No.10490

Abstract

With the development of the internet and information technology, assessing system based on social network becomes more and more important. This paper is committed to make prediction of books for user and intelligent recommendation through a complex network algorithms and factor analysis.

To begin with, we have established a directed layered complex network of users and books. The upper layer of the network describes the friend-relationship between the users, weighted by the numbers of users' public reading records. The second layer of user-book cross network is weighted by the users' evaluation. Then the Floyd algorithm is adopted to calculate the fitness index from the nodes of users to the nodes of books based on the complex network. In addition, other factors such as the influence of the node and relevant intensity are defined in order to measure the contact between users and books more accurately.

For question one and two, the factors which will affect the evaluation to the books are divided into users' evaluation feature, books' being evaluation feature and user-book link feature. The first two factors contain the average of the evaluation and the evaluation offset separately. The third factor includes nine indicators which are the influence of evaluation, relevant intensity and fitness degree etc. Utilizing SPSS, we implement factor analysis to extract and combine four common factors with the weight of variance contribution rate. Recommendations will be generated systematically in accordance with the maximum and minimum of books' evaluation.

For question three, the intelligent-recommended algorithm is designed for the application of optimal book recommendation. We are able to set up comprehensive indicators based on the four common factors in views of complex network algorithm and factor analysis, which will generate evaluation matrix of books. According to evaluation matrix and reading history, the optimal selections of the six users are 698573,698573,794171,702699,698573,776002 respectively.

Finally, we put forward certain development of the algorithm, including dimension-reduction clustering process of the complex network algorithm, standardization of users' evaluation and cold-starting recommendation, when assessing the advantages and disadvantages of the model.

Key words: Complex network algorithm The shortest path algorithm Fitness index Factor analysis Intelligent recommendation

1. Restatement of the problem

With the development of information technology and Internet, our lives have been filled with a large amount of information. Now, human are in the era of information overload rather than lack of information. In the process of dealing with information, information consumers and information producers are faced with challenges: the former find it more difficult to select their interested information from all kinds of information. Similarly, the core problem which information producers need to solve is how to make the information produced by them be recommended effectively and efficiently in order to meet the recognition of customers and maximize the benefit.

People began to research for models to solve this contradiction where comes from the recommendation. Gradually, it is widely used in products and applications of the Internet, including the searching, topic recommendation, a variety of products recommendation of electronic commerce, social network recommendation.

In this case scenario, we have obtained a large number of users and user behavior information from a famous online bookstore, also including the score data of books, label information of books and user social relationships. We are required to solve the following problems:

(1) analysis the factors which may influence the users when giving a score of the books one reads.

(2) establishing a model and making prediction of book scores which the users have not read it before in the predict.txt;

(3). give a recommendation of books to the users from the predict.txt

2. Assumption of model

(1) The evaluation of the books is objective.

(2) The social network is based on the users' similarity preference, that is to say the social network can accurately describe the preference feature of the users.

(3) The preference of users to the books is homogeneous.

3. Introduction of symbol

symbol	definition

e_i	The number of evaluation
N_{ij}^{b}	Directed-weight in the network
d_{ij}	Intensity of the node i and j
I^e	Users evaluation influence
$Path_{ij}$	Sequence point from i to j
X_1	Means of user evaluation
X_2	Evaluation offset for users
X_3	Means of book evaluation
X_4	Evaluation offset for books
X_6	Influence from user i to user j
X_7	Intensity of the user i
X_{10}	Revised means of users evaluation
F	Comprehensive indicator matrix
$Score(i,j)'$	Revised scores of evaluation

4. Problem analysis

It is a popular spot of current research to data mining based on the social network. Building up a social network forms a giant multilevel data network. And the significant meaning to research is how to effectively data mining and to maximize the benefit. To solve this problem, based on directed-weighted complex network model, we try to seek for the potential links between the data when analyzing the systematic evaluation problem solving of customer relationship and books information, and represent all comprehensive known factors in complex network, to realize the books scoring system more comprehensive, objective and intelligent.

In the prediction of books scoring system, we are supposed to consider all factors that might affect the book scores of users, which include score structure of individual user, score structure of the books available, matching degree of users and books. In this paper, authors carried on the analysis of the three factors above. When being in the analysis process, score structure of users and score structure of books are likely easier to measure, and it is more probably difficult to measure the matching degree of the user with the books which is the core of the system evaluation. The matching degree of users and books takes account of the distance between the user and books in the network, the index of influence of the user and the influence of the book itself.

In conclusion, the general idea of solving the problem is divided into three parts, which are constructions of social networks, index analysis and scoring forecast and intelligent recommendation.

4.1 Construction of complex networks

Construction of complex network system is the foundation of social score prediction model. Network architecture of the traditional directed weighted graph used in this paper has been implemented the double network layers. The nodes on the network layer consist of users and books. Besides them, the one-way relationship between users and books was involved in the network architecture as well. All of them formed the double directed weighted graph, and we calculated in accordance with the improved algorithm of complex networks.

4.2 Directed-weighted network of users

The network layer of readers is constructed in accordance with the records between the users and the weight is based on the number of same reading history between users, which means the preference similarity between two readers. Network diagram is mainly reflected the mutual influence between the readers through the attention generated by this behavior in the reading process. In practice, the more the same hobbies among readers, the greater the impact they have; the more the same recorded reading history in the reader’s network layer, the greater the reader node weights, the smaller the distance between nodes.

4.3 Cross Network between users and books

The information given in the scenario involved the users’ reading history and evaluation of records, according to which the directed-weighted network can be established between users and books. However, through a simple analysis of the data given in question, we can find that the number of the evaluation records of the entire system is much greater than the total number of users’ reading records(ie, some readers have not read the books but gave a score to the book).Therefore these data are not allowed to directly evaluate the applications and system must be installed to eliminate the false evaluation and make use of the remaining valid data for cross-network evaluation reader of books. All information about books is located in a new network layer and do not set up directed links.

The cross network of users and books is set up based on the users’ evaluation of books. Users and books are linked by the specific scores, forming a cross double network. The influence of books is definite represented by the number of evaluation records, defining the junction strength of the book.

4.4 Floyd shortest path algorithm

Based on the reader hierarchical network above, using data mining methods to analyze the directed weighted network. Floyd algorithm is required to calculate the shortest weighted distance of the users' node in the first layer and the books' node in the second layer in order to show book reader preferences.

4.5 Factor Analysis and Prediction

In this paper, a complex network model is used for solving the optimal fitness of books preferences for different audiences. However, through behavioral analysis, the users' evaluation will also depend on the habits of users, the content of books, other indicators of book influence. Therefore, we need to take all the possible impact of various factors into account. Due to possible collinear between factors and overlapping relationships, it is difficult to quantify a comprehensive evaluation index. Therefore we use factor analysis to extract common factors to determine weight system to predict the readers of books and evaluation.

After the previous algorithm and prediction and evaluation, problem three is probably solved by ranking the evaluation results before recommendation.

5 Establishment of model

5.1 Model I: complex network model

Books and readers' directed-weighted network is divided by two layers according to types of node-based social connection; the network can be set up via the shortest path algorithm to solve the optimal fitness to book preference. And fitness index is the most important factor in predicting the evaluation, while another fundamental factor the strength index of each node will be defined by social networks.

5.1.1 User-layer network

The weighted adjacency matrix of users' network can be explained by U^R

$$U^R = \begin{pmatrix} U_{11}^R & \cdots & U_{1j}^R \\ \vdots & \ddots & \vdots \\ U_{i1}^R & \cdots & U_{ij}^R \end{pmatrix}$$

Where U_{ij}^R indicates the strength of adjacency relationships between the ith user and jth user; $U_{ij}^R = \infty$ means the strongest intensity relationship between the node; $U_{ij}^R = 0$ represents that there is no links between the target nodes.

In the users network layer, the intensity of the users node is required to represent the dynamic degree of the users and influence of the users. This are separated into evaluation intensity and reading intensity, expressed by $e_i\ and\ v_i$.

Definition: matrix $\mathrm{E} = (e_1\ e_2 \cdots e_n)^T$ and matrix $\mathrm{V} = (v_1\ v_2 \cdots v_n)^T$

Where e_i is the number of the evaluation of the user. And v_i is the number of users reading records. e_i and v_i can explained the influence and dynamic degree of the users to some extent.

5.1.2 The cross network layer of users and books

Similarly, a cross-linking network layer between users and books was established based on the links between users and books.

The adjacency matrix is explained by U^B

$$U^B = \alpha \begin{pmatrix} U_{11}^b & \cdots & U_{1\,j}^b \\ \vdots & \ddots & \vdots \\ U_{i\,1}^b & \cdots & U_{i\,j}^b \end{pmatrix}$$

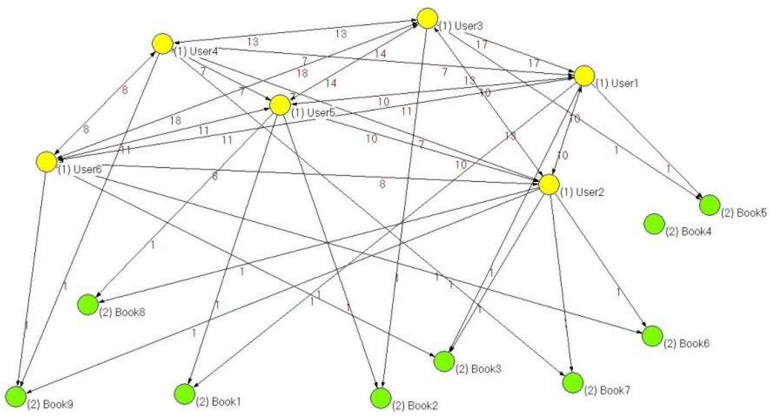
Where $U_{i\,j}^b$ is the adjacency weight between the user and the books.

$\mathrm{N}_{\mathrm{I}\,\mathrm{j}}^{\mathrm{b}} = \mathrm{b}_{\mathrm{I}\,\mathrm{j}} \cdot \mathrm{s}_{\mathrm{ij}}$ is defined, where $\mathrm{N}_{\mathrm{i}\,\mathrm{j}}^{\mathrm{b}}$ makes up the directed weight of users-books network, which represents ith user's preference to the jth book.

$$\mathrm{b}_{\mathrm{i}\,\mathrm{j}} = \begin{cases} 0 & \qquad\qquad not\ be\ read \\ 1 & \qquad \text{ith user read jth book} \end{cases}$$

$\mathrm{s}_{\mathrm{i}\,\mathrm{j}}$ is expressed the score of jth books which ith user evaluates. If there is no evaluation, $\mathrm{s}_{\mathrm{i}\,\mathrm{j}}$ is acquiesced to the average of the books score. The node intensity of the book is explained by evaluation intensity e_i' and reading intensity v_i' separately.

Where e_i' is the revised number of the evaluation to the book. v_i' is number of the book's label. The revised number of e_i' is required because of the false evaluation given in the data. Evaluation intensity e_i' and reading intensity v_i' can reflect the popularity of the ith book.



Graph1-cross network

5.1.3 The shortest path algorithm (Floyd)

(1) Hybrid network

User network and user-books network can be combined into a layered directed-weighted network. With the equivalent of the nodes of books and user, a new network topology architecture adjacency matrix can be defined as

$$\mathrm{U} = \begin{bmatrix} U^R & U^B \\ 0 & 0 \end{bmatrix}$$

Where $\mathrm{U}_{\mathrm{i\,j}}$ indicates the weight from ith node to jth node. U^R is the node weight between users and U^B is the node weight between users and books.

The distance between the random nodes in the network is expressed by $\mathrm{C}_{\mathrm{i\,j}}$:

$$\mathrm{C}_{\mathrm{i\,j}} = \begin{cases} 0 & \mathrm{i} = \mathrm{j} \\ \dfrac{1}{\mathrm{U}_{\mathrm{i\,j}}} & \mathrm{U}_{\mathrm{i\,j}} \neq 0 \text{ and } \mathrm{i} \neq \mathrm{j} \\ \infty & \mathrm{U}_{\mathrm{i\,j}} \neq 0 \end{cases}$$

So, the minimum distance of the nodes is $d_{ij} = \min\{\mathrm{C}_{ik_1} + \mathrm{C}_{k_1k_2} + \cdots + \mathrm{C}_{k_nj}\}$

d_{ij} can accurately measure the tightness between nodes i and j. If I represents the number of users, j means the number of books, then the d_{ij} (i∈users, j∈books) indicates the books' adaptation of user's preference.

(2) Floyd algorithm

Floyd algorithm is designed to solve the shortest path problem between nodes in the directed network. The main idea is to insert fixed point in the weighted adjacency matrix, using a recursive method to construct a sequence of n matrices D (1), D (2) ... D (n), D (n) which are the distance matrix of the network model. While introducing another point shows the shortest distance between two points on the matrix records.

There are two important attributes matrix D and Path in the Floyd algorithm.

Path ij is the subsequent-linked points from the path i to j, the algorithm is as follows:

Step1: Enter the weighted adjacency matrix C

Step2: initial value, $d_{ij} = c_{ij}, Path_{ij} = j, \mathrm{k} = 1.$

Step3: update D and Path

A group of i and j, if the D (i, k) + D (k, j) <D (i, j)

Then D (i, j) = D (i, k) + D (k, j)

Path (i, j) = Path (i, k)

Step4: K plus one, then return to Step3 until k = n +1 (n is total number of the nodes.).

According to the steps above, D and Path can be accurately solved.

5.1.4 Influence and intensity assessment of the nodes

In the hybrid network, the influence of each node can be described by the intensity and topology structure of the node in the network. For these complex networks, each node has two intensity e_i and v_i, so the influence of the nodes also have two types, which I^e is for the evaluation of influence, I^v is for the read influence.

$$\mathrm{I}_{\mathrm{j}}^{\mathrm{e}} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \frac{\mathrm{e}_{\mathrm{i}}}{\mathrm{d}_{\mathrm{ij}}} \quad (\mathrm{i} \neq \mathrm{j})$$

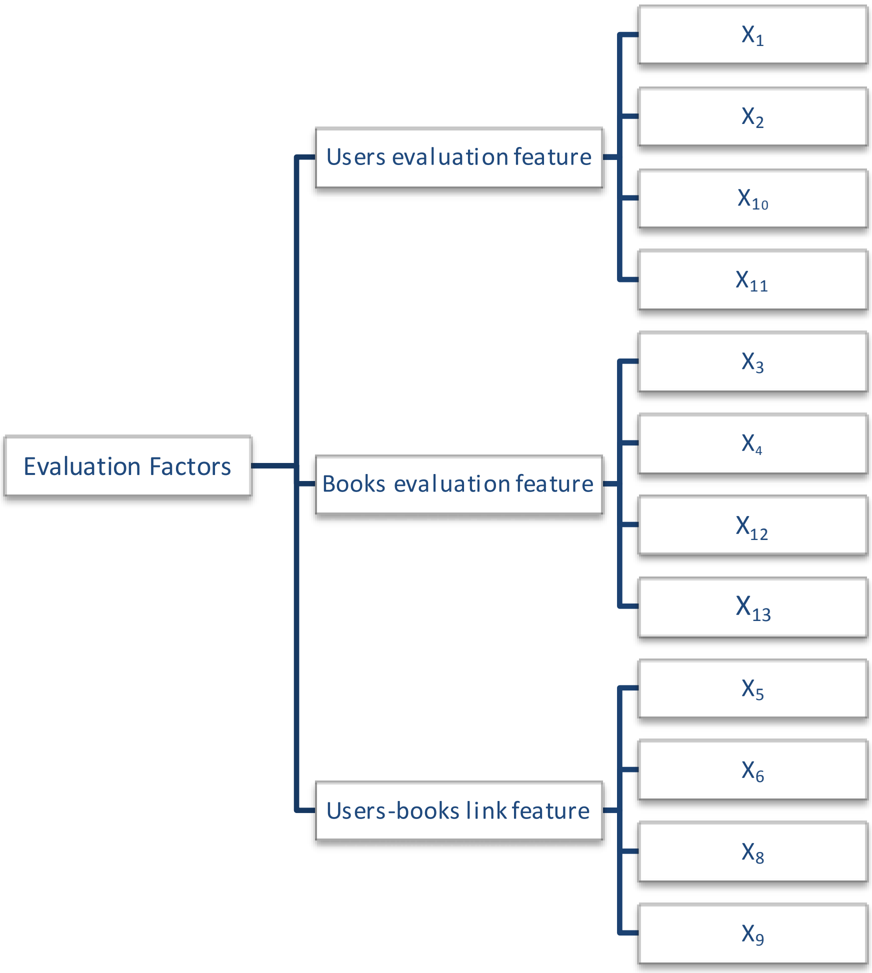
$$I_j^v = \sum_{i=1}^{n} \frac{p_i}{d_{ij}} \quad (\mathrm{i} \neq \mathrm{j})$$

The influence of the nodes considers the sum of the effects by another node passing through to the target node, while the intensity of the node means the external influence from the target node. The intensity of the user is explained by the effective evaluation number of books from the users. Similarly the intensity of the books is showed by the effective evaluation number of the book. The influence and the intensity of the node can be both applied to user node and book node.

5.2 Factor Analysis

5.2.1. Factor extraction

User Evaluation of books mainly affected by various factors, which is shown as below:



Graph-2 factors

Wherein, X (i, j) refers to the evaluation of user i in the j-books.

$X_1(\mathrm{i,j})$ is the average of all the ith user's evaluation.

$X_2(\mathrm{i,j})$ is the offset of all the ith user's evaluation.

$\mathrm{X}_3(i,j)$ is the average of all the evaluations of the jth book.

$X_4(i,j)$ is the offset of all the evaluations of the jth book.

$$X_2(\mathrm{i,j}) = \frac{\sum_{k=1}^{|T|} |R_{ik} - X_1(i,j)|}{|T|_i}$$

$$X_4(\mathrm{i,j}) = \frac{\sum_{k=1}^{|U|} |R_{kj} - X_3(i,j)|}{|U|_i}$$

Where R_{ik} represents the score of the kth book from the ith user (from evaluation history), and | T | and | U | are shown the total evaluation number of books and the users separately.

X_5(i,j)= d_i

$$X_6(i,j)=\frac{I_i^e-min(I^e)}{\max(I^e)-min(I^e)}+\frac{I_i^f-min(I^f)}{\max(I^f)-min(I^f)}$$

$X_6(i,j)$ is the influence from ith user to the jth user, consisting of the sum of the relative reading influence and relative evaluation influence.

$$X_7(i,j)=\frac{e_i-min(E)}{\max(E)-min(E)}+\frac{f_i-minF}{\max(F)-min(F)}$$

$X_7(i,j)$ is the intensity of the ith user, consisting of relevant reading intensity and relevant evaluation intensity.

$X_8(i,j) and X_9(i,j)$ are the influence and intensity of the books, which is defined as the way of $X_6(i,j) and X_7(i,j)$.

It is the nine most important factors affecting the results of the evaluation, the following will be the factor analysis.

5.2.2 The primary theory of factor analysis

Factor analysis is a comprehensive multivariate statistical analysis methods from the internal related dependencies about the variables, extracting to a few factors according the original complex relationship between variables. Using this method we can reclassify the raw data analytically extract by generalizing a number of integrated indicators which appeared close links. What's more, the information contained from these integrated indicators cannot overlap. These integrated indicators are be defined as the common factor.

The basic idea of factor analysis is to classify the variables in accordance with the high links between the variables, which also means that different types of the variables have lower links. In this way, each of the variables actually represents a basic structure, namely common factor. What we intend to study is to use the intended function and specific factors of minimum number of unpredictable factors to describe the each component of the original observations. This can relatively easily reflect most of the information of the original data with less number of factors, so as to extract the data.

The core of factor analysis was to implement factor analysis on a number of integrated indicators and extract common factors, then contribute the score function by summing up all the factors multiple with the variance contribution rate each factor as weights. Factor analysis represented as a matrix : X = AF + B,

$$\begin{cases} x_1 = a_{11}f_1 + a_{12}f_2 + a_{13}f_3 + \cdots + a_{1k}f_k + \beta_1 \\ x_2 = a_{21}f_1 + a_{22}f_2 + a_{23}f_3 + \cdots + a_{2k}f_k + \beta_2 \\ x_3 = a_{31}f_1 + a_{32}f_2 + a_{33}f_3 + \cdots + a_{3k}f_k + \beta_3 \\ \cdots \\ x_p = a_{p1}f_1 + a_{p2}f_2 + a_{p3}f_3 + \cdots + a_{pk}f_k + \beta_p \end{cases}$$

In model above, the vector $X = (x_1, x_2, x_3, \cdots, x_p)$ is observable random vector, ie, the original observed variables. Factor $F = (f_1, f_2, f_3, \cdots, f_k)$ is the common factors of $X = (x_1, x_2, x_3, \cdots, x_p)$ which are mutually independent theoretical unobservable variables. Specific meaning of common factors must be defined with the actual research questions. $A(\alpha_{ij})$ is the coefficients of the common factor $F = (f_1, f_2, f_3, \cdots, f_k)$, called factor loading matrix representing the weight from the ith variable to the jth common factors. α_{ij} is both the covariance and correlation coefficient of $x_i and f_j$. The larger the absolute α_{ij} indicates the greater loading capacity from common factor f_j to factor x_i. $B = (\beta_1, \beta_2, \beta_3, \cdots, \beta_p)$is the special factor of $X = (x_1, x_2, x_3, \cdots, x_p)$ that cannot be included by the pre-kth common factor. All the special factors and all the common factors are independent of each other.

5.2.3Mathematical meaning of the model

A factor loading matrix contains two statistics, which are the common degree of the variable and the variance contribution of the common factor.

(1) The common degree of the variable

The common degree of the variable is the square sum of i-th row element of matrix A.

$$h_i^2 = \sum_{j=1}^{k} \alpha_{ij}^2 \quad (i = 1,2,3 \ldots p)$$

(2) The degree of variance contribution

The degree of variance contribution is the square sum of j-th columns element of matrix A.

$$g_j^2 = \sum_{i=1}^{p} \alpha_{ij}^2 \ (j = 1,2,\ldots,k)$$

(3) Comprehensive index

We use the proportion of each factor’s contribution rate and the total contribution rate as weights. F is defined as follows:

$$F = [w_1, w_2, \cdots w_n] \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix} = \sum_{i=1}^{n} w_1 f_1$$

F is a comprehensive index, which is a reference index prediction and evaluation system.

$$f'(i,j) = \frac{f(i,j) - \min(f)}{\max(f) - \min(f)}$$

$f'(i,j)$is the revised index of F, which means the relative position of all the value of F.

Similarly, revised value scores is expressed as follows:

$$Score(i,j) = \min(Score_j) + f'(i,j)[\max(Score) - \min(Score)]$$

We use the modified comprehensive indicator scores as weights , meaning the value of a relative position of a specific fraction of the overall score, ranking between the maximum and minimum values the of all scores.

6. Model solution

6.1 Problem One: Evaluating Factors

With the establishment of a complex network models and factor analysis model, there are three types of main factors when users make a score to the book, which are evaluation feature of users, books being evaluation feature and the users-books link feature. Their detailed indicators are as follows:

6.1.1 Evaluation feature of users

(1) Average of evaluation: Average value of all the historical records generated by the target user.

(2) Evaluation Offset: Average value of the absolute deviation of the historical records generated by the target user.

(3) Revised average of evaluation: the ratio of the average of evaluation and the shortest distance between any two nodes in the complex networks.

(4)Revised evaluation offset: the ratio of the evaluation offset and the shortest distance between any two nodes in the complex networks.

6.1.2 The user-books' link feature

(1) User-books fitness degree: The degree of tightness between the target user node and target book node in the cross-network model.

(2) User influence: influence into the user to read and evaluate the impact of the sum of the degree of influence of the influence of other nodes is used to pass the network effects associated with the target node generated.

(3) Users intensity: the relative strength and relative evaluation of reading and

intensity, taken from the number of users and user evaluation of reading times.

(4) Book influence: The sum for the book to be read and evaluated the influence that other nodes used the sum of the impact associated with the pass-through effect on the target network node generated.

(5) Book intensity: the relative strength of the book and read the relative strength of the evaluation and, taken from the target number of books to be read and evaluated the number of users.

6.2 Question two: Evaluation forecast

6.2.1 Building complex networks

Establishing a complex hierarchical network model for data subject, due to the large number of users and the number of data books, it is not conducive to calculate deformity complete network. Since the model just use the shortest path between nodes evaluated, so the local topology has been constructed based on an assessment paths between nodes. We selected 7,245,481, 7,625,225, 4,156,658, 5,997,834, 9,214,078 and 251,537 six nodes to be evaluated and they are separated with no more than two nodes for all users of 1168, selected 34 books to be evaluated as a second layer nodes, it is as following:

Table1

Types of node	ID
users	1——1168
books	1168——1202

6.2.2 Data Preparation

Based on complex network model above, Floyd algorithm is used to calculate the shortest distance between the users and books, the result of data is calculated as follows:

Table2

No.	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
1	3.35	0.54	4.22	0.53	14.58	0.63	0.58	0.77	0.86	0.23	0.04	0.29	0.04
2	3.35	0.54	3.93	0.38	17.14	0.63	0.58	0.49	0.37	0.2	0.03	0.23	0.02
3	3.35	0.54	4.28	0.49	16.25	0.63	0.58	1	0.85	0.21	0.03	0.26	0.03
4	3.35	0.54	4.23	0.44	16.25	0.63	0.58	0.85	0.73	0.21	0.03	0.26	0.03
5	3.35	0.54	3.93	0.16	22.5	0.63	0.58	0.01	0.2	0.15	0.02	0.17	0.01
6	3.35	0.54	3.87	0.55	16.25	0.63	0.58	0.42	0.49	0.21	0.03	0.24	0.03
7	3.89	0.72	3.83	0.39	19.64	0.83	0.6	0.01	0.13	0.2	0.04	0.19	0.02
8	3.89	0.72	4.04	0.2	14.22	0.83	0.6	0.78	0.54	0.27	0.05	0.28	0.01
9	3.89	0.72	3.73	0.49	15.88	0.83	0.6	0.35	0.15	0.24	0.05	0.23	0.03
10	3.89	0.72	3.8	0.38	18.33	0.83	0.6	0.25	0.01	0.21	0.04	0.21	0.02

Note1: X1, X2, X3, ... X13 factors of 13 indicators specific meaning already

explained before;

Note 2: The table shows the evaluation portfolio numbered sequentially title sequence requirements;

Note 3: Complete the table is in the appendix.

6.2.3 Factor Analysis

Using SPSS for factor analysis, correlation coefficients and KMO and Bartlett test are to be implemented. The result is expressed in the table below:

Table3-KMO and Bartlett Test

Sampling sufficient degree of Kaiser-Meyer-Olkin measure	0.577
Approximate chi-square	789.969
Df	78
Sig.	0.000

In the appendix of the correlation matrix of the form can be obtained correlation coefficients were in the vast majority of more than 0.4, and KMO and Bartlett's test of sig is 0, so reject the null hypothesis that the correlation coefficient is 0. Correlation exists between the explanatory variables, factor analysis can be done deal.

Table4

indicators	initial	extraction	indicators	initial	extraction
x1	1.000	0.739	x8	1.000	0.729
x2	1.000	0.877	x9	1.000	0.893
x3	1.000	0.753	x10	1.000	0.924
x4	1.000	0.872	x11	1.000	0.929
x5	1.000	0.931	x12	1.000	0.984
x6	1.000	0.926	x13	1.000	0.895
x7	1.000	0.888			

The following table extracted factors to explain the total variance from SPSS.

Table5-Explained the total variance

Ingredients	Initial figure			Extract sum of square		
	total	variance %	accumulate %	total	variance %	accumulate %
1	5.085	39.115	39.115	5.085	39.115	39.115
2	3.442	26.479	65.594	3.442	26.479	65.594
3	1.422	10.935	76.529	1.422	10.935	76.529
4	1.391	10.702	87.231	1.391	10.702	87.231
5	.885	6.804	94.035			
6	.424	3.265	97.300			
7	.205	1.579	98.879			

8	.070	.538	99.417			
9	.051	.389	99.806			
10	.014	.111	99.917			
11	.006	.044	99.961			
12	.004	.032	99.992			
13	.001	.008	100.000			

Table6-Ingredients score coefficient matrix

	Ingredients			
	1	2	3	4
x1	-0.098	0.054	0.317	-0.363
x2	-0.024	-0.185	0.438	0.187
x3	0.070	0.227	-0.085	-0.026
x4	0.084	0.132	0.196	0.459
x5	-0.182	0.051	-0.039	0.137
x6	0.141	-0.150	-0.259	0.062
x7	0.153	-0.113	-0.256	0.005
x8	0.085	0.200	0.130	-0.126
x9	0.059	0.255	0.116	-0.041
x10	0.143	-0.037	0.183	-0.402
x11	0.097	-0.180	0.384	0.027
x12	0.189	-0.001	-0.013	-0.172
x13	0.147	0.085	0.140	0.330

In order to evaluate the users-books score, take the weight of each factor variance contribution rate factor adding to the total weight of each score. The result obtained as follows:

Table7

Users ID	Books ID	f1	f2	f3	f4	f	Revised F	Score
2515537	**900197**	1.20	0.78	0.75	1.28	1.03	0.99	4.98
2515537	**680158**	-0.16	-0.47	-0.46	1.07	-0.14	0.48	3.96
2515537	**770309**	0.77	1.05	0.29	1.16	0.85	0.91	4.82
2515537	**424691**	0.57	0.65	0.05	0.95	0.58	0.80	4.60
2515537	**573732**	-1.68	-1.10	-2.12	0.78	-1.26	0.00	4.00
2515537	**210973**	0.32	-0.24	0.25	2.08	0.36	0.70	4.40
4156658	**175031**	-0.79	-1.66	0.60	1.38	-0.61	0.28	2.56
4156658	**422711**	0.63	-1.17	1.74	-1.63	-0.06	0.52	4.04
4156658	**585783**	0.28	-1.62	1.91	1.21	0.02	0.56	4.12
4156658	**412990**	-0.53	-1.76	0.87	0.96	-0.54	0.31	4.00
4156658	**134003**	0.64	-1.51	2.03	-0.64	0.00	0.55	3.20
4156658	**443948**	-0.81	-1.83	0.24	-0.42	-0.94	0.14	3.14
5997834	**346935**	-1.30	0.95	0.86	-1.73	-0.40	0.37	3.74

5997834	**144718**	-0.88	2.07	1.36	0.10	0.42	0.73	4.46
5997834	**827305**	-1.48	0.62	0.90	-0.01	-0.36	0.39	4.39
5997834	**219560**	-0.73	1.73	1.49	-0.21	0.36	0.70	4.40
5997834	**242057**	-1.35	1.10	0.84	-0.28	-0.20	0.46	3.92
5997834	**803508**	-1.86	0.13	0.35	-0.93	-0.87	0.17	3.17
7245481	**794171**	1.05	-0.25	-0.74	-0.17	0.28	0.67	4.34
7245481	**381060**	1.62	0.47	-0.29	-0.32	0.80	0.89	4.78
7245481	**776002**	1.49	1.31	-0.40	0.30	1.05	1.00	5.00
7245481	**980705**	1.72	-0.14	-0.15	-0.46	0.66	0.83	4.66
7245481	**354292**	1.30	-0.65	-0.36	-1.44	0.16	0.62	4.24
7245481	**738735**	1.27	-0.54	-0.71	-2.14	0.05	0.57	4.14
7625225	**473690**	0.12	-0.28	-0.65	1.16	0.03	0.56	4.12
7625225	**929118**	-0.09	-0.31	-1.27	0.26	-0.26	0.43	4.43
7625225	**235338**	0.28	0.32	-0.85	0.29	0.15	0.61	4.22
7625225	**424691**	0.59	0.82	-0.58	0.80	0.54	0.78	4.56
7625225	**916469**	0.48	-0.17	-0.56	-0.30	0.05	0.57	4.14
7625225	**793936**	0.16	0.12	-1.06	-0.14	-0.04	0.53	4.06
9214078	**310411**	0.32	0.95	0.05	0.01	0.44	0.74	4.48
9214078	**727635**	-0.82	-0.74	-1.20	-1.12	-0.88	0.16	4.00
9214078	**724917**	-1.68	0.34	-1.66	0.51	-0.80	0.20	3.40
9214078	**325721**	-0.46	-0.22	-0.82	-1.49	-0.56	0.30	3.60
9214078	**105962**	0.10	0.53	-0.17	-0.31	0.15	0.61	4.22
9214078	**235338**	-0.32	0.73	-0.54	-0.56	-0.06	0.52	4.04

6.3 Intelligent Recommendation

6.3.1 Build Network

To carry out for the title of the six active users recommend, first we need to build complex hierarchical network based on the user's social connections. Because poor fitness in readers with books which are too far away from the user make the intelligent recommendation ineffective, so the local intelligence network topology to recommend is established. Adjacent to the target user to select a user node and all its related books nodes constitute a network of forest, of which books node 2434, the user node 78, as follows:

Table8

Node Type	ID
Users	1——78
Books	79——2512

Adjacency matrix of local network (local):

$$
Path(i,j) = \begin{bmatrix} 5 & 6 & 31 & 8 & 9 & 10 \\ 31 & 31 & 7 & 31 & 31 & 31 \\ 5 & 6 & 31 & 8 & 9 & 10 \\ 362 & 6 & 31 & 8 & 10 & 10 \\ 38 & 6 & 2 & 8 & 38 & 10 \\ 5 & 6 & 31 & 8 & 69 & 10 \end{bmatrix}
$$

The shortest distance matrix (partial):

$$
\min d(i,j) = \begin{bmatrix} 17.14 & 12.50 & 0.00 & 12.50 & 14.29 & 17.69 \\ 7.14 & 7.14 & 12.50 & 0.00 & 6.67 & 8.33 \\ 10.00 & 10.00 & 14.29 & 6.67 & 0.00 & 11.11 \\ 11.11 & 8.33 & 17.69 & 8.33 & 11.11 & 0.00 \\ 40.48 & 33.33 & 45.83 & 38.60 & 41.30 & 40.00 \\ 8.33 & 6.67 & 10.00 & 4.76 & 7.14 & 9.09 \end{bmatrix}
$$

Calculate 13 kinds of index data by the complex network and use the weight ratio of the second issue to forecast users for all books score.

Recommendation algorithm is recommended by the intelligent, the results are listed in the following table:

Table9-results

The user's ID	Recommendation first		Recommendation second		Recommendation third	
	Books ID	shortest path	**Books ID**	shortest path	**Books ID**	shortest path
2515537	698573	16.25	516012	18.75	709644	18.75
4156658	698573	15.88	516012	18.38	709644	18.38
5997834	794171	23.45	284550	34.32	284550	36.74
7245481	702699	24.56	962729	29.13	642256	31.78
7625225	698573	16.25	516012	18.75	709644	18.75
9214078	776002	27.25	551643	32.25	510372	32.25

7 .Model conclusion

7.1 Model advantage

(1) The model makes full use of adaptive social network information on books and user evaluation to analysis based on data association in the evaluation and recommendation process.

(2) The forecast process considering the user ratings, habits, books rated features, users and books related three a total of 13 indicators, which can reflect the strength of association between the user and the books. And each node is given strength value according to network importance to evaluate the books based on the social network.

(3) In the paper, the using of factor analysis method in the prediction model for score solves the problem of multi-collinearity among indexes by creating 4 public factors. And comprehensive index is weighted by its contribution to variance, accurately measuring the evaluation of readers of books.

(4) In intelligent recommendation model, the establishment of local topological structure of the network to facilitate evaluation of the specified user by integrated index of the topological graph. And then choose high evaluation of books as a recommendation, which makes the recommendation of intelligent objective and comprehensive.

7.2 Model shortcomings

(1) The model is the evaluation of the social network topology prediction and intelligent recommendation based on social networks. But the topology structure is more complex and difficult to achieve network wide most evaluation.

(2) 13 indexes in the model of factor analysis is not necessarily comprehensive coverage of all the factors. At the same time, the process of evaluation of the user books is also a change in psychology and behavior. Only with mathematical algorithms cannot completely simulate the scoring process.

(3) Score prediction, predictive value in the node original evaluation minimum value and maximum, but in the intelligent recommendation process, the evaluation value as sorted matter recommendation errors.

7.3 Model development

7.3.1 Shortest distance clustering method

For specific users and groups in the books, the whole complex networks assessed directly face the huge computation and a large amount of invalid redundant information, which need to have a mechanism to simplify the social network. The shortest distance clustering method can solve the difficult problem of network redundancy.

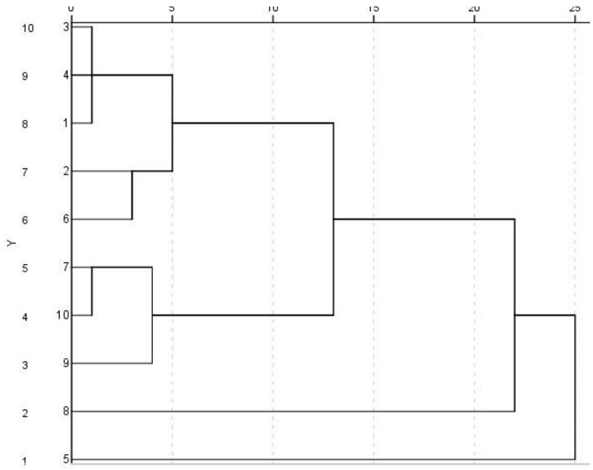
The main idea of the shortest distance clustering method is to classify the target data and to together the objects whose distance are far from the target to set as an object to the establishment of network, so as to reduce the dimension.

The shortest distance clustering method is found in the non-diagonal elements of the original m × m distance matrix. The classification of object Gp and Gq are merged into a new kind of Gr, then according to the calculation formulas:

$$d_{rk} = \min\{d_{pk}, d_{qk}\} \quad (d \neq p, q)$$

Calculate the distance between the original and new types, so you get a new (M - 1) distance matrix order. From the new distance matrix, select the smallest d_{ij} and merge Gi and Gj into the new class. Then commutate the distance between various new types and so on, until each classification of the object is classified as a class.

The clustering algorithm can make the shortest distance clustering dendrogram according to the following information, as follows:



7.3.2 Optimization model for evaluation and prediction

In the above model, we can evaluate to f=WF according to the factor analysis method.

$$Score(i,j) = \min\left(Score_j\right) + f'(i,j)[\max(Score) - \min(Score)]$$

Because of the difference of each book evaluation, difference between the highest and the lowest will cause the recommended standard not unified for each user. The calculation leads to the matching degree not the size of elements in Score matrix of complete response to users and books, which is likely to lead to recommend failure.

Therefore define: $Score(i,j) = \min\left(Score_j\right) + f'(i,j)[\max(Score) - \min(Score)]$. The score data adjusted higher availability. The correct forecast is the base of intelligent recommendation.

7.3.3 Cold start recommendation to expand

Cold start recommendation is recommended actions pointers to the system complex network in isolated node. Due to the short time, the new user does not form a social network and reading history. In the process of network analysis, it will exist as an isolated point. The user recommendation cannot be realized based on the shortest path algorithm of complex network model, but in the entire network of statistical power or books recommended as the basis of the highest intensity to present to the user.

Reference

[1] Ding Xue. Intelligent Recommendation Based on data mining. Journal of Wuhan University Book, 2010.

[2] Shen Wenjuan. Design function selection system intelligent recommendation. Technology square. 2013,11.

[3] Wang Ruiqin. Pattern recognition and artificial intelligence intelligent recommendation data sources and combined based on clustering, Kong Fansheng twenty-first No. sixth.2008.11.

[4] Feng Minyu. Optimization problems in complex network path of electronic research and application of . University of science and technology, 2013.

[5] Chen Genlang. Some of recommended technology based on social media, Zhejiang University.2012.

[6] Yao Qifu.. fund project of personalized intelligent recommendation based on Web usage mining, Service, 2006.

[7] Lin Zhao. The weighted Mahalanobis distance discriminant analysis method and the determination of weight -- Intelligent Tourism Information Service System of the recommended. economic mathematics, volume twenty-fourth issue second

[9] Tian Chao. Research and development system of computer intelligent recommendation based on the review analysis, SuperRank .2010.