

# 基于用户特征的分步协同推荐算法\*

黄文明<sup>a</sup>, 程广兵<sup>b</sup>, 邓珍荣<sup>a</sup>, 周先亭<sup>b</sup>

(桂林电子科技大学 a. 广西可信软件重点实验室; b. 计算机科学与工程学院, 广西 桂林 541004)

**摘要:** 协同过滤是解决信息过载问题的一种有效技术。针对基于内存的推荐面临着可扩展性问题、基于模型的推荐需要训练大量参数的问题进行了研究,从而提出了基于用户特征的 K-means 用户聚类算法,然后用分步协同过滤框架融合基于项目和基于用户的协同过滤给每一个聚簇训练一个模型。实验结果表明,提出的算法能极大地提高了推荐精度,同时一定程度上解决了基于模型和基于内存的推荐存在的不足。

**关键词:** 协同过滤; 用户特征; 聚类算法; 分步协同过滤框架; 模型

中图分类号: TP301.6 文献标志码: A 文章编号: 1001-3695(2017)04-1047-03

doi:10.3969/j.issn.1001-3695.2017.04.020

## Distributed collaborative recommendation algorithm based on user characteristics

Huang Wenming<sup>a</sup>, Cheng Guangbing<sup>b</sup>, Deng Zhenrong<sup>a</sup>, Zhou Xianting<sup>b</sup>

(a. Guangxi Key Laboratory of Trusted Software, b. School of Computer Science & Engineering, Guilin University of Electronic Technology, Guilin Guangxi 541004, China)

**Abstract:** Collaborative filtering was an effective technique addressing the information overload problem. Aiming at the memory based recommendation suffered from difficulty in scalability. Model based recommendation had a multitude of parameters to train conducted a study, thus this paper proposed K-means user clustering algorithm based on user characteristics. Then, by using distributed collaborative filtering framework mixed with item based and user based collaborative filtering to train a model to every cluster. The experimental result shows that the proposed algorithm can greatly improve the recommendation precision. And solves the limitations of model based and memory based recommendations to a certain extent at the same time.

**Key words:** collaborative filtering; user characteristics; clustering algorithm; distributed collaborative filtering framework; model

## 0 引言

互联网的快速发展带来了严重的信息过载问题。用户面对着海量的数据和资源,却不能有效地发现自己感兴趣的内容。推荐系统作为一种高效的解决方法,为用户推荐可能感兴趣的的项目。协同过滤(collaborative filtering)作为应用最广泛的推荐算法之一,它利用用户的历史评分信息来预测目标用户的偏好。根据处理方式不同,协同过滤又可以分为基于模型(model-based)的和基于内存(memory-based)的协同过滤。其中基于内存的协同过滤利用整个用户评分数据集来计算用户之间的相似度,寻找与目标用户偏好最相似的 K 个(Top-K)用户,然后根据相似用户给目标用户作出推荐。根据相似性对象不同,基于内存的协同过滤又可以分为基于用户的(user-based)协同过滤<sup>[1,2]</sup>和基于项目的(item-based)协同过滤<sup>[3,4]</sup>。基于内存的协同过滤思想简单,同时也能取得较高的推荐精度,但是算法为了得到目标用户的推荐结果必须扫描整个数据集,这带来了很大的可扩展性问题。随着用户和项目的大量增加,基于内存的协同过滤很不适合实时在线推荐系统<sup>[5,6]</sup>。

与基于内存的协同过滤不同是,基于模型的协同过滤利用用户评分数据集训练出一个模型,利用模型为目标用户进行预

测产生推荐。这其中主要包括朴素贝叶斯模型(naive Bayesian model)<sup>[7]</sup>、基于概率模型的隐含主题分析(probabilistic latent semantic analysis)<sup>[8]</sup>、基于矩阵分解模型的奇异值分解(singular value decomposition)<sup>[9]</sup>、聚类(clustering)分析模型<sup>[10-12]</sup>、图模型<sup>[13]</sup>等。对于诸如概率和矩阵分解等一类的潜在因子分析模型,模型推荐算法可以在尽可能短的时间内作出响应,同时缓解传统协同过滤的数据稀疏性问题。然而这些模型的训练需要的时间长<sup>[5]</sup>,参数过多且参数优化比较困难,推荐结果缺乏很好的解释性。同时基于模型的协同过滤对于数据的变动过于敏感,或者提出的模型不能很好地拟合数据,导致推荐效果差。此外模型的更新较为困难。

为了使推荐系统同时具有基于内存的和基于模型的协同过滤两者优点,本文提出首先利用用户特征运用 K-means 算法给用户聚类,然后在聚类的基础上利用分步协同过滤框架有效地融合基于项目和基于用户的方法。在第一步分别用基于项目和基于用户的方法来预测用户对项目的评分,然后在第二步利用两种方法的预测结果为每个聚簇训练一个模型。

## 1 相关工作

基于内存的协同过滤主要分为两步:a)根据 Top-N 找出与

收稿日期: 2016-03-07; 修回日期: 2016-04-28 基金项目: 广西可信软件重点实验室基金资助项目(kx201106);2015年广西科技攻关项目(桂科攻 1598019-6)

作者简介:黄文明(1963-),男,江苏苏州人,教授,学士,主要研究方向为大数据处理、图形图像处理;程广兵(1989-),男,硕士研究生,主要研究方向为数据挖掘、机器学习(zhchengguangbing@163.com);邓珍荣(1977-),女,教授,硕士,主要研究方向为协议安全、软件架构等;周先亭(1990-),男,硕士研究生,主要研究方向为数据挖掘。

目标用户的相似用户或者相似项目集合;b)根据步骤 a)的结果进行预测。

### 1.1 基于用户的协同过滤

基于用户的协同过滤假设目标用户会喜欢与他兴趣偏好相似用户喜欢的项目,相似性用户是根据用户项目的评分数据集计算出来的,与目标用户相似度较高的用户对预测得分的贡献也就相应的较大。目前比较流行的相似度计算方法主要有皮尔森相关相似度(Pearson correlation coefficient)<sup>[14]</sup>、余弦相似度(cosine similarity)<sup>[2]</sup>和欧氏距离(Euclidean distance)。本算法采用的相似度定义如式(1)所示。

$$s(u, v) = \frac{\sum_{i \in I(u) \cap I(v)} (r_{ui} - \bar{r}_u) \times (r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i \in I(u) \cap I(v)} (r_{ui} - \bar{r}_u)^2} \sqrt{\sum_{i \in I(u) \cap I(v)} (r_{vi} - \bar{r}_v)^2}} \quad (1)$$

其中: $s(u, v)$ 表示用户  $u, v$  之间基于皮尔森相关系数的相似度; $I(u)$ 表示用户  $u$  评分的所有项目集合; $I(v)$ 表示用户  $v$  评分的所有项目集合; $r_{ui}$ 表示用户  $u$  对项目  $i$  的评分; $r_{vi}$ 表示用户  $v$  对项目  $i$  的评分; $\bar{r}_u, \bar{r}_v$ 分别表示用户  $u, v$  在所有项目上的评分均值。

通过计算用户的相似度,根据 Top- $N$  得到目标用户  $u$  的最近邻居集合  $N(u)$ ,则用户  $u$  对未评分项目  $i$  的预测得分为

$$P(r_{ui}) = \bar{r}_u + \frac{\sum_{v \in N(u)} (r_{vi} - \bar{r}_v) \times s(u, v)}{\sum_{v \in N(u)} s(u, v)} \quad (2)$$

### 1.2 基于项目的协同过滤

基于项目的协同过滤与基于用户的协同过滤思想一样,只不过用户之间的相似度替换为项目之间的相似度。本算法采用的相似度定义式(3)所示。

$$s(i, j) = \frac{\sum_{u \in U(ij)} r_{ui} \times r_{uj}}{\sqrt{\sum_{u \in U(i)} r_{ui}^2} \sqrt{\sum_{u \in U(j)} r_{uj}^2}} \quad (3)$$

其是: $s(i, j)$ 表示项目  $i, j$  之间余弦相似度; $U(i)$ 表示所有对项目  $i$  有过评分的用户集合; $U(j)$ 表示所有对项目  $j$  有过评分的用户集合; $U(ij)$ 表示同时对项目  $i, j$  有过评分的用户集合。

通过计算项目之间的相似度,根据 Top- $N$  得到带预测项目  $i$  的最近邻居集合  $N(i)$ ,则用户  $u$  对未评分项目  $i$  的预测得分如式(4)所示。

$$P(r_{ui}) = \bar{r}_i + \frac{\sum_{j \in N(i)} (r_{uj} - \bar{r}_j) \times s(i, j)}{\sum_{j \in N(i)} s(i, j)} \quad (4)$$

其中: $\bar{r}_i, \bar{r}_j$ 分别表示项目  $i, j$  在所有用户上的评分均值。

以上两种算法都是基于内存的协同过滤,在推荐精度上都能取得不错的效果。但是随着系统用户和项目的增加,每次为目标用户作出推荐时都需要扫描整个评分数据集,因此给系统带来了可扩展性问题。

## 2 研究内容

为了解决基于内存的协同过滤面临的可扩展性问题,本文提出根据用户的特征运用 K-means 算法给用户聚类;为了减小用户经过聚类之后带来推荐精度下降的问题,本文提出了一个分步协同过滤框架,可以有效地融合基于项目和基于用户的协同过滤,给每个聚簇的用户训练一个模型,从而解决模型推荐精度不高的问题。本文提出的算法流程如图 1 所示。

### 2.1 基于用户特征的 K-means 聚类

由于用户的行为受用户的特征影响,用户的特征很多,如

性别、年龄、职业等。例如在电影推荐网站上,男性比较喜欢战争和动作电影,而女性比较喜欢爱情和生活片。不同的年龄也会有不同的兴趣爱好,小孩子比较喜欢动画片,青少年则比较喜欢励志的电影。然而大多数的推荐网站在进行推荐时大多会忽略用户的特征。本文首先在特定的场景中选择影响用户行为的特征;然后根据专家意见给选择的属性表示成离散的数值形成用户特征数据集  $UF = \{u_i | i \in 1, 2, \dots, m\}$ ,  $m$  为用户数,其中  $u_i = \{f_j | j \in 1, 2, \dots, n\}$ ,  $n$  为选择的特征数;最后运用 K-means 算法给用户聚类。假设用户可以聚成  $K$  类,最终聚类结果可以表示成  $C = \{C_u^1, C_u^2, \dots, C_u^K\}$ 。本文根据欧氏距离来判断用户所属哪一个簇。聚类算法如下:

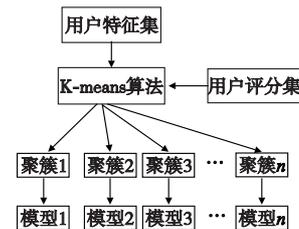


图1 本文算法流程

算法 基于用户特征的 K-means 聚类算法

输入:目标用户  $U_a$ , 聚类个数  $K$ 。

输出:目标用户  $U_a$  所在的聚类用户。

a)从用户特征数据集中随机选择  $K$  个用户特征向量作为质心。

b)计算其余的用户向量距离  $K$  个质心的距离,把用户加入距离最近的质心所在的簇中。

c)每个用户都已经属于其中的一个簇,然后根据每个质心所包含的数据向量的集合,重新计算得到新的质心。

d)如果新计算的质心与原来的质心之间的距离达到每一个设置的阈值,算法中止;否则需要迭代步骤 b) ~ d)。

### 2.2 分步协同过滤框架

Cacheda 等人<sup>[5]</sup>在实验中得出,基于模型的推荐算法在一定程度上解决了传统的协同过滤算法面临的可扩展性问题,随之而来的是推荐精度的下降。为了使推荐算法同时具有两者优点的问题,在本节中提出一种有效的框架对基于项目的协同过滤和基于用户的协同过滤算法进行有效的混合,从而可以有效地提高用户经过 K-means 算法聚类之后的推荐精度。在框架的第一步中,本文的训练集采用 5 折交叉验证的方式来产生,并根据该训练集分别构造基于项目和基于用户的预测函数,然后根据训练好的预测函数在相应的交叉验证数据集上进行测试,将预测函数产生的结果作为第二步的训练数据;在第二步采用最小二乘法训练一个融合模型。为了叙述方便,将原始的用户评分数据定义为第一步训练数据,将第一步中两个预测函数产生的预测结果定义为第二步训练数据。

#### 2.2.1 第一步框架

第一步中用 5 折交叉验证的方式来预测产生第二步的训练数据。定义  $R$  表示用户原始的评分矩阵,将整个原始评分数据集随机地分成五等份,其中第  $j$  份用  $R_j$  表示,并令  $R^j$  表示评分矩阵中除  $R_j$  以外的其他评分数据。 $R_j$  被用做训练数据,而  $R^j$  用做对应的测试数据。令  $P_u$  表示基于用户的预测函数, $P_i$  表示基于项目的预测函数。本文用式(5)构造第二步的训练数据。

$$TR_{\text{step-2}} = \{P_u^j(r_{ui}), P_i^j(r_{ui}), (r_{ui}) | r_{ui} \in R_j, j = 1, 2, \dots, 5\} \quad (5)$$

其中: $r_{ui}$ 表示一个测试评分; $P_u^j(r_{ui})$ 表示基于用户的预测函数对  $r_{ui}$  的预测值; $P_i^j(r_{ui})$ 表示基于项目的预测函数对  $r_{ui}$  的预测值。

### 2.2.2 第二步框架

第二步的任务是基于第一步产生的训练数据对基于项目和基于用户的预测函数进行线性融合。式(6)为融合后的模型式。

$$P(r_{ui}) = \theta^{(1)} \times P_U(r_{ui}) + \theta^{(2)} \times P_I(r_{ui}) \quad (6)$$

其中: $P(r_{ui})$ 表示经过线性融合后的预测函数, $\theta^{(1)}$ 和 $\theta^{(2)}$ 表示基于项目和基于用户的预测值所占的权重。

由于本文的问题具有实际的意义,利用线性回归进行模型融合时,将所有的权重参数约束为非负,所以本文的问题最终表示为式(7)所示的带约束的二次优化问题。其中 $\theta$ 是模型参数; $|TR_{step-2}|$ 表示训练集大小; $(x^{(k)}, r^{(k)})$ 表示第 $k$ 个训练样本。

$$\min_{\theta} \frac{1}{2} \sum_{k=1}^{|TR_{step-2}|} (\theta^T x^{(k)} - r^{(k)})^2 \quad \text{s. t. } \theta^{(1)} \geq 0, \theta^{(2)} \geq 0 \quad (7)$$

其中 $\theta = (\theta^{(1)}, \theta^{(2)})^T, x^{(k)} = (P_U(r^{(k)}), P_I(r^{(k)}))^T$ 。通过运用拉格朗日乘子和KT条件,解得

$$\begin{cases} \theta^{(1)} = \frac{ce - bd}{c^2 - ab} \\ \theta^{(2)} = \frac{cd - ac}{c^2 - ab} \\ \theta^{(1)} \geq 0, \theta^{(2)} \geq 0 \end{cases} \quad (8)$$

$$\begin{cases} \theta^{(1)} = 0 \\ \theta^{(2)} = \frac{e}{b} \end{cases} \quad (9)$$

$$\begin{cases} \theta^{(1)} = 0 \\ \theta^{(2)} = \frac{e}{b} \end{cases} \quad (10)$$

其中: $a = \sum_j P_U^2(r^{(j)})$ ;  $b = \sum_j P_I^2(r^{(j)})$ ;  $c = \sum_j P_U(r^{(j)})P_I(r^{(j)})$ ;  $d = \sum_j r^{(j)}P_U(r^{(j)})$ ;  $e = \sum_j r^{(j)}P_I(r^{(j)})$ 。

## 3 实验验证和相关分析

### 3.1 数据集

实验数据采用的是 MovieLens 公开数据集,此数据包含 943 个用户对 1 682 部电影的 10 万条评分记录,并且每个用户至少对 20 部电影的评分。本文实验用到了历史评分信息和用户特征信息两个数据集。其中用户特征信息如表 1 所示。本文选取了其中的三个属性并根据专家意见给每个属性表示成数值形式:age(1~3)、gender(0~1)、occupation(1~8),用 K-means 聚类之后随机地把在一个簇中的评分数据 80% 用做训练集,20% 用做测试集。

表 1 MovieLens 数据集用户特征

| 属性      | 属性描述 | 属性         | 属性描述 |
|---------|------|------------|------|
| user id | 用户编号 | occupation | 用户职业 |
| age     | 用户年龄 | zipcode    | 邮政编码 |
| gender  | 用户性别 |            |      |

### 3.2 评价指标

本文用平均绝对误差(MAE)<sup>[14]</sup>和均方误差(RMSE)来评定预测结果。MAE和RMSE越小,代表预测精度越高如式(11)(12)所示。

$$MAE = \frac{\sum_{u,i} |P(r_{ui}) - r_{ui}|}{N} \quad (11)$$

$$RMSE = \sqrt{\frac{\sum_{u,i} (P(r_{ui}) - r_{ui})^2}{N}} \quad (12)$$

其中: $P(r_{ui})$ 表示用户 $u$ 对项目 $i$ 的预测得分; $r_{ui}$ 表示用户 $u$ 对项目 $i$ 的实际得分; $N$ 表示测试集的大小。

### 3.3 算法结果对比

为了验证本文提出算法的有效性,设计了基于用户的协同过滤(UBCF)、基于项目的协同过滤(UBCF)、先用用户特征聚类再根据基于用户的协同过滤(FUBCF)、先用用户特征聚类再根据基于项目的协同过滤(FIBCF)和本文提出的先用用户特征聚类再用分步协同过滤框架(FIUCF)的对比实验。比较这五种算法运用5折交叉验证在 MovieLens 数据集上共进行 100 次实验,统计的平均 RMSE 和 MAE 如图 2 和 3 所示。Cacheda 等人<sup>[5]</sup>经过实验后得出在 MovieLens 数据集上聚成八个簇时算法在精确度和扩展性上得到比较好的折中,本文的实验聚簇也设定八个。

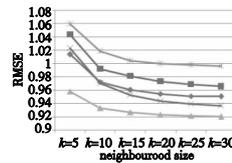


图2 RMSE对比

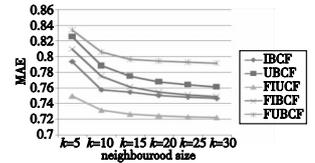


图3 MAE对比

由图可知,本文提出的算法比基于内存的和基于模型的推荐在 RMSE 和 MAE 上有较大的提高,本文提出的算法的确能提高系统的推荐质量。不失一般性,对于诸如购物推荐、书籍推荐的电子商务网站,可首先根据用户特征聚类,在聚类之后运用分步协同过滤框架给每一个聚簇训练一个模型,可有效地解决算法的可扩展性和推荐精度问题。从图 2 和 3 还可以看出,基于内存的推荐性能要优于基于模型的推荐,同时也验证了基于项目的协同过滤推荐性能要优于基于用户的协同过滤。

本文使用了余弦相似性作为基于项目的协同过滤(UBCF)相似性度量方法。如图 4 所示,为了探究不同的相似性度量方法在 IBCF 上起的作用,特设计了三组对比实验,即基于皮尔森相关相似度(PCC)的协同过滤、基于余弦相似性(CS)的协同过滤和基于欧氏距离(ED)的协同过滤。由图 4 可知,基于 CS 的协同过滤要优于其他两种相似性度量方法。主要原因是用户评分的差异,有的用户倾向于打高分,有的用户倾向于打低分。

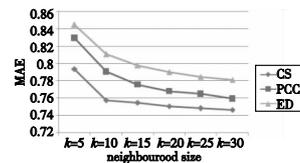


图4 不同的相似性度量在MAE上对比

## 4 结束语

本文提出的先基于用户特征聚类,后用机器学习的方法融合基于项目和基于用户的协同过滤给每一个聚簇训练一个模型,极大地提高了算法的推荐性能,同时在一定程度上解决了传统的算法面临的可扩展性问题。然而当聚类的初始点选择不好或者聚簇个数选择不合适时,比较容易造成局部优化,导致推荐性能的下降。后期工作中,将对如何自动确定聚簇个数和聚类的初始点展开研究。

### 参考文献:

[1] Patra B K, Launonen R, Ollikainen V, et al. A new similarity measure using Bhattacharyya coefficient for collaborative filtering in sparse data[J]. Knowledge-Based Systems, 2015, 82(2): 163-177.

- [6] Green K W, Inman R A, Birou L M, *et al.* Total JIT (T-JIT) and its impact on supply chain competency and organizational performance [J]. *International Journal of Production Economics*, 2014, 147 (1):125-135.
- [7] Ryan S M. Capacity expansion for random exponential demand growth with lead times[J]. *Management Science*, 2004, 50(6):740-748.
- [8] Huang Kai, Ahmed S. The value of multistage stochastic programming in capacity planning under uncertainty[J]. *Operations Research*, 2009, 57(4):893-904.
- [9] Chao Xiuli, Chen Hong, Zheng Shaohui. Dynamic capacity expansion for a service firm with capacity deterioration and supply uncertainty [J]. *Operations Research*, 2009, 57(1):82-93.
- [10] Álvarez-Socarr S A M, Berrones A, Moreno G J, *et al.* Practice summary: enhancing forecasting and capacity planning capabilities in a telecommunications company [J]. *Interfaces*, 2013, 43(4):385-387.
- [11] 施文, 刘志学, 刘丹. 零部件循环取货越库物流系统关键因子研究: 仿真试验设计与分析[J]. *管理科学学报*, 2013, 16(6):22-34.
- [12] Chuah K H, Yingling J C. Routing for a just-in-time supply pickup and delivery system [J]. *Transportation Science*, 2005, 39(3):328-339.
- [13] Alegre J, Laguna M, Pacheco J. Optimizing the periodic pick-up of raw materials for a manufacturer of auto parts[J]. *European Journal of Operational Research*, 2007, 179(3):736-746.
- [14] Shakeri M, Low M Y H, Turner S J, *et al.* A robust two-phase heuristic algorithm for the truck scheduling problem in a resource-constrained crossdock[J]. *Computers & Operations Research*, 2012, 39(11):2564-2577.
- [15] Liao T W, Egbelu P J, Chang P C. Simultaneous dock assignment and sequencing of inbound trucks under a fixed outbound truck schedule in multi-door cross docking operations[J]. *International Journal of Production Economics*, 2013, 141(1):212-229.
- [16] Shi Wen, Liu Zhixue, Shang J, *et al.* Multi-criteria robust design of a JIT-based cross-docking distribution center for an auto parts supply chain[J]. *European Journal of Operational Research*, 2013, 229(3):695-706.
- [17] 于辉, 陈飞平. 基于供应链协同的汽车制造企业入厂物流模式选择[J]. *系统工程理论与实践*, 2011, 31(7):1230-1239.
- [18] Kaneko J, Nojiri W. The logistics of Just-in-time between parts suppliers and car assemblers in Japan[J]. *Journal of Transport Geography*, 2008, 16(3):155-173.
- [19] Schittekat P, Sörensen K. OR practice—supporting 3PL decisions in the automotive industry by generating diverse solutions to a large-scale location-routing problem[J]. *Operations Research*, 2009, 57(5):1058-1067.
- [20] Jr Reeves K A, Caliskan F, Ozcan O. Outsourcing distribution and logistics services within the automotive supplier industry[J]. *Transportation Research Part E: Logistics and Transportation Review*, 2010, 46(3):459-468.
- [21] Kelton W D, Sadowski R P, Sturrock D T. *Simulation with ARENA* [M]. 4th ed. New York: McGraw-Hill, 2007.
- [22] Cachon G P, Terwiesch C. *Matching supply with demand* [M]. 3rd ed. New York: McGraw-Hill, 2011.
- [23] Kleijnen J P C. *Design and analysis of simulation experiments* [M]. 2nd ed. New York: Springer, 2015.
- (上接第 页)
- [2] Zhao Zhidan, Shang Mingsheng. User-based collaborative-filtering recommendation algorithms on Hadoop[C]//Proc of the 3rd International Conference on Knowledge Discovery and Data Mining. 2010: 478-481.
- [3] Barragúns-Martínez A B, Costa-Montenegro E, Burguillo J C, *et al.* A hybrid content-based and item-based collaborative filtering approach to recommend TV programs enhanced with singular value decomposition[J]. *Information Sciences*, 2010, 180(22):4290-4311.
- [4] Wang Jing, Yin Jian. Optimized item-based collaborative filtering recommendation algorithm[J]. *Journal of Chinese Computer Systems*, 2010, 31(12):2337-2342.
- [5] Cacheda F, Carneiro V, Fernandez D, *et al.* Comparison of collaborative filtering algorithms: limitations of current techniques and proposals for scalable, high-performance recommender systems[J]. *ACM Trans on the Web*, 2011, 5(1):161-171.
- [6] Sehgal M S B, Gondal I, Dooley L. Stacked regression ensemble for cancer class prediction[C]//Proc of IEEE International Conference on Industrial Informatics. 2005:831-835.
- [7] Zhao Jiashu, Huang J X, Hu Xiaohua. BPLT + : a Bayesian-based personalized recommendation model for health care[J]. *BMC Genomics*, 2013, 14(4):S6.
- [8] Zhang Tao, Zhang Qi, Liang Dequn. A novel probabilistic latent semantic analysis based image blur metric[C]//Proc of the 12th IEEE International Conference on Dependable, Autonomic and Secure Computing. [S. l]: IEEE Press, 2014:310-315.
- [9] Ge Shien, Ge Xinyang. An SVD-based collaborative filtering approach to alleviate cold-start problems[C]//Proc of the 9th International Conference on Fuzzy Systems and Knowledge Discovery. [S. l]: IEEE Press, 2012:1474-1477.
- [10] Kanimozhi M, Logeshwari K, Saranya K, *et al.* Clustering for collaborative filtering application in Web recommendations[J]. *International Journal of Computer Science & Mobile Computing*, 2013, 2(4):217-222.
- [11] Huang Guoyan, Li Youchao, Gao Jianpei, *et al.* Collaborative filtering recommendation algorithm based on user clustering of item attributes[J]. *Computer Engineering & Design*, 2010, 31(5):1038-1041.
- [12] Xue Guirong, Lin Chenxi, Yang Qiang, *et al.* Scalable collaborative filtering using cluster-based smoothing[C]//Proc of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2005:114-121.
- [13] Gao Huiji, Tang Jiliang, Hu Xia, *et al.* Exploring temporal effects for location recommendation on location-based social networks[C]//Proc of the 7th ACM Conference on Recommender Systems. New York: ACM Press, 2013:93-100.
- [14] Herlocker J L, Konstan J A, Terveen L G, *et al.* Evaluating collaborative filtering recommender systems[J]. *ACM Trans on Information Systems*, 2004, 22(1):5-53.
- [15] Koren Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model[C]//Proc of the 14th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2008:426-434.
- [16] 冯旻远. 综合用户特征的协同过滤推荐算法的研究[D]. 南京: 南京邮电大学, 2014.
- [17] 吴湖, 王永吉, 王哲, 等. 两阶段联合聚类协同过滤算法[J]. *软件学报*, 2010, 21(5):1042-1054.