

基于情境相似度和二次聚类的协同过滤推荐算法

蔡海尼 覃梦秋 文俊浩 熊庆宇 黎懋靛

(重庆大学软件学院 重庆 400044)

摘要 随着移动互联网规模的不断扩大,传统推荐系统因较少考虑多种情境因素和用户置信度对用户偏好预测的综合影响,造成了推荐算法预测结果的偏差。针对此问题,将情境信息引入个性化推荐的过程中,提出一种基于情境相似度和二次聚类的协同过滤算法。该算法首先根据用户情境的相似度对用户进行初始聚类,再基于评分矩阵计算用户评分置信度,将用户分为核心用户和非核心用户;然后根据核心用户评分对初始聚类的簇心进行调整,并对簇中非核心用户进行重聚类,形成新的聚簇;最终根据情境相似度对用户偏好进行预测。该算法可以在一定程度上降低评分矩阵中的噪点对聚类结果的影响,提高了推荐结果的准确性。基于实际数据集的仿真实验表明,该算法与传统协同过滤算法相比能够有效提高用户偏好预测的准确性,增加协同过滤推荐算法的精确度。

关键词 推荐系统,情境相似度,协同过滤,核心用户,二次聚类

中图分类号 TP312 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.4.043

Collaborative Filtering Recommendation Algorithm Based on Context Similarity and Twice Clustering

CAI Hai-ni QIN Meng-qiu WEN Jun-hao XIONG Qing-yu LI Mao-liang

(School of Software Engineering, Chongqing University, Chongqing 400044, China)

Abstract Aiming at solving the problem of personalized service recommendation in the field of mobile telecommunication network, this paper introduced the context information to the process of personalized recommendation, and proposed a collaborative filtering algorithm based on context similarity and twice clustering. Firstly, according to user context similarity, the users are clustered, and user rating confidence is calculated based on the rating matrix to distinguish core users from non-core users. Secondly, the center of clusters formed by initial clustering can be adjusted according to the rating of core users, and non-core users will be clustered again to form a new cluster. Finally, according to the context similarity, user preferences will be predicted. To some extent, this algorithm can reduce the influence of noise data from rating matrix on clustering results, and reduce the deviation of the recommendation. The experiment based on the simulation data set shows that, the algorithm improves the accuracy of user preferences effectively, and increases the accuracy of collaborative filtering recommendation.

Keywords Recommendation system, Context similarity, Collaborative filtering, Core users, Twice clustering

1 引言

随着互联网和移动通信技术的迅猛发展,以及通信技术和传感器技术的广泛应用,人们越来越频繁地通过互联网和移动设备来获取信息,情景感知服务在这样的背景下产生并极大地改变着人们的生活。情境也可以称为上下文,是指任何可以用来刻画实体特征的信息,其实体可以是人、位置或是与用户和应用交互相关的物理的或虚拟的对象,包括用户和应用本身^[1]。情境感知(context-awareness)也称为上下文感知,最早是源于普适计算(ubiquitous computing),由 B. N. Schilit 于 1994 年提出^[2]。它通过使用传感器、无线通讯

设备或者相关的信息获取设备和技术,使计算机或智能移动设备感知到当前的情境和环境。

推荐系统通过识别和预测用户偏好来提供个性化服务,是满足人们多元化推荐需求、解决信息超载问题的有效工具。推荐算法是推荐系统的核心,根据其特点和适用领域,推荐算法可以归纳为以下几类:基于内容的推荐算法、协同过滤推荐算法、基于规则的推荐算法和混合推荐算法等。其中协同过滤推荐算法(Collaborative Filtering, CF)是当前个性化推荐领域中应用最广泛的算法之一;而 Katsuhiko 等提出了基于聚类的协同过滤算法^[3],在一定程度上提高了推荐系统的鲁棒性和可扩展性。

到稿日期:2015-03-29 返修日期:2015-07-07 本文受国家自然科学基金(61379158,71301177),中央高校基本科研业务费科研专项(CDJXS11181161)资助。

蔡海尼(1972-),女,博士,副教授,主要研究方向为服务计算、面向服务的软件工程, E-mail: hainicai@cqu.edu.cn;覃梦秋(1989-),女,硕士生,主要研究方向为个性化服务推荐、软件架构, E-mail: qinmengqiu@cqu.edu.cn;文俊浩(1969-),男,教授,博士生导师,主要研究方向为软件过程与管理、数据挖掘与商务智能等;熊庆宇 男,教授,博士生导师,主要研究方向为智能感知与智能计算、普适计算与嵌入式系统等;黎懋靛(1991-),男,硕士生,主要研究方向为 Web 数据挖掘、个性化服务推荐。

近年来,国内外学者在 Web 服务个性化方面进行了大量的研究。Arkady Zaslavsky^[4]提出移动助理在情境感知应用中遇到的挑战,包括情境发现、提取、解析、应用等。Dong Liu^[5]等人通过分析服务的上下文环境,提出上下文相关的服务推荐框架。Esbensen H^[6]等人将项目按照情境进行分类,再引入协同过滤算法进行用户偏好的计算。这些研究在某些特定的领域有着比较好的效果,但随着移动互联网规模的不断扩大,传统推荐系统较少考虑多种情境因素对用户偏好的综合影响,因而影响了推荐算法的有效性。

本文从评分矩阵出发,提出一种基于情境相似度和二次聚类的协同过滤算法,对用户偏好进行预测。本算法从情境感知的角度出发,在基于聚类的协同过滤算法的聚类过程中将核心用户和非核心用户区分开来,从而削弱了非核心用户对算法结果的影响,有效提高了推荐算法的精确度。

2 相关研究

2.1 协同过滤算法

协同过滤算法是个性化推荐领域应用最广泛的推荐算法。传统的协同过滤算法主要是寻找与目标用户兴趣爱好相似的用户集,并根据用户集中喜爱的项目,对目标用户的需求进行预测和推荐^[7]。常见的协同过滤推荐算法主要包括基于用户的协同过滤算法(User-based Collaborative Filtering)和基于项目的协同过滤算法(Item-based Collaborative Filtering)^[8,9]。相似度计算是协同过滤推荐算法的核心,其主要的度量方法包括余弦相似性、关联度(Pearson 相关系数)和修正的余弦相似性^[10]。其中 Pearson 相关系数最常用,主要是通过和目标用户 u 相似的用户 v 对项目 i 的评分来预测用户 u 对项目 i 的喜好。计算方法如式(1)所示。

$$sim(u, v) = \frac{\sum_{i \in I_{uv}} (R_{(u,i)} - \bar{R}_u)(R_{(v,i)} - \bar{R}_v)}{\sqrt{\sum_{i \in I_{uv}} (R_{(u,i)} - \bar{R}_u)^2 \times \sum_{i \in I_{uv}} (R_{(v,i)} - \bar{R}_v)^2}} \quad (1)$$

其中, $sim(u, v)$ 表示用户的相似性系数, $R_{(u,i)}$ 和 $R_{(v,i)}$ 表示用户 u 和 v 对项目 i 的评分, I_{uv} 表示用户 u 和 v 共同评过分的项集合, \bar{R}_u 和 \bar{R}_v 分别表示用户 u 和 v 对项目的平均评分。

评分预测计算式^[7]如下所示。

$$P_{(u,i)} = \frac{\sum_{u' \in U_u} [sim(u, u') \times R_{(u',i)}]}{\sum_{u' \in U_u} |sim(u, u')|} \quad (2)$$

其中, $P_{(u,i)}$ 表示用户 u 对 i 的预测评分, U_u 表示用户 u 的最近邻聚集。

2.2 基于聚类的推荐算法

聚类是指通过一些聚类算法将用户或者项目聚集成若干具有共同特性的类别,然后计算每个用户在聚类内部的相似度,以此生成类成员内部的推荐向量,最终形成推荐^[11]。K-means 聚类算法是一种比较经典和有效的聚类算法,其主要步骤如下:

(1) n 维数据集合的基础上,随机选择 k 个元素,形成 k 个聚类的质心;

(2) 根据不同聚集类别中生成的聚类中心点,分别计算剩下元素到 k 个簇心的相异度(如情境的距离),将这些元素划

分到相异度最低的簇;

(3) 根据聚类的结果重新计算簇心,再将剩下元素重新划分到簇中,重复多次,直到聚类结果趋于稳定。

K 均值算法在广泛应用于聚类分析的同时,也存在一些不可避免的问题,比如聚类数 k 的选择、聚类初始质心的选择,同时该算法对于用户数据集中的非核心用户很敏感,少量的这类数据将会对聚类的结果产生较大的影响,原因在于当使用误差平方和标准时,这些少量的远离高密度数据密集区的数据会使平均值得到很大的偏离,从而影响最终的聚类准确度。

2.3 情境相似度计算方法

一般来说情境可以分为基本信息情境、环境情境和平台情境,因此情境信息包含多个情境变量。如用户情境包括其姓名、性别、活动状态,所处位置、天气、温度,所用设备的带宽、网络信息等等。由于情境信息的多样性和复杂性,情境信息包含的数据类型也有所不同。计算情境相似度时,为了量化需要,通常将情境分为数值型和区间型。

数值型情境相似度计算公式如式(3)所示。

$$sim(C_i, C_i') = 1 - \frac{|c_i - c_i'|}{c_i} \quad (3)$$

区间型情境相似度计算公式如式(4)所示。

$$sim(C_i, C_i') = 1 - \sqrt{(c_i - c_i')^2 + (c_i' - c_i')^2} \quad (4)$$

其中, c_i 和 c_i' 表示被量化成数值型的情境的值, c_i' 和 c_i' 表示区间型情境在量化后情境值的起始值, c_i' 和 c_i' 表示区间型情境在量化后情境值的结束值。 $sim(C_i, C_i')$ 表示两个情境之间的相似度,其值越接近 1 则情境相似度越高,反之则相似度越低。

3 基于情境相似度和二次聚类的协同过滤算法

3.1 情境量化数据模型

本文算法基于“U-I-C”三维数据模型,即由用户(User)、项目(Item)、情境(Context) 3 个维度构成的数据模型。情境 $Context = \{C_1, C_2, C_3, \dots, C_n\}$, $C = \{c_1, c_2, c_3, \dots, c_n\}$ 。其中 C_i 表示情境信息的某一属性(如地点、时间、网络状态等), c_i 为情境属性中具体的属性值。如 C_i 为地点,则 c_i 为具体的地理坐标。

3.2 算法设计思想

基于情境相似度和二次聚类的协同过滤算法(Contest-similarity and Twi-clustering based Collaborative Filtering, CTCF)将情境信息引入到协同过滤算法中,先使用用户情境相似度进行初始聚类,再通过评分矩阵计算出用户的评分置信度,然后根据评分置信度将核心用户和非核心用户区分开来。在此基础上,利用核心用户重新计算聚簇质心,再根据调整后的质心让类中的非核心用户重新分布,形成新的聚簇,最后根据情境相似度进行 Top-N 推荐。该算法可以更加准确地选择与目标用户情境相似度高的邻居用户,且削弱了低置信度用户评分对聚类结果的影响,以此达到提高推荐效果的目的。CTCF 主要有初始聚类、置信度计算、区分核心用户、质心调整、非核心用户重聚类、Top-N 推荐等 6 个步骤,主要流程如图 1 所示。

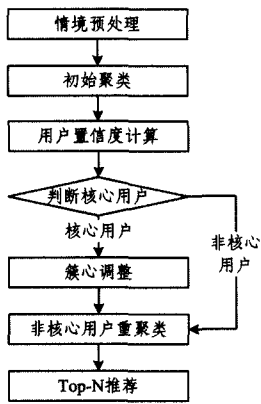


图1 CTCF 主要流程

3.3 算法过程描述

基于上述数据模型和算法思想,本节将对算法的输入、输出以及详细过程进行描述。

输入:“U-I-C”情境量化模型、用户评分矩阵

输出:基于情境相似度和用户置信度的 Top-N 推荐

算法具体过程如下:

1)情境预处理

根据用户情境属性值建立的矩阵模型,用离差标准化的方法,利用式(5)对原始数据进行线性变换,使结果落到 $[0, 1]$ 区间内,再用标准化后的情境属性值替换原始矩阵中的值。

$$c'_{ij} = \frac{c_{ij} - \min}{\max - \min} \quad (5)$$

2) 初始聚类

在上一步的基础上,采用 K-means 聚类算法,使用式(3)和式(4)计算用户情境距离,将数据集迭代划分为 k 个聚簇。划分后的每个簇至少包含一个数据对象,每个数据对象仅属于一个簇,簇内对象情境相似度较高,簇间对象的情境相似性较低。

3)用户置信度计算

评分置信度是评判评分是否准确的标准。较高的预测评分置信度说明该预测评分的可信度较高,较低的预测评分置信度说明该预测评分的可信度较低。基于此,将用户所有的预测评分置信度的均值作为用户置信度 B_u 。本文使用重采样的方法计算置信度,即从评分矩阵中随机抽取子集 $R^{(n)}$,再使用 Pearson 相关性计算公式计算出用户 u 对项目 i 的预测评分 $P_{u,i}^{(n)}$,然后重复采样计算用户 u 对项目 i 的预测评分置信度 $B_{u,i}$,最后根据 $B_{u,i}$ 计算用户 u 的置信度 B_u 。计算公式如式(6)、式(7)所示。

$$B_{u,i} = \frac{1}{stdev(\{P_{u,i}^{(n)}; n=1, \dots, N\})} \quad (6)$$

$$B_u = \frac{\sum_{i \in I_u} B_{u,i}}{|I_u|} \quad (7)$$

4)判断核心用户

利用上一步计算出的用户评分置信度,筛除评分置信度低于某一阈值 a 的用户。得到核心用户集合 U_c 以及非核心用户集合 U_e ,如下所示。

$$U_c = \{u \in U, B_u \geq a\}, U_e = \{u \in U, B_u < a\}$$

5)簇心调整

根据初始聚类的结果以及上一步得到的核心用户集合 U_c ,可以得到每一个聚簇中的核心用户,从而通过此部分用

户算出新的簇心 c_k 。此步骤可以有效降低非核心用户对聚类结果的影响。

6)非核心用户重聚类

在得到调整后的簇心之后,对于每一个 $u \in U_e$,计算出 u 与 k 个簇中心的距离 $D_{u,k}$,然后将其归到距离最近的簇中,直至形成新的聚类。计算过程如下所示。

算法1 非核心用户重聚类

输入:用户簇心集合 C ,非核心用户集合 U

输出:全部用户重新聚类集合 UC

```

1. for  $u \leftarrow 1$  to  $U$ 
2.     tempD =  $D_{u1}$ 
3.     tempP = 1
4.     for  $c \leftarrow 1$  to  $C$ 
5.         if  $D_{uc} \leq \text{tempD}_{u1}$ 
6.             then tempD =  $D_{uc}$ 
7.                 tempP =  $c$ 
8.         end for
9.          $U_u \in C_{\text{tempP}}$ 
10.    end for
11.    for  $c \leftarrow 1$  to  $C$ 
12.        for  $u \leftarrow 1$  to  $U$ 
13.            if  $U_u \in C_c$ 
14.                then  $UC_c += U_u$ 
15.            end for
16.        end for

```

7)Top-N 推荐

根据聚类结果以及服务情境的相似度,生成类成员内部的推荐向量,对其进行排序后形成最终的 Top-N 推荐结果。

4 实验结果分析

4.1 数据集

本文使用 Restaurant & consumer data 数据集来测试算法的性能。该数据集来源于一个真实的研究型推荐系统,常被用于协同过滤算法研究。该数据集包含基本的用户情境信息,如性别、年龄、地理位置、用餐偏好、历史消费记录等等,以及一定区域内 239 个用户对 131 个餐厅的 3162 条评分。

每一次实验将随机选出 10% 的数据作为测试集 (R-Test),其他 90% 的数据将作为训练集 (RTrain) 应用于基于二次聚类的协同过滤算法,将得出的结果与 RTest 进行比较,以验证算法的准确度和效率。

4.2 评价指标

本文采用平均绝对偏差 (Mean Absute Error, MAE)^[12] 作为衡量推荐系统预测准确度的评价标准。MAE 通过计算预测值和实际值之间的偏差来衡量预测是否准确,从而达到评价推荐系统的目的,它是现阶段最常用的评价度量之一。MAE 值越小,推荐的质量越高;反之,推荐的质量越低。

定义预测的用户评分集合为 $\{p_1, p_2, \dots, p_N\}$,对应的实际用户评分集合为 $\{r_1, r_2, \dots, r_N\}$,则平均绝对偏差 MAE 定义为:

$$MAE = \frac{\sum_{i=1}^N |p_i - r_i|}{N} \quad (8)$$

4.3 实验及结果分析

本实验将传统的基于用户相似度的 K-means 聚类协同

过滤算法(User-based Collaborative Filtering)与本文提出的算法进行比较。在个性化推荐的实际情况中经常会遇到数据稀疏性问题,从而影响推荐算法的准确度。本实验分别选取不同的稀疏度和不同邻居数(knn)进行对比实验。在实验过程中,用于判断用户置信度的阈值 a 可以根据实际需求进行预设,为了计算方便,本实验取用户的置信度的均值作为阈值 a 。实验所得 MAE 值如表 1 所列。

表 1 算法结果比较(MAE 值越小则算法效果越好)

knn	Density=0.1		Density=0.15		Density=0.2	
	User-based CF	CTCF	User-based CF	CTCF	User-based CF	CTCF
20	0.99	0.87	0.95	0.85	0.89	0.81
30	0.75	0.60	0.72	0.57	0.62	0.54
40	0.45	0.42	0.42	0.40	0.40	0.39
50	0.52	0.48	0.49	0.47	0.46	0.43
60	0.62	0.62	0.58	0.54	0.53	0.49
70	0.68	0.66	0.62	0.60	0.58	0.53

在数据稀疏度为 0.1 的情况下,本文算法与传统基于用户的协同过滤算法的效果对比如图 2 所示。

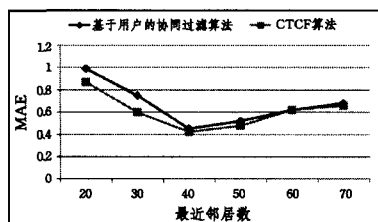


图 2 不同邻居数情况下本文算法与传统算法的 MAE 比较 (density=0.1)

由图 2 可知,面对数据稀疏性问题,在不同邻居数的情况下,本文算法相比传统的基于用户的协同过滤算法有较低的 MAE 值。实验结果表明本文算法在一定程度上比传统的基于用户的协同过滤算法有更高的准确度和有效性,可以更有效地降低非核心用户对于算法结果的影响,更准确地进行用户偏好的预测,提高个性化服务推荐的质量。从实验结果数据的变化趋势上来看,随着最近邻居数的增加,MAE 先呈下降趋势,后呈上升趋势,在最近邻居数为 40 左右时达到最低点。因此,在本次实验中,采用最近邻居数为 40 来进行聚类可以得到较好的聚类效果,从而更精确地预测用户的偏好。

在固定邻居数的情况下,通过分别选取稀疏度为 0.1、0.15、0.2 的数据,对本文算法和传统基于用户的协同过滤算法进行对比实验,结果如图 3 所示。实验结果表明,实验数据越稀疏,算法精确度越低,但相比传统的基于用户的协同过滤算法,本文提出的算法仍具有较低的 MAE 值。因此,该算法对于不同的数据稀疏度,有较高的有效性和稳定性,在一定程度上能有效缓解在实际推荐系统中出现的数据稀疏性问题。

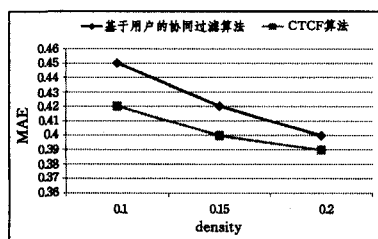


图 3 不同稀疏度情况下本文算法与传统算法的 MAE 比较 (knn=40)

结束语 本文在情境感知和个性化推荐系统相关研究的基础上,提出一种基于情境相似度和二次聚类的协同过滤算法(CTCF)。将该算法与传统的基于用户的协同过滤算法相比较,并采用 MAE 来检测算法的性能。实验结果表明,相对于传统的基于用户的协同过滤推荐算法,CTCF 算法在准确度、命中率等方面有比较明显的优势,在一定程度上降低了非核心用户对于聚类结果的影响,并提高了推荐算法的性能。然而,尽管该算法能相对有效地对情境相似度高的用户进行聚类,但是在具体的情境感知推荐系统中各情境参数所占的权重不同且难以定量,这在一定程度上影响了该聚类的稳定性。因此,未来的工作将更多地考虑情境参数权重对推荐结果的影响,以及如何更准确、快速地对核心用户进行聚类,以提高推荐系统的质量。

参考文献

- [1] Dey A K, Abowd G D, Salber D. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications[J]. Human-Computer Interaction, 2001, 16(2): 97-166
 - [2] Schilit B, Adams N, Want R. Context-aware computing applications [C]//IEEE Workshop on Mobile Computing System and Applications. Computer Society Press, 1994:89-101
 - [3] Honda K, Ichihashi H. Component-wise robust linear fuzzy clustering for collaborative filtering[J]. International Journal of Approximate Reasoning, 2004, 37(2): 127-144
 - [4] Zaslavsky A. Mobile Agents: Can They Assist with Context Awareness[C]//Proceedings of the 2004 IEEE International Conference on Mobile Data Management 2004. Berkeley, CA, USA, 2004:199-211
 - [5] Liu D, Meng X W, Chen J L. A Framework for Context-Aware Service Recommendation[C]//Proceedings of the 10th International Conference on Advanced Communication Technology. 2008:2131-2134
 - [6] Esbensen H. Computing near-optimal solutions to the Steiner-problem in a graph using a genetic algorithm[J]. Networks, 1995, 26(4): 173-185
 - [7] Adomavicius G, Tuzhilin A. Towards 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
 - [8] Linden G, Smith B, York J. Amazon.com recommendations: Item-to-item collaborative filtering [J]. Internet Computing, IEEE, 2003, 7(1): 76-80
 - [9] Resnick P, Iacovou N, Suchak M, et al. GroupLens: an open architecture for collaborative filtering of netnews[C]//Proceedings of the 1994 ACM Conference on Computer Supported cooperative Work. ACM, 1994: 175-186
 - [10] Xu Hai-ling, Wu Xiao, Li Xiao-dong, et al. Comparison study of internet recommendation system[J]. Journal of Software, 2009, 20(2): 350-362(in Chinese)
- 徐海玲, 吴潇, 李晓东, 等. 互联网推荐系统比较[J]. 软件学报, 2009, 20(2): 350-362

速变化性,本文提出了一种松弛的优化均衡流式图划分算法。首先,定义了一种同时包含划分内部代价和划分之间的割的代价的目标函数作为图划分的整体框架;接下来,在图划分框架的基础上通过最大化和最小化两种优化函数分析了均衡图划分问题,并给出了二者之间的关系;最后,针对流式图数据,提出了一种贪婪的图最优 k 划分算法。实验表明,本文提出的图划分算法与相关算法相比较不仅均衡性好,而且通信开销小,在基于该算法进行图划分时上层应用的计算性能得到了明显的提高。

参 考 文 献

- [1] Kwak H, Lee C, Park H, et al. What is Twitter, a social network or a news media? [C]// Proceedings of the 19th International Conference on World Wide Web. ACM, 2010; 591-600
- [2] Xu Da-chuan, Han Ji-ye, Du Dong-lei. Improved Approximation Algorithm for Graph Partition Problem[J]. Acta Mathematica Applicatae Sinica, 2006, 28(4): 587-597 (in Chinese)
徐大川, 韩继业, 杜东雷. 关于图划分问题的改进的近似算法[J]. 应用数学学报, 2006, 28(4): 587-597
- [3] Malewicz G, Austern M H, Bik A J C, et al. Pregel: a system for large-scale graph processing[C]// Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. ACM, 2010; 135-146
- [4] Kang U, Tsourakakis C E, Faloutsos C. Pegasus: A peta-scale graph mining system implementation and observations [C]// Ninth IEEE International Conference on Data Mining, 2009 (ICDM609). IEEE, 2009; 229-238
- [5] Low Y, Bickson D, Gonzalez J, et al. Distributed GraphLab: a framework for machine learning and data mining in the cloud [J]. Proceedings of the VLDB Endowment, 2012, 5(8): 716-727
- [6] Szell M, Lambiotte R, Thurner S. Multirelational organization of large-scale social networks in an online world[J]. Proceedings of the National Academy of Sciences, 2010, 107(31): 13636-13641
- [7] Garey M R, Johnson D S, Stockmeyer L. Some simplified NP-complete problems[C]// Proceedings of the sixth Annual ACM Symposium on Theory of Computing. ACM, 1974; 47-63
- [8] Zhou Shuang, Bao Yu-bin, Wang Zhi-gang, et al. BHP: BSP Model Oriented Hash Graph Data Partition with Load Balancing [J]. Journal of Frontiers of Computer Science & Technology, 2014, 8(1): 40-50 (in Chinese)
周爽, 鲍玉斌, 王志刚, 等. BHP: 面向 BSP 模型的负载均衡 Hash 图数据划分[J]. 计算机科学与探索, 2014, 8(1): 40-50
- [9] Ke Q, Prabhakaran V, Xie Y, et al. Optimizing data partitioning for data-parallel computing[C]// HotOS XIII. May 2011
- [10] Berger-Wolf T Y, Saia J. A framework for analysis of dynamic social networks[C]// Proceedings of the 12th ACM SIGKDD In-

ternational Conference on Knowledge Discovery and Data Mining. ACM, 2006; 523-528

- [11] Naaman M, Boase J, Lai C H. Is it really about me? message content in social awareness streams [C]// Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work. ACM, 2010; 189-192
- [12] Stanton I, Kliot G. Streaming graph partitioning for large distributed graphs[C]// Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2012; 1222-1230
- [13] Krauthgamer R, Naor J S, Schwartz R. Partitioning graphs into balanced components[C]// Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms. Society for Industrial and Applied Mathematics, 2009; 942-949
- [14] Lin Jiao, Chen Wen-guang, Li Qiang, et al. A New Data Clustering Algorithm for Parallel Whole-Genome Shotgun Sequence Assembly[J]. Journal of Computer Research and Development, 2006, 43(8): 1323-1329 (in Chinese)
林皎, 陈文光, 栗强, 等. 基于图划分的全基因组并行拼接算法[J]. 计算机研究与发展, 2006, 43(8): 1323-1329
- [15] Karypis G, Kumar V. A fast and high quality multilevel scheme for partitioning irregular graphs[J]. SIAM Journal on scientific Computing, 1998, 20(1): 359-392
- [16] Prabhakaran V, Wu M, Weng X, et al. Managing Large Graphs on Multi-Cores with Graph Awareness [C]// USENIX Annual Technical Conference. 2012; 41-52
- [17] Newman M E J. The structure and function of complex networks[J]. SIAM Review, 2003, 45(2): 167-256
- [18] Tsourakakis C, Bonchi F, Gionis A, et al. Denser than the densest subgraph: Extracting optimal quasi-cliques with quality guarantees [C]// Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2013; 104-112
- [19] Stanton I, Kliot G. Streaming graph partitioning for large distributed graphs[C]// Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2012; 1222-1230
- [20] Zafarani R, Cole W D, Liu H. Sentiment propagation in social networks: a case study in livejournal [M]// Advances in Social Computing. Springer Berlin Heidelberg, 2010; 413-420
- [21] Leng Yong-lin, Shen Hua, Lu Fu-yu. Distributed Storage of RDF Directed Graph Based on P-Rank Algorithm [J]. Journal of Chongqing University of Technology (Natural Science), 2015, 29(1): 91-95 (in Chinese)
冷泳林, 申华, 鲁富宇. 基于 P-Rank 的 RDF 有向图的分布式存储 [J]. 重庆理工大学学报 (自然科学), 2015, 29(1): 91-95

(上接第 213 页)

- [11] Chen Ke-han, Han Pan-pan, Wu Jian. User Clustering Based Social Network Recommendation [J]. Chinese Journal of Computers, 2013, 36(2): 349-359 (in Chinese)
陈克寒, 韩盼盼, 吴健. 基于用户聚类的异构社交网络推荐算法 [J]. 计算机学报, 2013, 36(2): 349-359
- [12] Herlocker J L, Konstan J A, Terveen L G, et al. Evaluating collaborative filtering recommender systems [J]. ACM Transactions on Information Systems, 2004, 22(1): 5-53
- [13] Niu Wen-jia, Li Gang, Zhao Zhi-jun. Multi-granularity context

model for dynamic Web service Composition [J]. Journal of Network and Computer Applications, 2011(34): 312-326

- [14] Matthias B, Gernot B. Improving the recommendation of mobile services by interpreting the user's icon arrangement [C]// Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services. Bonn, Germany, 2009; 1-9
- [15] Mazurowski M A. Estimating confidence of individual rating predictions in collaborative filtering recommender systems [J]. Expert Systems with Applications, 2013, 40(10): 3847-3857