

基于异质网络的意见领袖社区发现

张伟哲, 王佰玲, 何 慧, 谭卓鹏

(哈尔滨工业大学计算机科学与技术, 黑龙江哈尔滨 150001)

摘 要: 针对意见领袖社区发现问题, 通过将论坛中主题及其回复关系建模为异质网络, 准确表示社区结构. 提出意见领袖社区影响力概念及其量化方法, 在此基础上设计了一种基于异质网络的意见领袖社区发现算法. 通过采集天涯论坛的大量数据, 验证了该社区挖掘方案能够较准确地挖掘论坛中的意见领袖社区.

关键词: 网络社区; 社区发现; 意见领袖; 异质网络

中图分类号: TP393 **文献标识码:** A **文章编号:** 0372-2112 (2012) 10-1927-06

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2012.10.003

Public Opinion Leader Community Mining based on the Heterogeneous Network

ZHANG Wei-zhe, WANG Bai-ling, HE Hui, TAN Zhuo-peng

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China)

Abstract: In order to discover the public opinion leader community in a web forum, a heterogeneous network is constructed to represent the relationship between the topics and their replies. By quantifying the influences of forum users, a forum opinion leader community mining algorithm (FOLCM) is put forward. The algorithm is evaluated with real large-scale web data from the Tianya forum. The results demonstrate the effectiveness of the forum opinion leader community mining algorithm.

Key words: network community; community mining; public opinion leader; heterogeneous network

1 引言

互联网中具有共同兴趣和需求的人们, 汇聚成网络空间中的“虚拟社区”, 并由此催生了新兴的计算机科学研究领域——“计算社会科学”. David Lazer 等 15 位美国学者^[1]于 2009 年 2 月在《Science》联合发表具有里程碑意义的文章“Computational Social Science”, 指出: “计算社会科学”正在高速发展, 人们对数据的收集与利用将达到前所未有的深度和广度. 国内学者多称“计算社会科学”为“社会计算”. 中国科学院王飞跃等人^[2]系统地总结了社会计算的研究内容. 哈尔滨工业大学刘挺^[3]阐述社会计算的主要内容为社交网络服务 (Facebook)、群体智慧 (Wiki)、社会网络分析、内容计算和人工社会.

社会信息网络的形成过程中必然演化出社区, 分析与处理大型社会信息网络的一个基本问题就是如何发现、度量和利用这些社区. 中科院计算所程学旗等人认为社区结构研究大体上包括: 社区发现、社区演化分析, 及社区结构与网络动力学和网络压缩表示等功能性特

征之间的关系. 其中, 社区发现是基础和核心的研究问题^[3]. 因此, 在网络中挖掘自发而成的、有用的、相对稳定的社区, 对网络信息的获取与挖掘、信息的推荐以及网络的演化预测等具有重要的研究价值^[4].

从相异的维度可以划分出多种类型的社区, 但其中最为重要、具有普遍意义的一类社区为“意见领袖社区”. 所谓意见领袖, 是指论坛社区中较具威望, 能“一呼百应”的人. 他们的想法往往能影响社区内其他成员, 甚至引领整个社区的舆论导向. 本文的主要研究内容是从论坛社区中挖掘意见领袖社区. 目前的社区挖掘方案大多都是基于同质网络, 然而真实的社会网络中却存在着大量异质 web 对象, web 对象之间的关系也多种多样, 因此, 基于同质网络的社区挖掘方案难以准确地区分各类 web 对象与对象之间的关系.

本文鉴于目前大多 web 社区挖掘方案基于同质网络的现状, 提出了一种基于异质网络的意见领袖社区发现方案, 设计了一种 web 论坛意见领袖社区挖掘算法 FOLCM, 用以挖掘论坛社区中具有较高舆论导向能力的

意见领袖群体.采用大量真实的 web 论坛数据进行了实验,验证了提出的社区挖掘方案能够较准确地挖掘互联网论坛中的意见领袖社区.

本文第 2 节分析社区发现的研究现状.第 3 节建模异质网络并提出意见领袖社区发现问题的形式化描述.第 4 节提出一种基于异质网络的意见领袖社区发现算法.第 5 节对社区发现算法进行实验验证与性能分析.第 6 节进行总结.

2 相关工作

许多社区挖掘方案基于聚类 (clustering) 算法^[5-10].其中, Yanchun Zhang 等人^[5]通过分析 web 页面社区, web 用户的访问模式与 web 页面的共簇 (co-clustering), 使用 web 聚类方法进行 web 社区挖掘. Bo Yang 等人^[6]针对当前大部分社区挖掘方案都属于集中式 (centralized) 的现状提出了一个基于 multi-agent 的分散式 (decentralized) 算法, 并提出一种自聚集 (self-aggregation) 与自组织 (self-organization) 机制, 使一组自主 agents 协作挖掘网络, 处理集中式社区挖掘方案难以处理的分布式网络. Di Jin^[7]提出了一种新型的蚁群算法, 通过在网络之上建立马尔可夫随机游走模型挖掘网络社区结构. Jiyang Chen 等人^[8]指出以往的社区挖掘方法往往对所有类型的网络都应用同一种结构化方法, 但不同类型的网络往往有不同的结构特征, 他们提出一种新型的社区挖掘方法——Max-Min Modularity, 同时考察相连双方以及由领域专家定义的标准, 然后使用一种层次聚类算法在网络中挖掘社区.

目前大多数社区挖掘方案都是将 web 社区建模为同质网络, 而 Deng Cai 等人^[11]系统地分析了从异质社会网络中挖掘隐藏的社区的问题, 认为不同关系在不同的环境中具有不同的重要性, 提出了一种学习社区成员关系的优化线性组合的方法, 用于更好地满足用户的期望. Lei Tang 等人^[12]提出了在动态的多模式网络 (multi-mode net) 中进行社区挖掘的想法, 通过使用时间信息对多模式网络进行分析, 并详细研究了一个 spectral framework 及其扩展性. Yizhou Sun 等人^[13]对海量异质信息网络中发现动态社区的课题进行了研究. Tsuyoshi Murata 等人^[14]将在线问答社区作为异质网络进行了分析.

3 异质网络建模及问题描述

定义 1 [异质网络] 一个异质网络 H 是一个八元组 $H = (\sum_V, \sum_A, V, A, s, t, 1_V, 1_A)$, 其中 1) V 是结点集合, A 是边的多重集合; 2) \sum_V 和 \sum_A 分别是结点与边标签 (Label) 的有限字母表集合; 3) $s: A \rightarrow V$ 和 $t: V$

$\rightarrow A$ 是两个映射, 表示一条边的源结点和目的结点; 4) $1_V: V \rightarrow \sum_V$ 和 $1_A: A \rightarrow \sum_A$ 是两个映射, 用于描述结点与边的标签^[15].

意见领袖社区发现问题: 通过将论坛或博客建模为一个异质网络 $H = (\sum_V, \sum_A, V, A, s, t, 1_V, 1_A)$, 从中发现影响力较高的用户群体 C , C 具备如下特征:

$$P = \{p \in V \mid 1_V(p) = user\}$$

$$f: P \rightarrow R \quad (1)$$

$$C = \{p \in P \mid \forall p_1 \in C, p_2 \in (P \setminus C), f(p_1) > f(p_2)\}, |C| = N$$

其中 P 集合是指所有的论坛用户, R 是实数集, 意见领袖社区 C 是 P 的一个子集, 并且 C 中元素与非 C 中元素满足条件 $C = \{p \in P \mid \forall p_1 \in C, p_2 \in (P \setminus C), f(p_1) > f(p_2)\}, |C| = N$, 此处 N 是预设的意见领袖社区大小, f 是用户集合 P 到实数集 R 的一个映射, 即 $f(p)$ 即为用户 p 的社区影响力.

4 意见领袖社区发现算法 (FOLCM)

4.1 社区影响力指数

社区影响力是一个抽象概念, 比较不同社区成员的影响力相对高低, 需要对社区影响力进行量化. 本节提出“社区影响力指数”的概念. 将社区成员的社区积分作为获取数据前社区成员的社区影响力指数. 然后通过分析社区成员之间的交互行为, 根据社区成员在社区内与其他成员的回复关系, 得到社区成员最终的社区影响力指数. 提出以下计算天涯用户社区影响力指数公式:

1) 发表主题作者的影响力:

$$influence_X = \epsilon P_X + (1 - \epsilon) \sum_{\forall t_x \in T_x \forall Y} \sum_{replies \text{ in } t_x} repliesCount_{YX_{t_x}} * influence_Y \quad (2)$$

其中 $influence_X$ 是用户 X 的社区影响力指数, P_X 是 X 的社区积分, T_X 是 X 所发的主题列表, $repliesCount_{YX_{t_x}}$ 表示 Y 在主题 t_x 内回复 X 的次数, ϵ 是阻尼因子, 代表论坛用户归一化后的社区积分在其影响力指数中所占比重.

2) 发表回复作者的影响力:

$$influence_Y = \epsilon P_Y + (1 - \epsilon) \sum_{\forall r_y \in R_y \forall Z} \sum_{replies \text{ in } r_y} repliesCount_{ZY_{r_y}} * influence_Z \quad (3)$$

其中 $influence_Y$ 是用户 Y 的影响力指数, P_Y 是 Y 的社区积分, R_Y 是 Y 回复的主题列表, $repliesCount_{ZY_{r_y}}$ 是用户 Z 在主题 r_y 内回复 Y 的次数, ϵ 是阻尼因子, 代表论坛用户归一化后的社区积分在其影响力指数中所占比重.

4.2 意见领袖社区发现算法

(FOLCM Forum Opinion Leaders Community Mining), 算法伪码如下:

基于天涯论坛用户社区影响力指数计算公式, 本节提出一种针对异质社区网络意见领袖社区挖掘算法

算法 1 意见领袖社区发现算法 FOLCM

```

输入: 论坛某板块的主题以及回复集合 PostList, 论坛内用户的社区积分集合 PointList, 阻尼因子 eps, 预设的领袖社区大小 n
输出: 舆论领袖社区集合  $A_n$  ( $n$  是预设的领袖社区大小)
1  sort(PostList, time_forward) // sort the posts with time forwarding
2  let A be the set of all author in PointList; // A is the set of author
3  for each author a in A do
4      a.influence = eps * PointList.a.point // set the initial influence of authors
5  for each post p in PostList do
6      author = p.author
7      ReplyList = post.ReplyList
8      for each reply r in ReplyList do
9          sender = r.sender
10         receiver = r.receiver
11         SenderInfluence = A.sender.influence //the influence of post author
12         A.receiver.influence += (1 - eps) * SenderInfluence // update the influence of reply receiver
13         if receiver != author // update the influence of post author
14             A.author.influence += (1 - eps) * SenderInfluence
15     sort(A, A.influence) and select the first  $N^{\text{th}}$  leaders as  $A_n$ 
16 return  $A_n$ 

```

设主题总数为 P , 回复总数为 R , 用户总数为 A , 对 FOLCM 算法的时间复杂度与空间复杂度进行分析:

时间复杂度分析: FOLCM 算法第 1 行对主题进行排序, 基于关键字比较的排序算法时间复杂度下界是 $O(P \lg P)$, 则第 1 行实际复杂度为 $O(P \lg P)$, 第 3-4 行执行时间复杂度为 $O(A)$, 第 5 行的 for 循环内语句执行 P 次, 第 8 行的 for 循环内语句环执行次数取决于每个主题的回复数, 则第 9-14 行执行时间复杂度为 $O(P + R)$. 因此, 总的时间复杂度为 $O(P \lg P + A + P + R)$, 由于实验数据中大多主题都有回复, 而且回复数大多大于 2, 所以 $R \geq A, R \geq P$, 则 $O(P \lg P + A + P + R) = O(P \lg P + R) = O(\max\{P \lg P, R\})$.

空间复杂度分析: 由于 FOLCM 算法本身只是执行计算, 没有额外申请空间, 因此 FOLCM 算法的空间复杂度为 $O(1)$.

5 实验

本节首先给出实验数据集来源以及验证方案, 而后将本文提出的意见领袖社区发现算法 (FOLCM) 与 Google 搜索、人工验证进行交叉比对, 验证本文算法在意见领袖社区影响力和召回率等方面的有效性.

5.1 实验环境及验证方案

实验数据来源: 天涯社区法律论坛板块, 包含法律论坛内的 1045 个主题以及相应主题下的所有回复, 与发表主题和回复的 14131 个天涯用户的社区积分数据.

实验验证方案主要包括以下三种:

- 使用 Google 搜索引擎在天涯论坛网页中依次搜索社区影响力指数 Top 20 天涯用户 ID, 对比意见领袖社区成员的社区影响力指数和 Google 搜索结果数
- 考察意见领袖社区 Top N ($1 \leq N \leq 20$) 相对于 Google 搜索结果数排名的召回率
- 从天涯社区获取意见社区影响力指数 Top 20 天涯用户的社区职务以及职业信息, 人工核实意见领袖的社区影响力

5.2 实验结果与分析

5.2.1 FOLCM 与 Google 搜索的影响力指数对比

Google 搜索引擎在天涯论坛网页中依次搜索社区影响力指数 Top 20 天涯用户 ID, 分别记录搜索结果数, 对比用户的社区影响力指数, 对比结果如图 1 示. 图中蓝色和红色曲线分别代表归一化后的天涯用户社区影响力指数和 Google 搜索结果数, 递减排序.

两条曲线均呈现下降趋势, 且 FOLCM 算法的挖掘

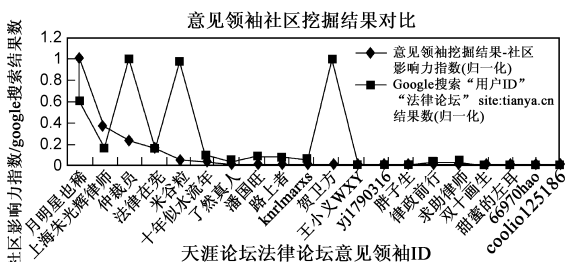


图1 意见领袖社区发现结果对比Google搜索结果

结果基本与 Google 搜索结果相吻合,验证了算法的有效性.但图中存在三个 Google 搜索结果数很高的异常点,通过查看其个人信息发现,三人中两人是版主,一人是知名网友.三个人的注册时间最短的也有一年半,最高的甚至接近七年,在论坛中积累的主题回复都很多,因此无论社区头衔或资历均很高,搜索结果数很高.但三人近期活跃度有所下降,在实验采集的 1045 个主题中,影响力较低.

5.2.2 FOLCM 相对于 Google 搜索结果的召回率

本文召回率是指 FOLCM 挖掘结果相对于 Google 搜索结果的比率.通过将意见领袖社区挖掘结果 Top20 的天涯用户 ID 取出,使用 Google 搜索引擎在天涯社区站内进行搜索,记录搜索所得结果数,按照搜索结果数由高到低进行排名.两种方案的结果对比如表 1.

表 1 意见领袖社区发现的 Top20 天涯用户排名与 Google 搜索得到的天涯用户排名

| 按照社区影响力指数递减 | | 按照 Google 搜索用户 ID 结果数递减 | |
|-------------|--------------|-------------------------|--------------|
| 1-10 名 | 11-20 名 | 1-10 名 | 11-20 名 |
| 月明星也稀 | 贺卫方 | 贺卫方 | 求助律师 |
| 上海朱光辉律师 | 王小义 WXY | 仲裁员 | 了然真人 |
| 仲裁员 | yjl790316 | 米谷粒 | 律政前行 |
| 法律在宪 | 胖子生 | 月明星也稀 | 胖子生 |
| 米谷粒 | 律政前行 | 法律在宪 | 双十画生 |
| 十年似水流年 | 求助律师 | 上海朱光辉律师 | yjl790316 |
| 了然真人 | 双十画生 | 十年似水流年 | 王小义 WXY |
| 潘国旺 | 甜蜜的左耳 | 潘国旺 | 甜蜜的左耳 |
| 路上者 | 66970hao | 路上者 | 66970hao |
| karlmarxs | coolio125186 | karlmarxs | coolio125186 |

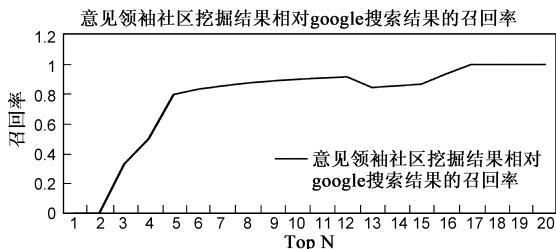


图2 意见领袖社区发现结果对比Google搜索结果的召回率

根据表 1 计算 Top N (1 ≤ N ≤ 20) 的召回率如图 2

所示:

- 意见领袖挖掘结果 Top 5 相对于 Google 搜索结果 Top 5 天涯用户的召回率达到 0.8
- 意见领袖挖掘结果 Top 10 相对于 Google 搜索 Top 10 天涯用户的召回率达到 0.9
- 意见领袖挖掘结果 Top 17 相对于 Google 搜索 Top 17 天涯用户的召回率达到 100%

5.2.3 人工核实验证

由于 Google 的搜索结果数排名并不能完全代表天涯用户在天涯社区内的影响力高低排名,为进一步确认 FOLCM 算法稳定性,从天涯社区获取意见社区影响力指数 Top 20 天涯用户的社区职务以及职业信息(如表 2),人工考察意见领袖社区挖掘结果.

表 2 社区影响力指数 Top 20 天涯用户的社区职务以及职业信息

| 天涯用户 ID | 社区影响力指数(归一化) | Google 搜索结果数(归一化) | 社区职务 | 职业 |
|--------------|--------------|-------------------|--------|--------|
| 月明星也稀 | 1 | 0.598993289 | 法坛助理版主 | 律师 |
| 上海朱光辉律师 | 0.368028 | 0.156040268 | | 律师 |
| 仲裁员 | 0.234426 | 0.998322148 | 法坛版主 | 律师 |
| 法律在宪 | 0.162447 | 0.161073826 | | 律师 |
| 米谷粒 | 0.0480202 | 0.964765101 | 法坛版主 | 律师 |
| 十年似水流年 | 0.0248758 | 0.090604027 | | 法律行业人士 |
| 了然真人 | 0.0130475 | 0.02852349 | | 律师 |
| 潘国旺 | 0.0125108 | 0.088926174 | | 律师 |
| 路上者 | 0.0113539 | 0.082214765 | | 律师 |
| karlmarxs | 0.0101 | 0.052013423 | | |
| 贺卫方 | 0.0100048 | 1 | 法坛知名网友 | 法学学者 |
| 王小义 WXY | 0.01 | 0.003355705 | | 律师 |
| yjl790316 | 0.00873842 | 0.006711409 | | |
| 胖子生 | 0.00710838 | 0.011744966 | | |
| 律政前行 | 0.00681841 | 0.02852349 | | 律师 |
| 求助律师 | 0.0062774 | 0.033557047 | | |
| 双十画生 | 0.00617309 | 0.008389262 | | 律师 |
| 甜蜜的左耳 | 0.0054501 | 0.003355705 | | |
| 66970hao | 0.00482109 | 0.003355705 | | |
| coolio125186 | 0.00435207 | 0.003355705 | | |

法律论坛中就职业和专业背景而言,法学学者或其他法律相关行业人士的意见最有说服力,更容易成为意见领袖. FOLCM 挖掘结果中该类用户排名最为靠前,且大多集中在列表中排名较高的地方.从表 2 中可见,社区影响力指数 Top 5 的天涯用户中有三人是版主,Top 10 天涯用户中有 9 位律师,Top 20 天涯用户中大部分都是律师或法律相关行业人士.由上述上面实验结果分析,可以得出结论:本文提出的意见领袖社区挖掘方案能够比较准确地挖掘论坛板块社区中的意见

领袖社区.

5.2.4 与同质网络方法对比

同质网络中 HITS 算法^[16]和 PageRank 算法^[17]通过对网页之间的超级链接引用进行分析对网页流行度进行排名的算法,但上述算法并不直接适用于意见领袖社区发现,因为:1)本文的研究对象是论坛成员,而 HITS 算法与 PageRank 的研究对象是 web 页面. HITS 算法中有 Hubs 与 Authorities 页面两类网页,如果对应到论坛成员中:Hubs——往往引述他人观点的人,Authorities——有威信并能获得大多数人支持的人,但总是引述他人观点的人虽然了解的知识甚深,但不向他人传达自身的想法,是难以获得多数人认可的;2)PageRank 的理论基础不适用于本文. PageRank 算法的基础——随机游走模型(The Random Surfer Model)为马尔可夫过程,页面是状态、状态之间的转换是超级链接,它们是等可能的.而论坛用户之间的关系只涉及在论坛主题之间的相互回复.虽然用户会通过点击一个 web 页面 A 内的超级链接到达另一个页面 B 继续浏览,但论坛用户 A 却几乎不会因为要回复用户 B 而继续回复 B 回复过的用户;3)从网络结构上看,异质网络中包含的 web 对象有论坛用户、论坛主题、主题内回复和回复之间的回复,而 HITS 与 PageRank 的研究对象则是同质网络,所有的 web 页面即是对象.而且,从更高层次上来说,PageRank 计算的是对浏览 web 页面的人来说页面的重要度,而本文要研究的是论坛用户在论坛内的重要度.因此同质网络方法无法直接适用于本问题.

本文通过对 PageRank 与 HITS 算法针对舆论领袖发现进行了重新设计. FOLCM 算法结果 Top 5 天涯用户相对于人工筛选的召回率达到 80%、Top 10 召回率达到 90%、Top 17 召回率接近 100%; PageRank 筛选结果 Top 5 天涯用户相对于人工筛选的召回率达到 70%、Top 10 召回率达到 88%、Top 17 召回率接近 92%; HITS 筛选结果 Top 5 天涯用户相对于人工筛选的召回率达到 50%、Top 10 召回率达到 68%、Top 17 召回率接近 77%. 可见本文提出的算法要优于基于同质网络的 PageRank 与 HITS 的舆论领袖筛选算法.

6 结束语

本文提出用户“社区影响力指数”的概念,量化用户社区影响力.通过研究 web 论坛社区中意见领袖演化模式,提出了意见领袖社区发现算法 FOLCM,以 web 论坛的用户社区积分数据为基础,以投票思路计算用户当前的社区影响力指数.对已经计算得到用户社区影响力指数的社区网络,提出意见领袖选取的方案和实验结果评价方案.通过真实数据进行实验与分析,提出的意见领袖社区发现算法能够较准确地挖掘论坛中的

意见领袖社区.

参考文献

- [1] Lazer D, Pentland A S, Adamic L, Aral S, et al. Life in the network: the coming age of computational social science [J]. Science, 2009, 323(5915): 721 - 723.
- [2] 王飞跃, 曾大军, 毛文吉. 社会计算的意义、发展与研究状况 [J]. 科研信息化技术与应用, 2010, 1(2): 3 - 15.
WANG Fei-yue, ZENG Da-jun, MAO Wen-ji. Social Computing: Its Significance, Development and Research Status [J]. E-Science Technology & Application, 2010, V1(2): 3 - 15. (in Chinese)
- [3] 程学旗, 沈华伟. 复杂网络的社区结构 [J]. 复杂性系统与复杂性科学, 2011, 8(1): 57 - 70.
CHENG Xue-qi, SHEN Hua-wei. Community structure of complex networks [J]. Complex Systems and Complexity Science, 2011, 8(1): 57 - 70. (in Chinese)
- [4] 刘挺. 方兴未艾的社会计算 [OL]. (2011-3-12) [2011-6-26]. http://blog.sina.com.cn/s/blog_4cbec5e90100pyvz.html.
LIU Ting. Growing Social Computing [OL]. (2011-3-12) [2011-6-26]. http://blog.sina.com.cn/s/blog_4cbec5e90100pyvz.html. (in Chinese)
- [5] Zhang Y C, Xu G D. Using web clustering for web communities mining and analysis [A]. IEEE//ACM International Conference on Web Intelligence and Intelligent Agent Technology [C]. USA: IEEE Computer Society, 2008. 20 - 31.
- [6] Yang B, Huang J, Liu D Y, et al. A multi-agent based decentralized algorithm for social network community mining [A]. International Conference on Advances in Social Network Analysis and Mining [C]. USA: IEEE Computer Society, 2009. 78 - 82.
- [7] Jin D, Liu D Y, Yang B, et al. Ant colony optimization with markov random walk for community detection in graphs [A]. Advances in Knowledge Discovery and Data Mining [C]. USA: IEEE Computer Society, 2011. 123 - 134.
- [8] Chen J, Zaiane O, and Goebel R. Detecting communities in social networks using max-min modularity [A]. SIAM International Conference on Data Mining [C]. USA: SIAM Society, 2009. 978 - 989.
- [9] 鲁明羽, 沈抖, 郭崇慧, 陆玉昌. 面向网页分类的网页摘要方法 [J]. 电子学报, 2006, 34(8): 1475 - 1480.
LU Ming-yu, SHEN Dou, GUO Chong-hui, LU Yu-chang. Web-page summarization methods for web-page classification [J]. Acta Electronica Sinica, 2006, 34(8): 1475 - 1480. (in Chinese)
- [10] 陈毅恒, 秦兵, 宋凡, 刘挺, 李生. 基于 ontology 抽取优化初始选择的检索结果聚类 [J]. 电子学报, 2008, 36(12A): 166 - 170.

- CHEN Yi-heng, QIN Bing, SONG Fan, LIU Ting, LI Sheng. Search result clustering based on centroid optimization by ontology extraction [J]. Acta Electronica Sinica, 2008, 36 (12A): 166 – 170. (in Chinese)
- [11] Cai D, Shao Z, He X F, et al. Community mining from multi-relational networks [A]. Principles and Practice of Knowledge Discovery in Databases [C]. USA: IEEE Computer Society, 2005. 8.
- [12] Tang L, Liu H, Zhang J P, et al. Community evolution in dynamic multi-mode networks [A]. ACM SIGKDD international conference on Knowledge discovery and data mining [C]. USA: ACM Society, 2008. 677 – 685.
- [13] Sun Y Z, Tang J, Jiawei Han, et al. Community evolution detection in dynamic heterogeneous information networks [A]. The Eighth Workshop on Mining and Learning with Graphs [C]. USA: IEEE Computer Society, 2010. 137 – 146.
- [14] Murata T, Ikeya T. Analysis of online question-answering forums as heterogeneous networks [A]. International Conference on Weblogs and Social Media [C]. USA: IEEE Computer Society, 2008. 210 – 211.
- [15] Zhao Q K, Bhowmick S S, Zheng X. Characterizing and predicting community members from evolutionary and heterogeneous networks [A]. ACM conference on Information and knowledge management [C]. USA: ACM Society, 2008. 309 – 318.
- [16] Chakrabarti S, Dom B, Raghavan P, et al. Automatic resource compilation by analyzing hyperlink structure and associated text [J]. Computer Networks and ISDN Systems. 1998, 30(1-7): 65 – 74.
- [17] Page L, Brin S, Motwani R, et al. The PageRank Citation Ranking: Bringing Order to the Web [R]. Stanford InfoLab. 1999: 1 – 17.

作者简介



张伟哲 男, 1976 年生于黑龙江省哈尔滨市. 副教授, 博士生导师. 计算机学会高性能计算专委会委员, 计算机学会青年科技论坛 YOCSEF 哈尔滨分论坛副主席, 中国计算机学会会员, IEEE&IEEE Computer 会员. 主要从事网络计算、分布式系统和网络安全方面的研究.

E-mail: wzhang@hit.edu.cn