评委一评分，签名及备注	队号： 10490	评委三评分，签名及备注
评委二评分，签名及备注	选题： B	评委四评分，签名及备注
题目：基于复杂网络和因子分析的智能推荐		

随着互联网的飞速发展，基于社交网络的评价系统在网络推荐中扮演者越来越重要的角色。本文致力于通过复杂网络算法和因子分析法进行用户-书籍评分预测和智能推荐。

首先，本文建立了用户-书籍有向分层复杂网络。第一层用户关系网络以用户之间的社交关系为基础，以用户间的公共阅读记录数目作为权重形成有向网络层；第二层则以用户评分为权重形成用户与书籍之间的有向交叉网络。基于这一复杂网络，我们定义了网络中结点之间的距离为所经过边权重的倒数和，采用 Floyd 最短路径算法求解结点之间的最短距离，并且定义书籍结点和用户结点的距离为用户和书籍的适应度指标。其次，我们根据可能影响用户对书籍的评分的因素，还定义了网络结点的强度，网络结点影响力，相对强度，适应度等指标以便更好地衡量用户与书籍间的有向联系。结点强度分为阅读强度和评价强度

针对问题一和问题二，本文把影响用户对书籍评价的因素分为用户评价特征，书籍受评特征和用户书籍联系特征三大类；其中用户评价特征和书籍受评特征分别包括用户评分均值，用户评分偏移均值，书籍受评均值，书籍受评偏移量等四项指标，而用户-书籍联系主要是基于网络评价产生的强度影响力，相对强度和适应度等九项指标。随后采用 SPSS 对上述 13 项因素进行因子分析，提取出四项公共因子，以方差贡献率为权重把四项公共因子合成综合指标。在对综合指标做了无量纲化运算后，结合书籍评价的极大值与极小值区间产生用户对书籍的评价预测。

针对问题三，本文建立了智能推荐算法进行最佳书籍推荐。在问题二的基础上，依靠复杂网络算法和因子分析法，建立起以四项公共因子为基础的综合指标，用来生成用户对书籍的评价矩阵。然后依照评分矩阵和阅读记录对未读书籍进行择优筛选，产生了最优推荐结果，题目中六位读者第一推荐书籍 ID 分别为：698573,698573,794171,702699,698573,776002。

最后，我们评价模型的优缺点，提出了复杂网络算法的降维聚类处理，用户评分的标准化以及冷启动推荐等拓展性算法。

关键字：复杂网络算法 最短路径算法 适应度指标 因子分析 智能推荐算法

基于复杂网络和因子分析的智能推荐

1. 问题重述

形色各异的信息随着信息技术和互联网的发展已经充斥着我们的生活，人类也早已从信息匮乏的时代走向信息过载的时代。在处理信息的过程中，信息消费者还是信息生产者都遇到了很大的挑战：信息消费者从大量信息中找到自己感兴趣的信息开始变得越来越困难；同样的，信息生产者需要解决的最大的问题则是如何让自己生产的信息脱颖而出在大量信息检索和推荐的过程中脱颖而出，醒目的呈现在需求者的面前，得到消费者的认可。

人们开始探索各种可能解决这一矛盾的工具，推荐由此应运而生。在互联网的产品和应用中被广泛采用，包括大家经常使用的相关搜索、话题推荐、电子商务的各种产品推荐、社交网络上的交友推荐等。

在本题中，我们首先获得了一个著名网上书店的大量的用户及用户行为信息，包括对于书籍的评分数据，书籍的标签信息以及用户的社交关系等大量的数据信息。我们需要解决如下的问题：

（1）分析影响用户对书籍评分的因素；

（2）建立一个模型，预测 predict.txt 附件中的用户对未看过书籍的评分；

（3）针对 predict.txt 附件中的用户，给每个用户推荐 3 本没看过的书籍。

2. 模型假设

假设一：用户对书籍的评价是客观且仅仅对其有偏好的书籍评价

假设二：用户社交网络的形成是基于用户对相似度的人群的偏好，即用户社交网络能较为准确地反映用户群的偏好特征。

假设三：用户对书的偏好受其他读者的影响，用户群体的偏好是趋同的。

3.符号说明

符号	意义
e_i	用户 *i* 对于书籍的评价次数
$\mathrm{N}_{\mathrm{ij}}^{\mathrm{b}}$	用户—书籍网络的有向权重
d_{ij}	i 和 j 结点间的紧密程度
I^e	用户评价影响力
$Path_{ij}$	i 到 j 路径上的后继点
X_1	用户 i 的评价均值
X_2	用户 i 的评价偏移量
X_3	第 j 本书籍的被评价的均值
X_4	第 j 本书籍的被评价的偏移量
X_6	用户 i 对用户 j 的影响力

X_7	用户 i 的强度
X_{10}	用户评价均值的调整量
F	综合指标矩阵
$Score(i,j)'$	评价分数的修正值

4. 问题分析

基于社交网络关系的数据挖掘是当前研究的热点。社交网络的构建使得互联网中多因素的数据有机的结合起来，形成了一个巨型多层次数据网络，如何有效的管理这些数据网络和利用这些数据网络的信息创造最大的效益是数据挖掘的深层意义所在。本题基于有向加权复杂网络模型，在分析解决客户关系和书籍信息的系统评价问题中，尝试寻找各个已知数据之间的联系，把所有已知因素综合性地表现在复杂网络算法中，以实现书籍评分系统的全面化、客观化和智能化。

在书籍评分的预测中，我们必须考虑所有可能影响用户对书籍评分的因素，主要包括用户个人的评分结构，用户对该书籍现有的评分结构，用户与书籍的匹配程度。本文对上述三种因素进行分析，在分析过程中，用户评分结构与书籍的评分结构比较容易衡量，而用户与书籍的匹配程度则是该系统评价的核心，也是算法设计中的重点和难点，它是基于用户社交网络和书籍的联系网络进行提取的。用户与书籍的匹配程度主要考虑用户与书籍在网络中的距离，用户的影响力，书籍本身的影响力等指标。

所以本题的总体思路分为社交网络的构建、指标分析以及预测和智能推荐三部分，其中，社交网络构建是解决该问题的基础。

4.1 复杂网络构建

基于社交关系的复杂网络系统是构建评分预测模型的基础，本文采用的网络构架在传统的有向加权图的基础上实现合理的网络分层，每层网络结点包括读者和书籍，除了读者与读者的网络，书籍与书籍的网络，还包括读者与书籍之间单向关系，由此组成双层有向加权图，并依照改进的复杂网络算法进行计算。

4.2 读者有向加权网络

读者层的网络依照读者之间的关注记录进行构建，并以读者之间相同的阅读历史记录次数作为两读者之间的有向权重，其含义为两个读者的偏好相似度。网络图主要体现在阅读过程中，读者之间通过关注这个行为产生的相互影响关系。并且在实际中，读者之间相同爱好越多，其产生的影响越大，在读者层网络中体现的共同阅读记录越多，读者结点权重值越大，结点间距离越小。

4.3 读者书籍交叉网络

材料中的信息包括读者阅读的历史记录和评价记录，依据这两项数据可建立读者与书籍间的有向加权网络图。但通过对题中所给数据的简单分析可以发现，整个系统中评价记录总数远远大于阅读记录（即有些读者没有阅读该书籍但是却对该书籍进行评分），因此，这些评价数据是无法直接应用的，必须剔除不合适评价逻辑的虚假评价（无阅读记录单有评价记录），将剩余的有效数据进行读者书籍的交叉网络评价。所有书籍的信息位于不同于读者有向加权网络的新的网络

层，书籍之间暂不建立有向关联关系。

读者书籍的交叉网络拓扑图以读者对书籍的评分为基础进行构建，评分反映了读者对该书籍的喜爱程度。通过评分，将读者与各书籍联系起来，进而形成交叉的双层网络图。而书籍自身的影响力通过书籍被评价次数来反映，定义其为书籍结点强度。

4.4 Floyd 最短路径算法

基于上述读者分层网络拓扑结构，采用数据挖掘的方法分析该有向加权网络，利用 Floyd 最短距离算法计算第一层用户结点和第二层书籍结点的最短网络加权路径，以此来表现书籍对读者喜好的适应度大小。

4.5 因子分析法及预测

本文采用复杂网络模型求解出书籍对不同读者喜好的最佳适应度。但通过行为学分析，读者对书籍的评价还取决于读者的评分习惯，书籍的内容，书籍自身的影响力等指标。因此我们需要对所有可能影响到评价的各种因素进行综合考虑。由于因素间可能的共线和重叠关系，导致综合评估指标难以量化，所以我们采用因子分析法提取公共因子，通过计算各因子的贡献率来确定各个因素的权重体系，预测出读者对书籍和评价。

经过前面的算法和预测评价，问题三的解决方法主要是对评价结果进行排序后为用户进行书籍推荐。

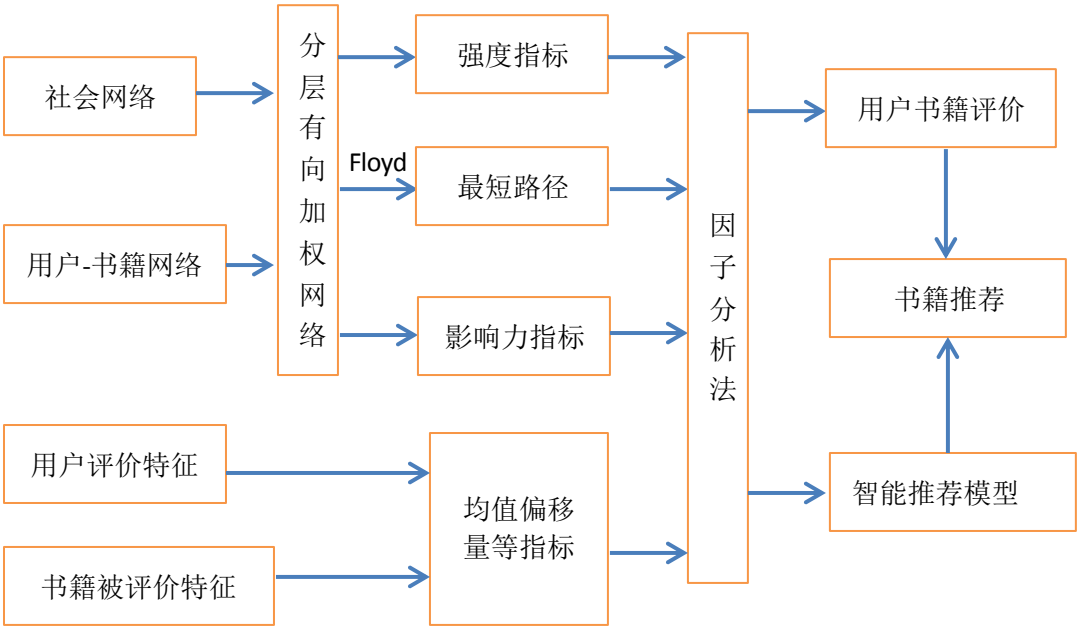


图 1 系统流程图

5. 模型建立

5.1 模型一：复杂网络模型

基于社交关联关系建立书籍和读者的有向加权网络，根据结点的两大种类进行分层，设定网络系统可以通过最短路径算法来求解出书籍对读者的最佳适应度指标。适应度指标是预测评价系统中较为重要的决定因素。预测评价的另一基本因素结点强度与影响力指标可以通过社交网络来定义。

5.1.1 用户层网络

记用户间网络拓扑结构的有向加权邻接矩阵为U^R

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

其中，$\mathrm{U_{ij}^R}$表示第 i 个用户对第 j 个用户邻接关系的强弱程度。

定义：$\mathrm{U_{ij}^R}=\infty$表示结点自身之间联系强度最强；

当$\mathrm{U_{ij}^R}=0$时表示结点之间无联通。

在读者用户网络中，用户的结点强度用来反映用户的活跃情况和用户自身的影响力，共分为评价强度和阅读强度，分别用e_i和v_i来表示。

定义：矩阵$\mathrm{E}=(e_1\ e_2\cdots e_n)^T$ 和矩阵$\mathrm{V}=(v_1\ v_2\cdots v_n)^T$

其中

e_i：用户的评价次数

v_i：用户的阅读记录次数

e_i和v_i在一定程度上可反映用户的影响力及活跃程度。

5.1.2 用户——书籍交叉层网络

同理，依照用户和书籍之间的关联情况建立用户和书籍之间的关联网络。

设邻接矩阵为U^B

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

矩阵中U_{ij}^b为用户书籍间的邻接权数。

同理我们定义：

$$\mathrm{N_{Ij}^b}=\mathrm{b_{Ij}}\cdot\mathrm{s_{ij}}$$

$\mathrm{N_{ij}^b}$构成用户—书籍网络的有向权重，它反映用户 i 对书籍 j 的偏好程度。

$$\mathrm{b_{ij}}=\begin{cases}0 & \text{第 i 个用户未阅读第 j 本书}\\ 1 & \text{第 i 个用户已阅读第 j 本书}\end{cases}$$

$\mathrm{s_{ij}}$表示第 i 个用户对第 j 个书籍的评分，无评价则默认为是平均值。

书籍结点的强度分别表示为评价强度e_i'和阅读强度v_i'。

其中e_i'为第 i 本书籍被评价的次数（校正后）；$\mathrm{v_i'}$为第 i 本书籍的书签数目。

评价强度e_i'和阅读强度v_i'能反映出第 i 本书籍的受欢迎程度，其中e_i'必须为校正后的评价次数，因为在题目中的数据存在虚假评价（前文中已解释）。

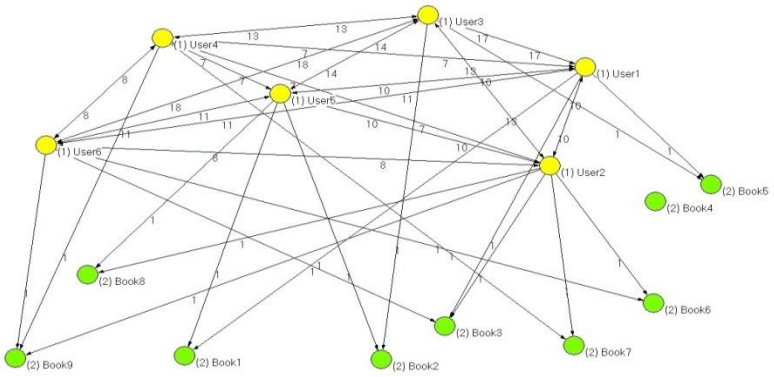


图 2 交叉网络示意图

5.1.3 最短路径算法（Floyd）

（1）混合网络

用户网络和用户书籍网络可以组合成一个分层的有向加权网络，把书籍和用户等效成同一类结点，则新的网络拓扑架构的邻接矩阵为

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

$\mathrm{U_{ij}}$表示 i 结点到 j 结点的权重；

其中，$\mathrm{U^{R}}$表示用户的之间的结点权重；$\mathrm{U^{B}}$表示用户与每一个书籍之间的结点权重。

定义$\mathrm{C_{ij}}$为网络任意相邻两点间的距离。

$$\mathrm{C_{ij}}=\begin{cases} 0 & \mathrm{i=j} \\ \dfrac{1}{\mathrm{U_{ij}}} & \mathrm{U_{ij}\neq 0\ 且\ i\neq j} \\ \infty & \mathrm{U_{ij}\neq 0} \end{cases}$$

则任意两点之间的最短距离为

$$\mathrm{d_{ij}=\min\left\{C_{ik_1}+C_{k_1k_2}+\cdots+C_{k_nj}\right\}}$$

$\mathrm{d_{ij}}$可以准确衡量 i 和 j 结点间的紧密程度，如果 i 为用户编号，j 为书籍编号，则$\mathrm{d_{ij}}$（i∈users，j∈books）表示书籍为用户偏好的适应度。

（2）Floyd 算法

Floyd 算法是求解有向图结点间最短路径的一种方法，它的思想是在加权邻接矩阵中用插入定点的方法依次递推地构造出 n 个矩阵 D(1),D(2)…D(n)，D(n)为网络模型的距离矩阵。同时引入一个后继点矩阵记录两点间的最短距离。

Floyd 算法两个重要属性矩阵为 D 和 Path。

其中$\mathrm{Path_{ij}}$ 为 i 到 j 路径上的后继点，算法如下：

Step1：输入加权邻接矩阵 C

Step2：赋初值，令$\mathrm{d_{ij}=c_{ij}}$，$\mathrm{Path_{ij}=j}$，k=1.

Step3：更新 D 和 Path；

对一组 i 和 j，如果满足 D(i,k) + D(k,j) <D(i,j)；

则 D(i,j)=D(i,k) + D(k,j) ；

Path(i,j)=Path(i,k)；

Step4：K 值加 1 返回 Step3 直至 k=n+1（n 为总结点数）。

依据上述步骤，可准确求解出 D 和 Path。

5.1.4 结点影响力和结点强度评定

在混合网络中，各个结点的影响力可以通过网络中每个结点的强度和网络的拓扑结构来进行描述。对上述复杂网络而言，每个结点均有两种强度e_i和v_i，所以结点影响力也有两种，记I^e为评价影响力，I^v为阅读影响力。

$$I_j^e = \sum_{i=1}^{n} \frac{e_i}{d_{ij}} \quad (i \neq j)$$

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

结点的影响力主要考虑了其他结点通过网络传递效应后对目标结点影响程度的总和，而结点的强度则主要为目标结点对外部的影响，用户的强度采用用户对书籍的有效评价次数来表示，书籍的强度则采用书籍被有效评价的次数。

结点影响力和结点强度指标既适用于用户结点，也适用于书籍结点。

5.2 因子分析法

5.2.1. 因素提取

用户对书籍的评价主要受各种因素制约，如下图所示

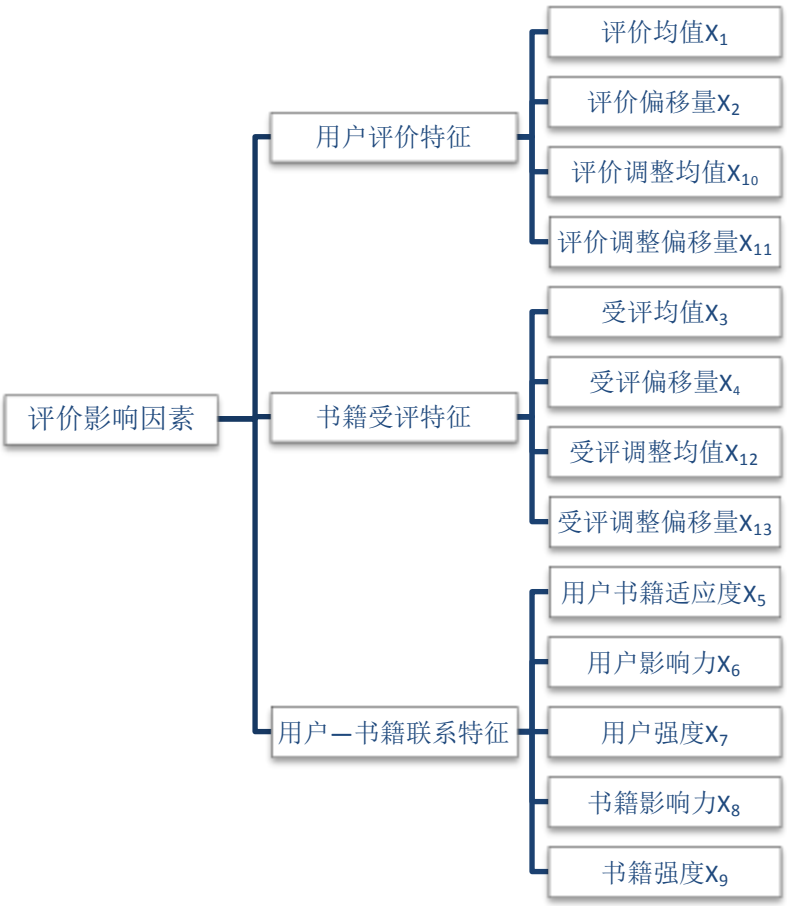


图 3 因素分解图

其中，

X_1（i，j）：第 i 个用户的所有评价的均值；

X_2（i，j）：第 i 个用户的所有评价的偏移量；

X_3（i，j）：第 j 本书籍的被评价的均值；

X_4（i，j）：第 j 本书籍的被评价的偏移量；

且

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

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

其中R_{ik}表示第 i 个用户对第 k 本书籍打分（来自评价历史），$|T|$和$|U|$分别指评价书籍的总个数和评价用户的总数。

$$X_5(i,j) = d_i$$

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

X_6 为用户 i 对用户 j 的影响力，定义为相对阅读影响力和相对评价影响力之和。

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

X_7 为用户 i 的强度，定义为相对阅读强度和相对评价强度之和；

X_8和X_9分别是书籍的影响力和强度，其定义方式如同X_6X_7；

X_{10} 是用户评价均值的调整值，定义为 $X_{10} = {X_1}/{d_{i,j}}$；

X_{11} 是用户评价偏移量的调整值，定义为$X_{11} = {X_2}/{d_{i,j}}$；

X_{12}是书籍被评价均值的调整值，定义为$X_{12} = {X_3}/{d_{i,j}}$；

X_{13}是书籍被评价偏移量的调整值，定义为$X_{13} = {X_4}/{d_{i,j}}$。

上述十三大因素是影响评价结果的最主要的原因，下面将对其进行因子分析。

5.2.2 因子分析法基本理论

因子分析法是从研究变量内部相关的依赖关系出发，把一些具有错综复杂关系的变量归结为少数几个综合因子的一种多变量统计分析方法。采用这种方法我们就可以对原始数据进行分类归并性的分析提取，将相关度较为密切的变量归纳为多个综合指标，同时令这些综合指标所综合的信息相互不重叠。我们则把这些

综合指标定义为公共因子。

因子分析法的基本思路是对研究变量进行分类，将关联度较高，联系比较紧密的人为的归在一起，相应的，不同类变量之间的关联度较低。在这样的方法下，每一类变量实际上就代表了一个基本结构，即公共因子.我们需要研究的就是试图用最少个数的不可测的所谓公共因子的拟定函数与特殊因子之和来描述原来观测的每一分量。这样，就能相对容易地以较少的几个因子反映原资料的大部分信息，从而达到浓缩数据，提取数据的作用和目的。

因子分析法的核心是对若干综合指标进行因子分析并提取公共因子，再以每个因子的方差贡献率作为权数与该因子的得分乘数之和构造得分函数。因子分析法的数学表示为矩阵：X=AF+B，即：

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

模型中，向量$\mathrm{X} = (\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3, \cdots, \mathrm{x}_\mathrm{p})$是可观测随机向量，即原始观测变量。$\mathrm{F} = (\mathrm{f}_1, \mathrm{f}_2, \mathrm{f}_3, \cdots, \mathrm{f}_\mathrm{k})$是$\mathrm{X} = (\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3, \cdots, \mathrm{x}_\mathrm{p})$的公共因子，即各个原观测变量的表达式中共同出现的因子，是相互独立的不可观测的理论变量。公共因子的具体含义必须结合实际研究问题来界定。$\mathrm{A}(\alpha_{\mathrm{ij}})$是公共因子$\mathrm{F} = (\mathrm{f}_1, \mathrm{f}_2, \mathrm{f}_3, \cdots, \mathrm{f}_\mathrm{k})$的系数，称为因子载荷矩阵，$\alpha_{\mathrm{ij}}(\mathrm{i} = 1,2,\ldots,\mathrm{p};\mathrm{j} = 1,2,\ldots,\mathrm{k})$称为因子载荷，是第 i 个原有变量在第 j 个因子上的负荷，或可将α_{ij}看作第 i 个变量在第 j 公共因子上的权重。α_{ij}是 $\mathrm{x_i}$和f_j的协方差，也是$\mathrm{x_i}$和$\mathrm{f_j}$的相关系数，表示$\mathrm{x_i}$对$\mathrm{f_j}$的依赖程度或相关程度。α_{ij}的绝对值越大，表明公共因子$\mathrm{f_j}$ 对于$\mathrm{x_i}$的载荷量越大。$\mathrm{B} = (\beta_1, \beta_2, \beta_3, \cdots, \beta_\mathrm{p})$是$\mathrm{X} = (\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3, \cdots, \mathrm{x}_\mathrm{p})$的特殊因子，是不能被前 k 个公共因子包含的部分，这种因子也是不可观测的。各特殊因子之间以及特殊因子与所有公共因子之间都是相互独立的。

5.2.3 模型的数学含义

因子载荷矩阵 A 中包含了两个统计量，分别是变量共同度和公共因子的方差贡献度。

（1）变量共同度

变量共同度是因子载荷矩阵 A 的第 i 行的元素的平方和。

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

它衡量全部公共因子对$\mathrm{x_i}$的方差所做出的贡献，反映全部公共因子对变量 $\mathrm{x_i}$ 的影响。$\mathrm{h_i^2}$越大，表明 X 对于 F 每一分量的依赖程度大。

（2）方差贡献度

方差贡献度因子载荷矩阵中各列元素的平方和。

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

$\mathrm{g_j^2}$称为公共因子$\mathrm{F} = (\mathrm{f}_1, \mathrm{f}_2, \mathrm{f}_3, \cdots, \mathrm{f}_\mathrm{k})$对$\mathrm{X} = (\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3, \cdots, \mathrm{x}_\mathrm{p})$的方差贡献，表示第 j 个公共因子$f_j$对于 x 的每一个分量$x_i$ (i=1,2,...,p)所提供的方差的总和，是衡量公共因子相对重要性的指标。

（3）综合指标

我们采用上述各因子的贡献率占总贡献率的比重作为权重 w，对上述个影响因子加权，定义 F，如下。

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

f(i,j)为综合指标，它是做预测评价系统的参考指标。

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

f' 为 f 的修正指标，数值的含义为 f(i,j)在所有 f 取值的相对位置。

同理定义：评价分数的修正值，表达式如下：

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

我们采用修正的综合指标作为分数修正值的权数，数值的含义为某个具体分数在整体评分中的一个相对位置，数值范围在所有评分的最大与最小值之间。

5.3 智能推荐模型

定义矩阵 History 表示用户对书籍的阅读情况。

$$History(i,j)=\begin{cases}0 & \text{用户 i 未阅读书籍 j}\\1 & \text{用户 i 已阅读书籍 j}\end{cases}$$

$$History=\begin{bmatrix}H_{11}&\cdots&H_{1j}\\\vdots&\ddots&\vdots\\H_{i1}&\cdots&H_{ij}\end{bmatrix}$$

定义矩阵 Score 为预测得分矩阵。

Score(i,j) 表示用户 i 对书籍 j 的评价值。

记 M 为一新矩阵，其中$\mathrm{m(i,j)}=\mathrm{score(i,j)}\times\mathrm{history(i,j)}$

智能推荐系统致力于为用户推荐 K(k=1,2,3,...,n)个从未阅读过的书籍。推荐系统的主要思想为，对所有的可能推荐的书籍的打分项进行排序，根据排序结果系统自动进行推荐。

详细步骤如下：

Step1：赋值 i，j，k=1,...max

Step2：查询 M 矩阵第 i 列最大值列表，并设为第 K 个推荐结果。

Step3：第 i 行第 C 列清零，并使 k=k+1 直至 k>max

Step4：返回 Step2

6. 模型求解

6.1 问题一：评分因素

经过建立复杂网络模型和因子分析模型，影响用户评分的因素主要有三大类，分别是用户评价特征、书籍受评特征、用户—书籍联系特征。它们的细分指标解释如下：

1、用户评价特征

（1）评价均值：目标用户所产生的历史评价记录的均值。

（2）评价偏移量：目标用户所产生的各历史评价记录的离差绝对值的平均数。

（3）评价调整均值：评价均值与复杂网络中任意两个结点的最短距离的比值。

（4）评价调整偏移量：评价偏移量与复杂网络任意两结点的最短距离的比值。

2、书籍受评特征

（1）受评均值：目标书籍被用户评价的历史记录的均值。
（2）受评偏移量：目标书籍被用户评价的历史记录的离差绝对值的平均数。
（3）受评调整均值：受评均值与复杂网络中任意两个结点的最短距离的比值。
（4）受评调整偏移量：受评偏移量与复杂网络任意两结点的最短距离的比值。
3、用户—书籍联系特征
（1）用户—书籍适应度：在交叉网络模型中目标用户与目标书籍两结点之间的紧密程度，即两结点之间的最短距离。
（2）用户影响力：分为用户阅读影响力和评价影响力，该影响力即为其他结点用过网络关联传递效应后对目标结点所产生的影响程度总和。
（3）用户强度：为相对阅读强度和相对评价强度之和，取自于用户阅读次数和用户评价次数。
（4）书籍影响力：为书籍被阅读和被评价的影响力的总和，即其他结点用过网络关联传递效应后对目标结点所产生的影响程度总和。
（5）书籍强度：为书籍的相对阅读强度和相对评价强度之和，取自于目标书籍被用户阅读次数和评价次数。

6.2 问题二：评价预测

6.2.1 复杂网络构建

针对题目中数据建立分层复杂网络模型，由于数据用户数和书籍数目较大，不利于畸形完整网络的计算。由于模型中只是利用结点间的最短路径进行评估，所以局部拓扑图的构建已经是以进行结点间路径的评估。我们选取 7245481 ,7625225, 4156658, 5997834, 9214078 和 251537 六个待评价结点以及与它们相隔结点不超过两个的所有用户共 1168 个，选取待评价的 34 本书籍为第二层结点，对它们进行统一编号如下表：

表 1 编号分类

结点类型	编号
用户	1——1168
书籍	1168——1202

6.2.2 数据准备

基于上述复杂网络模型，利用 Floyd 算法求解出用户书籍的最短距离，并基于此可求得结点影响力以及结点的相对强度等指标，部分的计算数据如下所示（完整表格见附录）：

表 2 因素一览表

编号	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

注 1：$X_1,X_2,X_3,\ldots X_{13}$ 为 13 个相关因素的指标，具体含义前文中已经解释过；

注 2：表格中编号依次表示题目中要求的评价组合序列；

注 3：完整表格见附录。

6.2.3 因子分析

对上述 13 列数据利用 SPSS 进行因子分析，首先对其进行相关系数及 KMO 和 Bartlett 检验，相关矩阵及相关矩阵的逆矩阵的详细结果见附录。KMO 和 Bartlett 的检验如下表所示：

表 3 KMO 和 Bartlett 的检验

取样足够度的 Kaiser-Meyer-Olkin 度量。		0.577
Bartlett 的球形度检验	近似卡方	789.969
	Df	78
	Sig.	0.000

在附录的相关矩阵的表格中可以得到相关系数绝大多数均在 0.4 以上，具有相关性，且 KMO 和 Bartlett 的检验中的 sig 值为 0，所以拒绝相关系数为 0 的原假设。说明变量间存在相关性，可以做因子分析处理。

表 4 公因子方差

指标	初始	提取	指标	初始	提取
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			

由上表可知，公共因子对各项指标方差的反应程度都达到了 0.85 以上，用四项公因子来表示 13 项指标是可行的且程度较高。

下表为 SPSS 提取因子解释的总方差。

表 5 解释的总方差

成份	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
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	100.00			

上表中第一列为特征值（主成分的方差），第二列为各个主成分的贡献率，第三列为累积贡献率，由上表看出前 4 个主成分的累计贡献率就达到了 87.231%>85%，所以选取主成分个数为 4。选 x1 为第一主成分，x2 为第二主成分，x3 为第三主成分。且这四个主成分的方差和占全部方差的 87.231%，即基本上保留了原来指标的信息。这样由原来的 13 个指标变为了 4 个指标。

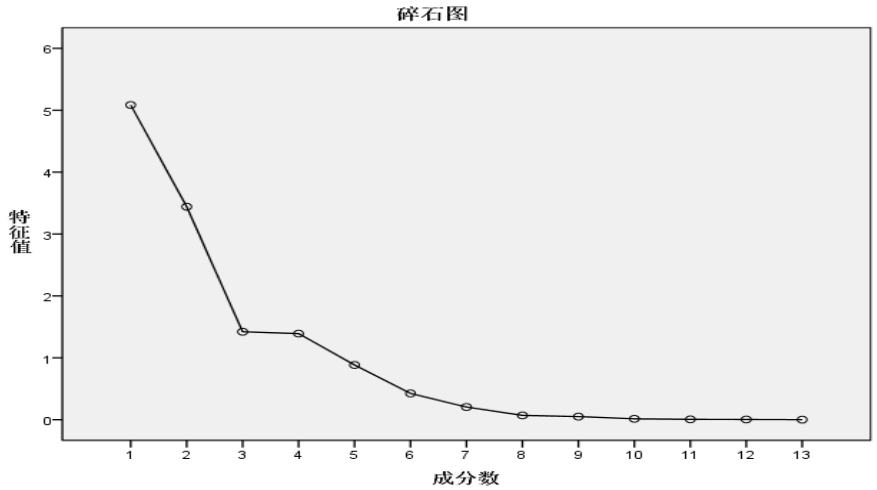


图 4 碎石图

上图为因子分析法碎石图，可见因子数达到 4 以上时，特征值小于 1，特征值的变化曲线趋于平缓，所以由碎石图也可大致确定出主成分个数为 4。与按累计贡献率确定的主成分个数是一致的。

表 6 成份得分系数矩阵

	成份			
	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

上述表格代表了各指标与公共因子的线性关系，上述数据构成了 A 矩阵的元素，由上述表格可求得各公共因子见表 7。为了评价各项用户-书籍评分，取各因子的方差贡献率为权重对因子加总得到各评分如下表所示：

表 7 公共因子和评分结果表

用户 ID	书籍 ID	f1	f2	f3	f4	f	f 修正后	评价得分
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 智能推荐

6.3.1 构建网络

针对题目中的六个用户进行推荐活动，首先需要构建基于用户社交联系的复杂分层网络，有于距离太远的书籍和用户适应度较差，对智能推荐效果不佳，所以建立局部的拓扑网络进行智能推荐。选取与目标用户相邻的用户结点以及所有与其关林的书籍结点构成一个网络，其中书籍结点 2434 个，用户结点 78 个，详情如下：

表 8 编号分配

结点类型	编号
用户	1——78
书籍	79——2512

网络的局部邻接矩阵（局部）

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

最短距离矩阵（局部）：

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

通过进行复杂网络计算，根据 13 类因素指标并沿用问题二中的权重比例预测出用户对所有书籍的评分。

利用智能推荐算法进行推荐，结果如下表：

表 9 推荐结果表

用户 ID	推荐一		推荐二		推荐三	
	书籍 ID	最短路径	书籍 ID	最短路径	书籍 ID	最短路径
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. 模型总结

7.1 模型优点

（1） 该模型充分利用社交网络信息对书籍和用户的适应性进行评价，能够在评价和推荐过程中基于数据关联分析。

（2） 本文在预测评分过程中综合考虑，用户打分习惯，书籍受评特征，用户和书籍关联三方面共计 13 种指标，较全面地反映了用户与书籍间的关联强度，且对每一个结点根据网络重要性赋予强度值，立足于社交网络联系来评价书籍。
（3） 本文在评分预测模型中采用因子分析法，解决了指标间多重共线性的问题生成了 4 项公共因子，并以其对方差的贡献率为权重加权得到综合指标，较准确地衡量了读者对书籍的评价。
（4） 智能推荐模型中，建立局部网络拓扑结构先便利整个拓扑图对指定用户利用综合指标评价，再优先选择评价较高的书籍作为推荐结果，实现了推荐的智能，客观，全面。

7.2 模型缺点

（1）模型是基于社交网络拓扑结构进行的评价预测和智能推荐，但社交网络拓扑结构较为复杂且很难实现全网的最有评估。
（2）该模型因子分析法的 13 项指标不一定能全面覆盖所有因素，同时，用户对书籍评价的过程也是一个心理学和行为学变化的过程，仅仅用数学算法并不能完全模拟出评分过程。
（3）评分预测中，预测值位于该结点原有评价最低值与最高值之间，但在智能推荐过程中，以评价值作为排序条件回事推荐出现误差。

7.3 模型拓展

7.3.1 最短距离聚类法

对于特定的用户和书籍群体而言，在整个复杂网络中直接进行了评估要面临庞大的运算和大量无效的冗余信息，需要一种机制来简化社交网络。最短距离聚类法可以解决网络冗余这一难题。

最短距离聚类法的主要思想是把要研究目标数据全部分类，把距离研究目标较远的对象集合当成一个对象再建立网络，从而达到降维的作用。

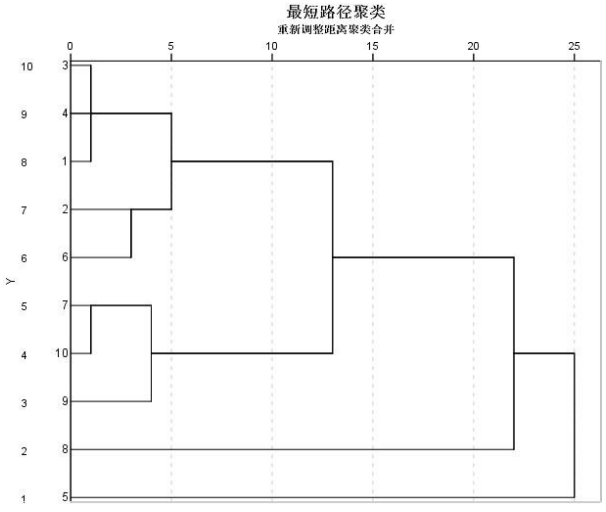


图 5 聚类树图

最短距离聚类法是在原来的 m×m 距离矩阵的非对角元素中找出，把分类对象 Gp 和 Gq 归并为一新类 Gr，然后按计算公式：

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

计算原来各类与新类之间的距离，这样就得到一个新的（m－1）阶的距离矩阵；再从新的距离矩阵中选出最小者d_{ij}，把 Gi 和 Gj 归并成新类；再计算各类与新类的距离，这样一直下去，直至各分类对象被归为一类为止。

综合上述聚类算法，可以根据本题所给信息作出最短距离聚类谱系图，如下：

7.3.2 评价预测模型优化

在上述模型中，我们根据因子分析法能够准确评估到 f=WF。

而$Score(i,j) = \min(Score_j) + f'(i,j)[\max(Score) - \min(Score)]$，由于每个书籍评价存在差异性，对于每个用户来说最高与最低评分的差异会导致推荐时标准的不统一，在计算中会导致 Score 矩阵中元素大小的排序不能完全反应用户与书籍的匹配程度，进而很有可能导致推荐失败。

因此定义$Score(i,j) = \min(Score_j) + f'(i,j)[\max(Score) - \min(Score)]$使得经过调整后的评分数据可用性更高。而评分的正确预测是进行智能推荐的基础。

7.3.3 冷启动推荐拓展

冷启动推荐是指针对系统复杂网络中孤立结点的推荐行为，由于部分新增用户短时间内并未形成社交网络和阅读历史，在进行网络分析的过程中，它会作为一个孤立点而存在，对这些用户进行推荐是无法基于复杂网络模型下的最短路径算法来实现的，而应在全网中统计影响力或者强度最高的书籍作为基础的推荐呈现给用户。

8.参考文献

[1] 丁雪.基于数据挖掘的图书智能推荐系统研究[J].武汉大学，2010

[2] 沈文娟.智能推荐功能选课系统的设计与实现[J].科技广场，2013,11

[3] 王瑞琴.孔繁盛.基于多数据源和联合聚类的智能推荐[J].模式识别与人工智能,第 21 卷第 6 期,2008 年 12 月

[4] 冯旻昱.复杂网络中路径优化问题的研究与应用[J].电子科技大学，2013

[5] 陈根浪.基于社交媒体的推荐技术若干问题研究[D].浙江大学，2012

[6] 姚奇富.基于 Web 访问挖掘的个性化智能推荐服务[J].基金项目,2006

[7] 赵琳等.加权马氏距离判别分析方法及其权值确定——旅游信息服务系统的智能推荐[J].经济数学，第 24 卷第 2 期

[9] 田超等.SuperRank 基于评论分析的智能推荐系统[J].计算机研究与发展,2010

9.附录

附录一：数据整理程序

```
k=1;
for i=1:6
    for j=1:25
     if S(i,j)>0
         M(k,1)=S(i,j);
         k=k+1;
     end
    end
end
for i=1:2434
    for m=1:1168
        if H(i,1)==MR(m,1)
            H(i,1)= MR(m,2);
        end
    end
    for n=1:34
       if H(i,2)==MB(n,1)
            H(i,2)= MR(n,2);
        end
    end
end
for i=1:1100
    for j=1:214
        for m=1:1168
        if S(i,j)==MR(m,1)
            S(i,j)= MR(m,2);
        end
        end
    end
end
for i=1:1100
    for j=1:214
       if S(i,j)>1168
            S(i,j)= 0;
       end
    end
end
for i=1:311
    for m=1:1168
        if SC(i,1)==MR(m,1)
            SC(i,1)= MR(m,2);
```

```
        end
    end
    for n=1:34
       if SC(i,2)==MB(n,1)
            SC(i,2)= MR(n,2);
        end
    end
end
 for i=1:1100
    for j=2:214
       if S(i,j)>0
           U(S(i,1),S(i,j))=1;
       end
    end
 end
 for i=1:2434
     U(H(i,1),H(i,2))=1;
 end
for i=1:1202
for j=1:1202
F(i,j)=Q(i,j)+Q(j,i);
end
end
>> for i=1:1202
for j=1:1202
Q(i,j)=F(i,j);
end
end
%-- 2014/5/27 1:05 --%
for i=1:6
for j=1:40
if D(i,j)>0
R(i,7)=R(i,7)+H1(1,j)/D(i,j);
R(i,8)=R(i,8)+H1(2,j)/D(i,j);
end
end
end
for i=1:34
for j=1:6
if D(i,j)>0
B(i,7)=R(i,7)+H2(j,1)/D(j,i+6);
R(i,8)=R(i,8)+H2(j,2)/D(j,i+6);
end
end
```

```
end
for i=1:34
for j=1:6
if D(j,i+6)>0
B(i,7)=R(i,7)+H2(j,1)/D(j,i+6);
R(i,8)=R(i,8)+H2(j,2)/D(j,i+6);
end
end
end
for i=1:6
for j=1:40
if D(i,j)>0
R(i,7)=R(i,7)+H1(1,j)/D(i,j);
R(i,8)=R(i,8)+H1(2,j)/D(i,j);
end
end
end
for i=1:34
for j=1:6
if D(j,i+6)>0
B(i,7)=B(i,7)+H2(j,1)/D(j,i+6);
B(i,8)=B(i,8)+H2(j,2)/D(j,i+6);
end
end
end
)
size(B)
[a,b]=size(B)
DF=Tianke(x);
[a,b]=size(x);
for i=1:b
c=max(B(:,b));
d=min(B(:,b));
for j=1:a
if c>d
B(a,b)=(B(a,b)-d)/(c-d);
end
end
end
for i=1:36
X(i,5)=D1(X(i,3),X(i,4));
end
for i=1:6
[hk(i,1),hk(i,2)]=found(max(jk(i,:)));
```

```
end
for i=1:6
[hk(i,1),hk(i,2)]=found(jk(i,:)=max(jk(i,:)));
for i=1:6
end
for i=1:6
[hk(i,1),hk(i,2)]=found(jk(i,:)==max(jk(i,:)));
end
for i=1:6
[hk(i,1),hk(i,2)]=find(jk(i,:)==max(jk(i,:)));
end
[c,d]=find(jk(1,:)==max(jk(1,:)));
[c,d]=find(jk(1,:)==min(jk(1,:)));
min(jk(1,:))
[c,d]=find(jk(2,:)==min(jk(2,:)));
min(jk(2,:))
[c,d]=find(jk(3,:)==min(jk(3,:)));
min(jk(3,:))
[c,d]=find(jk(4,:)==min(jk(4,:)));
min(jk(4,:))
[c,d]=find(jk(5,:)==min(jk(5,:)));
min(jk(5,:))
[c,d]=find(jk(6,:)==min(jk(6,:)));
min(jk(6,:));
[c,d]=find(jk(6,:)==min(jk(6,:)));
min(jk(6,:));
[c,d]=find(jk(6,:)==min(jk(6,:)));
function [A] = Tianke( B )
%UNTITLED2 Summary of this function goes here
%   Detailed explanation goes here
[a,b]=size(B);
for i=1:b
    c=max(B(:,b));
    d=min(B(:,b));
    for j=1:a
    if c>d
    B(a,b)=(B(a,b)-d)/(c-d);
    end
    end
end
end
```

附录二：Floyd 算法

```
function [D,path,min1,path1]=floyd(a,start,terminal)
```

```
D=a;n=size(D,1);path=zeros(n,n);
for i=1:n
   for j=1:n
      if D(i,j)~=inf
         path(i,j)=j;
end,  end, end
for k=1:n
   for i=1:n
      for j=1:n
         if D(i,k)+D(k,j)<D(i,j)
            D(i,j)=D(i,k)+D(k,j);
            path(i,j)=path(i,k);
end, end, end,end
if nargin==3
   min1=D(start,terminal);
   m(1)=start;
   i=1;
   path1=[ ];
   while   path(m(i),terminal)~=terminal
      k=i+1;
      m(k)=path(m(i),terminal);
      i=i+1;
   end
   m(i+1)=terminal;
   path1=m;
end
```

附录三：SPSS 分析结果一（聚类）

```
CLUSTER     VAR00001 VAR00002 VAR00003 VAR00004
  /METHOD BAVERAGE
  /MEASURE=SEUCLID
  /PRINT SCHEDULE
  /PLOT DENDROGRAM VICICLE.
```

聚类

附注

创建的输出		27-MAY-2014 15:28:34
注释		
	活动的数据集	数据集 0
	过滤器	<none>
输入	权重	<none>
	拆分文件	<none>
	工作数据文件中的 N 行	10
缺失值处理	对缺失的定义	用户定义的缺失值作为缺失数据对待。

	使用的案例	统计是在所使用的变量不带有缺失值的案例基础上进行的。
语法		CLUSTER VAR00001 VAR00002 VAR00003 VAR00004 /METHOD BAVERAGE /MEASURE=SEUCLID /PRINT SCHEDULE /PLOT DENDROGRAM VICICLE.
资源	处理器时间	00:00:00.37
	已用时间	00:00:00.40

[数据集 0]

案例处理汇总 [a,b]

案例					
有效		缺失		总计	
N	百分比	N	百分比	N	百分比
10	100.0	0	.0	10	100.0

a. 平方 Euclidean 距离 已使用

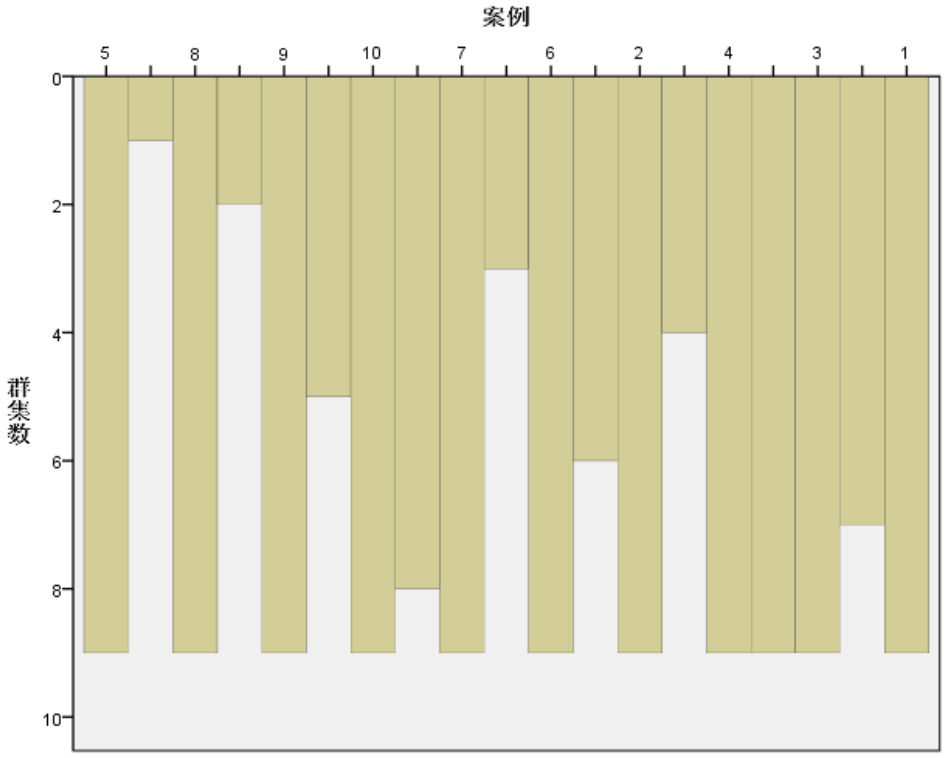
b. 平均联结（组之间）

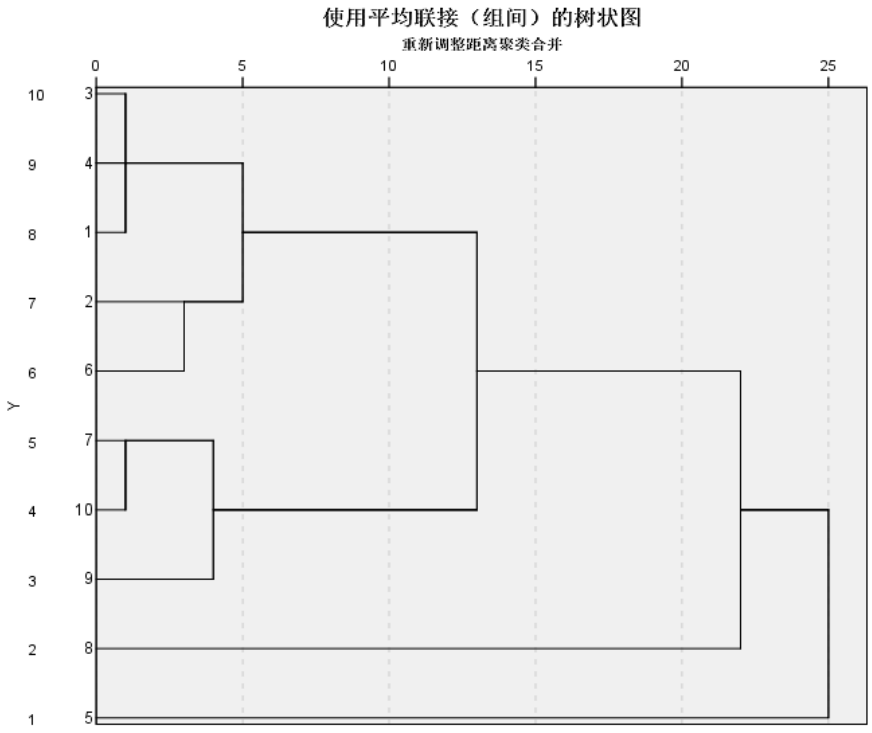
平均联结（组之间）

聚类表

阶	群集组合		系数	首次出现阶群集		下一阶
	群集 1	群集 2		群集 1	群集 2	
1	3	4	.305	0	0	3
2	7	10	.327	0	0	5
3	1	3	.744	0	1	6
4	2	6	1.809	0	0	6
5	7	9	2.364	2	0	7
6	1	2	3.065	3	4	7
7	1	7	7.277	6	5	8
8	1	8	12.710	7	0	9

9	1	5	14.818	8	0	0





附录二：SPSS 分析结果二

FACTOR

/VARIABLES x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13

/MISSING LISTWISE

/ANALYSIS x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13

/PRINT UNIVARIATE INITIAL CORRELATION SIG DET KMO INV REPR AIC EXTRACTION FSCORE

/PLOT EIGEN ROTATION

/CRITERIA MINEIGEN(1) ITERATE(25)

/EXTRACTION PC

/ROTATION NOROTATE

/SAVE REG(ALL)

/METHOD=CORRELATION.

因子分析

附注

创建的输出		27-MAY-2014 03:57:54
注释		
输入	活动的数据集	数据集 0
	过滤器	<none>
	权重	<none>
	拆分文件	<none>
	工作数据文件中的 N 行	36
缺失值处理	对缺失的定义	MISSING=EXCLUDE：用户定义的缺失值作为缺失对待。
	使用的案例	LISTWISE：统计量基于对所使用任何变量都不含缺失值的案例。

语法		FACTOR /VARIABLES x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13 /MISSING LISTWISE /ANALYSIS x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13 /PRINT UNIVARIATE INITIAL CORRELATION SIG DET KMO INV REPR AIC EXTRACTION FSCORE /PLOT EIGEN ROTATION /CRITERIA MINEIGEN(1) ITERATE(25) /EXTRACTION PC /ROTATION NOROTATE /SAVE REG(ALL) /METHOD=CORRELATION.
资源	处理器时间	00:00:00.39
	已用时间	00:00:00.42
	所需的最大内存	23148 (22.605K) 字节
已创建的变量	FAC1_1	成份得分 1
	FAC2_1	成份得分 2
	FAC3_1	成份得分 3
	FAC4_1	成份得分 4

[数据集 0]

描述统计量

	均值	标准差	分析 N
x1	3.7776	.34292	36
x2	.5169	.10085	36
x3	4.0407	.15678	36
x4	.3657	.11113	36
x5	17.0675	2.88195	36
x6	.6552	.32117	36
x7	.5381	.29235	36
x8	.5966	.27149	36
x9	.5161	.24700	36
x10	.2259	.03449	36
x11	.0310	.00752	36
x12	.2433	.04160	36
x13	.0222	.00795	36

相关矩阵[a]

		x1	x2	x3	x4	x5	x6	x7
相关	x1	1.000	.010	-.077	-.162	.434	-.619	-.507
	x2	.010	1.000	-.482	-.102	.048	.075	-.060
	x3	-.077	-.482	1.000	.370	-.128	-.046	.074
	x4	-.162	-.102	.370	1.000	-.198	.024	.104
	x5	.434	.048	-.128	-.198	1.000	-.679	-.696
	x6	-.619	.075	-.046	.024	-.679	1.000	.925
	x7	-.507	-.060	.074	.104	-.696	.925	1.000
	x8	-.079	-.343	.677	.327	-.321	-.094	-.053
	x9	.024	-.441	.823	.453	-.152	-.304	-.184
	x10	.127	-.074	.077	.060	-.816	.426	.544
	x11	-.240	.762	-.301	.015	-.587	.489	.403
	x12	-.400	-.192	.348	.235	-.954	.653	.730
	x13	-.334	-.136	.367	.905	-.567	.328	.427
Sig.（单侧）	x1		.477	.328	.172	.004	.000	.001
	x2	.477		.001	.276	.390	.331	.364
	x3	.328	.001		.013	.229	.394	.334
	x4	.172	.276	.013		.123	.445	.273
	x5	.004	.390	.229	.123		.000	.000
	x6	.000	.331	.394	.445	.000		.000
	x7	.001	.364	.334	.273	.000	.000	
	x8	.323	.020	.000	.026	.028	.294	.380
	x9	.445	.004	.000	.003	.188	.036	.141
	x10	.230	.333	.327	.363	.000	.005	.000
	x11	.079	.000	.037	.466	.000	.001	.007
	x12	.008	.130	.019	.084	.000	.000	.000
	x13	.023	.215	.014	.000	.000	.025	.005

相关矩阵[a]

		x8	x9	x10	x11	x12	x13
相关	x1	-.079	.024	.127	-.240	-.400	-.334
	x2	-.343	-.441	-.074	.762	-.192	-.136
	x3	.677	.823	.077	-.301	.348	.367
	x4	.327	.453	.060	.015	.235	.905
	x5	-.321	-.152	-.816	-.587	-.954	-.567

	x6	-.094	-.304	.426	.489	.653	.328
	x7	-.053	-.184	.544	.403	.730	.427
	x8	1.000	.792	.270	-.037	.446	.371
	x9	.792	1.000	.125	-.254	.304	.421
	x10	.270	.125	1.000	.491	.824	.387
	x11	-.037	-.254	.491	1.000	.467	.233
	x12	.446	.304	.824	.467	1.000	.601
	x13	.371	.421	.387	.233	.601	1.000
	x1	.323	.445	.230	.079	.008	.023
	x2	.020	.004	.333	.000	.130	.215
	x3	.000	.000	.327	.037	.019	.014
	x4	.026	.003	.363	.466	.084	.000
	x5	.028	.188	.000	.000	.000	.000
	x6	.294	.036	.005	.001	.000	.025
Sig.（单侧）	x7	.380	.141	.000	.007	.000	.005
	x8		.000	.056	.416	.003	.013
	x9	.000		.234	.067	.036	.005
	x10	.056	.234		.001	.000	.010
	x11	.416	.067	.001		.002	.086
	x12	.003	.036	.000	.002		.000
	x13	.013	.005	.010	.086	.000	

a. 行列式 = 3.263E-012

相关矩阵的逆矩阵

	x1	x2	x3	x4	x5	x6	x7	x8
x1	151.186	-45.573	-53.373	.608	-36.154	-3.029	23.083	-1.697
x2	-45.573	78.559	2.896	-4.969	15.848	6.643	-13.731	5.454
x3	-53.373	2.896	30.153	-9.251	1.306	-1.725	-6.415	.494
x4	.608	-4.969	-9.251	73.925	5.256	-6.235	6.225	-5.584
x5	-36.154	15.848	1.306	5.256	49.055	11.550	-19.411	-2.684
x6	-3.029	6.643	-1.725	-6.235	11.550	16.926	-14.858	-1.200
x7	23.083	-13.731	-6.415	6.225	-19.411	-14.858	22.432	2.151
x8	-1.697	5.454	.494	-5.584	-2.684	-1.200	2.151	4.616
x9	5.006	-.441	-6.960	.480	2.377	4.365	.811	-2.198
x10	-258.275	80.481	92.832	-8.114	60.782	10.347	-42.197	3.676
x11	57.323	-94.207	-4.022	5.135	-18.674	-8.982	17.874	-6.579
x12	210.675	-11.098	-102.970	48.628	1.001	2.653	13.566	-7.171

x13	-7.608	7.203	13.593	-87.026	-2.596	8.611	-10.286	6.632

相关矩阵的逆矩阵

	x9	x10	x11	x12	x13
x1	5.006	-258.275	57.323	210.675	-7.608
x2	-.441	80.481	-94.207	-11.098	7.203
x3	-6.960	92.832	-4.022	-102.970	13.593
x4	.480	-8.114	5.135	48.628	-87.026
x5	2.377	60.782	-18.674	1.001	-2.596
x6	4.365	10.347	-8.982	2.653	8.611
x7	.811	-42.197	17.874	13.566	-10.286
x8	-2.198	3.676	-6.579	-7.171	6.632
x9	9.363	-7.515	1.042	8.121	-2.039
x10	-7.515	447.824	-101.422	-370.464	21.957
x11	1.042	-101.422	115.144	17.261	-8.063
x12	8.121	-370.464	17.261	434.720	-66.268
x13	-2.039	21.957	-8.063	-66.268	104.916

KMO 和 Bartlett 的检验

取样足够度的 Kaiser-Meyer-Olkin 度量。		.577
Bartlett 的球形度检验	近似卡方	789.969
	df	78
	Sig.	.000

反映像矩阵

		x1	x2	x3	x4	x5	x6
反映像协方差	x1	.007	-.004	-.012	5.443E-005	-.005	-.001
	x2	-.004	.013	.001	-.001	.004	.005
	x3	-.012	.001	.033	-.004	.001	-.003
	x4	5.443E-005	-.001	-.004	.014	.001	-.005
	x5	-.005	.004	.001	.001	.020	.014
	x6	-.001	.005	-.003	-.005	.014	.059
	x7	.007	-.008	-.009	.004	-.018	-.039
	x8	-.002	.015	.004	-.016	-.012	-.015
	x9	.004	-.001	-.025	.001	.005	.028
	x10	-.004	.002	.007	.000	.003	.001
	x11	.003	-.010	-.001	.001	-.003	-.005

	x12	.003	.000	-.008	.002	4.692E-005	.000
	x13	.000	.001	.004	-.011	-.001	.005
	x1	.287[a]	-.418	-.790	.006	-.420	-.060
	x2	-.418	.424[a]	.060	-.065	.255	.182
	x3	-.790	.060	.438[a]	-.196	.034	-.076
	x4	.006	-.065	-.196	.531[a]	.087	-.176
	x5	-.420	.255	.034	.087	.776[a]	.401
	x6	-.060	.182	-.076	-.176	.401	.724[a]
反映像相关	x7	.396	-.327	-.247	.153	-.585	-.762
	x8	-.064	.286	.042	-.302	-.178	-.136
	x9	.133	-.016	-.414	.018	.111	.347
	x10	-.993	.429	.799	-.045	.410	.119
	x11	.434	-.991	-.068	.056	-.248	-.203
	x12	.822	-.060	-.899	.271	.007	.031
	x13	-.060	.079	.242	-.988	-.036	.204

反映像矩阵

		x7	x8	x9	x10	x11	x12	x13
	x1	.007	-.002	.004	-.004	.003	.003	.000
	x2	-.008	.015	-.001	.002	-.010	.000	.001
	x3	-.009	.004	-.025	.007	-.001	-.008	.004
	x4	.004	-.016	.001	.000	.001	.002	-.011
	x5	-.018	-.012	.005	.003	-.003	4.692E-005	-.001
	x6	-.039	-.015	.028	.001	-.005	.000	.005
反映像协方差	x7	.045	.021	.004	-.004	.007	.001	-.004
	x8	.021	.217	-.051	.002	-.012	-.004	.014
	x9	.004	-.051	.107	-.002	.001	.002	-.002
	x10	-.004	.002	-.002	.002	-.002	-.002	.000
	x11	.007	-.012	.001	-.002	.009	.000	-.001
	x12	.001	-.004	.002	-.002	.000	.002	-.001
	x13	-.004	.014	-.002	.000	-.001	-.001	.010
	x1	.396[a]	-.064	.133	-.993	.434	.822	-.060
	x2	-.327	.286[a]	-.016	.429	-.991	-.060	.079
	x3	-.247	.042	-.414[a]	.799	-.068	-.899	.242
反映像相关	x4	.153	-.302	.018	-.045[a]	.056	.271	-.988
	x5	-.585	-.178	.111	.410	-.248[a]	.007	-.036
	x6	-.762	-.136	.347	.119	-.203	.031[a]	.204
	x7	.627	.211	.056	-.421	.352	.137	-.212[a]

x8	.211	.757	-.334	.081	-.285	-.160	.301
x9	.056	-.334	.823	-.116	.032	.127	-.065
x10	-.421	.081	-.116	.429	-.447	-.840	.101
x11	.352	-.285	.032	-.447	.548	.077	-.073
x12	.137	-.160	.127	-.840	.077	.609	-.310
x13	-.212	.301	-.065	.101	-.073	-.310	.658

a. 取样足够度度量 (MSA)

公因子方差

	初始	提取
x1	1.000	.739
x2	1.000	.877
x3	1.000	.753
x4	1.000	.872
x5	1.000	.931
x6	1.000	.926
x7	1.000	.888
x8	1.000	.729
x9	1.000	.893
x10	1.000	.924
x11	1.000	.929
x12	1.000	.984
x13	1.000	.895

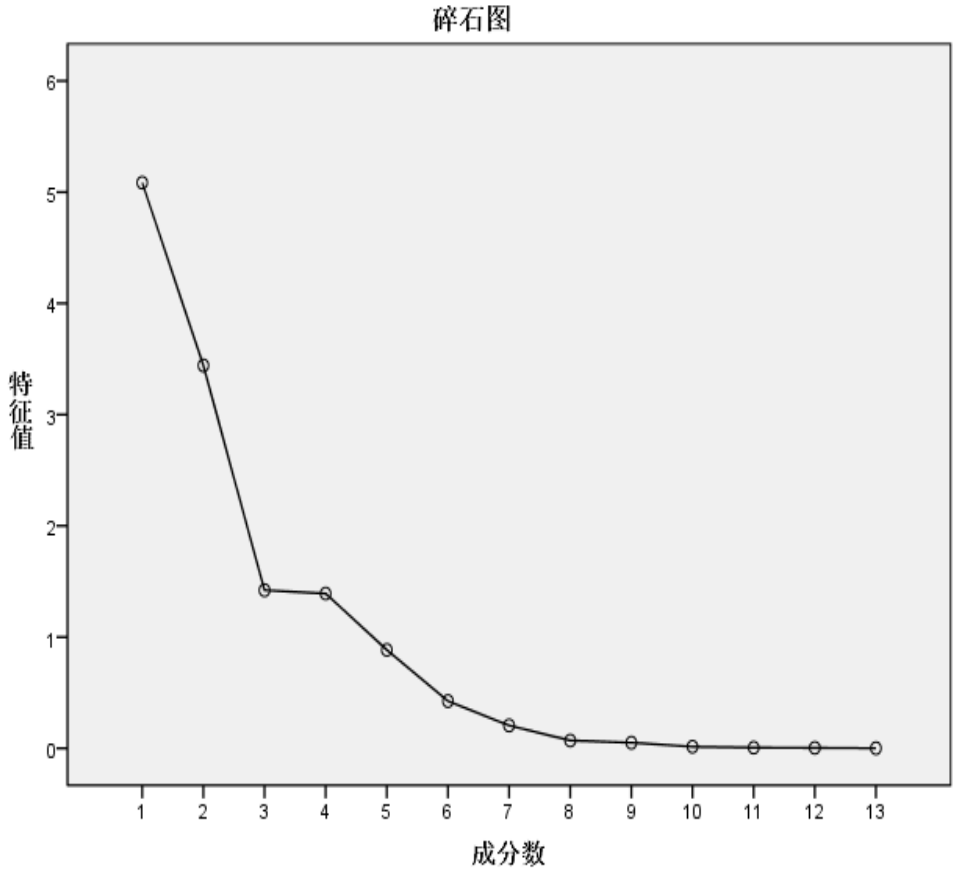
提取方法：主成份分析。

解释的总方差

成份	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
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			

提取方法：主成份分析。



成份矩阵 a

	成份			
	1	2	3	4
x1	-.496	.186	.451	-.505
x2	-.123	-.637	.623	.261
x3	.356	.781	-.121	-.036
x4	.425	.454	.279	.638
x5	-.928	.176	-.056	.191

x6	.718	-.517	-.368	.087
x7	.778	-.388	-.364	.007
x8	.434	.690	.185	-.175
x9	.299	.879	.165	-.057
x10	.726	-.128	.261	-.559
x11	.495	-.621	.546	.038
x12	.962	-.002	-.018	-.240
x13	.747	.294	.199	.459

提取方法 :主成份。[a]

a. 已提取了 4 个成份。

再生相关性

		x1	x2	x3	x4	x5	x6	x7
再生的相关性	x1	.739[a]	.092	-.067	-.323	.371	-.662	-.627
	x2	.092	.877[a]	-.626	-.002	.017	.035	-.074
	x3	-.067	-.626	.753[a]	.450	-.192	-.107	.018
	x4	-.323	-.002	.450	.872[a]	-.208	.024	.058
	x5	.371	.017	-.192	-.208	.931[a]	-.720	-.768
	x6	-.662	.035	-.107	.024	-.720	.926[a]	.894
	x7	-.627	-.074	.018	.058	-.768	.894	.888[a]
	x8	.085	-.423	.677	.438	-.325	-.128	.001
	x9	.118	-.509	.775	.536	-.143	-.305	-.168
	x10	.016	.009	.147	-.033	-.817	.443	.515
	x11	-.134	.684	-.376	.105	-.592	.479	.427
	x12	-.365	-.190	.351	.250	-.938	.678	.754
	x13	-.458	-.035	.455	.800	-.565	.352	.399
残差[b]	x1		-.082	-.009	.160	.063	.044	.119
	x2	-.082		.144	-.101	.032	.041	.013
	x3	-.009	.144		-.079	.065	.060	.056
	x4	.160	-.101	-.079		.010	.000	.046
	x5	.063	.032	.065	.010		.041	.072
	x6	.044	.041	.060	.000	.041		.031
	x7	.119	.013	.056	.046	.072	.031	
	x8	-.165	.080	-.001	-.110	.004	.035	-.054
	x9	-.094	.068	.047	-.083	-.009	.001	-.016
	x10	.112	-.084	-.069	.094	.002	-.017	.029
	x11	-.106	.078	.075	-.090	.005	.010	-.024

	x12	-.035	-.002	-.003	-.016	-.016	-.026	-.024
	x13	.125	-.100	-.088	.105	-.002	-.024	.028

再生相关性

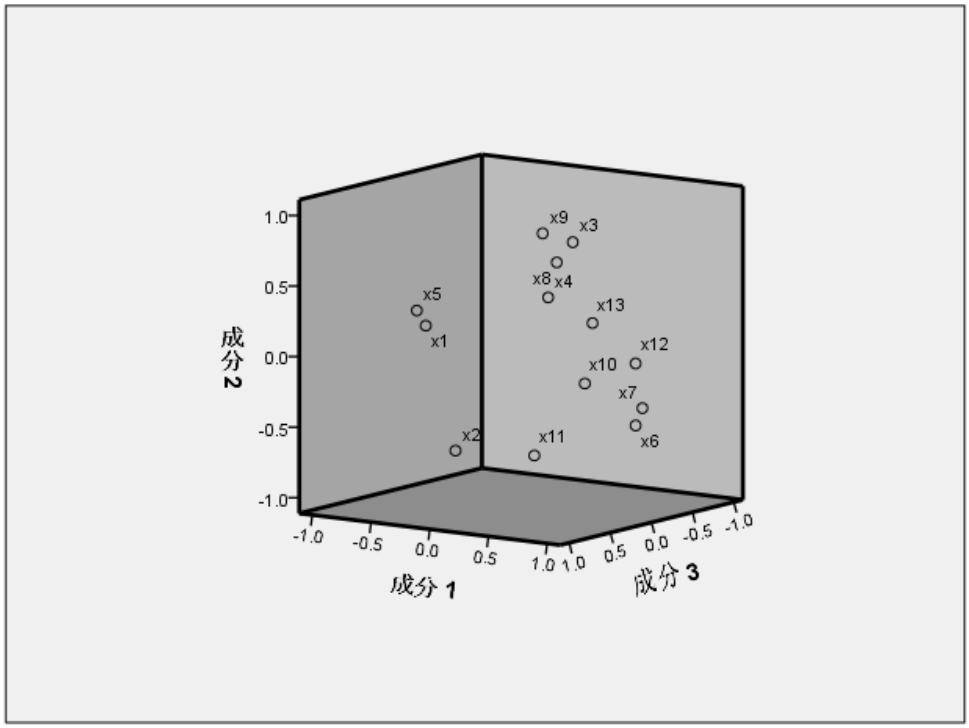
		x8	x9	x10	x11	x12	x13
再生的相关性	x1	.085[a]	.118	.016	-.134	-.365	-.458
	x2	-.423	-.509[a]	.009	.684	-.190	-.035
	x3	.677	.775	.147[a]	-.376	.351	.455
	x4	.438	.536	-.033	.105[a]	.250	.800
	x5	-.325	-.143	-.817	-.592	-.938[a]	-.565
	x6	-.128	-.305	.443	.479	.678	.352[a]
	x7	.001	-.168	.515	.427	.754	.399
	x8	.729	.777	.373	-.119	.454	.483
	x9	.777	.893	.179	-.310	.297	.488
	x10	.373	.179	.924	.560	.828	.300
	x11	-.119	-.310	.560	.929	.459	.314
	x12	.454	.297	.828	.459	.984	.605
	x13	.483	.488	.300	.314	.605	.895
残差[b]	x1	-.165	-.094	.112	-.106	-.035	.125
	x2	.080	.068	-.084	.078	-.002	-.100
	x3	-.001	.047	-.069	.075	-.003	-.088
	x4	-.110	-.083	.094	-.090	-.016	.105
	x5	.004	-.009	.002	.005	-.016	-.002
	x6	.035	.001	-.017	.010	-.026	-.024
	x7	-.054	-.016	.029	-.024	-.024	.028
	x8		.015	-.103	.082	-.009	-.112
	x9	.015		-.054	.055	.008	-.068
	x10	-.103	-.054		-.069	-.004	.087
	x11	.082	.055	-.069		.008	-.081
	x12	-.009	.008	-.004	.008		-.004
	x13	-.112	-.068	.087	-.081	-.004	

提取方法：主成份分析。

a. 重新生成的公因子方差

b. 将计算观察到的相关性和重新生成的相关性之间的残差。有 39 (50.0%) 个绝对值大于 0.05 的非冗余残差。

成分图



成份得分系数矩阵

	成份			
	1	2	3	4
x1	-.098	.054	.317	-.363
x2	-.024	-.185	.438	.187
x3	.070	.227	-.085	-.026
x4	.084	.132	.196	.459
x5	-.182	.051	-.039	.137
x6	.141	-.150	-.259	.062
x7	.153	-.113	-.256	.005
x8	.085	.200	.130	-.126
x9	.059	.255	.116	-.041
x10	.143	-.037	.183	-.402
x11	.097	-.180	.384	.027
x12	.189	-.001	-.013	-.172
x13	.147	.085	.140	.330

提取方法 :主成份。

构成得分。

成份得分协方差矩阵

成份	1	2	3	4
1	1.000	.000	.000	.000
2	.000	1.000	.000	.000
3	.000	.000	1.000	.000
4	.000	.000	.000	1.000

提取方法 :主成份。

构成得分。

附录五：结果

用户 ID	书籍 ID	f1	f2	f3	f4	f	f 修正后	评价得分
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

用户 ID		2515537	4156658	5997834	7245481	7625225	9214078
推荐一	中间结点	3379860	3379860	1136254	1548351	3379860	3780484
	书籍 ID	698573	698573	794171	702699	698573	776002
	最短路径	16.25	15.88	23.45	24.56	16.25	27.25
推荐二	中间结点	3379860	3379860	5913342	1548351	3379860	3780484
	书籍 ID	516012	516012	284550	962729	516012	551643
	最短路径	18.75	18.38	34.32	29.13	18.75	32.25
推荐三	中间结点	3379860	3379860	8981825	1548351	3379860	3780484
	书籍 ID	709644	709644	284550	642256	709644	510372
	最短路径	18.75	18.38	36.74	31.78	18.75	32.25