

## 融合朋友关系和标签信息的张量分解推荐算法

丁小焕\*, 彭甫榕, 王 琼, 陆建峰

(南京理工大学 计算机科学与工程学院, 南京 210094)

(\*通信作者电子邮箱 xiaohuan849@sina.com)

**摘 要:**针对大众标注网站项目推荐系统中存在数据矩阵稀疏性影响推荐效果的问题,考虑矩阵奇异值分解(SVD)能有效地平滑数据矩阵中的数据,以及朋友圈能够反映出一个人的兴趣爱好,提出了一种融合朋友关系和标签信息的张量分解推荐算法。首先,利用高阶奇异值分解(HOSVD)方法对用户-项目-标签三元组信息进行潜在语义分析和多路降维,分析用户、项目、标签三者间关系;然后,再结合用户朋友关系、朋友间相似度,修正张量分解结果,建立三阶张量模型,从而实现推荐。该模型方法在两个真实数据集上进行了实验,结果表明,所提算法与高阶奇异值分解的方法比较,在推荐的召回率和精确度指标上分别提高了2.5%和4%,因此,所提算法进一步验证了结合朋友关系能够提高推荐的准确率,并扩展了张量分解模型,实现用户个性化推荐。

**关键词:**张量分解;高阶奇异值分解;朋友关系;标签;推荐

**中图分类号:** TP311; TP391.1 **文献标志码:** A

### Tensor factorization recommendation algorithm combined with social network and tag information

DING Xiaohuan\*, PENG Furong, WANG Qiong, LU Jianfeng

(School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing Jiangsu 210094, China)

**Abstract:** The item recommendation precision of social tagging recommendation system was affected by sparse data matrix. A tensor factorization recommendation algorithm combined with social network and tag information was proposed, in consideration of that Singular Value Decomposition (SVD) had good processing properties to deal with sparse matrix, and that friends' information could reflect personal interests and hobbies. Firstly, Higher-Order Singular Value Decomposition (HOSVD) was used for latent semantic analysis and multi-dimensional reduction. The user-project-tag triple information could be analyzed by HOSVD, to get the relationships among them. Then, by combining the relationship of users and friends with the similarity between friends, the result of tensor factorization was modified and the third-order tensor model was set up to realize the item recommendation. Finally, the experiment was conducted on two real data sets. The experimental results show that the proposed algorithm can improve respectively recall and precision by 2.5% and 4%, compared with the HOSVD method. Therefore, it is further verified that the algorithm combining with the relation of friends can enhance the accuracy of recommendation. What's more, the tensor decomposition model is expanded to realize the user personalized recommendation.

**Key words:** tensor factorization; Higher-Order Singular Value Decomposition (HOSVD); social network; tag; recommendation

## 0 引言

随着 Web 2.0 技术的发展,允许用户自由地对系统项目资源赋予个性化标签。这些标签信息间接地反映了用户的兴趣,成为推荐系统中一种可靠有用的信息。传统的推荐算法如协同过滤推荐算法<sup>[1-2]</sup>,主要是基于二维的用户-项目评分矩阵实现推荐,在利用用户-项目-标签三元组信息时,主要解决的思路是将三维关系转换为二维关系。如在文献[3]中,提出利用标签信息扩充用户-项目二维矩阵的协同过滤推荐模型;文献[4]中提出将三元组信息拆分为三个二维矩阵(用户-项目、项目-标签、用户-标签)一起矩阵分解的推荐算法。但是拆分三维矩阵的方法存在丢失三者整体间的相互关系的缺点。近年来,有学者提出将三元组信息整体考虑的推荐模

型,文献[5-6]利用高阶奇异值分解(Higher-Order Singular Value Decomposition, HOSVD)方法,建立张量分解模型,获得用户、标签与项目三者潜在关系,实现推荐。文献[7]以传统的 Tucker 和 ParaFac 张量分解模型为基础,进行社会标签系统中的标签推荐预测。本文选择高阶奇异值分解(HOSVD)<sup>[8]</sup>的方法处理三维数据。因为,在推荐系统中,数据的稀疏性是一直都存在的一个难点,矩阵奇异值分解(Singular Value Decomposition, SVD)<sup>[9]</sup>对数据矩阵分解后将无用数据去除,达到降噪和降低稀疏性的目的,对二维矩阵稀疏问题具有良好处理的特性<sup>[5]</sup>。

通过不同的算法优化,虽然稀疏问题能够得到一定的缓解,但是无法从根源上解决。例如:当用户历史信息太少(甚至可能没有)而不能学习其兴趣爱好时,这时是没有办

收稿日期:2015-01-20;修回日期:2015-03-26。

基金项目:江苏省“六大人才高峰”高层次人才项目;江苏省研究生科研创新计划项目(KYLX0382)。

作者简介:丁小焕(1991-),女,浙江衢州人,硕士研究生,主要研究方向:数据挖掘; 彭甫榕(1987-),男,贵州遵义人,博士研究生,主要研究方向:数据挖掘、推荐系统; 王琼(1981-),女,江苏南京人,副教授,博士,主要研究方向:模式识别、智能系统; 陆建峰(1969-),男,江苏淮安人,教授,博士,主要研究方向:智能系统、数据挖掘。

法进行个性化推荐的,因此,寻找一个比较有意义的辅助信息成为解决稀疏性问题的另一个思路。随着社交网络发展,朋友关系逐渐被人们重视,并被引入到推荐系统中。正所谓“物以类聚,人以群分”,朋友之间的兴趣爱好存在着一定的相似性。在引入朋友关系以后,一个客户即使没有任何的历史信息,也可以根据他朋友的喜好为其推荐商品,因此,基于朋友之间的兴趣爱好存在相似性的假设,引入朋友关系到推荐系统中,可以解决推荐系统中存在着稀疏性问题。文献[10]中,假设网络中的朋友关系会影响用户在网络中的行为,将评分矩阵和朋友关系矩阵联合在一起进行概率矩阵分解(Probabilistic Matrix Factorization, PMF),计算用户和物品的特征向量,该方法比单纯依靠评分矩阵的 PMF 精确度提高了 10%,并一定程度上缓解用户冷启动的问题。文献[11]提出基于移动用户社交化关系挖掘,将用户的好友数据加入到协同过滤算法中,提高了用户偏好预测和推荐结果的准确度。本文认为,朋友关系信息是项目推荐系统中另一个重要的信息,不仅能够解决数据的稀疏性,还能预测用户偏好,提高推荐的准确率。

因此,本文结合朋友关系和标签信息研究张量分解推荐算法,利用三元组信息,计算朋友间的相似度,结合朋友关系修正张量分解的结果,建立推荐模型。

## 1 基于标签信息的张量分解模型

### 1.1 高阶奇异值分解(HOSVD)算法

张量即一个多维矩阵,用  $\mathbf{A} \in \mathbf{R}^{I_1 \times I_2 \times \dots \times I_N}$  来表示一个  $N$  阶张量,其中  $a_{i_1 i_2 \dots i_N}$  表示张量中的元素。本文将用户、项目、标签的三元组数据表示成一个三阶张量  $\mathbf{A}_{uii}$ 。接着,定义三阶张量  $\mathbf{A}$  的张量展开,即将张量  $\mathbf{A}$  沿第  $n$  模式展开成矩阵的操作:

$$\begin{aligned} \mathbf{A}_1 &\in \mathbf{R}^{I_1 \times I_2 I_3} \\ \mathbf{A}_2 &\in \mathbf{R}^{I_2 \times I_1 I_3} \\ \mathbf{A}_3 &\in \mathbf{R}^{I_3 \times I_1 I_2} \end{aligned}$$

其中  $\mathbf{A}_1, \mathbf{A}_2$  和  $\mathbf{A}_3$  叫作张量  $\mathbf{A}$  的 1-模(1-mode)、2-模(2-mode)和 3-模(3-mode)的矩阵展开。

令  $\mathbf{U} \in \mathbf{R}^{I_1 \times I_2 \times \dots \times I_N}$ , 则张量  $\mathbf{A}$  与矩阵  $\mathbf{U}$  的第  $n$  模式的 Tucker 积,可得到一个  $(I_1 \times I_2 \times \dots \times I_{n-1} \times J_n \times I_{n+1} \times \dots \times I_N)$  阶张量:

$$(\mathbf{A} \times_n \mathbf{U})_{i_1 i_2 \dots i_{n-1} i_{n+1} \dots i_N} = \sum_{j_n} a_{i_1 i_2 \dots i_{n-1} j_n i_{n+1} \dots i_N} U_{j_n i_n} \quad (1)$$

高阶奇异值分解(HOSVD)[8]是在矩阵奇异值分解(SVD)的概念上延伸出来的。在 SVD[9]方法中,对二阶矩阵  $\mathbf{F} \in \mathbf{R}^{I_1 \times I_2}$  可以分解成如下形式:

$$\mathbf{F} = \mathbf{s} \times_1 \mathbf{u}^{(1)} \times_2 \mathbf{u}^{(2)} = \mathbf{u}^{(1)} \times \mathbf{s} \times (\mathbf{u}^{(2)})^T \quad (2)$$

其中:  $\mathbf{u}^{(1)} = (\mathbf{u}_1^{(1)} \mathbf{u}_2^{(1)} \dots \mathbf{u}_{I_1}^{(1)})$  是一个  $(I_1 \times I_1)$  的西矩阵; 同样  $\mathbf{u}^{(2)} = (\mathbf{u}_1^{(2)} \mathbf{u}_2^{(2)} \dots \mathbf{u}_{I_2}^{(2)})$  是一个  $(I_2 \times I_2)$  的西矩阵,  $\mathbf{s}$  是一个  $(I_1 \times I_2)$  的对角矩阵,并且

$$\begin{aligned} 1) \mathbf{s} &= \text{diag}(\partial_1, \partial_2, \dots, \partial_{\min(I_1, I_2)}) \\ 2) \partial_1 &\geq \partial_2 \geq \dots \geq \partial_{\min(I_1, I_2)} \geq 0 \end{aligned}$$

将 SVD 扩展到 HOSVD 的三阶张量  $\mathbf{A}$  时:

$$\mathbf{A}' = \mathbf{S} \times_1 \mathbf{u}^{(1)} \times_2 \mathbf{u}^{(2)} \times_3 \mathbf{u}^{(3)} \quad (3)$$

其中:  $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}$  和  $\mathbf{u}^{(3)}$  是展开矩阵  $\mathbf{A}_1, \mathbf{A}_2$  和  $\mathbf{A}_3$  在列空间向量上的正交化处理, 1-模、2-模和 3-模的奇异值向量;  $\mathbf{S}$  是具有完整正交性的核心张量,类似于 SVD 中的奇异值矩阵,控制着各个子空间矩阵间的相互作用,但不是对角结构;  $\mathbf{A}'$  为张量  $\mathbf{A}$  的重新构造,将会包含新的数据信息。

### 1.2 张量分解模型

本文将三元组信息(用户、项目和标签)初始化处理构成

一个三阶张量  $\mathbf{A}_{uii} \in \mathbf{R}^{U \times T \times I}$ , 其中:  $U$  为用户的总数,  $I$  为商品的总数,  $T$  为标签的总数,  $a_{uii}$  是张量  $\mathbf{A}_{uii}$  的元素。

$$a_{uii} = \begin{cases} 0, & \text{当用户 } u \text{ 对项目 } i \text{ 标注了标签 } t \\ 1, & \text{其他} \end{cases} \quad (4)$$

基于用户-项目-标签三元组信息的 HOSVD 算法具体过程如下所示。

输入:  $\mathbf{A}_{uii}$ 。

输出: 重构后的  $\mathbf{A}'_{uii}$ , 将包含有新的数据信息。

算法步骤如下。

步骤 1 将  $\mathbf{A}_{uii}$  作张量展开, 得到张量  $\mathbf{A}_{uii}$  的 1-模、2-模和 3-模的张量展开的矩阵形式, 分别为  $\mathbf{A}_1, \mathbf{A}_2$  和  $\mathbf{A}_3$ 。

步骤 2 对展开的矩阵  $\mathbf{A}_1, \mathbf{A}_2$  和  $\mathbf{A}_3$  作 SVD:

$$\mathbf{A}_n = \mathbf{U}^{(n)} \cdot \mathbf{s}^{(n)} \cdot (\mathbf{V}^{(n)})^T; \quad 1 \leq n \leq 3 \quad (5)$$

其中:  $\mathbf{U}^{(n)}$  为  $\mathbf{A}_n$  的左奇异矩阵,  $\mathbf{V}^{(n)}$  为  $\mathbf{A}_n$  的右奇异矩阵。

步骤 3 低阶近似处理  $\mathbf{U}^{(n)}$  [6], 得到新的  $\mathbf{U}^{(n)}$ , 记作  $\mathbf{u}^{(n)}$ 。

步骤 4 计算核心张量  $\mathbf{S}$ :

$$\mathbf{S} = \mathbf{A}_{uii} \times_1 (\mathbf{u}^{(1)})^T \times_2 (\mathbf{u}^{(2)})^T \times_3 (\mathbf{u}^{(3)})^T \quad (6)$$

步骤 5 计算重构张量  $\mathbf{A}'_{uii}$ :

$$\mathbf{A}'_{uii} = \mathbf{S} \times_1 \mathbf{u}^{(1)} \times_2 \mathbf{u}^{(2)} \times_3 \mathbf{u}^{(3)} \quad (7)$$

低阶近似处理  $\mathbf{U}^{(n)}$ :

在矩阵降维过程中, 低阶近似处理是为了过滤掉一些小的奇异值, 而这些小的奇异值往往是噪声, 这个过程有利于提高计算速度和精度。定义参数  $c_n$  ( $1 \leq n \leq 3$ ) 为  $n$ -模展开矩阵  $\mathbf{A}_n$  降维后的维度数, 则低阶近似处理后:

$$\mathbf{A}_n = \mathbf{u}_{i_n \times c_n}^{(n)} \cdot \mathbf{s}_{c_n \times c_n}^{(n)} \cdot (\mathbf{v}_{c_n \times j_n}^{(n)})^T \quad (8)$$

其中:  $\mathbf{A}_n$  为一个  $(I_n \times J_n)$  的矩阵,  $\mathbf{u}_{i_n \times c_n}^{(n)}$  为步骤 2 得到的  $\mathbf{U}^{(n)}$  的前  $c_n$  列矩阵,  $\mathbf{v}_{c_n \times j_n}^{(n)}$  为步骤 2 得到的  $\mathbf{V}^{(n)}$  的前  $c_n$  行矩阵,  $c_n$  的值为:

$$\sum_{i=0}^{c_n} s_{ii}^{(n)} = \theta \times \sum_{i=0}^{\min(I_n, J_n)} s_{ii}^{(n)} \quad (9)$$

其中:  $s^{(n)}$  为步骤 2 得到的结果,  $\theta$  一般的取值为 70%。

综上过程, 得到了 HOSVD 处理后的重构的  $\mathbf{A}'_{uii}$ , 以此反映用户、项目和标签相互间的关联, 其中每个元素可以表示为四元组  $\{u, t, i, p\}$ ,  $p$  为用户  $u$  用标签  $t$  标注项目  $i$  的概率值。

## 2 融合朋友关系和标签信息的推荐算法

在现实生活中, 熟人推荐是促成用户产生消费行为的一种重要形式, 同时社交网络中的朋友关系是现实中人际关系的真实反映, 因此推荐系统作出项目的决策时, 不仅要考虑用户个人的兴趣爱好, 同时还需要考虑用户的朋友们的喜好, 特别是当用户  $u$  同时对多个项目都感兴趣, 而他的朋友们也喜欢其中某个项目  $i$ , 那么系统应该将项目  $i$  首先推荐给用户  $u$ 。本文将把朋友关系结合到 HOSVD 方法中去, 利用朋友关系修正第 1 章中得到的概率  $p$  值, 建立推荐模型, 提高推荐的精度, 同时考虑到每个朋友对用户影响程度不同, 提出再次利用用户-标签-项目三元组信息计算用户与朋友间的相似度, 作为朋友对该用户影响率。

### 2.1 朋友间相似度的计算

用户间成为朋友其中大部分原因是因为两者有着共同的兴趣与爱好, 因此本文算法计算朋友间相似度来衡量两者的共性, 认为当相似度值越大时, 该朋友对用户的影响也就越大。

用户对项目标注标签的过程中, 无论是项目还是标签都

反映了用户的爱好。基于用户-项目-标签三元组信息来计算朋友间的相似度,避免了传统单一地使用用户项目矩阵计算用户相似度。用户选择的项目直接反映了用户的兴趣,建立“用户-项目”二维 0-1 矩阵  $M^i$ , 计算朋友间的余弦相似度:

$$Sim_i(u, f) = \frac{M_u \cdot M_f}{\|M_u\| \|M_f\|} \quad (10)$$

其中用户  $u$  与  $f$  为朋友关系。

另外,用户对项目的标注的标签信息,标签使用的次数在某种程度上也间接地反映了用户的偏好程度。建立“用户-标签”矩阵  $M^t$ , 其中矩阵中每个元素为:

$$M_{ut}^t = n_{ut}/N_u \quad (11)$$

其中:  $n_{ut}$  为用户  $u$  使用标签  $t$  的次数,  $N_u$  为用户  $u$  标注标签的总次数。同样,根据式(10)可以得到朋友间的余弦相似度  $Sim_t(u, f)$ 。

然后设置权重  $\beta$ , 线性组合上面的得到的“用户-项目”相似度和“用户-标签”相似度:

$$Sim(u, f) = \beta \cdot Sim_i(u, f) + (1 - \beta) \cdot Sim_t(u, f) \quad (12)$$

## 2.2 融合朋友信息

第1章中,得到了重构张量  $A'_{uit}$ , 反映用户、项目和标签相互间关联关系,其中拥有原训练集没有的新数据,得到用户  $u$  用标签  $t$  标注新项目  $i$  的概率值记作  $P_{fi}$ , 在本节中需要修正这些概率值。因为在现实情况下,每个用户有各自的兴趣爱好,同时,其朋友的喜好也会在不同程度上影响用户的选择,而影响程度的大小值,本文以用户与朋友间的相似度来衡量,定义一个更加实际的推荐模型,来反映这两个因素的共同影响:

$$P(i_{ut}) = \alpha \cdot P_{uit} + (1 - \alpha) \cdot \left( \sum_{f=0}^{F_u} P_{fi} \cdot Sim(u, f) \right) / F_u \quad (13)$$

其中:  $P(i_{ut})$  表示用户  $u$  在标签  $t$  下,选择项目  $i$  的概率值;  $F_u$  为用户  $u$  的朋友中对商品  $i$  在标签  $t$  下也感兴趣的朋友个数;  $f$  为用户  $u$  的朋友,并用  $\alpha$  来平衡用户的兴趣和其朋友的喜好。

最后,得到修正后的用户  $u$  在标签  $t$  下对项目  $i$  的喜好程度  $P(i_{ut})$ 。

## 2.3 融合朋友关系和标签信息实现推荐

在多数情况下,用户的兴趣是多方面的,例如在作电影推荐时,某个用户可能喜欢看的主要有警匪片、科幻片、恐怖片,如果针对具体的标签推荐项目,将更符合用户的兴趣,以此可以提高推荐系统的可信度。本文提出的模型中,在具体标签  $t$  下给用户推荐项目,实现了相应更加准确的个性化推荐。在推荐过程中本文选择前  $N$  个概率值高的项目推荐给用户(top- $N$  方法)。具体算法步骤如下。

输入:  $A_{uit}$  朋友关系矩阵  $F$  以及  $N$  的值。

输出: 推荐结果。

步骤1 用 HOSVD 算法处理原三阶张量  $A_{uit}$ , 获得重构后张量  $A'_{uit}$ 。

步骤2 再次利用  $A_{uit}$  信息,以及朋友关系矩阵  $F$ , 采用式(12)计算朋友间相似度,得到朋友相似度矩阵。

步骤3 利用式(13),结合朋友相似度矩阵,修正重构的  $A'_{uit}$ , 得到用户  $u$  在标签  $t$  下对项目  $i$  的喜好程度  $P(i_{ut})$ 。

步骤4 在标签  $t$  下,为用户  $i$  在推荐前  $N$  个  $P(i_{ut})$  较大的项目  $i$ 。

## 3 实验分析

### 3.1 测试数据集

本文选用 Last.fm 和 Delicious<sup>[17]</sup> 两个真实数据集对本文提出的推荐算法进行实验分析,各项统计数据如表1所示(其中“朋友数”列为每个用户平均拥有的朋友数)。

表1 实验所用数据集的各项统计数据

数据集	用户数	标签数	项目数	朋友数
Last.fm	87	81	221	2.5
Delicious	350	138	274	2.1

Last.fm 来自 Last.fm 网站的网页数据,网站为用户提供多媒体服务,同时允许用户相互间加朋友和为多媒体服务添加标签。Delicious 来自 Delicious.com 在线网页书签网站,用户可以对各种网页链接进行标注,用户间可以相互交流。本实验将数据集划分为训练集和测试集,其中 75% 作为训练集,25% 作为测试集。

### 3.2 评价准则

本文采用精确度(Precision)、召回率(Recall)来测量推荐算法的准确性和有效性,其定义如下:

$Precision = \text{命中的数目} / \text{推荐的数目}$

$Recall = \text{命中的数目} / \text{测试集的大小}$

精确度又称精度或查准率,是检索到的项目中相关项目所占的比例,当精确度越大,表明算法的推荐结果越准确。召回率又称查全率,表明检索到的相关项目占全部相关项目的比例,其值越高,算法的推荐效果越好。

### 3.3 实验结果分析

#### 3.3.1 朋友关系对推荐算法性能的影响

本文提出结合朋友关系实现推荐的方法——HOSVDF (Higher-Order Singular Value Decomposition with Friends), 在 HOSVD 张量分解推荐算法上增加了朋友信息,考虑到朋友对用户的兴趣爱好的影响,更加贴合于实际。实验将与直接使用 HOSVD 方法<sup>[6]</sup>进行比较,其中参数  $\alpha$  与  $\beta$  取到最佳值(具体取值参照 3.3.2 节)。实验分别在两个数据集上测试评估,按照 top- $N$  值分别取 1, 2, 3, 4, 5, 6, 7, 8, 9 和 10 进行计算分析,实验结果如图1所示。

观察图1(a)与图1(b),本文提出的算法 HOSVDF 的精确度和召回率都高于没有融合朋友关系的 HOSVD 方法。其中比较两者最好的结果, HOSVDF 算法在推荐的召回率和精确度指标上分别提高了 2.5% 和 4%。

精确度越高,说明推荐算法其推荐的结果中符合用户兴趣爱好的项目越多,推荐越准确;召回率越高,说明推荐结果中符合用户喜好的项目占全部用户所喜欢的项目的比例越高,推荐的效果更好。因此,实验结果表明本文算法 HOSVDF 的推荐更加准确有效。实验中发现在 Delicious 数据库中的实验效果差一点,主要因为是该数据库的朋友关系比较稀疏。其次,该数据库的数据比 Last.fm 数据库的数据要稀疏,朋友间的相似性关系较弱,因此当朋友关系越密集,本文的方法越有效。

本文提出的 HOSVD 算法验证了朋友关系可以辅助提升张量分解推荐算法的准确率,提高推荐质量。

#### 3.3.2 算法参数的选择

在本文模型中,有两个重要的参数,分别为相似度关系权重  $\beta$  和朋友影响权重  $\alpha$ , 其对本文的方法 HOSVDF 的结果有何影响,分别作了以下两个实验。



### 1) 相似度关系权重 $\beta$ 。

在本文的方法中,线性组合了“用户-项目”相似度和“用户-标签”相似度。其中,固定  $\alpha$  的值,将权重  $\beta$  设为 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 分别进行实验。

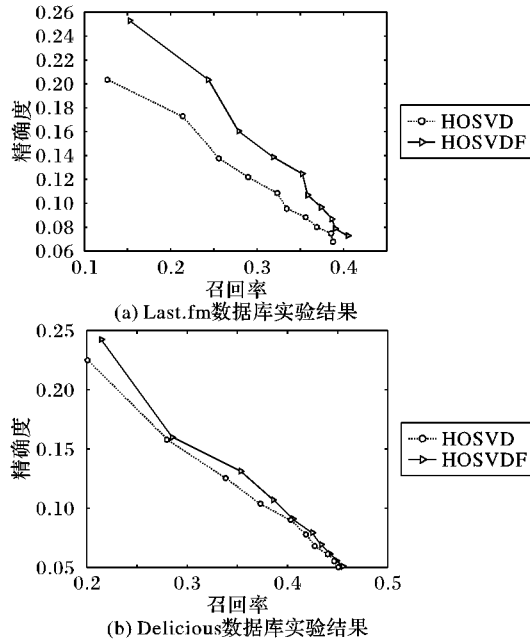


图1 HOSVD与HOSVDF算法性能比较

为了显示清楚,图2只给出了两个数据集中最好的结果与附近取值的比较实验结果,其中横轴是召回率,纵轴是精确度。从图2可以看到,不同的权重值  $\beta$  对应不同的推荐效果,其中,在推荐效果最佳时,Last.fm数据库中,  $\beta$  取 0.2,说明该数据库中用户标注的标签更能反映用户的爱好;相应地,在Delicious中  $\beta$  值为 0.6,用户消费的项目更能体现用户爱好。

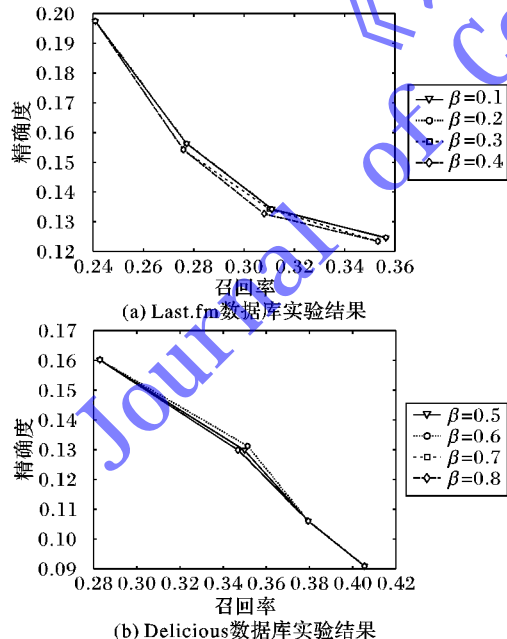


图2 参数相似度关系权重  $\beta$  对推荐结果的影响

在两个数据集上  $\beta$  取值不一致,说明用户消费的项目与标注的标签都能反映用户的兴趣,但在不同数据集上项目与标签所反映用户兴趣的程度不同,因此本文线性组合“用户-项目”相似度和“用户-标签”相似度是有必要的。

### 2) 朋友影响权重 $\alpha$ 。

在本文提出的模型中,参数  $\alpha$  决定了朋友与用户的权重关系。为了得到朋友影响权重比,将  $\alpha$  分别设为 0.1, 0.2, ..., 0.8, 0.9, top- $N$  取 1 到 10 进行实验,其中  $\beta$  的取值按照以上实验的结果分别设为 0.2 和 0.6,在两个数据库上实验结果如图 3 所示。

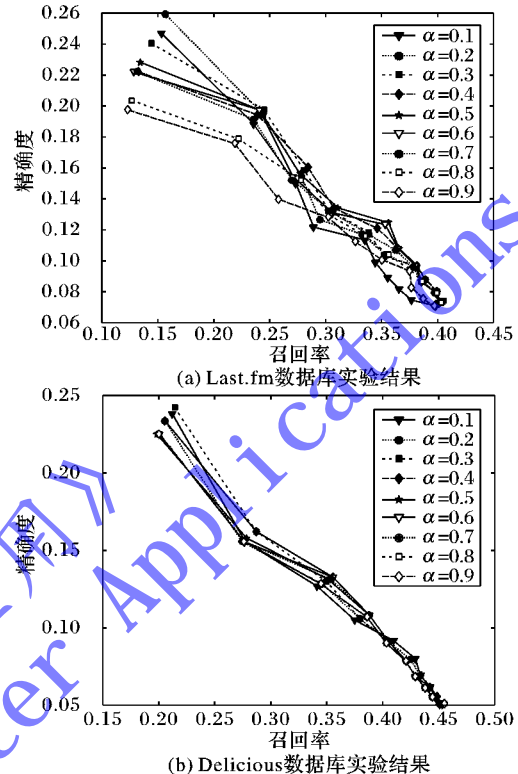


图3 参数朋友影响权重  $\alpha$  实验结果

图3中,每条折线从左上角到右下角标注的点,分别是  $N$  取 1 到 10 的实验结果值。观察图3,当不同  $N$  取值时,  $\alpha$  的最优值也不同,同时实验发现,  $\alpha$  的取值集中在  $[0.1, 0.5]$  上,使得两个数据集上都能得到最好的推荐结果,因此,当  $N$  确定时,算法只需在  $[0.1, 0.5]$  调节朋友影响权重参数  $\alpha$  的值,获取最好的推荐结果。

另外,参数  $\alpha$  取值小于 0.5,说明用户的喜好受其朋友的影响权重较大,进一步验证了本文的想法。

## 4 结语

利用标签信息和朋友关系是解决个性化推荐的一个新趋势,本文结合这两种信息进行了张量分解的推荐算法研究。利用 HOSVD 张量分解的重构三元组信息,一定程度上解决了数据稀疏性问题,并结合朋友关系修正得到的结果,实现推荐。其中考虑朋友对用户决策的影响时,本文考虑了用户与朋友间的相似度,以及个人与其朋友的影响权重比例。通过实验验证了本文提出的算法的有效性,一定程度上提高了推荐的准确度。

### 参考文献:

- [1] BREESE J S, HECKERMAN D, KADIE C. Empirical analysis of predictive algorithms for collaborative filtering [C]// Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence. San Francisco: Morgan Kaufmann Publishers, 1998: 43-52.
- [2] KARYPIS G. Evaluation of item-based top- $N$  recommendation algorithms [C]// Proceedings of the Tenth International Conference on Information and Knowledge Management. New York: ACM, 2001: 247-254.

- [3] TSO-SUTTER K H L, MARINHO L B, SCHMIDT-THIEME L. Tag-aware recommender systems by fusion of collaborative filtering algorithms [C]// Proceedings of the 2008 ACM Symposium on Applied Computing. New York: ACM, 2008: 1995–1999.
- [4] ZHOU T C, MA H, KING I, *et al.* TagRec: leveraging tagging wisdom for recommendation [C]// CSE'09: Proceedings of the 2009 International Conference on Computational Science and Engineering. Washington, DC: IEEE Computer Society, 2009: 194–199.
- [5] SUN L, LI S. Social tagging recommendation system based on  $K$ -means cluster and tensor decomposition [J]. Journal of Jiangsu University of Science and Technology: Natural Science Edition, 2012, 26(6): 597–601. (孙玲芳, 李烁朋. 基于  $K$ -means 聚类与张量分解的社会化标签推荐系统研究[J]. 江苏科技大学学报: 自然科学版, 2012, 26(6): 597–601.)
- [6] SYMEONIDIS P, NANOPOULOS A, MANOLOPOULOS Y. A unified framework for providing recommendations in social tagging systems based on ternary semantic analysis [J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(2): 179–192.
- [7] LIAO Z, WANG C, LI X. *et al.* Tag recommendation and new user tag recommendation algorithms based on tensor decomposition [J]. Journal of Chinese Computer Systems, 2013, 34(11): 2472–2476. (廖志芳, 王超群, 李小庆, 等. 张量分解的标签推荐及新用户标签推荐算法[J]. 小型微型计算机系统, 2013, 34(11): 2472–2476.)
- [8] de LATHAUWER L, de MOOR B, VANDEWALLE J. A multilinear singular value decomposition [J]. SIAM Journal on Matrix Analysis and Applications, 2000, 21(4): 1253–1278.
- [9] SYMEONIDIS P, NANOPOULOS A, PAPADOPOULOS A, *et al.* Scalable collaborative filtering based on latent semantic indexing [C]// ITWP 2006: Proceedings of the 2006 IJCAI Workshop on Intelligent Techniques for Web Personalization. Boston: [s. n.], 2006: 1–9.
- [10] MA H, YANG H, LYU M R, *et al.* SoRec: social recommendation using probabilistic matrix factorization [C]// Proceedings of the 17th ACM Conference on Information and Knowledge Management. New York: ACM, 2008: 931–940.
- [11] HUANG W, MENG X, WANG L. A collaborative filtering algorithm based on users' social relationship mining in mobile communication network [J]. Journal of Electronics and Information Technology, 2011, 33(12): 3002–3007. (黄武汉, 孟祥武, 王立才. 移动通信网中基于用户社会化关系挖掘的协同过滤算法[J]. 电子与信息学报, 2011, 33(12): 3002–3007.)
- [12] YU H, LI J. Collaborative filtering recommendation algorithm using social and tag information [J]. Journal of Chinese Computer Systems, 2013, 34(11): 2467–2471. (于洪, 李俊华. 结合社交与标签信息的协同过滤推荐算法[J]. 小型微型计算机系统, 2013, 34(11): 2467–2471.)
- [13] MA H, KING I, LYU M R. Learning to recommend with social trust ensemble [C]// Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2009: 203–210.
- [14] YAN Z, ZHOU J. User recommendation with tensor factorization in social networks [C]// Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing. Washington, DC: IEEE Computer Society, 2012: 3853–3856.
- [15] WU L, CHEN E, LIU Q, *et al.* Leveraging tagging for neighborhood-aware probabilistic matrix factorization [C]// Proceedings of the 21st ACM International Conference on Information and Knowledge Management. New York: ACM, 2012: 1854–1858.
- [16] LIU H, HU Z, MIAN A, *et al.* A new user similarity model to improve the accuracy of collaborative filtering [J]. Knowledge-Based Systems, 2014, 56: 156–166.
- [17] HetRec 2011. Datasets [EB/OL]. [2014-06-08]. <http://groups.plens.org/datasets/hetrec-2011/>.

(上接第 1978 页)

- [8] TOMITA J, NAKAWATASE H, ISHII M. Calculating similarity between texts using graph-based text representation model [C]// Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management. New York: ACM, 2004: 248–249.
- [9] LIANG T, LAI H. Discovering user interests from Web browsing behavior: an application to Internet news services [C]// HICSS 2002: Proceedings of the 35th Annual Hawaii International Conference on System Sciences. Piscataway: IEEE, 2002: 2718–2727.
- [10] LI Y, FENG B, WANG F. Page interest estimation based on the user's browsing behavior [C]// ICIC'09: Proceedings of the Second International Conference on Information and Computing Science. Piscataway: IEEE, 2009: 258–261.
- [11] CHEN D, LYU L, SHANG M, *et al.* Identifying influential nodes in complex networks [J]. Physica A: Statistical Mechanics and Its Applications, 2012, 391(4): 1777–1787.
- [12] KITSACK M, GALLOS L K, HAVLIN S, *et al.* Identification of influential spreaders in complex networks [J]. Nature Physics, 2010, 6(11): 888–893.
- [13] BRIN S, PAGE L. The anatomy of a large-scale hypertextual Web search engine [J]. Computer Networks and ISDN systems, 1998, 30(1): 107–117.
- [14] JIN D, YANG B, LIU J, *et al.* Ant colony optimization based on random walk for community detection in complex networks [J]. Journal of Software, 2012, 23(3): 451–464. (金弟, 杨博, 刘杰, 等. 复杂网络簇结构探测——基于随机游走的蚁群算法[J]. 软件学报, 2012, 23(3): 451–464.)
- [15] ESTER M, KRIEGER H P, SANDER J, *et al.* A density-based algorithm for discovering clusters in large spatial databases with noise [C]// KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. Menlo Park: AAAI Press, 1996: 226–231.
- [16] WANG H, SONG Y, CHANG M W, *et al.* Learning to extract cross-session search tasks [C]// Proceedings of the 22nd International Conference on World Wide Web. New York: ACM, 2013: 1353–1364.
- [17] YIN X, TAN W. Semi-supervised truth discovery [C]// Proceedings of the 20th International Conference on World Wide Web. New York: ACM, 2011: 217–226.
- [18] YANG Q, HAO H, NENG X. The research on user interest model based on quantization browsing behavior [C]// ICCSE 2012: Proceedings of the 7th International Conference on Computer Science & Education. Piscataway: IEEE, 2012: 50–54.
- [19] ALHARBI A, SMITH D, MAYHEW P. Web searching behavior for academic resources [C]// Proceedings of the 2013 Science and Information Conference. Piscataway: IEEE, 2013: 104–113.
- [20] BASHIR M B, ABD LATIFF M S, ABDULHAMID S M, *et al.* Grid-based search technique for massive academic publications [C]// ICT-ISPC 2014: Proceedings of the Third ICT International Student Project Conference. Piscataway: IEEE, 2014: 173–176.