

# 基于 Weighted-slope One 的用户聚类推荐算法研究

郑丹,王名扬,陈广胜

(东北林业大学 信息与计算机工程学院,黑龙江 哈尔滨 150040)

**摘要:**针对传统协同过滤推荐算法存在的数据稀疏性以及实时性差的问题,提出一种基于 Weighted-slope One 的用户聚类推荐算法。该算法首先利用 Weighted-slope One 算法的思想对初始的用户-评分矩阵进行有效填充,降低数据的稀疏性;然后,结合初始聚类中心优化改进的  $K$ -means 方法对用户进行聚类,生成相似用户集合,以缩小目标用户搜索最近邻的范围;最后,结合目标用户所属的聚类,利用基于用户的协同过滤算法搜索最近邻居,为目标用户推荐对应的产品。仿真实验结果表明,改进算法可以显著降低数据的稀疏度,同时提升推荐的准确性和实时性。

**关键词:**协同过滤;高维稀疏矩阵;Weighted-slope One;  $K$ -means;聚类中心

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2016)04-0051-05

doi:10.3969/j.issn.1673-629X.2016.04.011

## Research on User Clustering Recommendation Algorithm Based on Weighted-slope One

ZHENG Dan, WANG Ming-yang, CHEN Guang-sheng

(College of Information and Computer Engineering, Northeast Forestry University, Harbin 150040, China)

**Abstract:** Aiming at the problems of data sparseness and the poor real-time performance of traditional collaborative filtering recommendation algorithm, a new user clustering recommendation algorithm based on Weighted-slope One algorithm is proposed. Firstly, the zero items in user-item matrix are filled by Weighted-slope One algorithm. This operation can effectively reduce the data sparseness. Secondly, the users are clustered by the optimized  $K$ -means algorithm. This operation can effectively find the nearest neighbor of the target user. Finally, the corresponding products are recommended to the target users according to their nearest neighbors which are found by the users' collaborative filtering recommendation algorithm. Experimental results show that the improved algorithm can significantly reduce the data sparseness, and improve the real time performance and the accuracy of recommendation.

**Key words:** collaborative filtering; high-dimensional sparse matrix; Weighted-slope One;  $K$ -means; clustering center

### 1 概述

随着信息网络的不断发展,令人眼花缭乱的商品、繁杂且真假难辨的信息都在急剧增长。对于用户来说,想要在这些信息中快速发现、收集和维护自己所需的信息记录需要花费大量的时间和精力,这就是“信息超载”现象。在电子商务领域,通过合理地整合信息对用户实现个性化推荐是解决信息超载问题的一种有效方案。个性化推荐作为一种依靠用户购买记录、用户偏好等为基础来实现推荐目标产品推荐的算法,在电子商务领域有着广泛的应用。在目前的电子商务

推荐系统中,协同过滤是使用最广泛、最成功的推荐算法。它通过构建用户-项目的评分矩阵,搜索与目标用户具有相似偏好的邻居用户,根据邻居用户对某项目的评分值来预测目标用户对项目的评分值,从而产生推荐结果<sup>[1]</sup>。

尽管协同过滤算法在商品的个性化推荐当中取得了巨大的成功,但是,近年来随着电子商务平台上用户以及产品数量的不断增长,数据稀疏性成为限制传统协同过滤算法应用的瓶颈,并由此带来了推荐精度低、实时性差和冷启动等问题<sup>[2]</sup>。由于用户评分的项目相对于所有项目来说只占到很小的一部分,所以当用户

收稿日期:2015-07-14

修回日期:2015-10-20

网络出版时间:2016-03-22

基金项目:中央高校基本科研业务费专项资金项目(2572014DB05);中国博士后科学基金面上基金(2012M520711);国家自然科学基金资助项目(71473034)

作者简介:郑丹(1991-),女,硕士,研究方向为数据挖掘;王名扬,博士,副教授,研究方向为数据挖掘、社交网络挖掘;陈广胜,教授,研究方向为生物质材料材质预测、新型复合材料智能化检测、技术创新方法推广应用。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20160322.1521.068.html>

的评分项目之间没有交集的时候,很难判断两个用户之间是否具有相似偏好,从而导致推荐精度的下降。同时,用户、产品数量的激增是用户-项目评分矩阵过大,使搜索目标用户的最近邻居范围变大,推荐实时性受到严重影响。冷启动问题是针对“小白”用户来说的。“小白”用户指的是刚进入电子商务平台的用户,没有任何的评价和购买记录,无法去度量他与其他用户的相似度。冷启动问题实际上是数据稀疏性的一个极端表现<sup>[3]</sup>。究其本质,形成这两个问题的主要原因还是数据稀疏性问题,因此,有必要对数据稀疏性进行深入研究。

近年来,针对数据稀疏性带来的问题,为了提升推荐效果,学者们将主成分分析、聚类分析、奇异值分解等算法引入到传统的协同过滤推荐算法中,通过降维,缩小目标用户搜索最近邻居的范围,使得推荐的精度和实时性有了明显提升。文献[4]提出使用主成分分析对维数进行约减,然后使用传统的  $K$ -means 聚类技术对用户进行聚类,生成用户簇类,以缩小目标用户搜索最近邻居的范围,有效缓解了数据稀疏性,提升了推荐质量。文献[5]提出使用奇异值分解算法对用户-项目的评分矩阵进行处理,然后使用协同过滤算法对目标用户产生推荐,有效缓解了数据稀疏性的问题,同时推荐精度也得到了提升。文献[6]提出了一种基于项目聚类的协同过滤算法,通过对各项目进行聚类,在与目标项目最相似的若干个类中搜索最近邻居产生推荐列表。但是,直接使用主成分分析、SVD 奇异值分解、聚类分析等来对原始的用户-评分矩阵进行处理的能力是有限的。因为,当用户-评分矩阵数据的稀疏程度很大时,这些方法在第一步进行处理时的效果就不太理想。

基于以上启发,文中提出一种基于 Weighted-slope One 的用户聚类推荐算法。首先借鉴 Weighted-slope One 算法的思想对初始的用户-项目评分矩阵进行填充,使初始的用户-评分矩阵的稀疏性大大降低。其次,对于已经填充的用户-评分矩阵使用一种初始聚类中心优化改进的  $K$ -means 算法对用户进行聚类,生成相似用户集合。最后,使用协同过滤算法实现对目标用户的推荐。具体推荐过程为:首先判断目标用户所属的簇类,在目标用户所属的相似用户集合内实现推荐。实验仿真结果表明,文中所提出的方法不但有效降低了数据的稀疏性,同时由于是在目标用户所属簇类内为其实现系统过滤推荐,因此可以有效地降低协同过滤搜索目标用户最近邻居的范围;并且由于都是目标用户的相似用户,因此文中提出的算法在降低计算复杂度的同时,还能有效提升推荐系统的实时响应速度。

## 2 基于 Weighted-slope One 算法降低用户-评分矩阵的稀疏性

Slope One 算法是于 2005 年由 Daniel Lemire 教授提出的一个基于项目的推荐算法<sup>[7]</sup>。算法以  $f(x) = x + b$  作为预测器,使用线性回归预测用户对未评分项目的打分,参数  $b$  为用户对两个项目评分的平均偏差<sup>[8]</sup>。Weighted-slope One 是 Slope One 的一个递进算法,它均衡了每个评分项目对于目标项目的影响。由于 Weighted-slope One 在预测用户对项目的打分值时充分考虑了用户共同评分的差异度以及不同项目被用户同时评论的次数,因此,所得预测的评分与真实的用户评分是比较相近的。但是,文献[9]研究表明 Slope One 在用户-项目评分矩阵比较密集的时候推荐的效果较好,当数据比较稀疏时,其推荐的效果并不好。鉴于此,文中并不直接使用 Weighted-slope One 算法进行推荐,而是借鉴 Weighted-slope One 算法的这种对目标用户打分思想来对初始高维的用户-评分矩阵进行填充,对填充后的矩阵使用初始聚类中心优化的  $K$ -means 算法进行聚类,生成相似用户集合。当对目标用户进行推荐时,首先确定目标用户所属的簇类,然后在该簇类内搜索目标用户的最近邻居,产生推荐。

使用 Weighted-slope One 对用户  $u$  的未评分项目进行填充降低用户-评分矩阵的稀疏性的过程为:首先,寻找用户  $u$  和用户  $v$  评分过的项目集合  $I_{uv} = I_u \cap I_v$ ,利用式(1)计算用户  $u$  的未评分项目  $j$  对于项目  $i$  的偏移量  $\text{dev}_{ji}$ ,根据式(2)得到目标用户对于项目  $j$  的评分值。

这是经典的 Slope One 算法计算目标用户对项目评分的方法,但是对项目进行评分的用户数量在最终的预测评分时显然也有影响。因此,使用 Weighted-slope One 算法修正最终目标用户  $u$  对于项目的预测评分,即根据式(3)得出目标用户对项目  $j$  的最终评分。

$$\text{dev}_{ji} = \sum_{u \in (I \cap j)} \frac{u_j - u_i}{\text{card}(i \cap j)} \quad (1)$$

$$p_{uj} = \frac{1}{\text{card}(R(u) \cap D(j))} \sum_{i \in R(u) \cap D(j)} (u_i + \text{dev}_{ji}) \quad (2)$$

$$p_{uj} = \frac{1}{\sum_{i \in R(u) \cap D(j)} \text{card}(i \cap j)} \sum_{i \in R(u) \cap D(j)} ((u_i + \text{dev}_{ji}) * \text{card}(i \cap j)) \quad (3)$$

其中:  $u_i, u_j$  分别表示用户  $u$  对项目  $i$  和项目  $j$  的评分值;  $\text{card}(i \cap j)$  表示项目  $i$  和项目  $j$  同时被评分过的用户数目;  $R(u)$  表示目标用户  $u$  已经评价过的项目的集合;  $D(j)$  表示与目标项目  $j$  计算得到的平均偏移量  $\text{dev}_{ji}$  的集合。

通过使用 Weighted-slope One 算法对初始的用户

-评分矩阵进行处理,矩阵中的大部分零元素就会被填充,从而有效降低了整个矩阵的稀疏程度。

### 3 初始聚类中心优化的 $K$ -means 算法

$K$ -means 算法是一种基于划分的聚类分析方法,由于其理论可靠、算法实现简单、收敛速度快,并且又容易实现对大规模数据的处理,因此是应用最广泛的聚类算法之一。但是由于传统的聚类算法对聚类中心敏感,随机地选取任意  $k$  个点作为初始聚类中心,并迭代地改变中心使聚类准则收敛,经常会导致算法陷入局部最优,产生完全不同的聚类结果。为了避免这种初始聚类中心随机选取带来的聚类结果不稳定性,学者们将数据分布的稠密程度作为指标引入到初始聚类中心的选取上,希望所选取的初始聚类中心能够在一定程度上代表数据本身的分布情况。

文献[10]通过划分高、低密度区域选择高密度区域中具有代表性的数据对象作为初始的聚类中心。文献[11]通过建立数据对象分布的密度集合,计算最近两点的垂直中点方法来确定初始聚类中心,迭代形成最终簇类。通过考察密度特性,这些算法都有效地改善了  $K$ -means 的稳定性。但算法中还存在着参数选择较为主观的现象,比如对于密度参数  $\theta$  选得过大,可导致本不属于同一类的样本点被划分到同一类的情况;如果过小,会导致相反的情形,导致噪声数据也被列入候选聚类中心的情况,由此导致聚类性能的下降。

基于此,文中提出一种基于距离-期望密度参数的聚类中心选择方法。该方法结合每个聚类簇大致期望的样本平均个数来确定密度  $\theta$  的大小。同时,在满足期望密度的前提下,计算每个密度区域的半径,作为期望密度参数,结合数据样本间的距离,寻找初始的聚类中心。由于对用户聚类依靠的是用户对产品的评分相似性,与密度类似,偏好同一类产品的用户一般会具有相同的喜好,所以将改进的  $K$ -means 聚类算法直接应用于用户聚类产生相似用户是十分有意义的。

(1) 相关定义。

定义1:密度参数。

取空间中的任一数据样本  $x_i$ ,以其为中心,寻找距离  $x_i$  最近的  $\theta$  个数据样本,把  $x_i$  与  $\theta$  个样本之间的最大距离称之为样本  $x_i$  的密度参数,用  $\varepsilon$  表示。

定义2:期望密度。

$$E_\theta = a * n / k \quad (4)$$

其中: $n$  为数据集的样本总数; $k$  为设定的聚类个数; $a$  为期望密度比例系数。

$a$  可以根据经验选取,文中取  $a = 0.6$ 。即,如果某个数据样本被选作聚类中心,则离该对象最近的  $\theta$  个样本应该占到以该对象为中心所形成的聚类簇覆盖样

本的60%。

定义3:平均相异度。

数据样本之间的平均相异度定义为:

$$d(x_i) = \sum_{j \neq i} d(x_i, x_j) / t \quad (5)$$

其中: $x_i, x_j$  为数据集中两个数据样本; $t$  为数据对象  $x_j$  的个数; $d(x_i, x_j)$  为数据样本  $x_i$  与  $x_j$  间的距离,文中用欧几里得距离公式进行计算:

$$d(x_i, x_j) = \sqrt{\sum_{p=1}^m |x_{ip} - x_{jp}|^2} \quad (6)$$

(2) 改进算法的思想。

算法的基本思想是:给定数据集  $D$ ,通过数据集的样本总数  $n$  和要生成簇的总个数  $k$ ,确定数据样本的期望密度  $E_\theta$ ;根据期望密度  $E_\theta$ ,确定每个数据样本的期望密度参数  $E_\varepsilon$ ;将  $E_\varepsilon$  最小的样本点  $x_i$  作为第一个初始聚类中心添加到集合  $C$  中;根据式(5)计算  $D - C$  中每个数据对象与集合  $C$  中元素的平均相异度。在此基础上,计算并选取满足  $\left\{ \frac{\sum_{j \in D-C} d(x_j)}{E_{\varepsilon_i}} \right\}_{\max}$  的样本点

添加到初始聚类中心集合  $C$  中,作为下一个初始聚类中心;直至集合  $C$  中的元素个数为  $k$  时算法停止。最后,以集合  $C$  中的所有元素作为初始聚类中心,执行  $K$ -means 聚类算法,得到最终的  $k$  个簇类。

聚类中心优化的  $K$ -means 算法中,选取具有较低期望密度参数且彼此相距最远的  $k$  个对象作为初始的聚类中心,这些中心点能够很好地反映数据对象的空间分布特性,同时能保证选出的聚类中心所覆盖的区域样本密度较高。这些都使得算法具有较高的稳定性,同时使得算法能够获得全局最优的聚类效果。

### 4 基于用户的协同过滤推荐

协同过滤算法是于1992年由Goldberg等学者提出的,该算法是以项目属性或者用户评分相似性作为个性化推荐的基础。协同过滤算法可以分为基于用户的协同过滤和基于项目的协同过滤<sup>[12]</sup>。基于用户的协同过滤算法实现推荐时基于如下事实:如果用户对一些项目的评分比较相似,就认为他们在其他项目的评分上具有相似关系<sup>[13]</sup>。

算法的执行流程为:首先,计算目标用户与推荐系统中所有用户的相似性,找到与目标用户最相似的邻居用户集合;其次,根据最近邻居集合中用户对项目的评分去预测目标用户对未评分项目的打分;最后,根据预测的打分情况为目标用户推荐打分值最高的 top  $M$  个项目<sup>[14-15]</sup>。

算法执行过程中先明确的几个计算单元如下:

(1) 用户相似度计算。



用户之间的相似度是通过用户对项目的历史评分来进行度量的<sup>[16]</sup>。文中使用余弦相似性来产生用户最近邻居,其公式如下:

$$\text{sim}(u, v) = \cos(\vec{u}, \vec{v}) = \frac{\vec{u} \times \vec{v}}{\|\vec{u}\| \times \|\vec{v}\|} \quad (7)$$

其中:  $\vec{u}$  表示用户  $u$  对各个项目的评分,未评分的项目用 0 表示;  $\vec{v}$  表示用户  $v$  对各个项目的评分,同样未评分的项目用 0 表示;  $\|\vec{u}\|$  表示用户  $u$  对项目评分向量的模;  $\|\vec{v}\|$  表示用户  $v$  对项目评分向量的模。

## (2) 项目评分预测

在预测项目评分时,经常选取前  $N$  个与目标用户相似性最大的用户作为目标用户的最近邻居,根据这些邻居用户对目标用户未评分项目的评分情况来预测产生目标用户对其的评分值。假设  $N_u$  表示邻居用户的集合,根据式(8)可以计算目标用户对于某项目的预测评分。

$$p_{uj} = \bar{R}_u + \frac{\sum_{l \in N_u} \text{sim}(u, l) \times (R_{l,j} - \bar{R}_l)}{\sum_{l \in N_u} |\text{sim}(u, l)|} \quad (8)$$

其中:  $\bar{R}_u$  表示目标用户的已评分项的评分均值;  $\text{sim}(u, l)$  表示邻居用户  $l$  与目标用户  $u$  的相似性;  $R_{l,j}$  表示邻居用户  $l$  对项目  $j$  的评分;  $\bar{R}_l$  表示邻居用户  $l$  对各项目的平均评分。

针对预测评分的值给出目标用户的推荐列表。

## 5 基于 Weighted-slope One 的用户聚类推荐算法

设用户集合为  $U = \{u_1, u_2, \dots, u_m\}$ , 使用改进的  $K$ -means 算法生成的用户集合表示为  $U' = \{C_1, C_2, \dots, C_k\}$ 。其中,  $k$  表示生成簇类的个数,  $C_k$  表示第  $k$  个簇类。基于 Weighted-slope One 的用户聚类推荐算法的具体执行步骤如下:

输入: 用户-评分矩阵  $D$ , 聚类个数  $k$ , 最近邻居及大小  $N$ , 推荐列表的个数  $M$ ;

输出: 目标用户的推荐列表以及对应的评分。

Step1: 根据 Weighted-slope One 算法利用式(3)对初始的用户-评分矩阵  $D$  中的 0 元素进行填充, 生成填充矩阵  $D^c$ ;

Step2: 利用改进的  $K$ -means 算法基本思想对  $D^c$  进行处理, 生成  $k$  个聚类输出, 各聚类的中心为  $c_i, i = 1, 2, \dots, k$ ;

Step3: 通过式(6)计算目标用户与聚类中心  $c_i$  的距离, 将目标用户分配到距离最小的聚类  $C_i (i = 1, 2, \dots, k)$  中;

Step4: 在目标用户所属的簇类中, 根据式(7)计算目标用户与其他用户的相似性, 搜索出相似性最大的前  $N$  个用户作为目标用户的最近邻居集合, 记为  $N_u$ ;

Step5: 利用式(8)计算各个项目的预测评分, 列出 top  $M$  个预测得分最大的项目推荐给目标用户。

## 6 仿真实验与结果分析

### 6.1 数据集

仿真实验中, 选取了 MovieLens 站点的电影数据集作为实验数据。数据集中包含了 943 个用户 1 682 部电影共 10 000 条评价得分, 其中每个用户至少对 20 部电影进行了评分, 评分值为 1~5 之间的整数。评分值越大说明用户对该项目的偏好性越强, 0 值表示用户未对该电影进行评分。

### 6.2 评测指标

在推荐算法中经常使用平均绝对误差 (MAE) 来衡量推荐结果的好坏。MAE 是通过计算预测的用户评分与实际的用户评分之间的偏差来度量预测的准确性<sup>[12]</sup>。假设目标用户对项目的实际评分为  $\{p_i | i = 1, 2, \dots, n\}$ , 其中  $n$  表示项目的总个数, 使用推荐算法预测得到的目标用户对项目的评分为  $\{q_i | i = 1, 2, \dots, n\}$ , 那么 MAE 的计算公式为:

$$\text{MAE} = \frac{\sum_{i=1}^n |p_i - q_i|}{n} \quad (9)$$

从式(9)可以看出, MAE 值越小说明推荐结果的准确性越高。

### 6.3 仿真结果分析

为保证所选取的相似用户集合最为合理, 首先针对用户簇类个数进行系统过滤算法的实现, 计算其对应的 MAE 值, 以选择最合理的用户簇类个数来对目标用户产生最终的推荐结果。

图 1 就是在不同的聚类个数下 MAE 的变化情况。

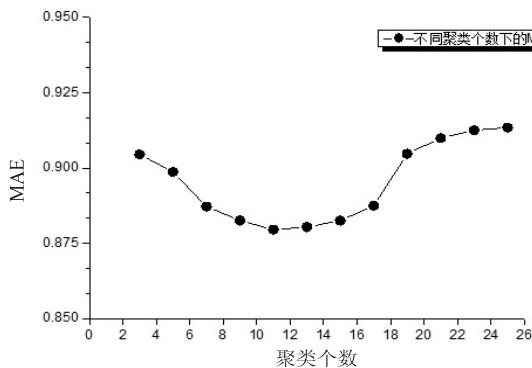


图 1 不同聚类数目下的 MAE 值

由图 1 可以看出, 随着聚类个数的增加, MAE 的值先降后升, 这说明在用户数固定的情况下, 需要找到一个最优的聚类数目能够使推荐的 MAE 值达到最小。从图 1 中观察可知, 当聚类个数为 11 时, MAE 的值最

低,显然对于文中所选取的数据集而言,令  $k = 11$  是最合理的聚类个数。

在确定了最优的聚类个数后,将文中所提出的基于 Weighted-slope One 算法与传统的协同过滤算法、基于  $K - \text{means}$  用户聚类的协同过滤算法进行了比较。最近邻居的个数从 5 增至 50,间隔为 5。实验对比结果如图 2 所示。

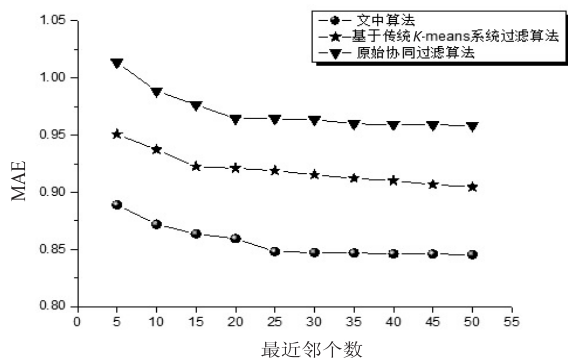


图2 推荐结果的 MAE 值比较

根据图 2,随着目标用户最近邻居个数的不断增加,三种算法的 MAE 值呈明显的下降趋势。这说明随着最近邻居个数的不断增加,可以有效提升推荐结果的准确率。但从 MAE 下降的速率来看,当最近邻居的个数上升到某个个数之后,MAE 下降的速率越来越慢,最后趋近于平滑。这说明最近邻居的个数并不是越多越好,而是随着最近邻居数目的增多,其为 MAE 减小所做的贡献越来越小。在实际应用中,当考虑推荐系统实时性、存储、计算的复杂度时,合理选择最近邻居的个数对于推荐精度和维护成本都至关重要。

除此之外,在不同的最近邻居数目下对比三种算法可得,文中提出的算法较其他算法均有最小的 MAE 值,传统的协同过滤算法的推荐效果最不理想,添加了用户聚类的推荐效果次之。但是由于数据稀疏性和初始聚类中心随机选取可能带来的不稳定聚类结果,基于  $K - \text{means}$  的协同过滤推荐的精度也不高。而文中提出的算法无论是对数据稀疏性还是聚类结果都进行了对应的处理,可以确保在数据稀疏性降低的同时有稳定的聚类结果。显然,通过文中填充高维稀疏矩阵,实现用户聚类,为目标用户在聚类内给出推荐结果的方法不但有效降低了数据的稀疏性,同时在聚类内搜索目标用户的最近邻居缩小了搜索范围,加快了推荐算法的响应速度,推荐 MAE 值得到了提升。

## 7 结束语

文中分析陈述了协同过滤推荐算法所存在的问题以及学者为提升推荐准确率所做的研究工作。在此基础上提出了一种基于 Weighted-slope One 的用户聚类

推荐算法。通过借鉴 Weighted-slope One 进行评分预测的思想对初始的用户-项目矩阵中的 0 元素进行填充,降低数据稀疏性,使用改进的  $K - \text{means}$  聚类算法对用户进行聚类,在簇类内查找目标用户的最近邻居最终给出预测评分以及推荐结果。仿真实验证明,文中提出的算法可以有效地提升推荐的准确性,同时由于其缩小了搜索最近邻居的空间,推荐算法的实时性也得到了提升。

## 参考文献:

- [1] 王桂芬. 电子商务个性化推荐系统中协同过滤算法的研究与应用[D]. 南昌:南昌大学,2012.
- [2] 李华,张宇,孙俊华. 基于用户模糊聚类的协同过滤推荐研究[J]. 计算机科学,2012,39(12):83-86.
- [3] 孙小华. 协同过滤系统的稀疏性与冷启动问题的研究[D]. 杭州:浙江大学,2005.
- [4] 郁雪,李敏强. 一种结合有效降维和  $K - \text{means}$  聚类的协同过滤推荐模型[J]. 计算机应用研究,2009,26(10):3718-3720.
- [5] Vozalis M G, Margaritis K G. Applying SVD on item-based filtering[C]//Proc of 5th international conference on intelligent systems design and applications. [s. l.]: IEEE, 2005: 464-469.
- [6] 邓爱林,左子叶,朱扬勇. 基于项目聚类的协同过滤推荐算法[J]. 小型微型计算机系统,2004,25(9):1665-1670.
- [7] 李朝阳. 基于 Slope One 算法的协作过滤个性化推荐系统设计及实现[D]. 武汉:华中科技大学,2010.
- [8] 王毅. 基于 Hadoop 的 Slope One 及其改进算法实现[D]. 成都:西南交通大学,2011.
- [9] 董丽,邢春晓,王克宏. 基于不同数据集的协作过滤算法评测[J]. 清华大学学报:自然科学版,2009,49(4):590-594.
- [10] 袁方,周志勇,宋鑫. 初始聚类中心优化的  $k - \text{means}$  算法[J]. 计算机工程,2007,33(3):65-66.
- [11] 周炜奔,石跃祥. 基于密度的  $K - \text{means}$  聚类中心选取的优化算法[J]. 计算机应用研究,2012,29(5):1726-1728.
- [12] 马宏伟,张光卫,李鹏. 协同过滤推荐算法综述[J]. 小型微型计算机系统,2009,30(7):1282-1288.
- [13] 李涛,王建东,叶飞跃,等. 一种基于用户聚类的协同过滤推荐算法[J]. 系统工程与电子技术,2007,29(7):1178-1182.
- [14] 李晓艳,张子刚,张逸石. 集成  $k - \text{means}$  聚类和有监督特征选择的混合式协同过滤推荐[J]. 管理学报,2013,10(9):1362-1367.
- [15] 张海燕,丁峰,姜丽红. 基于模糊聚类的协同过滤推荐方法[J]. 计算机仿真,2005,22(8):144-147.
- [16] 贺银慧,陈端兵,陈勇,等. 一种结合共同邻居和用户评分信息的相似度算法[J]. 计算机科学,2010,37(9):184-186.