

一种松弛的优化均衡流式图划分算法研究

殷晓波¹ 罗恩²

(安徽理工大学计算机科学与工程学院 淮南 232001)¹ (中南大学信息科学与工程学院 长沙 410083)²

摘要 在大规模图数据的分布式处理中,往往需要将图数据进行划分并放置在不同的节点上。如果数据划分得不均衡,那么部分节点可能会成为分布式系统的瓶颈。为了提高图数据划分的均衡性,并且有效地应对图数据的快速更新,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。然后,在图划分框架的基础上通过最大化 and 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,提出一种贪婪的图最优 k 划分算法。该划分算法以最大化优化函数为基础,通过最大化顶点放置产生的目标函数增加值进行节点划分块的选取。实验表明,提出的图划分算法与相关算法相比,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。

关键词 图,均衡图划分,分布式计算,启发式算法

中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.4.047

Relaxed Optimal Balanced Streaming Graph Partitioning Algorithm

YIN Xiao-bo¹ LUO En²

(College of Computer Science and Engineering, Anhui University of Science and Technology, Huainan 232001, China)¹

(School of Information Science and Engineering, Central South University, Changsha 410083, China)²

Abstract During the distributed processing of large scale graph data, it usually needs to partition the whole graph into different nodes. If the partition isn't balanced, some of the nodes may be the bottleneck of the distributed system. In order to achieve a balanced partition and handle the quick updates of graph efficiently, this paper proposed a relaxed optimal balanced streaming graph partitioning algorithm. Firstly, we defined an objective function including costs of both inter-partitions and intra-partition as the general graph partition framework. Secondly, we analyzed the graph partition problem according to a maximal and a minimal optimization functions based within the proposed framework, and gave their relationships. Finally, we proposed a greedy optimal graph k -partitions algorithm upon the streaming graph. The experiments show that, the proposed graph partitioning algorithm have better balance and lower communication cost than related works, and upper applications upon this algorithm have better performance.

Keywords Graph, Balanced graph partitioning, Distributed computing, Heuristic algorithm

随着互联网技术的发展,各种在线服务所要处理的图数据规模越来越大,例如 Facebook 的用户已突破 10 亿并且用户间的好友关系已经超过 1400 亿个^[1]。面对如此大规模的图数据,服务提供商往往将整个图数据划分为若干个小的划分,采用分布式系统对小的划分数据进行可扩展计算并将计算结果进行整合^[2]。为了保证分布式系统的性能,各个划分之间的数据量要保持平衡从而充分提高分布式系统的加速比。此外,不同的划分往往分布在不同的处理节点上,为了使不同处理节点之间的通信量尽可能小,要满足各个数据划分之间占用的总的图数据的边数最小化。

现有的分布式图数据处理系统(如 Pregel^[3]、Pegasus^[4]和 GraphLab^[5])在对图数据进行划分时往往对图的节点采用哈希进行划分,即将图的每个节点随机均匀地划分到 k 个划分上。这种方法可以保证不同划分之间的节点数平衡,然而

却忽视了图数据的整体结构。当将整个图随机划分为 k 个数据块时,图中边的割所占的边总数为图中边总数的 $1-1/k$ 。由于现实应用中的图往往是稀疏的,因此需要一种有效的均衡图划分方法^[6]。均衡的图划分问题是最小化划分过程中割所占的边的数量。在理论计算机科学中,均衡图划分问题是 NP 难问题,已有的近似算法如文献^[7-9]。

在大规模图数据的处理中,面临的另外一个挑战是动态图的高效处理。在线社会网络中每天都有新的用户加入和离开,以及用户间好友关系的变化^[10]。此外,由消息转发构成的图往往也是动态的^[11]。面对规模如此大的图数据,图划分问题已然十分困难,再加上图数据的动态变化,使得均衡图划分问题变得十分棘手。动态图中的均衡图划分问题又称为流式图划分^[12]。当节点或边进入或离开时,节点或边的放置必须进行实时计算,因而需要高效的流式图划分算法。现有的

图划分方法大致可以分为严格的数学方法和启发式方法两类。严格的数据方法^[13]通过对问题进行分析找出近似解,这种方法没有好的可扩展性,不能应用于大规模图数据的处理。启发式方法^[12,14-16]效率高并且可扩展性好,在实际中往往被采用。

本文研究了流式图数据下的均衡 k 划分算法,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,在该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,提出了一种针对流式图数据贪婪的图最优 k 划分算法。

1 图划分模型

$G=(V,E)$ 为无向图,图中节点和边的数量分别为 $|V|=m$ 和 $|E|=n$ 。对于顶点的子集 $S \subseteq V$,令 $e(S,S)$ 为两个端点都在 S 中的边的集合, $e(S,V \setminus S)$ 为图中割 $(S,V \setminus S)$ 穿过的边的集合。对于节点 v ,令 $t_S(v)$ 为三角 (v,w,z) 的个数,其中 $w,z \in S$ 。令 $S_i (i \in \{1, \dots, k\})$ 为顶点集合,那么图 G 的顶点划分 $P=(S_1, S_2, \dots, S_k)$ 满足 $S_i \cap S_j = \emptyset (i \neq j)$, 并且 $\bigcup_{i=1}^k S_i = V$ 。令 $\partial e(P)$ 为 G 中横跨不同划分的边的集合,那么有 $\partial e(P) = \bigcup_{i=1}^k e(S_i, V \setminus S_i)$ 。

1.1 图划分框架

对于每一个划分 S_i ,它的大小 $\sigma(S_i)$ 可以通过如下 3 种情况进行计算:(a) 内部边的数量 $|e(S_i, S_i)|$; (b) 边的数量 $|e(S_i, S_i)| + |e(S_i, V \setminus S_i)|$; (c) 顶点的数量 $|S_i|$ 。

本文提出一种同时包含划分内部代价和顶点划分之间的割的代价的目标函数作为图的划分框架,该目标函数包含划分之间的代价 $c_{OUT}: \mathbb{N}^k \rightarrow \mathbb{R}_+$ 和划分内部的代价 $c_{IN}: \mathbb{N}^k \rightarrow \mathbb{R}_+$,这两个代价函数在整数范围内都是增函数。给定图的划分 $P=(S_1, S_2, \dots, S_k)$,那么全局代价目标函数为

$$f(P) = c_{OUT}(|e(S_1, V \setminus S_1)|, \dots, |e(S_k, V \setminus S_k)|) + c_{IN}(\sigma(S_1), \dots, \sigma(S_k)) \quad (1)$$

计算划分之间代价 $c_{OUT}(|e(S_1, V \setminus S_1)|, \dots, |e(S_k, V \setminus S_k)|)$ 时,如果每个划分坐落在不同的计算节点上,那么不同划分之间的通信将成为网络的瓶颈。为了便于分析,假设图的划分 P 穿过的每条边的代价为固定的常数,那么 c_{OUT} 与割的边数成正比。在计算划分内部的代价 $c_{IN}(\sigma(S_1), \dots, \sigma(S_k))$ 时,令 $c_{IN} = \sum_{i=1}^k c(\sigma(S_i))$,其中代价函数 $c(x)$ 为 x 的凸的增函数,并且 $c(0)=0$ 。由于需要对各个划分之间的代价进行平衡,因此当各个划分内部的代价取相同的值时总的代价是最小的。于是,上述图划分问题可以进行如下定义。

定义 1(图的最优 k 划分) 给定图 $G=(V,E)$,寻找顶点集合 V 的划分 $P^*=(S_1^*, \dots, S_k^*)$,使得 $\forall P, |P|=k$ 都有 $f(P^*) \leq f(P)$ 。

当图数据以流的形式到达时,图的扩张方式包括边的随机到达、深度优先和宽度优先 3 种方式。当新的边或者顶点到达时,流式图划分算法确定如何对新到来的顶点进行放置,并且保证已经放置好的顶点不再移动。

1.2 均衡图划分

均衡图划分问题是图的最优 k 划分的一个特例。在均衡图划分中,划分之间的代价等于割所穿过的边的数量,划分内部的代价为每个划分的顶点数量。于是,定义 1 转化为如下

优化问题:

$$|S_i| \leq \frac{n}{k}, \forall i \in \{1, \dots, k\} \quad (2)$$

$$\min_{P=(S_1, \dots, S_k)} |\partial e(P)|$$

本文的思想是对式(1)中的约束条件进行松弛化,令 $|S_i| = \frac{n}{k}, \forall i \in \{1, \dots, k\}$,此时 $c_{IN}(P)$ 达到最小值,并且在目标函数上加入松弛变量 $c_{IN}(P)$ 。于是,式(2)所示的优化问题可转化为:

$$\min_{P=(S_1, \dots, S_k)} |\partial e(P)| + c_{IN}(P) \quad (3)$$

为了使不同划分的内部代价之间保持均衡,那么函数 $c_{IN} = \sum_{i=1}^k c(|S_i|)$ 中的代价函数 $c(x)$ 是一个超模的增函数,即当 $0 \leq y \leq x$ 时, $c(x+1) - c(x) \geq c(y+1) - c(y)$ 。本文选取 $c(x) = \alpha x^\gamma$ 作为代价函数,其中参数 $\alpha > 0, \gamma \geq 1$ 。通过对参数 γ 的选取,该代价函数可以控制目标函数中划分块大小的不均衡性。在 $\gamma=1$ 的极端情况下,该代价函数等价于被划分所穿过的边的数量,因此忽略了数据块大小产生的不均衡性。当 γ 逐渐增大时, x 的指数大于 1,那么 x 的变化将带来越来越大的代价,因而导致代价在不同划分之间产生更大的不均衡性。在参数 α 的选取上,本文令 $\alpha = mk^{\gamma-1}/n^\gamma$,那么式(3)所示的优化函数进一步转化为:

$$\min_{P=(S_1, \dots, S_k)} \frac{\sum_{i=1}^k e(S_i, V \setminus S_i)}{m} + \frac{1}{k} \sum_{i=1}^k \left(\frac{k|S_i|}{n} \right)^\gamma \quad (4)$$

根据图的模块化理论^[17],图的最优 k 划分同样可以看作一个等价的关于边的最大化问题。给定图 $G=(V,E)$ 和 $S \subseteq V$,将函数 $h: 2^V \rightarrow \mathbb{R}$ 定义为

$$h(S) = |e(S,S)| - c(|S|) \quad (5)$$

其中对于所有的 $v \in V$ 满足 $h(\emptyset) = h(\{v\}) = 0$ 。给定 $k \geq 1$ 以及 V 上的划分 $P=(S_1, S_2, \dots, S_k)$,定义目标函数为

$$g(P) = \sum_{i=1}^k h(S_i) \quad (6)$$

那么对于所有满足 $|P|=k$ 的划分 P ,图 G 的 k 划分问题可以转化为最大化函数 $g(P)$ 。通过进一步推导,可得

$$\begin{aligned} g(P) &= \sum_{i=1}^k (|e(S_i, S_i)| - c(|S_i|)) \\ &= m - \left(\sum_{i=1}^k |e(S_i, V \setminus S_i)| - c(|S_i|) \right) \\ &= m - f(P) \end{aligned} \quad (7)$$

当代价函数 $c(x) = \alpha x^\gamma$ 的参数 $\alpha > 0$ 并且 $\gamma=2$ 时,通过最大化 $g(P)$ 可以得到如下的优化目标函数:

$$\max_{P=(S_1, \dots, S_k)} \sum_{i=1}^k (|e(S_i, S_i)| - p \binom{|S_i|}{2}) \quad (8)$$

其中 $p=2\alpha$ 。如果图 G 为满足概率为 p 的 Erdős-Rényi 随机图,那么式(8)中的每一项为该划分块中定点数的期望值与真实值之间的差^[18]。

1.3 流式图划分算法

在流式图数据中,本文提出一种贪婪的图最优 k 划分算法对顶点的划分进行放置,该算法的基本思想如下:在式(8)定义的最大化目标函数中,当新的顶点到达时,通过放置顶点使得目标函数的增加值最大。如果当前的顶点划分为 $P=(S_1, S_2, \dots, S_k)$,那么新到达的顶点 v 在满足如下条件时被放置在划分块 S_i : 对于所有的 $j \in \{1, \dots, k\}$,有

$$g(S_1, \dots, S_i \cup v, \dots, S_j, \dots, S_k) \geq g(S_1, \dots, S_i, \dots, S_j \cup v, \dots, S_k) \quad (9)$$

令 $\delta g(v, S_i) = g(S_1, \dots, S_i \cup v, \dots, S_j, \dots, S_k) - g(S_1, \dots, S_k)$, 那么上述贪婪的图最优 k 划分算法就是对每一个到达的节点 v 找到一个划分块 S_i , 使得 $\forall j \in \{1, \dots, k\}, \delta g(v, S_i) \geq \delta g(v, S_j)$ 。

下面进一步分析如何均衡边的割和顶点的个数来优化目标函数。如果令 $\delta c(x) = c(x+1) - c(x)$, $N(v)$ 表示顶点 v 的邻居节点集合, 那么可得:

$$\delta g(v, S_i) = |N(v) \cap S_i| - \delta c(|S_i|) \tag{10}$$

其中, $|N(v) \cap S_i|$ 表示划分 S_i 中 v 的邻居节点要尽可能多, $-\delta c(|S_i|)$ 表示将 v 加入到 S_i 中时产生的代价要尽可能小。

当代价函数 $c(x) = \alpha x^\gamma$ 时, 可得 $\delta c(x) = \alpha \gamma x^{\gamma-1}$ 。如果令 $\gamma = 1$, 那么式(10)所示的贪婪规则变为 $\delta g(v, S_i) = |N(v) \cap S_i| - \alpha$, 其含义为将顶点加入到含有邻居节点数最多的划分块中, 该贪婪规则将导致划分块间的顶点数严重失衡^[19]。当代价函数 $c(x) = x^2/2$ 时, 式(10)所示的贪婪规则变为 $\delta g(v, S_i) = |N(v) \cap S_i| - |S_i| = |S_i \setminus N(v)|$, 其含义为将顶点加入到含有最少数量非邻居的划分块中^[16]。

2 实验结果与分析

2.1 数据准备与评价指标

为了验证本文提出的算法的性能, 实验采用了公开的 LiveJournal 数据集和合成数据集两种。LiveJournal 数据集的详细信息参见文献[20]。由于真实的图数据往往服从幂率分布, 本文采用文献[17]所示的算法生成一系列不同规模的图数据。

为了对图划分算法的代价进行度量, 本文定义了 λ 和 ρ 两个指标用于衡量划分结果。 λ 为划分 P 的割所穿过的边占图中所有边的比例, ρ 为规范化的最大负载量, 它们的计算公式分别为:

$$\lambda = \frac{P \text{ 穿过的边的数量}}{\text{边的总数}} = \frac{|E(P)|}{m} \tag{11}$$

$$\rho = \frac{\text{最大负载量}}{n/k} \tag{12}$$

由于划分块的最大负载量为 n , 并且至少存在一个划分块含有 n/k 及以上数目的顶点, 因此 $k \geq \rho \geq 1$ 。

2.2 图划分代价测试

在该组实验中, 通过合成数据集对本文提出的算法的参数 γ 进行评估。在选取合适的参数后将本文提出的算法与 METIS 图划分算法^[19]的代价进行了对比。

在本文算法的参数评估中, 将 γ 的取值从 1 增大到 4, 并且每次增加 0.25。图 1 和图 2 分别为代价评价指标 λ 和 ρ 随着 γ 的变化图。在图 1 中, λ 随着 γ 的增加不断增大, 当 γ 增加到 1.25 后, ρ 的增加变缓; 在图 2 中, ρ 随着 γ 的增加不断减小, 当 γ 增加到 1.25 后, ρ 的减小变缓。因此, 在后续的实验, 我们取参数 $\gamma = 1.5$ 。

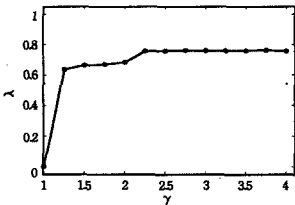


图 1 本文算法中 λ 随着 γ 的变化

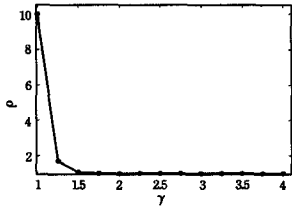


图 2 本文算法中 ρ 随着 γ 的变化

在选取参数 $\gamma = 1.5$ 后, 通过实验对比了本文算法和 METIS 算法的 λ 和 ρ 。我们随机生成了 6 个大规模图数据, 每个图的节点规模都在六百万以上。对于不同的图数据, 分别将节点划分为 2^m 个数据块 ($m = 2, \dots, 7$), 实验结果见表 1。从表 1 中可以看出, 不论 m 取何值, 本文算法的 λ 值都低于 METIS 算法, 并且 ρ 值都高于 METIS 算法。这表明本文算法与 METIS 算法相比较具有更低的划分代价和更好的负载均衡性。

表 1 算法关于合成数据集的性能比较

节点数量	划分个数	本文算法		METIS ^[19]	
		$\lambda(\%)$	ρ	$\lambda(\%)$	ρ
6185314	4	62.50	1.03	65.20	1.02
6714510	8	82.20	1.02	82.50	1.02
6483201	16	92.90	1.03	93.20	1.02
6364819	32	96.30	1.04	96.70	1.02
6308013	64	98.20	1.02	98.90	1.02
6279566	128	98.40	1.03	98.80	1.02

2.3 整体性能测试

在算法的整体性能测试中, 我们在 LiveJournal 数据集上分别采用本文提出的图划分算法和传统的 Hash 划分算法计算了所有节点的 PageRank 值。

首先, 观察了两种方法在划分个数分别为 4、8 和 16 这 3 种情况下的运行时间和产生的通信开销, 结果见表 2。在该组实验中, 仍然采用参数 $\gamma = 1.5$ 。从表 2 可以看出, 在采用了本文提出的图划分算法后, 在计算 LiveJournal 数据集中节点的 PageRank 值时所需的运行时间和通信开销都大大降低了。接下来, 在本文提出的图划分算法中, 分别取不同边的数量的图数据, 通过观察计算 PageRank 值所需的时间变化来衡量该算法的可扩展性。从图 3 所示的实验结果可知, 当待划分的图中包含的边的数量增加时, 算法的运行时间近似呈线性趋势增长, 这表明该划分算法具有很好的可扩展性。

表 2 算法关于 LiveJournal 数据集的性能比较

划分个数	运行时间(s)		通信开销(MB)	
	Hash	本文	Hash	本文
4	32.27	25.49	321.41	196.9
8	17.26	15.14	285.35	180.02
16	10.64	9.05	222.28	148.67

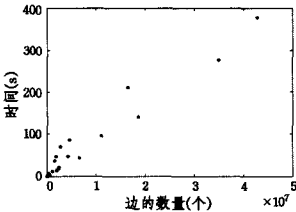


图 3 采用本文算法计算 PageRank 值时运行时间随边的规模变化

结束语 大规模图数据的均衡划分是图数据分布式处理存在的必要前提。如果数据划分得不均衡, 那么部分节点可能会成为分布式系统的瓶颈。针对流式图数据的海量性和快

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