一种基于偏好的个性化标签推荐系统 *

许棣华[1,2a], 王志坚[1], 林巧民[2b], 黄卫东[2c]
(1. 河海大学 计算机与信息学院, 南京 210098; 2. 南京邮电大学 a. 计算机学院; b. 教育科学与技术学院; c. 经济与管理学院, 南京 210003)

摘 要: 建立一个基于用户偏好模型的标签推荐系统,从该系统产生的标签集合中选择出能降低一般性概念描述的模糊性的标签子集,推荐给用户。实验表明,该系统具有较高的可靠性和精准度。

关键词: 标签; 用户模型; 模糊度; 个性化推荐
中图分类号: TP391 **文献标志码**: A **文章编号**: 1001-3695(2011)07-2573-03
doi:10.3969/j.issn.1001-3695.2011.07.048

Personalized tag recommendation system based on preferences

XU Di-hua[1,2a], WANG Zhi-jian[1], LIN Qiao-min[2b], HUANG Wei-dong[2c]
(1. *College of Computer & Information, Hohai University, Nanjing* 210098, *China*; 2. *a. College of Computer, b. College of Educational Science & Technology, c. College of Economics & Management, Nanjing University of Posts & Telecommunications, Nanjing* 210003, *China*)

Abstract: Firstly, proposed a tag recommendation system based on user profile of preference. Then from the recommended tag set, selected some certain tags which could reduce the ambiguity of general concept description, and recommend could them to users finally. The experiments show that the final tags recommendation from the proposed system has better reliability and precision.

Key words: tag; user profile; ambiguity; personalized recommendation

社会标签(social tagging)是 Web2.0 的主要应用之一。它首先由社会大众用户(users)使用标签(tags)为项(item)进行标注(annotating),然后把所有具有相同的标签信息进行归类整理,从而以标签为中心形成的一种全新的信息分类方式。信息专家 Wal 将其命名为分众分类法(Floksonomy,由 Folks 和 Taxonomy 合成),即集合众人之力产生的分类法[1]。与传统的基于内容、文本分词的分类方法不同,在社会标签中,用户对信息内容的高度概括,带有用户的主观认知特点,标签所用的词语虽然可能在全文中的词频不高,但却比词频高的其他类型关键字更加能够反映出整个信息的特征。因此分众分类法被认为是一种更有效的、以人为本的分类方法,在信息检索等领域中起着重要的作用。

随着 Web2.0 的发展,社会网络中的标签数据越来越多。一方面社会标签系统允许用户自由地对网络资源添加自定义标签, 对网络资源进行自组织、分类及与他人共享;另一方面社会网络中的标签数据处在无控制状态,大量标签数据存在冗余性和概念上的模糊性,不利于标签系统的进一步应用。而标签推荐系统在如何简化标注过程中,提供及时、准确、能反映用户意愿的标签和减少标签数据的无控制性、冗余性、模糊性等诸多方面起着举足轻重的作用。

与传统的推荐系统类似,标签推荐技术也可粗略分为三种:基于内容(content-based)的标签推荐、协同式(collaborative)的标签推荐和混合(hybrid)方式。a)基于内容的标签推荐通过分析资源的内容或元信息(meta-information)来提取关键字作为标签推荐给用户,如 Mishne[2] 提出的一种简单的基于内容的标签推荐系统。该系统根据用户提供的新书签(bookmark)确认与它相似的书签,对应的标签被聚集,生成一个标签的列表,排在前面的标签最终被推荐给用户。b)协同式的标签推荐利用了用户、资源、标签之间的关系为某一资源推导出最合适的标签集合。FolkRank[3] 是一种基于图的标签推荐系统,它是 PageRank 算法的变体,并且被应用在分众分类法的用户—资源—标签图上。它的基本思想是:一个资源如果被重要用户用重要的标签标注,那么这个资源也是重要的,对用户、标签也是同样的道理。c)混合方式结合了基于内容的标签推荐和协同式的标签推荐两种方式的特点,如 Song 等人[4] 提出的一种结合聚类和混合模型的标签推荐方法。被标注的文件用两个二部图表示成一个三元组(词,文件,标签),这些图聚类为若干主题(topics),新文件的标签推荐是基于主题分布的后验概率,并根据类内的标签排列给出标签推荐。

1 个性化标签推荐

标签推荐的个性化需求也在与日俱增,越来越多的用户希望系统向他们推荐能反映个人爱好和意愿的、具有个性化特征的标签,以便在日后能更迅速、更准确地搜索到他们需要的资源。Chirita 等人[5] 提出了一种能为网页自动生成个性化标签的 P-TAG 系统,该系统能根据页面内容和用户桌面上的数据来产生个性化的关键字;Mishne[2] 提出了一种基于用户对相似

收稿日期: 2010-12-19; **修回日期**: 2011-01-25 **基金项目**: 国家自然科学基金资助项目(70871061);国家"863"计划资助项目(2007AA01Z178);江苏省自然科学基金资助项目(BK2010524); 中国博士后基金资助项目(20100471289);南京邮电大学"青兰"计划资助项目(NY206034,NY207099)

作者简介:许棣华(1974-),女,江苏如皋人,讲师,博士研究生,主要研究方向为机器学习、数据挖掘、软件新技术(dxu@njupt.edu.cn);王志坚(1958-),男,江苏南京人,教授,博导,博士,主要研究方向为软件自动化、软件构件复用;林巧民,讲师,博士研究生;黄卫东,教授,博士.

Weblog 推荐标签的方法；Symeonidis 等人[6] 使用奇异值分解(SVD)将维数约简应用到个性化标签推荐；Garg 等人[7] 提出了一种交互式方法，在用户为一个新资源输入一个标签后，算法基于所有用户资源的同现标签进行标签推荐；Shepitsen 等人[8] 提出一种基于标签空间的层次聚类的推荐系统，将用户模型和标签聚类应用到个性化推荐中。

本文所提出的个性化标签推荐系统基于如下假设：任何一个项(item)在进入推荐系统前，都有一个一般性概念描述(general concept description，GCD)，如名称、物理特征等。这个一般性概念描述对所有用户来说都认同，但对不同用户可能有不同的含义，因此这个描述是具有模糊性(ambiguous)的。如何针对特定用户降低该描述的模糊性，是本文解决个性化标签推荐的关键。本文首先建立一个基于用户偏好的标签推荐系统，然后利用 KL-离散度(KL-divergence)度量方法和标签同现(tag co-occurrence)建立概率模型，从该系统产生的标签推荐集合中，选择出能降低一般性概念描述模糊性的标签子集，最终推荐给用户。

2　问题定义与系统描述

与传统的推荐系统不同，标签推荐系统不再体现用户与资源之间的二元关系，而是表现为用户、标签、资源三者之间的关系。因此，可将标签推荐系统形式化为一个四元组(U, R, T, F)。其中，U 表示用户集合，R 表示资源集合，T 表示标签集合，$F \rightarrow U \times R \times T$ 表示 U、R、T 三者之间的关系，$F=1$ 则表明三者关系存在，否则为三者关系不存在。

2.1　建立用户模型

偏好是一种以公开的形式说明意愿、意图的方式。本文将偏好问题理解为用户对标签的选择上的序(order)关系。用户模型的表示和构建要考虑到如何将用户偏好以一种结构式的方法加入到模型中。本文采用在标签系统中选择最能体现系统特征的项的集合，将项集合及对应的标签集合展示给用户进行评判(rating)，利用评判结果建立用户模型。

在标签系统中，有以下方式为用户选择具有代表性的项[9]：

a)随机选择方法。按照概率分布随机地选择项。

b)基于项的流行度(popularity)。按照项的被标签的次数由多到少地提供给用户评判，被标签的次数越多的对象，流行度越高。这种方法的缺点是流行度较低的对象总是具有更少的被选择机会。

c)基于项的信息熵(pure entropy)。如果一个项被不同性质的用户标注，则认为该项所含有的信息量多。基于信息熵的选择方法是指为用户选择所含信息量较大的项。这种方法的缺点是要考虑用户标注数据缺失等情况。

d)流行度与信息熵相结合。根据流行度与信息熵的乘积来选择项。这种方法克服了单纯地利用流行度或单纯地利用信息熵的缺点，既考虑到项的流行度，也考虑了项所包含的信息量。实验也表明，这种方法能大大减少用户负担，建立的用户模型的精确程度较高。本文将采用这种方法。

建立用户偏好模型的具体步骤如下：

a)使用流行度与信息熵相结合的方法选择两者乘积值较高的项。

b)去除不符合用户兴趣范围的项后询问用户，按用户的满意程度为项以及相应的标签集合进行评判。将用户对项的标签集合满意程度分为五级：{perfect，good，fair，poor，bad}，分别对应 $\alpha=\{1, 0.8, 0.5, 0.2, 0\}$。

c)根据所选项集和对应的用户集合，建立 uers-item 矩阵 UI，如表 1 所示。矩阵 UI 中的每个元素 $UI_{k,l}$ 表示用户 k 为对象 l 是否进行标注，若进行了标注，则 $UI_{k,l}=1$，否则为 0。对每个用户，根据标注的对象集合和相应的标签集合建立 item-tag 矩阵 IT，如表 2 所示。矩阵 IT 中每个元素 $IT_{i,j}$ 为该用户对第 i 个对象使用标签 j 的满意程度，$IT_{i,j} \in \alpha$。

表 1　user-item 矩阵

user	item					
	I_1	I_2	I_3	…	I_l	…
U_1	0	1	1	…	0	…
U_2	0	0	1	…	1	…
U_3	1	1	1	…	1	…
…		…		…	0	…
U_k	1	1	0	…	1	…
…		…		…		…

表 2　item-tag 矩阵

Item	Tag					
	T_1	T_2	T_3	…	T_l	…
I_1	0.2	0.8	0.2	…	0	…
I_2	0.2	0.5	0.5	…	0.8	…
I_3	1	1	0.5	…	0.8	…
…				…		…
I_i	0.5	0.2	0.8	…	0.6	…
…		…		…		…

d)用向量 P_u 表示用户 u 的用户模型，$P_u=(s_u^1, s_u^2, \cdots, s_u^k)$，其中 $s_u^i, i \in \{1, 2, \cdots, k\}$ 表示用户 u 对第 j 个标签的满意度总和，即 $s_u^i=\sum_i UI_{i,j}$。

2.2　标签推荐算法

本文的标签推荐算法，充分利用了协同式推荐和基于内容的推荐的技术特点：

a)利用 consine 函数计算用户相似度 $\mathrm{sim}(U, U_j)$：

$$\mathrm{sim}(U, U_j)=\frac{\sum_i S_U(i) \times \mathrm{iuf}(i) \times v_{S_j}(i) \times \mathrm{iuf}(i)}{\sqrt{\sum_k (S_U(k) \times \mathrm{iuf}(k))^2 \times (S_{U_j}(k) \times \mathrm{iuf}(k))^2}} \quad (1)$$

其中，$\mathrm{iuf}(k)=\log \frac{\text{用户总数}}{\mathrm{tag}_k \text{ 的总标注次数}}$，iuf 为逆向用户频率[10](inverse users frequency)，相当于基于内容的推荐中的 idf(inverse documents frequency)；$S_{U_j}(k)$ 为用户 j 对 tag_k 的总标注次数的满意程度的归一化表示，相当于基于内容的推荐中的 tf(term frequency)。

b)通过用户相似度计算，取与用户 u 最相似的 k 个用户组成用户集合 $U=\{u_1, u_2, \cdots, u_k\}$。

c)向用户 u 推荐 tag_k 集合为

$$\mathrm{rec}(U, \mathrm{tag}_k)=\{\mathrm{tag}_k \mid \sum_{j \in U} S_j(\mathrm{tag}_k) \times \mathrm{sim}(U, U_j) > \partial\} \quad (2)$$

其中 ∂ 为大于 0 的某个阈值。

2.3　标签模糊性

可以通过 2.2 节提出的方法向用户 u 对系统中的项 i 推荐标签集合 T，$T=\{\mathrm{tag}_1, \mathrm{tag}_2, \cdots, \mathrm{tag}_n\}$。本文假设项 i 在进入标签推荐系统前有一个一般性概念描述 GCD。一般性概念描述是指该项公认的描述，如名称、物理属性等。对 GCD 提取关键字，形成标签集合 T'。显然，集合 T' 不具备个性化特征。

定义 1　称一个标签 t 是模糊(ambiguous)的，如果这个标签对不同用户有不同的含义。

定义 2　称一个集合 T 是具有模糊性的，如果集合中至少有一个标签是模糊的。

标签集合 T' 是具有模糊性的。如何针对特定用户降低标签集合 T' 的模糊性，是本文解决个性化标签推荐的有效途径。本文对标签集合的模糊性是通过加权的 KL 离散度[11] 来度量，并利用标签同现对标签集合的模糊性度量进行建模[12]。假定，如果标签集合的模糊性是可降低的，则存在两个标签 t_i 和 t_j，分

别加入到集合 T' 中后,使得其概率分布产生很大不同,即 $P(t|\{T'\cup t_i\})$ 和 $P(t|\{T'\cup t_j\})$ 达到概率分布不同的最大化。

下面引入概率模型,利用同现标签的性质,在推荐标签集合 T 中选择能降低 T' 的模糊性的标签集合 T_R 推荐给用户。假设一个标签集合 $T^0=\{t_1,t_2,\cdots\}$,表达式 $I(T^0)$ 表示包含这个标签集合的项的数目,定义同现标签 t_i 和 t_j 的项的数目为 $I(t_i\cap t_j)$。因此有:当 t_j 出现时 t_i 的概率为

$$P(t_i|t_j)=\frac{I(t_i\cap t_j)}{\sum_k I(t_k\cap t_j)} \tag{3}$$

t_i 的先验概率为

$$P(t_i)=\frac{\sum_j I(t_i\cap t_j)}{\sum_{j,k} I(t_k\cap t_j)} \tag{4}$$

假设标签同现是条件独立的,有 $P(T|t_i)=\prod_{t\in T}P(t|t_i)$。基于以上假设和贝叶斯公式可以得到

$$P(t_i|T)=\frac{P(T|t_i)P(t_i)}{P(T)}=\frac{P(t_i)\prod_{t\in T}P(t|t_i)}{\sum_j P(t_j)\prod_{t\in T}P(t|t_j)} \tag{5}$$

为了方便,将 t_i 加入到 T' 中对应的条件概率记为 $p_i=p_i(t)=P(t)=P(t|\{T'\cup t_i\})$,$t\in T'$。$t_i$ 和 t_j 加入集合 T 后,p_i 和 p_j 分布的 KL-离散度为

$$\mathrm{KL}(p_i\|p_j)=\sum_t p_i(t)\log\left(\frac{p_i(t)}{p_j(t)}\right) \tag{6}$$

由于 $\mathrm{KL}(p_i\|p_j)$ 不等于 $\mathrm{KL}(p_j\|p_i)$,所以取

$$\overline{\mathrm{KL}}(p_i,p_j)=\frac{1}{2}(\mathrm{KL}(p_i\|p_j)+\mathrm{KL}(p_j\|p_i)) \tag{7}$$

当 $\overline{\mathrm{KL}}(p_i,p_j)>\beta$ 时,β 为系统定义的某个大于 0 的阈值,标签 t_i 和 t_j 就加入到集合 T_R(T_R 初始为空)被推荐给用户。

本文在选择能降低标签模糊性的标签时,既不是在整个系统的标签集合中进行选择,也不是从某个流行度最大(最常见的)的标签集合中进行选择[13],而是在 2.2 节所提出的标签推荐系统得到的标签集合 T 中进行选择。这样,不仅大大降低了计算复杂性,而且还保留了标签推荐集合的个性化特征。

3　实验及实验数据

从目前比较流行的 Web2.0 网站 CiteUlike 获取数据进行测试。CiteULike 是由著名的施普林格出版社(Springer)提供的一个免费协助用户存储、管理和分享学术文章的网站。从该网站获取 2006 年—2008 年标签数据,保留这些标签数据中能对应到 CiteSeer 上的论文,去除标签个数为 1 的论文。每一条 CiteULike 记录包含四个域:用户名、标签、论文标题,创建时间,共约 35 764 条记录。总的论文数约 10 532 篇,总标签数(不重复)约9 412 个,每篇论文的标签数约为 3.28。按照记录创建时间,排在前面 90% 的数据作为训练数据,排在后面 10% 的数据作为测试数据。本文按标签数据创建时间,对 2006 年、2006 年—2007 年、2006 年—2008 年分别进行了三次实验。实验数据集如表 3 所示。

表 3　实验数据集

时间/年	记录数	论文数	标签数(不重复)	论文标签率
2006	8 031	2 290	2 651	3.03
2006—2007	14 605	5 835	4 901	2.98
2006—2008	35 764	10 532	9 412	3.28

采用精准率(precision)、召回率(recall)、F1-score 作为评估尺度。标签推荐的精准率为正确推荐标签占推荐标签总数的百分比。

$$\mathrm{precision}(t_{\mathrm{rec}}(u,r))=\frac{|t(u,r)\cap t_{\mathrm{rec}}(u,r)|}{|t_{\mathrm{rec}}(u,r)|} \tag{8}$$

其中:$|t_{\mathrm{rec}}(u,r)|$ 表示用户 u 推荐给资源 r 的标签的个数;$|t(u,r)|$ 表示用户 u 分配给资源 r 实际的标签个数。

标签预测的召回率为正确推荐标签占标签总数的百分比。

$$\mathrm{recall}(t_{\mathrm{rec}}(u,r))=\frac{|t(u,r)\cap t_{\mathrm{rec}}(u,r)|}{|t(u,r)|} \tag{9}$$

则 F1-score 定义为

$$\mathrm{F1\text{-}score}(t_{\mathrm{rec}}(u,r))=\frac{2\times\mathrm{precisio}(t_{\mathrm{rec}}(u,r))\times\mathrm{recall}(t_{\mathrm{rec}}(u,r))}{\mathrm{precision}(t_{\mathrm{rec}}(u,r))+\mathrm{recall}(t_{\mathrm{rec}}(u,r))} \tag{10}$$

根据 2.2 节的标签推荐算法,对应 2006 年、2006 年—2007 年、2006 年—2008 年的三次实验,得到表 4 的实验结果数据;为了进一步推荐出降低标签模糊度的标签,采用 2.3 节中的方法得到表 5 中的实验结果数据。从实验结果数据来看,本文提及的标签算法是有效和可靠的,并且通过降低标签集合模糊度,可有效提高标签推荐系统的各项评价指标的值。

表 4　2.2 节的推荐算法的数据结果

实验次序	精准率	召回率	F1-score
1	0.291 5	0.371 4	0.326 6
2	0.322 8	0.453 1	0.377 0
3	0.397 4	0.563 8	0.466 2

表 5　2.3 节的推荐算法的数据结果

实验次序	精准率	召回率	F1-score
1	0.317 3	0.412 6	0.358 7
2	0.354 9	0.493 5	0.412 9
3	0.437 1	0.532 0	0.475 8

为了进一步验证本文提及的算法的有效性,与文献[13]当中提到的几种算法(VS + IG、LDA、SVM、PMM、MMSG)进行了比较,比较结果如图 1 所示。图 1 中的数据均来自于文献[13]和本文。从图 1 可见,本文提出的算法在精准率、召回率、F1-score 这几个指标上均优于其他大多数算法。

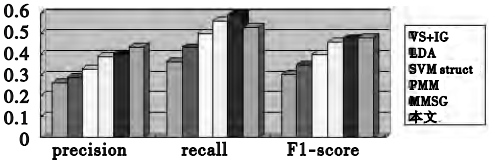

图1　几种算法比较实验结果数据

4　结束语

标签推荐系统在为用户简化标注过程、提高标注效率、为标签系统的进一步应用等方面起着积极作用。由于标签可以从不同用户兴趣角度揭示事物本质特征,因而研究个性化标签推荐是十分有意义的。实验证明,本文所提出的标签推荐系统不仅具有较高的可靠性和精准性,而且降低了标签推荐的计算复杂度,有效提高了标签推荐系统的性能。在研究中发现,标签同现现象在标签系统中非常普遍,已经有学者对标签同现问题进行了研究,如何利用标签同现来获得用户偏好,是笔者今后的工作重点。

参考文献:

[1] WAL T V. Folksonomy[EB/OL]. (2004)[2007-02-02]. http://vanderwal.net/folksonomy.html.

[2] MISHNE G. AutoTag: a collaborative approach to automated tag assignment for Weblog posts[C]//Proc of the 15th International Conference on World Wide Web. 2006:953-954.

[3] HOTHO A, JÄSCHKE R, SCHMITZ C, *et al*. Information retrieval in folksonomies[C]//Proc of the 3rd European Semantic Web Conference. 2006:411-426.

(下转第 2579 页)

把研究的重点转向了应用层面的容错技术上。

在解决单个计算节点失效的问题上,笔者所在计算中心与美国国家超级计算应用中心(NCSA)及伊利诺伊大学香槟分校(UIUC)计算机学院合作,引入 Charm ++[9] 并行编程框架,在此基础上开发出应用框架。Charm ++ 不仅具有处理器虚拟化、自主负载均衡[10] 的特点,还通过 adaptive MPI (AMPI)完全实现了消息传递的接口(message passing interface),与此同时还提供了很好的容错机制。通过该应用框架,使用 Charm ++ 与 adaptive MPI 编写的应用程序能够自动容忍单点故障错误。

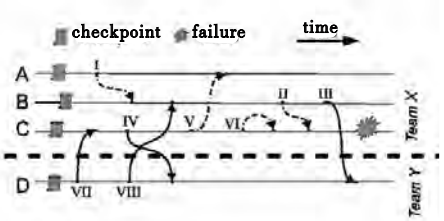

图6　Charm++的team-based消息日志容错协议

Charm ++ 现有的容错协议是基于同步检查点和消息日志的以及组(team-based)的容错模式[11]。如图 5 所示,将一个计算节点(A,B,C)放在一个组中,并且在这个组内部使用协同检查点技术,并不存储消息本身,而只存储对应的消息发送顺序等信息。当错误发生时,重启这个组内的所有进程,恢复其所有进程。而在组与组之间的消息(图 5 中 D 到 C、B,B 到 D 的消息),采用非协同检查点加消息日志的方式,将所有发送出去的消息存起来。

这样的设计使得协同检查点只存在于组内,避免了规模扩大时带来的瓶颈;另一方面,组内消息通信时不需要存储消息日志,而是只需要存储组与组之间的消息通信,大大减少了消息通信的存储量,降低了内存消耗与通信延迟,一定程度上提高了效率。同时,以组为单位重启的技术,也使得能够容忍一个组内多个进程同时失效的情况,提升了容错框架的应用范围和可靠性。笔者对于这一容错协议作出了一个改进,通过采用 send-determinism 机制[12],对现有的悲观消息日志方式进行改进。对于具有 send-determinism 特性的应用,在使用 Charm ++ 框架时,其容错可以不需要使用悲观消息日志的方式,而是通过一种乐观的消息日志来实现,同时增加了额外的操作来保证错误恢复时消息的顺序。该机制的采用,使得在正常情况下(没有错误发生)的程序执行效率相比之前的协议提高了将近 20%,基本上接近了不采用容错机制的情况。

3　结束语

三层架构超级计算环境自 2010 年 1 月正式上线以来,在多次服务器故障和分中心离线维护的情况下系统保持稳定运行。在总中心深腾 7000 集群因水冷机组故障需要停机维修的情况下,三层架构的超级计算环境仍然提供了稳定可靠的服务,一定程度上缓解了由于总中心集群停机给整个院超级计算带来的压力。实践证明,本文设计实现的容错框架保证了三层架构的超级计算环境可以提供稳定可靠的服务。下一步工作将会进一步完善计算节点可靠性设计,研究支持作业管理系统及 MPI 的系统级节点可靠性解决方案;同时,也将开展可靠性评价体系研究,以更好地度量容错方案。

参考文献:

[1] DAI Zhi-hui. A lightweight grid middleware based on OPENSSH-SCE [C]//Proc of the 6th International Conference on Grid and Cooperative Computing. Washington DC:IEEE Computer Society,2007:387-392.

[2] JIN Hai. Fault-tolerant grid architecture and practice[J]. Journal of Computer Science and Technology,2003,18(4):423-433.

[3] 石宣化,金海,羌卫中. 通用网格容错框架研究[J]. 华中科技大学学报:自然科学版,2006,34(7):42-45.

[4] 邱敏,桂小林. 实现可靠计算的容错网格结构[J]. 微电子学与计算机,2005,22(7):99-102.

[5] 刘亮,周兴社,刘秋让. 网格计算自主容错服务[J]. 计算机工程与应用,2004,40(3):4-6.

[6] 肖海力. ScGrid 冗余系统部署[R]. 北京:中科院计算机网络信息中心超级计算中心,2010.

[7] 戴志辉,曹荣强,肖海力. SCE 备份手册[R]. 北京:中科院计算机网络信息中心超级计算中心,2010.

[8] 杜云飞. 容错并行算法的研究与分析[D]. 长沙:国防科学技术大学,2008.

[9] KALE L V. Charm ++ tutorial [EB/OL]. http://charm.cs.uiuc.edu/tutorial/.

[10] ZHENG Geng-bin, MENESES E, BHATELE A, *et al*. Hierarchical load balancing for Charm ++ applications on large supercomputers [C]//Proc of the 39th International Conference on Parallel Processing Workshops. Washington DC:IEEE Computer Society,2010:436-444.

[11] MENESES E, MENDES C L, KALE L V. Team-based message logging: preliminary results[C]//Proc of the 10th IEEE/ACM International Conference on Clusters, Clouds, and Grids Computing. 2010:697-703.

[12] CAPPELLO F. Revisiting fault tolerant protocols for parallel HPC applications[C]//Proc of the 2nd Workshop of the Joint Laboratory for Petascale Computing. 2009.

(上接第 2575 页)

[4] SONG Yang, ZHUANG Zi-ming, LI Hua-jing, *et al*. Real-time automatic tag recommendation[C]//Procs of the 31st Annual International ACM Conference on Research and Development in Information Retrieval. New York:ACM,2008:515-522.

[5] CHIRITA P A, COSTACHE S, NEJDL W, *et al*. P-TAG: large scale automatic generation of personalized annotation tags for the Web [C]//Proc of the 16th International Conference on World Wide Web. New York:ACM,2007:845-854.

[6] SYMEONIDIS P,NANOPOULOS A,MANOLOPOULOS Y. Tag recommendations based on tensor dimensionality reduction[C]//Proc of ACM Conference on Recommender Systems. New York:ACM,2008:43-50.

[7] GARG N, WEBER I. Personalized, interactive tag recommendation for flickr[C]//Proc of ACM Conference on Recommender Systems. New York:ACM,2008:67-74.

[8] SHEPITSEN A, GEMMELL J, MOBASHER B, *et al*. Personalized recommendation in social tagging systems using hierarchical clustering [C]//Proc of ACM Conference on Recommender Systems. New York:ACM,2008:259-266.

[9] RASHID A M, ALBERT I, COSLEY D, *et al*. Getting to know you: learning new user preferences in recommender systems[C]//Proc of International Conference on Intelligent User Interfaces. 2002:127-134.

[10] BREESE J S, HECKEMAN D, KADIE C. Empirical analysis of predictive algorithms for collaborative filtering[C]//Proc of Conference on Uncertainty in Artificial Intelligence. 1998:43-52.

[11] KULLBACK S, LEIBLER R. On information and sufficiency[J]. The Annals of Mathematical Statistics,1951,22(1):79-86.

[12] WEINBERGER K Q, SLANEY M,Van ZWOL R. Resolving tag ambiguity[C]//Proc of the 16th International ACM Conference on Multimedia. New York:ACM. 2008:303-309.

[13] SONY Yang, ZHANG Lu, GILES C L. Automatic tag recommendation algorithms for social recommender systems[J]. ACM Trans on the Web (TWEB),2011,5(1):4.