

基于“问题-用户”的网络问答社区专家发现方法研究*

■ 龚凯乐 成颖

南京大学信息管理学院 南京 210023

摘要: [目的/意义]以网络问答社区为研究对象,提出基于“问题-用户”传播网络的专家发现方法,为建立用户激励机制、完善专家推荐方法提供借鉴。[方法/过程]通过分析开放问答模式的特点,以“问题”和“用户”为节点、“答题关系”为有向边,构建“问题-用户”权威值传播网络,利用答案质量改进加权的 HITS 算法。[结果/结论]提出的算法可以较好地兼顾用户的答题数量与答案质量,能够选择出活跃度高、知识渊博的用户作为专家。

关键词: 网络问答社区 专家发现 传播网络 链接分析 HITS 算法

分类号: G203

DOI:10.13266/j.issn.0252-3116.2016.24.016

1 引言

网络问答社区 (community question answering, CQA) 是 Web2.0 环境下的一种新型应用,通过用户在线问答与反馈实现信息分享与知识传递。因其具有交互性、开放性等特点,能够解决用户个性化的信息需求, Yahoo! Answers、百度知道、知乎等平台在备受网民欢迎的同时也引起了学界的广泛关注。由于 CQA 的核心是用户生成内容,具备专业知识并乐于分享的用户往往能够提供优质的答案,因此专家发现成为学界的研究热点。

目前,越来越多的 CQA 平台采用开放的问答模式^[1],不再设置“最佳答案”或“被采纳答案”以结束讨论,而是鼓励更多用户尽可能深入地探讨相关问题,从而使问题下的观点更加多元,答案更加准确、全面。本文围绕开放的问答模式,提出基于“问题-用户”传播网络的专家发现方法,具体工作包括:①分析不同问答模式的特点及其传播网络。②以问题和用户为节点、答题关系为边,构建“问题-用户”传播网络,利用答案质量改进加权的 HITS 算法以发现专家。通过对知乎“数据挖掘”领域的实证分析,证实了本算法的可行性、有效性。

2 相关研究

CQA 的专家发现方法主要分为指标评价、链接分析和用户建模 3 种,前二者适用于宏观地发现某一领域的专家,后者则主要应用于具体问题的专家推荐。

指标评价法通过统计用户提问数、答题数、获赞数等数据构建相应的指标体系,综合评判用户的权威度。由于其较为简单,主要作为用户建模方法的组成部分,单纯基于此的研究并不多见。J. Zhang 等^[2]提出的 Z-score 基于如下假设:用户在某一领域答题越多意味着其具备较强的专业知识,提问越多则表明其缺乏该领域知识。因此当用户在某一领域内答题数为 a 、提问数为 q 时利用式(1)计算出 z 值,降序确定专家。

$$z = (a - q) / \sqrt{a + q} \quad \text{公式(1)}$$

链接分析法主要利用 PageRank^[3]与 HITS^[4]算法,将用户类比为网页,并将链接关系移植到问答关系上,通过构建“用户-用户”传播网络并迭代计算用户的权威值对其排序。J. Zhang 等^[2]提出的 ExpertiseRank 算法借鉴了 PageRank 算法的权威值传递思想。L. Hong 等^[5]利用概率潜在语义分析 (probabilistic latent semantic analysis, PLSA) 构建了内容主题相关的 PageR-

* 本文系国家自然科学基金面上项目“融合范式视角下的链接分析理论集成框架及其实证研究”(项目编号:71273125)和中国科学技术信息研究所合作研究项目“用户行为视角的链接动机与系统可用性研究”的研究成果之一。

作者简介: 龚凯乐 (ORCID:0000-0001-9269-8669), 硕士研究生, E-mail: njugong@163.com; 成颖 (ORCID:0000-0002-0664-7206), 教授, 博士生导师, 博士。

收稿日期: 2016-07-13 **修回日期:** 2016-12-10 **本文起止页码:** 115-121 **本文责任编辑:** 徐健

ank 算法,并将其应用于专家发现。戴秋敏^[6]采用赞成、反对、选择最佳答案等用户反馈对 PageRank 算法中的问答关系加权,兼顾了答题的质量。在 HITS 算法的应用中,P. Jurczyk 等^[7]对每个用户设立基于提问的中心度(hub)和基于答题的权威值(authority),迭代计算权威值并排序。此后 J. Shen 等^[8]引入点赞、举报等用户反馈,提出加权的 HITS 专家发现算法。宣慧明^[9]利用语义相似度抽取与用户提问相关的问题集合,加大相似问题回答者的权重,基于 HITS 算法确定专家。

用户建模法综合运用了信息检索、指标评价、链接分析和机器学习等领域的研究成果。X. Liu 等^[10]将擅长解答某一问题的专家定义为回答过相似问题的用户,通过分析全体用户的答题记录抽取相应的词项形成专家概况(expert profile),将待回答问题作为提问式(query),专家概况作为文档(document),探寻查询相似度模型、关联模型和基于聚类的语言模型在专家发现中的应用。F. Riahi 等^[11]实证分析了 TF-IDF、LDA 和 STM 等多种方法在构建专家概况和专家推荐中的应用效果。J. Zhang 等^[12]在上述思路的基础上,综合专业知识水平、问题发布时间、答题人数等多个指标推荐专家。A. Pal 等^[13]引入答题能力与积极性,抽取用户的 6 个特征建模,通过 SVM 等机器学习算法预测潜在专家。D. R. Liu 等^[14]依据主题偏好、声望和权威度的线性组合对用户建模,其中主题偏好由专家概况与目标问题的文本相似度算出,声望取决于用户的历史答题数与最佳答案数,权威度则依靠链接分析算法求得。Y. Tian 等^[15]将 LDA 应用于构建用户概况,并基于用户过往答题所获得的投票信息计算其在每个主题中的声望,通过机器学习预测答题专家。G. Zhou 等^[16]同样采用 LDA 模型生成用户概况,基于提问者与答题者的用户概况相似度构建主题敏感的问答关系,通过 PageRank 算法发现专家候选人,并利用概率模型最终选定专家。林鸿飞等^[17]基于链接分析算法计算用户在每个主题中的权威度,利用 LDA 模型计算主题间的相似度和用户对所有主题的参与度,综合权威度与相近主题的参与度对用户排序。H. Li 等^[18]利用 tag-LDA 模型和 word2vec 提升了主题敏感的 PageRank 算法应用于专家排名的效果。X. Cheng 等^[19]针对 LDA 识别短文本的不足,提出一种基于标签词的主题模型,将其应用于判别用户所属的领域,并通过机器学习算法根据用户答题的反馈情况对其排序,列举出某一问题的潜在专家。S. Patil 等^[20]通过分析专家与非专家的用户行为,构建基于用户活动特征、答案质量特征、语言特征和时间特

征的专家判别模式,采用统计模型发现专家。

回顾相关文献可以发现,近期的研究热点主要集中在通过用户建模方法进行答题专家的推荐。在这类研究中,一般利用信息检索与自然语言处理技术定位用户的擅长领域,并通过问答关系构建“用户-用户”传播网络,采用 PageRank、HITS 等链接分析算法迭代计算用户的权威值,选定可能提供优质答案的专家。因此,无论是某个领域的专家发现,还是具体问题的专家推荐,链接分析方法均起到了重要作用。随着越来越多的 CQA 平台采用开放的问答模式,不再通过设置“最佳答案”或“被采纳答案”关闭问题,使得问题可以被持续研讨,使得部分问题中汇集了大量的用户智慧,这些特点为 CQA 专家发现的链接分析方法提供了新的思路。因此,本文拟围绕这些特点提出基于“问题-用户”的传播网络并利用答案质量加权改进 HITS 算法,以期拓展用户建模等领域的深入研究。

3 研究设计

3.1 不同问答模式的特点及其传播网络

3.1.1 封闭的问答模式及其传播网络 在百度知道等早期的 CQA 平台中,一个问题的生命周期始于提出,终于提问者满足其知识需求后选择出“最佳答案”或“被采纳的答案”,其后问题被自动关闭。这种较为封闭的问答模式面向提问者的知识需求,以提问者和答题者之间的知识传递为核心,每个问题往往仅有少数几个回答。故早期的专家发现研究从该问答模式中抽象出以“用户”为节点、“问答关系”为有向边的“用户-用户”权威值传播网络。若该模式下有问答关系如图 1 所示,则抽象出的“用户-用户”传播网络主要有两种,见图 2、图 3。

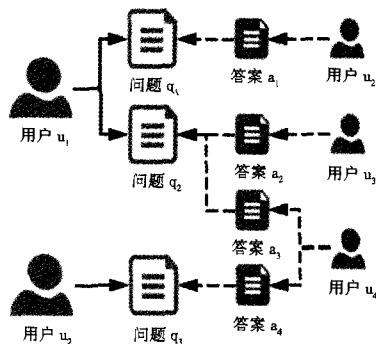


图 1 问答关系(a)

在图 2 的传播网络中,用户 u_i 回答了用户 u_j 的问题(特别是 u_i 的答案被 u_j 所采纳),表明 u_i 将其具有的知识传递给 u_j ,从而满足了 u_j 的知识需求,因此 u_i

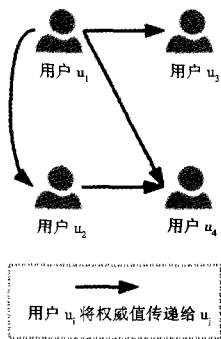


图2 “用户-用户”传播网络 I

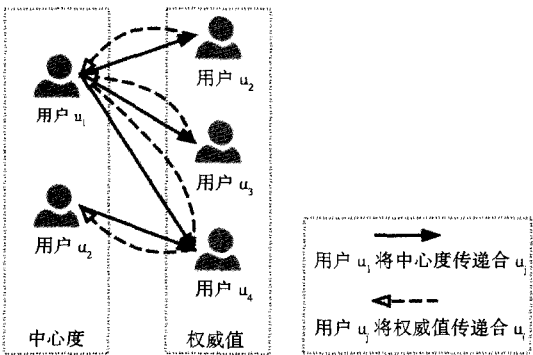


图3 “用户-用户”传播网络 II

相较于 u_j 在相关领域中拥有更为丰富的知识储备。故 u_j 将其权威值传递给 u_i ，促进了 u_i 的权威值增长，说明 u_i 更可能是专家。利用 PageRank 等链接分析算法在该网络中迭代传播用户的权威值并进行排序以发现专家^[2,5,6]。研究中发现该算法存在如下不足：问题在获得“最佳答案”后即关闭的较封闭的问答模式导致用户群体形成了闭环网络，该网络中成员间的频繁互动使其具有较高的权威值，但这并不能够真实体现其在整个问答社区中的影响力与知识水平^[21]。

在图3的传播网络中，每个用户具有中心度和权威值两个属性。用户 u_i 的出度越多，说明其能提出较“好”的问题，吸引其他用户回答，故中心度高；用户 u_j 回答了较多高中心度的用户所提出的问题，说明其能满足这些用户的知识需求，故权威值高。利用 HITS 算法在该网络中迭代传播中心度与权威值以完成专家发现^[7-9]。研究中发现该算法存在如下不足：不同于社交媒体中以吸引其他用户关注为目的的交互动机，在 CQA 中，用户的提问动机来源于知识需求。基于省力原则，用户一般在平台中搜索、浏览已有的答案，在知识需求仍无法满足的情况下则转向提问^[22]。因此，CQA 平台中存在许多提问寥寥或从未提问的用户，呈现出“去中心”^[23]的特点，从而使得该算法效果欠佳。

3.1.2 开放的问答模式及其传播网络 在知乎等开

放的 CQA 平台中，系统不再设置“最佳答案”或“被采纳的答案”，所有问题均被持续研讨。这种开放的问答模式面向整个社区用户的知识需求与兴趣，每个问题的回答人数、蕴含的知识总量不再取决于提问者的需求是否得到满足，而是由问题本身的话题效应^[24]、答题者的意见领袖效应^[25]、现有答案是否满足大众的认知需要等因素综合决定，故该问答模式中出现了部分汇集了大量用户充分讨论的问题。在该问答模式中，判断一个答案好坏的依据不再仅仅是提问者的采纳与否，而是其观点是否得到了社区中其他用户的支持或反对，具体表现为该答案获得的“点赞”数、“投票”数等。据此，本文将专家定义为答题积极且答案质量较优的用户。从开放的问答模式中可以抽象出以“问题”和“用户”为节点、“答题关系”为有向边的“问题-用户”权威值传播网络。若该模式下有如图4所示的问答关系，则可以抽象出“问题-用户”传播网络，如图5所示：

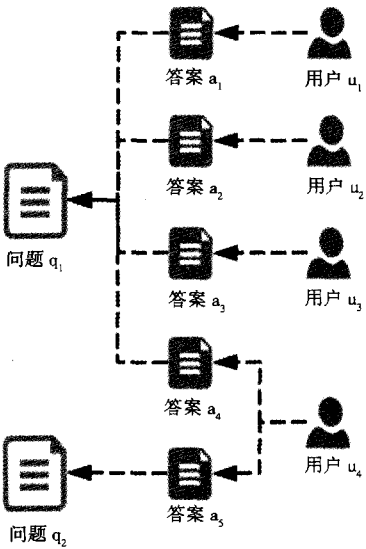


图4 问答关系 (b)

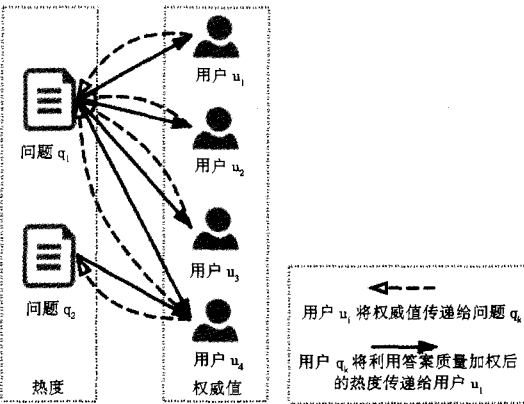


图5 “问题-用户”传播网络

其中,问题起到了汇集用户权威值并依据答案质量对权威值“再分配”的作用。例如问题 q_k 被 u_i, u_j 等用户回答,故 u_i, u_j 等用户将其权威值传递给问题 q_k ,使其热度升高(所谓热度是指回答该问题所有用户的权威值之和);反过来,问题 q_k 再将收集到的权威值(表现为升高后的热度)依据答案质量的优劣加权后传递给答题者。在同一个问题中,答案质优的用户不仅接收了其自身传递给问题的权威值,还通过该问题间接获得了其他用户的权威值;答案质量不高的用户虽然相较于没有答题的用户,其权威值有所升高,但由于其将自己的权威值通过问题热度间接地传递给了答案质优的用户,拉大了其与这些用户权威值的差距,表明答案质优的用户相较于自己在该领域具有更丰富的知识储备。基于该传播网络,答题积极且质优(特别是在高热度的问题中提供了优质答案)的用户,具有较高的权威值,依据权威值可以确定专家。

3.2 基于问题热度与用户权威值的 HITS 算法

为了实现问题热度根据答案质量对用户权威值的“再分配”,本文改进了加权的 HITS 算法。设问题集合为 $Q = \{q_1, q_2, \dots, q_n\}$,用户集合为 $U = \{u_1, u_2, \dots, u_m\}$,“问题-用户”关系集合为 $C = \{ \langle q_k, u_i \rangle, \dots \}$, $h(q_k)$ 是问题 q_k 的热度, $a(u_i)$ 是用户 u_i 的权威值, $h_{old}(q_k)$ 和 $a_{old}(u_i)$ 分别用于存放上一次迭代时 q_k 的热度和 u_i 的权威值, $w(\langle q_k, u_i \rangle)$ 是用户 u_i 在问题 q_k 中依据答案质量得出的加权系数。据此,本算法流程如下:

Step1: 依据公式(2)和公式(3)初始化每个问题的热度与每个用户的权威值。

$$h(q_k) = h_{old}(q_k) = 1 \quad \text{公式(2)}$$

$$a(u_i) = a_{old}(u_i) = 1 \quad \text{公式(3)}$$

Step2: 依据式(4)和式(5)更新每个问题的热度与每个用户的权威值。

$$h(q_k) = \sum_{(q_i, u_i) \in C} a_{old}(u_i) \quad \text{公式(4)}$$

$$a(u_i) = \sum_{(q_i, u_i) \in C} [h_{old}(q_k) \times w(\langle q_k, u_i \rangle)] \quad \text{公式(5)}$$

Step3: 依据公式(6)标准化每个问题的热度,并由公式(7)赋值给 h_{old} ;依据公式(8)标准化每个用户的权威值,并由公式(9)赋值给 a_{old} 。

$$h(q_k) = \frac{h(q_k)}{\max_{q_j \in Q} [h(q_j)]} \quad \text{公式(6)}$$

$$h_{old}(q_k) = h(q_k) \quad \text{公式(7)}$$

$$a(u_i) = \frac{a(u_i)}{\max_{u_j \in U} [a(u_j)]} \quad \text{公式(8)}$$

$$a_{old}(u_i) = a(u_i) \quad \text{公式(9)}$$

Step4: 若任意用户均满足公式(10),则执行 Step5,否则转至 Step2,其中 ε 取 0.000 000 000 1。

$$|a(u_i) - a_{old}(u_i)| < \varepsilon \quad \text{公式(10)}$$

Step5: 按权威值由高到低对用户排序,排名靠前的用户为该领域的专家。

当算法依据答案的获赞数对热度加权时,可由公式(11)设定加权系数,其中 $s(\langle q_k, u_i \rangle)$ 为用户 u_i 在问题 q_k 中答题所获得的点赞数, $s_sum(q_k)$ 为问题 q_k 中所有答案获得的点赞总数。假设用户浏览并评价所有答案的概率相同,则该加权系数反映了不同答案获得赞同的概率之差,即大众认同的差异。

$$w(\langle q_k, u_i \rangle) = \frac{s(\langle q_k, u_i \rangle)}{s_sum(q_k)} \quad \text{公式(11)}$$

当算法依据 CQA 平台中的答案排名机制对热度加权时,可由公式(12)设定加权系数,其中 $r(\langle q_k, u_i \rangle)$ 为用户 u_i 的答案在问题 q_k 中的排名,1是平滑系数,用以缩小前几名答案之间的差距。设定该加权系数的前提假设是答案的质量大致呈幂律分布。

$$w(\langle q_k, u_i \rangle) = \frac{1}{r(\langle q_k, u_i \rangle) + 1} \quad \text{公式(12)}$$

4 实验与结果分析

4.1 数据采集与处理

通过编写网络爬虫获取了知乎平台“数据挖掘”话题页面中截止 2016 年 5 月 16 日的的数据,其中问题 4 941 个、答案 11 540 个、精华答案 1 000 个、非匿名答题用户 7 456 个;同时爬取了每个答案在问题中的排名与获赞数,每个用户的全部提问信息、擅长话题、教育经历与职业经历。通过答案信息中包含的用户 ID,统计每个用户在该领域内的答题总数、获赞总数、答案平均获赞数与精华答案数,删去匿名答案。

4.2 专家发现效果的评价

本文采用定量与定性分析两种方式评价专家发现效果。

4.2.1 定量分析 定量分析首先需要人为确定专家评价列表,进而比较其与本研究结果的相关性。J. Zhang 等^[2]通过人工打分形成专家列表,采用 Spearman 和 Kendall 相关系数评价研究结果的准确性;P. Jurczyk 等^[7]基于答案所获票数、最佳答案比例、题主评分均值形成 3 种专家列表,通过 Pearson 相关系数和 P@k 衡量研究结果的好坏;J. Shen 等^[8]采取了类似的策略。

参考前人研究,本文将用户按领域内的答题总数、

获赞总数、答案平均获赞数与精华答案数由高到低分别排序,形成4种专家评价列表,从多种角度进行分析:

(1) 答题总数反映了用户在该领域内的活跃度。

(2) 获赞总数是用户所有答案获得的累积“赞”数,体现了社区中其他用户对该用户专业知识的累积认同。答题积极,回答了某些高热度的问题且答案质优的用户均具有较高的获赞总数。

(3) 答案平均获赞数由用户的获赞总数除以答题总数求出,体现了该用户每篇答案所能获得大众认同的均值,其弱化了用户的活跃度因素,答题少而精的用户排名较高。

(4) 精华答案是知乎评选出的优质答案,因此精华答案数类似科研评价中的h指数^[26],在反映用户答题数量的同时体现了部分质量因素。

定量分析的评价指标采用Spearman相关系数^[2],它由专家评价列表中的前50位用户的排名与其在本研究结果中的排名的相关性求得,结果如表1所示:

表1 专家发现效果的定量分析

Spearman 相关系数	答题总数	获赞总数	平均获赞数	精华答案数
答案获赞数加权	0.638 **	0.375 **	0.328 *	0.623 **
答案排名加权	0.733 **	0.222	0.220	0.379 **

注:*. 在0.05级别(双尾),相关性显著;** 在0.01级别(双尾),相关性显著

表1显示,本研究提出的算法采用答案获赞数加权时得出的专家列表与4种专家评价列表均显著相关,说明该算法既能选择答题数量多的用户,也能选择答案质量好的用户作为专家,从而较好地兼顾用户的活跃度及其专业知识。此外,该算法得出的专家列表与依据答题总数、精华答案数排序的专家评价列表亦具有较高的相关系数,说明本方法更倾向于选择答题活跃、乐于分享知识的用户作为专家。

4.2.2 定性分析 为了进一步验证专家发现的效果,本文列举了算法所得出的前10名专家的详细信息,进行定性分析。如表2所示:

表2 专家发现效果的定性分析

排名	用户ID	权威值	答题 总数	获赞 总数	平均 获赞数	精华 答案数	擅长话题	教育经历	职业经历
			排名	排名	排名	排名			
1	xiaozhibo	0.056 341	72	6 491	90.15	20	计算机软件	大连海事大学	南洋理工大学
			2	3	121	1		· 机器学习	· Research Fellow
2	xiaomingwang	0.045 306	34	10 417	306.38	13	互联网	南安普顿大学	携程
			9	1	37	2		· Data Mining	
3	hanhsiao	0.042 759	26	7 772	298.92	10	互联网	知行合一	运营商
			15	2	39	6		· 数据分析	· 数据挖掘
4	ji-lu - 15 - 70	0.023 298	37	580	15.68	12	互联网	大连大学	保密
			7	37	508	3		· 控制理论与控制工程	· 数据挖掘 工程师
5	xie-ke - 41	0.017 387	4	5 750	1 437.5	4	互联网	康奈尔大学	Twitter
			297	4	6	25		· PhD in IS 北京航空航天大学、Rutgers University · 计算机科学	· 搜索引擎 Betaworks、Social Flow · Data Scientist
6	yeka	0.017 280	1	2 467	2 467	1	出版业	四川大学	JUSTPUB
			2 721	8	2	317			· 创始人
7	maigo	0.015 463	24	789	32.88	10	高等教育	Carnegie Mellon University	Facebook
			18	26	285	5		· 语音识别	· Software Engineer Intern
8	li-eta	0.015 354	28	203	7.25	7	高新科技	还没有填写	中国科学技术大学
			13	131	910	8			
9	liuliudong	0.014 692	19	313	16.47	11	高等教育	清华大学	清华大学
			25	88	486	4		· 自然语言处理	· 助理研究员
10	kennethse	0.014 502	24	505	21.04	5		浙江大学	阿里巴巴
			19	45	401	14	互联网	· 计算机技术 北京大学 · 分析化学	· 广告算法专家 网易 · 数据库开发工程师 华为 · 安全软件 工程师

从表2发现,在答题总数等4项指标中,上述用户至少在某一指标上名列前茅,排名最靠前的用户在多项指标中均表现优秀,证实了本文提出的专家发现方法可以有效地挖掘出答题数量多、答案质量好的专家用户。在用户的擅长话题中,绝大多数是互联网,其他还包括计算机软件、高新科技等。查阅知乎的话题结构发现,它们与“数据挖掘”联系紧密。教育经历中,绝大多数用户毕业于国内外知名高校且学科背景与计算机有关。同时,他们所从事的职业广泛涉及数据挖掘领域。因此,用户的专业背景也能够为本算法的有效性与可靠性提供有力佐证。

5 结语

本文围绕网络问答社区中开放的问答模式,以“问题”和“用户”为节点、“答题关系”为有向边,构建“问题-用户”权威值传播网络,利用答案质量加权的HITS算法迭代传递用户权威值,实现答案热度对用户权威值的“再分配”,按照权威值对用户排序选定专家。通过对知乎“数据挖掘”领域的实证分析,证实了该方法的可行性与有效性。本算法可以为开放性网络问答社区的用户激励机制、专家推荐等方面的研究与应用提供借鉴。

参考文献:

- [1] 刘佩,林如鹏. 网络问答社区“知乎”的知识分享与传播行为研究[J]. 图书情报知识, 2015(6):109-119.
- [2] ZHANG J, ACKERMAN M S, ADAMIC L. Expertise networks in online communities: structure and algorithms[C]// International conference on World Wide Web. New York: ACM, 2007:221-230.
- [3] BRIN S, PAGE L. The anatomy of a large-scale hypertextual web search engine[C]// International conference on World Wide Web. Amsterdam: Elsevier Science Publishers B. V., 1998:107-117.
- [4] KLEINBERG J. Authoritative sources in a hyperlinked environment[J]. Journal of the Acm, 1999, 46(5):604-632.
- [5] HONG L, YANG Z, DAVISON B D. Incorporating participant reputation in community-driven question answering systems[C]// International conference on computational science and engineering. Washington DC: IEEE Computer Society, 2009:475-480.
- [6] 戴秋敏. 互动问答平台专家发现及问题推荐机制的研究[D]. 上海: 华东师范大学, 2014.
- [7] JURCZYK P, AGICHTEIN E. Discovering authorities in question answer communities by using link analysis[C]// Sixteenth ACM conference on information and knowledge management. New York: ACM, 2007:919-922.
- [8] SHEN J, SHEN W, FAN X. Recommending experts in Q&A communities by weighted HITS algorithm[C]// International forum on information technology and applications. Washington DC: IEEE Computer Society, 2009:151-154.
- [9] 宣慧明. 基于交流平台QA系统的专家发现[D]. 南京: 南京师范大学, 2013.
- [10] LIU X, CROFT W B, KOLL M. Finding experts in community-based question-answering services[C]// ACM international conference on information and knowledge management. New York: ACM, 2005:315-316.
- [11] RIAHI F, ZOLAKTAF Z, SHAFIEI M, et al. Finding expert users in community question answering[C]// International conference companion on World Wide Web. New York: ACM, 2012:791-798.
- [12] ZHANG J, ACKERMAN M S, ADAMIC L, et al. QuME: a mechanism to support expertise finding in online help-seeking communities[C]// Proceedings of the 20th annual ACM symposium on user interface software and technology. New York: ACM, 2007: 111-114.
- [13] PAL A, HARPER F M, KONSTAN J A. Exploring question selection bias to identify experts and potential experts in community question answering[J]. Acm transactions on information systems, 2012, 30(2):1-28.
- [14] LIU D R, CHEN Y H, KAO W C, et al. Integrating expert profile, reputation and link analysis for expert finding in question-answering websites[J]. Information processing & management, 2013, 49(1):312-329.
- [15] TIAN Y, KOCHHAR P S, LIM E P, et al. Predicting best answers for new questions: an approach leveraging topic modeling and collaborative voting[C]// Social informatics. Heidelberg: Springer Berlin, 2014: 55-68.
- [16] ZHOU G, ZHAO J, HE T, et al. An empirical study of topic-sensitive probabilistic model for expert finding in question answer communities[J]. Knowledge-based systems, 2014, 66(9):136-145.
- [17] 林鸿飞, 王健, 熊大平, 等. 基于类别参与度的社区问答专家发现方法[J]. 计算机工程与设计, 2014, 35(1):333-338.
- [18] LI H, JIN S, LI S. A hybrid model for experts finding in community question answering[C]// International conference on cyber-enabled distributed computing and knowledge discovery. Washington DC: IEEE Computer Society, 2015:176-185.
- [19] CHENG X, ZHU S, CHEN G, et al. Exploiting user feedback for expert finding in community question answering[C]// IEEE international conference on data mining workshop. Washington DC: IEEE Computer Society, 2015:295-302.
- [20] PATIL S, LEE K. Detecting experts on quora: by their activity, quality of answers, linguistic characteristics and temporal behaviors[J]. Social network analysis and mining, 2015, 6(1):1-11.
- [21] JIAO J, YAN J, ZHAO H, et al. ExpertRank: an expert user ranking algorithm in online communities[C]// International conference on new trends in information and service science. Washing-

ton DC: IEEE Computer Society, 2009:674-679.

[22] 陈晓宇,邓胜利,孙雅梦.网络问答社区用户信息行为研究进展及展望[J].图书情报知识,2015(4):71-81.

[23] 袁红,赵娟娟.问答社区中用户与资源互动研究[J].图书情报工作,2014,58(18):102-109.

[24] 贾佳,宋恩梅,苏环.社会化问答平台的答案质量评估——以“知乎”、“百度知道”为例[J].信息资源管理学报,2013,3(2):19-28.

[25] 王秀丽.网络社区意见领袖影响机制研究——以社会化问答社区“知乎”为例[J].国际新闻界,2014(9):47-57.

[26] 宋振世,周健,吴士蓉.h指数科研评价实践中的应用研究[J].图书情报工作,2013,57(1):117-121.

作者贡献说明:

龚凯乐:设计研究方案,采集数据,实证分析,撰写初稿;

成颖:修正研究方案,修改论文。

Finding Experts in Community Question Answering Based on “Question - User”

Gong Kaile Cheng Ying

School of Information Management, Nanjing University, Nanjing 210023

Abstract: [Purpose/significance] Taking Community Question Answering as the research object, this paper proposes an experts finding algorithm, which is based on “question - user” propagation network, to provide a reference for establishing incentive mechanism and improving experts recommendation. [Method/process] By analyzing the features of open question answering model, this paper takes “question” and “user” as nodes, “answering question” as directed edges, and then builds an “question - user” authority propagation network. It also improves the weighted HITS algorithm by quality of answers. [Result/conclusion] Experimental results show that the algorithm can achieve a balance between number and quality of users’ answers, then select active and knowledgeable users as experts.

Keywords: Community Question Answering experts finding propagation network link analysis HITS algorithm

《现代图书情报技术》更名启事

《现代图书情报技术》将于2017年正式更名为《数据分析与知识发现》。

《数据分析与知识发现》是由中国科学院主管、中国科学院文献情报中心主办的学术性专业期刊。办刊宗旨是聚焦各行各业中以大数据为基础,依靠复杂挖掘分析方法,进行知识发现与预测、支持决策分析和政策制定的研究与应用,致力于提供理论指导、技术支持和最佳实践。该刊内容定位于广泛吸纳计算机科学、数据科学、情报科学以及数字科研、数字教育和数字文化等领域的技术与方法,研究数据驱动的语义计算、内容分析、数据挖掘、知识发现、智能管理和决策支持等方面的技术、方法、系统以及支撑设施、政策与机制等,尤其是聚焦从海量、异构、分布、动态、甚至富媒体数据中挖掘和发现知识以支持研究、管理和决策的理论、方法和技术。

具体内容详见该刊官网(<http://www.infotech.ac.cn>)。