

基于项目近邻的约束概率矩阵分解算法

张天杰, 曹苏燕, 闫世洋

(南京邮电大学 计算机学院, 江苏 南京 210003)

摘要:在个性化推荐领域,协同过滤是目前最为成功应用最为广泛的推荐技术之一。约束概率矩阵分解算法便是一种基于模型的协同过滤算法,它能有效地面对推荐系统中遇到的海量数据问题,保证推荐的实时性。然而,传统的约束概率矩阵分解算法并没有考虑用户或者项目之间的关系,使得算法的推荐质量受到影响。为进一步提高算法推荐的质量,文中在约束概率矩阵分解算法模型的基础上引入项目近邻关系,通过结合从项目简介中提取的固有特征和用户评定的标签特征两方面信息来确定项目的最近邻居集合,并将该邻居集合融合到基于约束的概率矩阵分解模型中产生推荐。通过在真实的数据集上的验证结果表明,该算法能够更有效地预测用户对项目的评分,提高算法的推荐精度。

关键词:推荐系统;协同过滤;约束概率矩阵分解;项目近邻

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2016)10-0064-05

doi:10.3969/j.issn.1673-629X.2016.10.014

A Constrained Probabilistic Matrix Factorization Algorithm Based on Item-neighborhood

ZHANG Tian-jie, CAO Su-yan, YAN Shi-yang

(College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: Collaborative filtering is one of the most successful applications in the field of personalized recommendation. Constrained probabilistic matrix factorization is a model-based collaborative filtering algorithm which can effectively deal with the problem of scalability in large-scale recommendation system and guarantee the real-time of recommendation. However, the traditional one does not consider the relationship between the users or the items, which makes the quality of the algorithm affected. It takes the relationship of item-neighborhood into the algorithm model of constrained probability matrix factorization to improve the quality of the proposed algorithm. To guarantee the accuracy of item-neighborhood, the inherent features extracted from the item's summary and the tag of user marked for the item are used to get the set of the nearest neighbor for the items, then the item-neighborhood set is applied into the framework of the constrained probabilistic matrix factorization algorithm. The experiments on real datasets show that the proposed algorithm can predict the user's rating on the item more effectively, and improve the accuracy of the recommendation.

Key words: recommendation system; collaborative filtering; constrained probabilistic matrix factorization; item-neighborhood

0 引言

在推荐系统^[1-3]领域,协同过滤推荐(Collaborative Filtering)以其推荐策略简单、实用、适用对象多样的特性而备受青睐。根据使用策略的不同,协同过滤可以分为基于内存的协同过滤(Memory based)和基于模型的协同过滤(Model based)^[4]。基于内存的协同过滤主要包括基于用户^[5]和基于项目^[6]两类,两者的重点分别为寻找相似的用户和项目。然而,基于内存的协同过滤对于用户项目评分矩阵的稀疏性比较敏感,并且随着系统用户、项目量的增大,计算量呈非线性增

长趋势,不利于实时性推荐。基于模型的协同过滤推荐通过采用线下学习、周期性更新的方式对评分数据进行挖掘。目前,常见的基于模型的协同过滤方法有聚类模型方法^[7]、贝叶斯网络^[8]、矩阵分解^[9]等。相比较基于内存的协同过滤算法,这类算法具有更好的扩展性,在面对海量数据时推荐更具实时性。约束概率矩阵分解算法^[10](Constrained Probabilistic Matrix Factorization, CPMF)便是一种基于模型的协同过滤推荐的经典算法,其可以在线性时间内产生推荐,相比较了其他推荐策略能更有效地应对评分稀疏、海量数据等

收稿日期:2016-01-10

修回日期:2016-04-13

网络出版时间:2016-09-19

基金项目:国家“863”高技术发展计划项目(2006AA01Z201)

作者简介:张天杰(1992-),男,硕士研究生,研究方向为数据挖掘、大数据、云计算。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20160919.0842.054.html>

问题^[10]。

然而,传统的 CPMF 算法模型并未考虑用户或者项目之间的关系,算法仅使用了用户对项目的评分信息,忽略了很多额外的信息,从而影响了实际的推荐效果。近年来,学者们^[11-13]从评分矩阵信息以外的信息中挖掘用户或项目的关系并将其用于协同过滤算法中,从而实现原有算法推荐效果的提升。

考虑到项目的关系往往比用户的关系更具稳定性,不需要频繁更新,对于大型系统更具优势^[6],文中从项目角度出发,提出了一种基于项目近邻的约束概率矩阵分解算法(Item-neighborhood-based Constrained Probabilistic Matrix Factorization, ICPMF)。在 ICPMF 中,结合从项目简介中提取固有特征和用户对项目评定的标签特征两方面信息来挖掘项目之间相互影响的关系,将其量化为项目之间的相似度,通过相似度的大小来确定项目的近邻集合,并将该邻居集合融合到 CPMF 中实现推荐。

1 CPMF 算法概述

约束概率矩阵分解算法是由 Salakhutdinov 等^[10]提出的一种基于模型的协同过滤推荐策略,是一种潜在特征因子分解算法。该算法假设每个用户的兴趣只受几个潜在的特征影响,并假设这些特征向量符合均值为 0 的高斯先验分布,通过用户对项目的评分记录来学习获得每个用户和项目的潜在特征向量,最后利用得到的低维度的潜在特征向量矩阵计算出用户对项目的预测评分,进而产生推荐。CPMF 能够通过线下学习潜在特征,线上直接使用学习到的潜在特征向量计算出用户对某个项目的预测评分,从而可以快速地为用户产生要推荐的项目集合。实验结果表明,CPMF 在推荐准确性和抗稀疏性上和其他只使用用户-项目评分矩阵的协同过滤算法比较更具优势^[10]。

具体而言,假设在某系统中存在 N 个用户,其构成的集合为 $\{u_1, u_2, \dots, u_N\}$, 及 M 个要用于推荐的项目,其对应的集合为 $\{v_1, v_2, \dots, v_M\}$ 。设 $\mathbf{U} \in R^{D \times N}$, $\mathbf{V} \in R^{D \times M}$, 其中 \mathbf{U} 为用户潜在特征矩阵, \mathbf{V} 为项目潜在特征矩阵, D 为潜在特征向量的维度, \mathbf{U}_i 为用户 i 的特征向量, \mathbf{V}_j 为项目 j 的特征向量。考虑到评分项目集合相似的用户可能具有相似的兴趣, Salakhutdinov 等^[10]对用户的潜在特征向量进行改进,如式(1)所示。

$$\mathbf{U}_i = \mathbf{Y}_i + \frac{\sum_{k=1}^M I_{ik} \mathbf{W}_k}{\sum_{k=1}^M I_{ik}} \quad (1)$$

其中, $\mathbf{W} \in R^{D \times M}$ 为潜在相似约束矩阵, \mathbf{W}_k 为 \mathbf{W} 的第 k 列, $\mathbf{Y}_i \in R^{1 \times D}$, \mathbf{Y}_i 代表 \mathbf{Y} 的第 i 列; I 为指示函

数,即如果用户 i 对项目 k 有评分,则 I_{ik} 为 1, 否则 I_{ik} 为 0, 当用户 i 没有评分时 $\mathbf{U}_i = \mathbf{Y}_i$ 。

根据以上的定义,已观察到的评分数据的条件概率定义如式(2)。

$$p(R | \mathbf{Y}, \mathbf{V}, \mathbf{W}, \sigma^2) = \prod_{i=1}^N \prod_{j=1}^M [N(R_{ij} | g([\mathbf{Y}_i + \frac{\sum_{k=1}^M I_{ik} \mathbf{W}_k}{\sum_{k=1}^M I_{ik}}]^T \mathbf{V}_j), \sigma^2)]^{I_{ij}} \quad (2)$$

其中, $g(x) = 1/(1 + e^{-x})$ 用于将预测评分区间映射到 $[0, 1]$ 上; R_{ij} 对应用户 i 对项目 j 的评分数据,通过函数 $t(x) = (x - 1)/(K - 1)$ 将其从 1 到 K 映射到 $[0, 1]$ 区间上; $\mathbf{Y}_i, \mathbf{W}_k, \mathbf{V}_j$ 先验服从均值为 0 的高斯先验分布且相互独立,即如式(3)、(4)、(5)所示。

$$p(\mathbf{Y} | \sigma_Y^2) = \prod_{i=1}^N N(\mathbf{Y}_i | 0, \sigma_Y^2 \mathbf{I}) \quad (3)$$

$$p(\mathbf{W} | \sigma_W^2) = \prod_{k=1}^M N(\mathbf{W}_k | 0, \sigma_W^2 \mathbf{I}) \quad (4)$$

$$p(\mathbf{V} | \sigma_V^2) = \prod_{j=1}^M N(\mathbf{V}_j | 0, \sigma_V^2 \mathbf{I}) \quad (5)$$

由贝叶斯推理可得, $\mathbf{Y}, \mathbf{V}, \mathbf{W}$ 的后验概率分布如式(6)所示。

$$p(\mathbf{Y}, \mathbf{W}, \mathbf{V} | R, \sigma_Y^2, \sigma_V^2, \sigma_W^2, \sigma^2) \propto p(R | \mathbf{Y}, \mathbf{V}, \mathbf{W}, \sigma^2) p(\mathbf{Y} | \sigma_Y^2) p(\mathbf{W} | \sigma_W^2) p(\mathbf{V} | \sigma_V^2) \quad (6)$$

将式(3)、(4)、(5)带入式(6),对上述的后验概率取对数后并最大化后验,化简后等价于式(7)。

$$E = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^M I_{ij} (R_{ij} - g([\mathbf{Y}_i + \frac{\sum_{k=1}^M I_{ik} \mathbf{W}_k}{\sum_{k=1}^M I_{ik}}]^T \mathbf{V}_j))^2 + \frac{\lambda_Y}{2} \sum_{i=1}^N \mathbf{Y}_i^T \mathbf{Y}_i + \frac{\lambda_V}{2} \sum_{j=1}^M \mathbf{V}_j^T \mathbf{V}_j + \frac{\lambda_W}{2} \sum_{k=1}^M \mathbf{W}_k^T \mathbf{W}_k \quad (7)$$

其中, $\lambda_Y = \sigma^2/\sigma_Y^2, \lambda_V = \sigma^2/\sigma_V^2, \lambda_W = \sigma^2/\sigma_W^2$, 通过梯度下降法,结合用户-项目评分矩阵,即可求出上述目标函数的最优解 $\mathbf{Y}, \mathbf{V}, \mathbf{W}$ 。

CPMF 算法的图模型^[10]如图 1 所示。

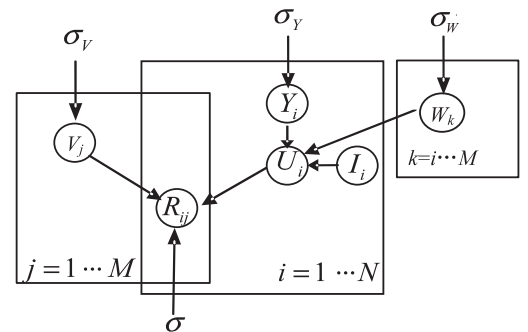


图 1 约束概率矩阵分解模型

2 ICPMF 算法描述

结合前文对于 CPMF 算法的描述,文中提出的 ICPMF 算法共包括两个过程:根据项目的简介和标签信息进行项目近邻选择;根据相似性假设将项目近邻关系融入到 CPMF 算法中进行特征向量学习,得到用户和项目的潜在特征向量。

ICPMF 算法的两个过程皆可由线下计算完成,并进行周期性更新。

2.1 项目近邻的选择

一般而言,项目简介是系统为用户介绍特定的项目内容而设,如图书有图书内容简介、电影有剧情简介等。项目简介一般由一段简短的文本描述,是对项目特征的一段客观描述,通过项目简介用户可以了解项目的内容和它所具有的一些特征。所以作为为用户导航的一种方式,项目简介中往往蕴含了项目所具有的一些固有特性,即客观特性。标签则一般由使用过某个项目的用户给出,是用来描述信息的关键词,可以实现对信息的分类^[14],对于每一个项目而言标签往往反映了用户对其使用过项目的主观感受,对于同样的项目而言,不同的用户可能会给出截然不同的标签,所以系统在收集用户为项目添加的标签信息时往往会附带标签的数量,从而能够更好地衡量标签对项目的价值。

考虑到以上情况,为了更好地构建项目的特征,进而寻找到更加合理的项目近邻关系,文中结合项目简介和用户对项目打的标签两方面数据对项目近邻进行计算和选择。具体的近邻选择过程包括:

- (1) 项目特征的选取和权重计算得到项目的特征向量;
- (2) 针对任一项目 $Item_j$,根据特征向量计算其与系统中任一其他项目 $Item_l$ 的项目间的相似度得到 $S_{(j,l)}$;
- (3) 选择相似度最大的 L 个项目作为项目 j 的近邻 N_j 。

具体步骤如下:

Step1:对项目 j 的简介内容进行中文分词(文中采用 IKAnalyzer 中文分词器),过滤停止词和虚词,并对剩余的词进行词频统计。

Step2:以 Step1 统计出的词作为项目的一组特征,并根据对应的词频,计算 TF-IDF^[15] 值作为词权重、计算项目标签的 TF-IDF 值作为标签的权重。TF-IDF 计算公式如下:

$$TF - IDF = \frac{TF(t, Item_j)}{|Item_j|} \times \log\left(\frac{N}{n_t + 1}\right) \quad (8)$$

其中, $TF(t, Item_j)$ 为特征 t (词或标签) 在项目 $Item_j$ 中的频度 (词为词频, 标签为被标记的个数); $|Item_j|$ 为项目 j 所有特征的频度之和; N 为训练集

中项目的个数; n_t 为训练集中出现 t 的项目数。

Step3:将 Step1 和 Step2 中统计和计算出的特征和权重,按照向量空间模型 (Vector Space Model, VSM)^[16] 法得到项目 j 的特征向量,即 $\vec{Item}_j = (w_{j1}, w_{j2}, \dots, w_{jn})$ 。其中, n 为向量空间模型的维度; w_{jn} 为第 n 个特征的权重 (TF-IDF 值)。

Step4:计算项目 j 与任一其他项目 l 的余弦相似度得到 $S_{(j,l)}$,如式(9)所示,并选取和项目 j 相似度最大的 L 个项目集合作为项目 j 的近邻 N_j 。

$$S_{(j,l)} = \cos(\vec{Item}_j, \vec{Item}_l) = \frac{\sum_{k=1}^n w_{jk}w_{lk}}{\sqrt{(\sum_{k=1}^n w_{jk}^2)(\sum_{k=1}^n w_{lk}^2)}} \quad (9)$$

2.2 ICPMF 算法模型

在得到项目的近邻和它们之间的相似度后,假设相似的项目应该有相似的潜在特征向量。基于这样的假设,则对于任意的项目 j ,它的潜在特征向量可近似地表示为: $\vec{V}_j = \sum_{l \in N_j} S_{(j,l)} V_l$ 。

其中, N_j 为项目 j 的邻居集合; $S_{(j,l)}$ 为项目 j 与项目 l 的相似度,即权重参数。

ICPMF 算法的模型图如图 2 所示。

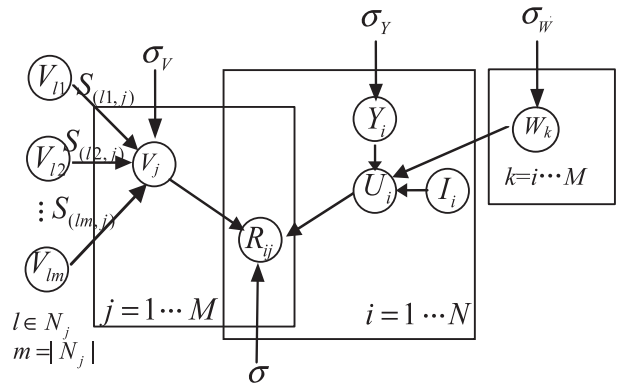


图 2 ICPMF 算法模型

具体地,对于项目的潜在特征先验概率可以表示为式(10),与式(7)类似通过贝叶斯推理,后验概率如式(11)所示,将式(2)、(3)、(4)、(10)带入后,进行对数处理并且最大化,除去常数项后等价于最小化公式(12)。

$$p(\mathbf{V} | \mathbf{S}, \sigma_s^2, \sigma_v^2) \propto p(\mathbf{V} | \sigma_v^2) \times p(\mathbf{V} | \mathbf{S}, \sigma_s^2) = \prod_{j=1}^M N(\mathbf{V}_j | \mathbf{0}, \sigma_v^2 \mathbf{I}) \times \prod_{j=1}^M N(\mathbf{V}_j | \sum_{l \in N_j} S_{(j,l)} \mathbf{V}_l, \sigma_s^2 \mathbf{I}) \quad (10)$$

$$p(\mathbf{Y}, \mathbf{W}, \mathbf{V} | \mathbf{R}, \mathbf{S}, \sigma_y^2, \sigma_v^2, \sigma_w^2, \sigma^2) \propto p(\mathbf{R} | \mathbf{Y}, \mathbf{V}, \mathbf{W}, \sigma^2) p(\mathbf{Y} | \sigma_y^2) p(\mathbf{W} | \sigma_w^2) p(\mathbf{V} | \mathbf{S}, \sigma_s^2, \sigma_v^2) \quad (11)$$

$$\begin{aligned}
 E = & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^M I_{ij} (R_{ij} - g([\mathbf{Y}_i + \frac{\sum_{k=1}^M I_{ik} \mathbf{W}_k}{\sum_{k=1}^M I_{ik}}]^T \mathbf{V}_j))^2 + \\
 & \frac{\lambda_Y}{2} \sum_{i=1}^N \mathbf{Y}_i^T \mathbf{Y}_i + \frac{\lambda_V}{2} \sum_{j=1}^M \mathbf{V}_j^T \mathbf{V}_j + \frac{\lambda_W}{2} \sum_{k=1}^M \mathbf{W}_k^T \mathbf{W}_k + \\
 & \frac{\lambda_S}{2} \sum_{j=1}^M (\mathbf{V}_j - \sum_{l \in N_j} S_{(j,l)} \mathbf{V}_l)^T (\mathbf{V}_j - \sum_{l \in N_j} S_{(j,l)} \mathbf{V}_l)
 \end{aligned} \tag{12}$$

其中, $\lambda_Y = \sigma^2 / \sigma_Y^2, \lambda_V = \sigma^2 / \sigma_V^2, \lambda_W = \sigma^2 / \sigma_W^2, \lambda_S = \sigma^2 / \sigma_S^2$, 结合用户-项目评分矩阵, 通过梯度下降法可求得式(12)的最优解, 得到潜在特征向量。

3 实验

3.1 实验数据

为了探究文中所提算法的推荐效果, 从豆瓣读书上抓取了真实的数据集作为实验数据。该数据集共包含三部分: 34 250 个用户对 10 897 本图书所做出的评分信息共 258 111 条, 其中评分区间为 [1, 5] 的整数; 用户对图书添加的标签共 23 084 个, 并附带了每个标签被用户添加的次数; 评分信息中每本图书的简介信息, 图书简介信息平均在 300 字左右。

实验中使用了开源的分词工具 IK Analyzer 2012 对图书简介信息进行分词并进行停止词过滤。实验采用 10 折交叉实验的方式, 将总数据集分成 10 份, 每次取 1 份作为测试集, 剩余作为训练集, 迭代 10 次后, 取 10 次评价标准的平均值作为算法最终的评测结果。

3.2 评价标准

文中使用均方根误差 (RMSE) 作为算法准确性的评价标准, 均方根误差越小, 表示预测的准确度越高。对于测试集中的一个用户 i 和项目 j , 令 R_{ij} 为用户 i 对项目 j 的实际评分, P_{ij} 为推荐算法给出的预测评分, T 表示测试数据集, $|T|$ 表示测试数据集中的评分数量, 那么 RMSE 可表示为:

$$\text{RMSE} = \sqrt{\frac{\sum_{R_{ij} \in T} (R_{ij} - P_{ij})^2}{|T|}}$$

3.3 实验结果及分析

实验共包括两部分: (1) 比较 CPMF、TCPMF、ICPMF 算法在不同的特征向量维度 D 下的推荐准确度情况, 其中 TCPMF 为只使用标签信息计算项目近邻的 ICPMF 算法, ICPMF 则融合了项目简介和标签两方面信息进行项目近邻计算。(2) 测试参数 λ_S 对于 ICPMF 的影响。实验中设定所使用到的参数 $\lambda_Y = \lambda_V = \lambda_W = 0.001$ 。

图 3 为三种算法在不同的特征维度 D 下的推荐准确度情况, 其中 TCPMF 和 ICPMF 的项目近邻个数 $L =$

25, 参数 $\lambda_S = 5$ 。

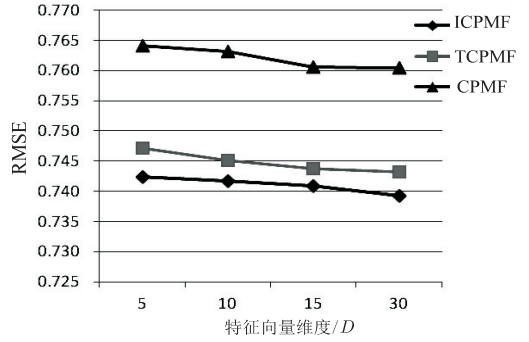


图 3 不同特征向量维度 D 下各算法结果比较

由图 3 可知, 融入了项目近邻信息的 ICPMF 和 TCPMF 的推荐准确度明显好于 CPMF, ICPMF 好于 TCPMF, 这说明文中将项目近邻关系融入 CPMF 算法进行推荐的有效性, 同时也表明在项目近邻计算时融合从项目简介中提取固有特征和标签特征这两方面信息进行项目近邻计算的合理性。

图 4 比较了不同的 λ_S 取值对 ICPMF 算法的影响, 实验中将 λ_S 的值分别设为 0.1, 0.5, 1, 5, 10, 25, 并设用户近邻个数 $L = 25$, 特征向量维度 $D = 10$ 。

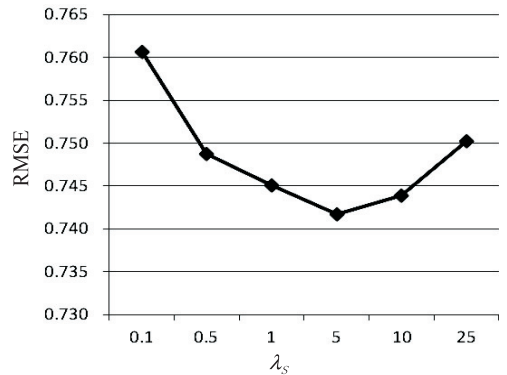


图 4 λ_S 对 ICPMF 的影响

由图 4 可知, λ_S 对 ICPMF 的推荐准确度产生了较明显的影响, 当 λ_S 在 [0.1, 5] 上变化时, ICPMF 的推荐准确率随着 λ_S 的增大而提高, 当 λ_S 大于 5 时算法的准确率随之下降, 由此可知项目近邻的引入对改进 CPMF 算法的有效性。

4 结束语

文中通过结合项目简介和用户为项目添加的标签两方面信息来挖掘系统中项目的近邻关系, 并将该近邻关系融入 CPMF 算法中进行评分预测。在真实的数据集上的实验结果表明, 提出的方法和传统的 CPMF 算法相比能够有效地提高推荐算法评分预测的准确度, 进而验证了文中项目近邻关系计算方法的合理性和将此近邻关系融入 CPMF 算法的有效性。然而, 实验中也发现, 训练过程中存在特征向量初始值设置问题以及如何防止模型过拟合问题, 这些值得进一步

研究。

参考文献:

- [1] Ricci F, Rokach L, Shapira B. Introduction to recommender systems handbook[M]. US:Springer,2011.
- [2] 王国霞,刘贺平.个性化推荐系统综述[J].计算机工程与应用,2012,48(7):66-76.
- [3] 任磊.推荐系统关键技术研究[D].上海:华东师范大学,2012.
- [4] 刘士琛.面向推荐系统的关键问题研究及应用[D].合肥:中国科学技术大学,2014.
- [5] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(6): 734-749.
- [6] Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms[C]//Proceedings of the 10th international conference on world wide web. [s. l.]: ACM, 2001; 285-295.
- [7] O'Connor M, Herlocker J. Clustering items for collaborative filtering[C]//Proceedings of the ACM SIGIR workshop on recommender systems. UC Berkeley: ACM, 1999.
- [8] Miyahara K, Pazzani M J. Collaborative filtering with the simple Bayesian classifier[M]//PRICAI 2000 topics in artificial intelligence. Berlin: Springer, 2000; 679-689.
- [9] Goldberg K, Roeder T, Gupta D, et al. Eigentaste: a constant time collaborative filtering algorithm[J]. Information Retrieval, 2001, 4(2): 133-151.
- [10] Mnih A, Salakhutdinov R. Probabilistic matrix factorization[C]//Proc of advances in neural information processing systems. [s. l.]: [s. n.], 2007; 1257-1264.
- [11] Ma H, Yang H, Lyu M R, et al. SoRec: social recommendation using probabilistic matrix factorization[C]//Proc of international conference on information & knowledge management. [s. l.]: ACM, 2008; 931-940.
- [12] 孙光福,吴乐,刘淇,等.基于时序行为的协同过滤推荐算法[J].软件学报,2013,24(11):2721-2733.
- [13] 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. [s. l.]: ACM, 2008; 1995-1999.
- [14] Du W H, Rau J W, Huang J W, et al. Improving the quality of tags using state transition on progressive image search and recommendation system[C]//Proc of IEEE international conference on systems, man, and cybernetics. [s. l.]: IEEE, 2012; 3233-3238.
- [15] Joachims T. A probabilistic analysis of the Rocchio algorithm with TFIDF for text categorization[R]. USA: Carnegie-Mellon University, 1996.
- [16] Salton G, Wong A, Yang C S. A vector space model for automatic indexing[M]. [s. l.]: Morgan Kaufmann Publishers Inc., 1997.
- (上接第 63 页)
- [3] Pei T, Zhang H, Li Z, et al. Survey of community structure segmentation in complex networks[J]. Journal of Software, 2014, 9(1): 89-93.
- [4] Newman M E J. The structure of scientific collaboration networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2001, 98: 404-409.
- [5] Li Z, Zhang S, Wang R S, et al. Quantitative function for community detection[J]. Physical Review E, 2008, 77: 036109.
- [6] Yang B, Liu D Y, Liu J, et al. Complex network clustering algorithms[J]. Journal of Software, 2009, 20(1): 54-66.
- [7] Gong M G, Cai Q, Chen X W, et al. Complex network clustering by multiobjective discrete particle swarm optimization based on decomposition[J]. IEEE Transactions on Evolutionary Computation, 2014, 18(1): 82-97.
- [8] Leskovec J, Lang K J, Mahoney M. Empirical comparison of algorithms for network community detection[C]//Proc of 19th international world wide web conference. Raleigh, NC, United States: [s. n.], 2010; 631-640.
- [9] Tong C, Niu J W, Dai B, et al. Complex networks clustering algorithm based on the core influence of the nodes[C]//Proc of 2012 IEEE 31st international performance computing and communications conference. [s. l.]: IEEE, 2012; 185-186.
- [10] Liu X, Li D, Wang S, et al. Effective algorithm for detecting community structure in complex networks based on GA and clustering[C]//Proc of 7th international conference on computational science. Beijing, China: [s. n.], 2007; 657-664.
- [11] Tong C, Niu J W, Dai B, et al. A novel complex networks clustering algorithm based on the core influence of nodes[J]. Scientific World Journal, 2014, 2014: 801854.
- [12] Liu Y H, Jin J Z, Zhang Y, et al. A new clustering algorithm based on data field in complex networks[J]. Journal of Supercomputing, 2014, 67(3): 723-737.
- [13] Danon L, Díaz-Guilera A, Duch J, et al. Comparing community structure identification[J]. Journal of Statistical Mechanics Theory & Experiment, 2005, 2005: P09008.
- [14] Lusseau D, Schneider K, Boisseau O J, et al. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations[J]. Behavioral Ecology & Sociobiology, 2003, 54: 396-405.
- [15] Newman M E. Finding community structure in networks using the eigenvectors of matrices[J]. Physical Review E, 2006, 74: 036104.
- [16] Girvan M, Newman M E J. Community structure in social and biological networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2002, 99: 7821-7826.