7. Performance with different values of s

4.3.2. *Global test*

We compared the performance of SRGRA and all baseline methods on all users. We verified their performance when the feature vector dimensions were 5, 10, and 20 respectively. The experimental results as listed in Table 3 indicate that:

First, the proposed method's MAE and RMSE values were smaller than those of the baseline methods, indicating its superiority to the baseline methods. The performance improvements, calculated by comparing our method with other methods, ranged from 0.58% to 20.27%.

Second, social recommendation algorithms (such as SR^{u+-}_{i+-} and Hell-TrustSVD) were superior to MF that only rely on user ratings. This is due to the utilization of social relationship information, which can effectively improve the accuracy of rating prediction.

Third, SRGRA was better than other implicit social relationships-based recommendation methods, because the heterogeneous social network had richer information that alleviated the sparsity problem, and SRGRA accurately identified implicit social relationships through grey relational analysis for more accurate prediction.

Table 3. Accuracy comparison (s=0.85)

Dimention	Metrics	MF	SR^{u+-}_{i+-}	Hell-TrustSVD	ITRA	SRGRA
5	MAE	0.9237	0.8193	0.7972	0.7534	**0.7365**
	Improve	**20.27%**	**10.11%**	**7.61%**	**2.24%**	
	RMSE	1.1574	1.0742	0.9964	0.9457	**0.9398**
	Improve	**18.80%**	**12.51%**	**5.68%**	**0.62%**	
10	MAE	0.9243	0.8199	0.7978	0.7538	**0.7494**
	Improve	**18.92%**	**8.60%**	**6.07%**	**0.58%**	
	RMSE	1.1579	1.0748	0.9968	0.9578	**0.9491**
	Improve	**18.03%**	**11. 70%**	**4.79%**	**0.91%**	
20	MAE	0.9239	0.8195	0.7974	0.7536	**0.7406**
	Improve	**19.84%**	**9.63%**	**7.12%**	**1.73%**	
	RMSE	1.1578	1.0746	0.9963	0.9521	**0.9402**
	Improve	**18.79%**	**12.51%**	**5.63%**	**1.25%**	

4.3.3. Rating sparsity test

We divided the rating data into six groups according to the number of user ratings, and verified the performance of various algorithms in these six groups of different rating sparsity (rating number) when the feature vector dimension was 10 and the threshold s=0.85. The experimental results are shown in Fig. 8, where the horizontal axis represents the number of user ratings, and the vertical axis represents MAE and RMSE. It can be seen from Fig. 8 that:

First, the accuracy of SRGRA was always significantly better than that of other algorithms under different rating sparsity. This is because the proposed method adopted grey relational analysis to identify implicit social relationships, which can improve the rating prediction accuracy, hence good robustness.

Second, in the case of sparse ratings (the number of ratings<50), the MAE and RMSE of SRGRA were lower than other algorithms because the utilization of user profile effectively alleviated the user cold-start problem.

Third, in the case of dense ratings (the number of ratings>200), the gap of MAE and RMSE between SRGRA and other algorithms was reduced, but our method still performed better than other algorithms. This is partly due to the fact that, with the increase of rating data, SRGRA can still accurately extract implicit social relationship information in social networks, thus improving the quality of rating prediction.

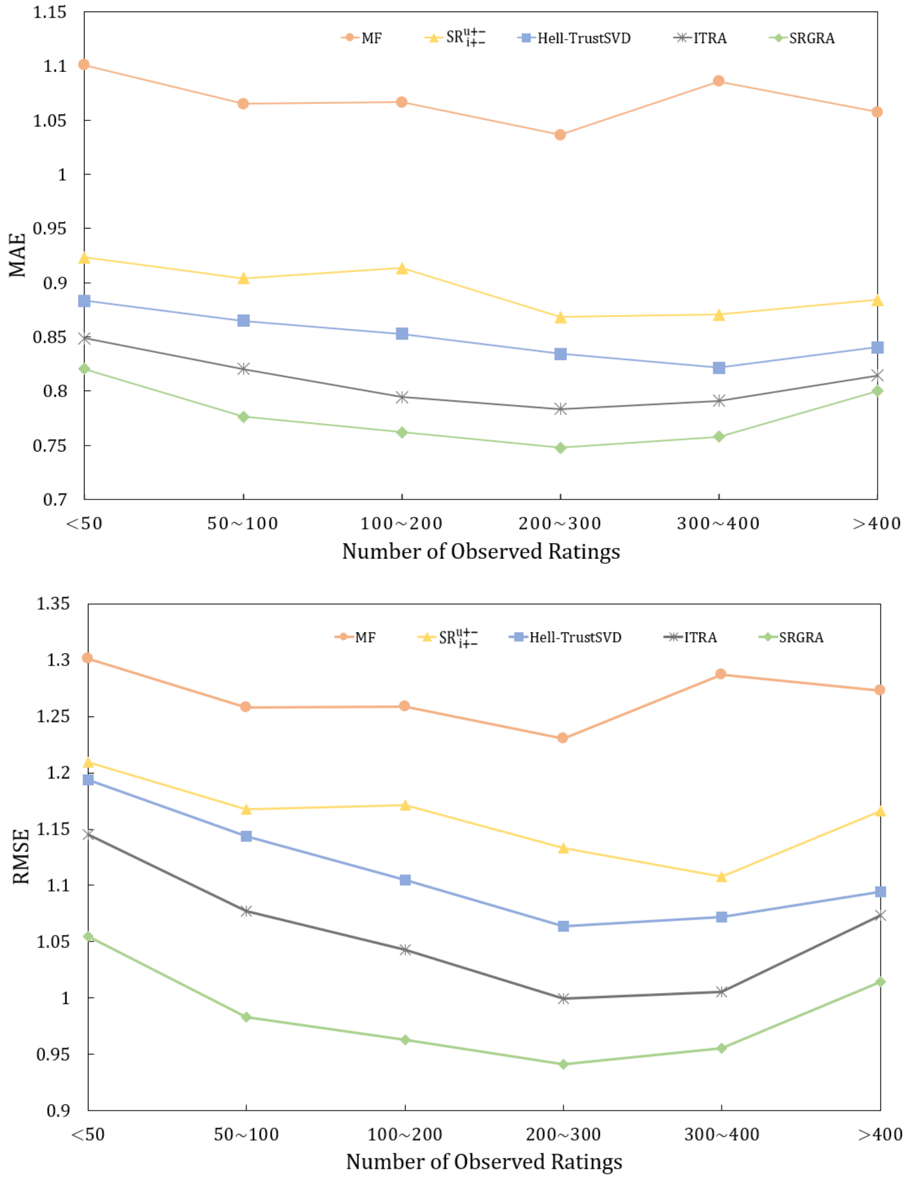

Fig. 8. The accuracy of each algorithm under different rating sparsity

4.3.4. Top-N recommendation test

We tested the performance of our method for items recommendation when the number of recommended items was 20, the dimension of the feature vector was 10 and the threshold s=0.85. The experimental results as shown in Fig. 9 indicate that:

First, the performance of all the five methods varied to different extents, but their Precision was always higher than Recall since the improvement of Precision could be at the expense of Recall, these two metrics cannot be simultaneously high.

Second, the four social recommendation algorithms were still superior to the traditional MF without social relationship information in terms of the Precision, Recall and F1-measure due to the utilization of social relationship information, which can improve the accuracy of rating prediction in the social recommendation and ranking precision.

Third, SRGRA outperformed other implicit social relationships-based recommendation methods in terms of the Precision, Recall and F1-measure because the user profile helped to improve the ranking performance.

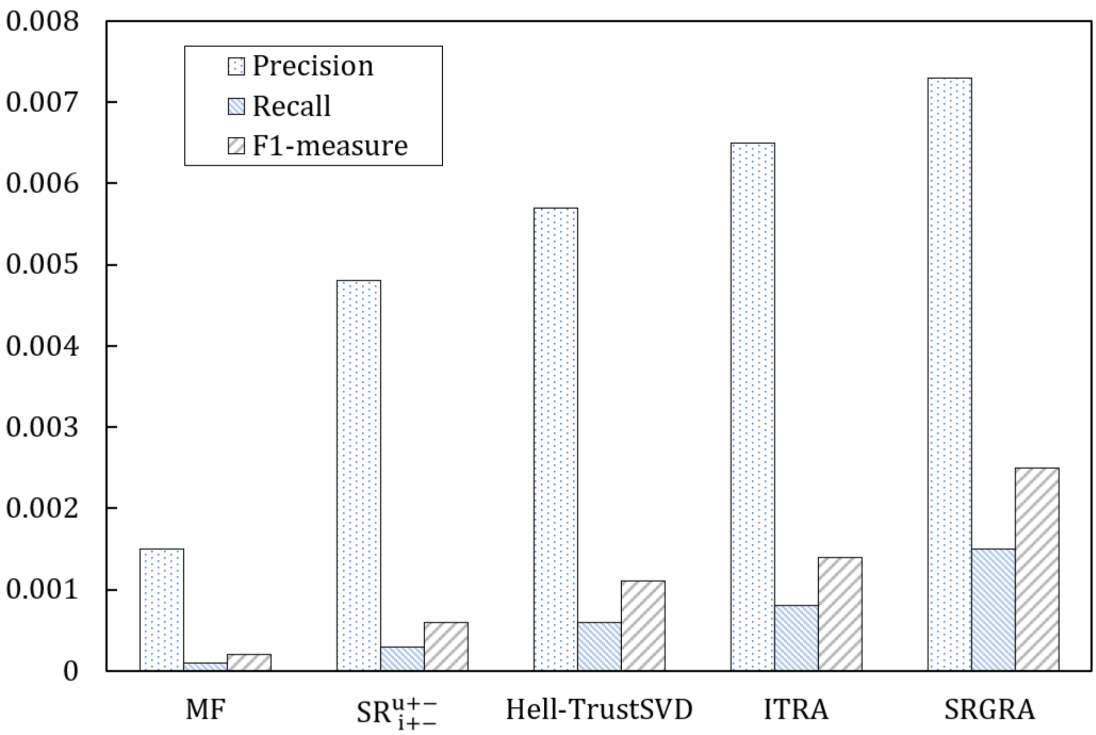

Fig. 9. Experimental results for Precision, Recall and F1-measure

4.3.5. *Efficiency evaluation test*

We compared the process-time of our proposed method with the baseline algorithms when the feature vector dimension was 10 and the threshold s=0.85. The experimental results as shown in Fig. 10 indicate that:

First, the process-time of MF was lower than that of the four social recommendation algorithms, because the four social recommendation algorithms considered the social relationship information, which increased the demand for process-time.

Second, SRGRA was faster than other social recommendation algorithms in terms of recommendation efficiency, because the proposed method can quickly identify implicit social relationships by adopting grey relational analysis.

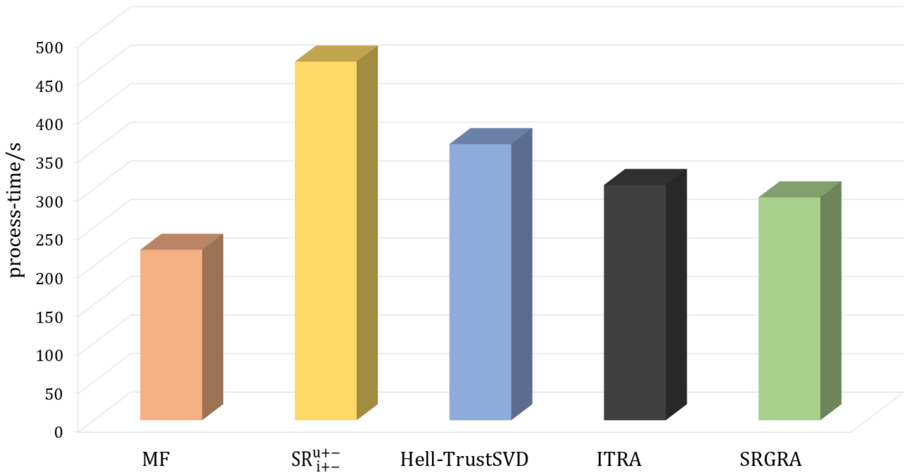

Fig. 10. The process-time of each algorithm

4.3.6. Remarks

The key insights from the above experimental results can be summarized as follows:

First, the proposed method has the highest rating prediction accuracy on the 100K

MovieLens dataset when the similarity threshold s is equal to 0.85. On one hand, if we set a small value of s, there will be more users who have implicit social relationships with the target user, so the correlation between these users and the target user is too low. On the other hand, if we set a large s, there will be too few users who have implicit social relationships with the target user. Both of these decline the accuracy of rating prediction in the social recommendation.

Second, SRGRA has better performance than the baseline algorithms in the recommendation process. This is because the proposed method is designed based on heterogeneous social networks, which can effectively alleviate the sparsity problem. In addition, based on the user profile, this paper adopts grey relational analysis to identify implicit social relationships, which can relieve the cold-start problem efficiently.

Third, the process-time of SRGRA is higher than the traditional MF due to the proposed method considered social relationship information, but it is faster than social recommendation algorithms (such as $SR^{\mathrm{u}+-}_{\mathrm{i}+-}$ and Hell-TrustSVD) in terms of recommendation efficiency, because the proposed method can quickly identify implicit social relationships by adopting grey relational analysis.

5. Conclusions and research implications

Aiming to solve the problem of recommendation performance degradation caused by sparse data in homogeneous networks, this paper proposes a social recommendation algorithm based on grey relational analysis considering the heterogeneous characteristics of traditional social networks. The experimental results show that the proposed method is superior to baseline algorithms in prediction accuracy and ranking precision, by effectively alleviating the data sparsity and cold start problems. The study offers fresh ideas of using grey relational analysis for the research on recommender systems using data from heterogeneous social networks.

5.1. *Implications for theory and practice*

Theoretically, this study provides additional insights into social recommendation techniques for the generation of accurate and efficient recommendation service for users. Specifically, the study is an advancement of existing social recommendation methods, especially the application of grey relational analysis in the recommender system. A successful application of grey relational analysis helps to overcome the algorithmic challenges of exploiting the opportunities presented in the heterogeneous social networks. The experimental results of the study show that heterogeneous social networks present rich information that can be leveraged to alleviate the sparsity problem, through the user-item bipartite graph, user social network graph and user-attribute bipartite graph. The implicit social relationships can be mined for solving the cold-start problem through the use of grey relational analysis based on the user profiles.

Practically, the method proposed in this study can help users to effectively find the target information, and also provide powerful technical support for many applications, such as online advertising and public management. High-quality recommendation algorithms such as the one we proposed in this study will play an important role in people's social life by generating accurate recommendation results. They are conducive to creating a good e-commerce environment, helping businesses to provide users with accurate services, thereby improving user loyalty. More specifically, the social recommendation algorithm provided in this study can help users find the target information quickly and prevent users from wasting too much time searching for information.

5.2. Methodological implications

This study illustrates that grey relational analysis can be used in a recommender system to calculate the similarity between users or items. The method measures the grey relational degree of sequences by calculating the similarities between the reference data sequences and multiple comparison data sequences. It has the advantage of low requirements for data size and its clustering can represent the commonness of internal items, alleviating the problem of data sparsity. A grey methodology is non-parametric and has no distribution assumptions. It has the unique capability of deriving useful information from limited, incomplete, or noisy data. Grey theory has been successfully applied in various fields to perform tasks such as data processing, systems analysis, modeling, prediction and decision making. Given its good performance, grey system theory can provide a valuable methodological approach to designing a recommender system and other expert systems.

5.3. Limitations and directions for future research

This study is limited to utilizing the data and methods for identifying implicit social relationships in heterogeneous social networks. There are several areas that have not been explored in the current study, which provides important directions for future research. First, social networks have dynamic characteristics, future work could focus on the recommendation algorithm in dynamic social networks, that is, introducing time factors on the basis of existing recommendation algorithms to describe dynamic user interests and social relationships. Second, in social networks, in addition to positive relationships between users, there are also negative relationships, such as distrust or dislike, which are valuable information that is less examined in the current study or others in the literature. Future studies may consider using negative social relationships to recommend items. Third, users usually have multiple social media accounts, which have not been considered in the social recommendation research, thus may lead to inaccurate or repetitive recommendations. Future studies may seek the fusion of cross-social media data in the social recommendation process, which is expected to generate recommendations with greater precision.

CRediT authorship contribution statement

Lijuan Weng: Writing - original draft, Formal analysis, Methodology, Software. **Qishan Zhang:** Conceptualization, Supervision, Validation, Funding acquisition. **Zhibin Lin:** Writing - review & editing, Supervision. **Ling Wu:** Conceptualization, Funding acquisition.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (Grant No. 62002063), the Natural Science Foundation of Fujian Province, China (Grant No. 2018J01791) and Social Science Planning Project of Fujian Province, China (Grant No. FJ2020C016)

References

Ahmadian, S., Joorabloo, N., Jalili, M., et al. (2020). A social recommender system based on reliable implicit relationships. *Knowledge-Based Systems, 192*, 105371.

Alabdulrahman, R., & Viktor, H. (2021). Catering for unique tastes: Targeting grey-sheep users recommender systems through one-class machine learning. *Expert Systems with Applications, 166*, 114061.

Al-Sabaawi, A. M A., Karacan, H., & Yenice, Y. E. (2020). Exploiting implicit social relationships via dimension reduction to improve recommendation system performance. *PloS ONE, 15(4)*, e0231457.

Altan, A., & Karasu, S. (2019). The effect of kernel values in support vector machine to forecasting performance of financial time series and cognitive decision making. *The Journal of Cognitive Systems, 1(4)*, 17-21.

Altan, A., & Karasu, S. (2020). Recognition of COVID-19 disease from X-ray images by hybrid model consisting of 2D curvelet transform, chaotic salp swarm algorithm and deep learning technique. *Chaos, Solitons and Fractals, 140*, 110071.

Altan, A., Karasu, S., & Bekiros, S. (2019). Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques. *Chaos, Solitons and Fractals, 126*, 325-336.

Bertani, R. M., Bianchi, R. A. C., & Costa, A. H. R. (2020). Combining novelty and popularity on personalised recommendations via user profile learning. *Expert Systems with Applications, 146*, 113149.

Camacho, L. A. G., & Alves-Souza, S. N. (2018). Social network data to alleviate cold-start in recommender system: A systematic review. *Information Processing & Management, 54*(4), 529-544.

Deng, J. (1989). Introduction to grey theory system. *The Journal of Grey System, 1*(1), 1-24.

Deng, S., Huang, L., & Xu, G. (2014). Social network-based service recommendation with trust enhancement. *Expert Systems with Applications, 41*(18), 8075-8084.

Gogna, A., & Majumdar, A. (2015). A Comprehensive Recommender System Model: Improving Accuracy for Both Warm and Cold Start Users. *IEEE Access, 3*, 2803-2813.

Guo, T., Luo, J., Dong, K., et al. (2018). Differentially private graph-link analysis based social recommendation. *Information Sciences, 463-464*, 214-226.

Harper, F. M., & Konstan, J. A. (2015). The MovieLens Datasets: History and Context. *ACM Transactions on Interactive Intelligent Systems 5*(4), Article 19.

Jakomin, M., Bosnić, Z., & Curk, T. (2020). Simultaneous incremental matrix factorization for streaming recommender systems. *Expert Systems with Applications, 160*, 113685.

Jiang, M., Cui, P., Chen, X., et al. (2015). Social Recommendation with Cross-Domain Transferable Knowledge. *IEEE Transactions on Knowledge and Data Engineering, 27*(11), 3084-3097.

Karasu, S., Altan, A., Bekiros, S., et al. (2020). A new forecasting model with wrapper-based feature selection approach using multi-objective optimization technique for chaotic crude oil time series. *Energy, 212*, 118750.

Kayacan, E., Ulutas, B., & Kaynak, O. (2010). Grey system theory-based models in time series prediction. *Expert Systems with Applications, 37*(2), 1784-1789.

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix Factorization Techniques for Recommender Systems. *Computer, 42*(8), 30-37.

Lai, C.-H., Lee, S.-J., & Huang, H.-L. (2019). A social recommendation method based on the integration of social relationship and product popularity. *International Journal of Human-Computer Studies, 121*, 42-57.

Li, Y., Liu, J., Ren, J., et al. (2020). A Novel Implicit Trust Recommendation Approach for Rating Prediction. *IEEE Access, 8*, 98305-98315.

Li, Z., Xiong, F., Wang, X., et al. (2019). Topological Influence-Aware Recommendation on Social Networks. *Complexity, 2019*(3), 1-12.

Li, Z., Xiong, F., Wang, X., et al. (2020). Mining Heterogeneous Influence and Indirect Trust for Recommendation. *IEEE Access, 8*, 21282-21290.

Liu, S., & Lin, Y. (2010). *Grey systems: theory and applications*: Springer Science & Business Media.

Ma, H. (2013). *An experimental study on implicit social recommendation*. Paper presented at the Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval. Retrieved from https://doi.org/10.1145/2484028.2484059

Mao, M., Lu, J., Zhang, G., et al. (2017). Multirelational Social Recommendations via Multigraph Ranking. *IEEE Transactions on Cybernetics, 47*(12), 4049-4061.

Mazhari, S., Fakhrahmad, S. M., & Sadeghbeygi, H. (2015). A user-profile-based friendship recommendation solution in social networks. *Journal of Information Science, 41*(3), 284-295.

Ngaffo, A. N., Ayeb, W. E., Choukair, Z. (2020). A time-aware service recommendation based on implicit trust relationships and enhanced user similarities. *Journal of Ambient Intelligence and Humanized Computing, (11)*, 1-19.

Pham, T. N., Li, X., Cong, G., et al. (2016). A General Recommendation Model for Heterogeneous Networks. *IEEE Transactions on Knowledge and Data Engineering, 28*(12), 3140-3153.

Qian, F., Zhao, S., Tang, J., et al. (2016). SoRS: Social recommendation using global rating reputation and local rating similarity. *Physica A: Statistical Mechanics and its Applications, 461*, 61-72.

Reafee, W., Salim, N., Khan, A. (2016). The Power of Implicit Social Relation in Rating Prediction of Social Recommender Systems. *PLoS ONE, 11(5)*, e0154848.

Škrinjarić, T. (2020). Dynamic portfolio optimization based on grey relational analysis approach. *Expert Systems with Applications, 147*, 113207.

Sun, G., Guan, X., Yi, X., et al. (2018). Grey relational analysis between hesitant fuzzy sets with applications to pattern recognition. *Expert Systems with Applications, 92*, 521-532.

Taheri, S. M., Mahyar, H., Firouzi, M., et al. (2017). *Extracting implicit social relation* *for social recommendation techniques in user rating prediction*. Paper presented at the Proceedings of the 26th International Conference on World Wide Web Companion. Retrieved from https://doi.org/10.1145/3041021.3051153

Tang, J., Hu, X., & Liu, H. (2013). Social recommendation: a review. *Social Network Analysis and Mining, 3*(4), 1113-1133.

Tao, W., & Dang, Y. (2018). Collaborative filtering recommendation algorithm based on grey incidence clustering. *Operations Research and Management Science, 27*(1), 84-88.

Walek, B., & Fojtik, V. (2020). A hybrid recommender system for recommending relevant movies using an expert system. *Expert Systems with Applications, 158*, 113452.

Wang, C., Song, Y., Li, H., et al. (2015). *KnowSim: A Document Similarity Measure on Structured Heterogeneous Information Networks.* Paper presented at the 2015 IEEE International Conference on Data Mining.

Wang, Y., Xia, Y., Tang, S., et al. (2017). Flickr group recommendation with auxiliary information in heterogeneous information networks. *Multimedia Systems, 23(6)*, 703-712.

Wu, H., Yue, K., Pei, Y., et al. (2016). Collaborative Topic Regression with social trust ensemble for recommendation in social media systems. *Knowledge-Based Systems, 97*, 111-122.

Wu, L., & Zhang, Q. (2015). *Identification of overlapping community structure with Grey Relational Analysis in social networks.* Paper presented at the 2015 IEEE International Conference on Grey Systems and Intelligent Services (GSIS).

Yu, J., Gao, M., Li, J., et al. (2018). *Adaptive Implicit Friends Identification over Heterogeneous Network for Social Recommendation*. Paper presented at the Proceedings of the 27th ACM International Conference on Information and Knowledge Management. Retrieved from https://doi.org/10.1145/3269206.3271725

Zhang, T.-w., Li, W.-p., Wang, L., et al. (2020). Social recommendation algorithm based on stochastic gradient matrix decomposition in social network. *Journal of Ambient Intelligence and Humanized Computing, 11*(2), 601-608.

Zhang, Z., Sun, R., Choo, K. R., et al. (2019). A Novel Social Situation Analytics-Based Recommendation Algorithm for Multimedia Social Networks. *IEEE Access, 7*, 117749-117760.

Zhao, W. X., Li, S., He, Y., et al. (2016). Exploring demographic information in social media for product recommendation. *Knowledge and Information Systems, 49*(1), 61-89.