

DOI: 10.7652/xjtuxb201911021

采用社会网络能力发现的众包服务节点选择算法

彭振龙^{1,2,3,4}, 桂小林^{2,3}

(1. 西安交通大学电子与信息学部, 710049, 西安; 2. 泉州师范学院陈守仁商学院, 362000, 福建泉州;
3. 西安交通大学陕西省计算机网络重点实验室, 710049, 西安;
4. 泉州师范学院云计算物联网电子商务智能福建省高校工程研究中心, 362000, 福建泉州)

摘要: 为了解决社会网络众包系统中由于潜在工作者未注册或注册用户的信息更新不及时而导致的服务节点选择不准确问题,提出了一种基于社会关系能力发现的网络众包服务节点选择算法。引入社会关系认知,以众包任务发布者或接受者个人为中心,利用社会网络的相关理论,提炼众包任务主题及相应的关键词,分析其与好友之间的交互信息中出现关键词的次数,结合其朋友圈中相应关键词的出现次数进行综合加权计算,量化朋友的能力大小,生成朋友能力矩阵,重构其社会关系。当众包任务生成或到达时,结合任务完成的质量因子、节点的诚信指数、匹配度等指标,有选择性地向符合能力要求的朋友进行任务转发。模拟微信交互的仿真实验表明:所提算法能较好地进行能力发现和选择合适的任务接受者,并且比随机分发平均节省了约 64.5%的时间,可以实现快速精准的任务分发与数据收集。

关键词: 社会网络;能力发现;众包;服务节点;任务分发

中图分类号: TP319 **文献标志码:** A **文章编号:** 0253-987X(2019)11-0148-08

Crowdsourcing Service Node Selection Algorithm Based on Social Network Ability Discovery

PENG Zhenlong^{1,2,3,4}, GUI Xiaolin^{2,3}

(1. Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China;
2. Tan Siu Lin Business School, Quanzhou Normal University, Quanzhou, Fujian 362000, China;
3. Key Laboratory of Computer Network of Shaanxi Province, Xi'an Jiaotong University, Xi'an 710049, China;
4. High Educational Engineering Research Center of Fujian Province for E-Commerce Intelligent Based on Cloud Computing and Internet of Things, Quanzhou Normal University, Quanzhou, Fujian 362000, China)

Abstract: To solve the problem of inaccurate crowdsourcing service node selection caused by the unsuccessful information update of the unregistered potential workers or the registered workers in the social network crowdsourcing system, a network crowdsourcing service node selection algorithm is proposed based on social relationship ability discovery. The social relationship cognition is introduced, and the crowdsourcing task publishers or recipients are supposed to be the tasks release centers. Following the relevant theories of social networks, the themes of crowdsourcing tasks and the corresponding keywords are refined, and the key words in the interaction information among friends are analyzed. The number of key words, combined with the

number of occurrences of the corresponding keywords in their friends circles, is comprehensively and weightedly calculated to quantify the size of a friend ability to generate a friend ability matrix, thus their social relationships are reconstructed. When a crowdsourcing task is generated or arrives, considering the quality factor of the task completion, the honesty index of the node, the matching degree, the task is selectively forwarded to the friend who meets the ability requirement. Simulation experiment of WeChat interaction shows that the proposed algorithm can better perform ability discovery, select appropriate task receiver, and save about 64.5% of time than random distribution to achieve fast and accurate task publishing and data collection.

Keywords: social network; capability discovery; crowdsourcing; service node; task distribution

众包系统的质量控制始终是一个严峻挑战^[1-2]。一个众包系统一般包括平台、网络、任务的发布者和任务的接受者。众包平台通常通过一定的激励机制,吸引任务接受者参与众包任务^[3-5]。由于众包系统的开放性、匿名性,导致可能存在很多的参与者不认真执行任务,而仅仅为了获得奖励,这一现象严重影响了众包系统的效率和可信度^[6-8]。控制众包服务质量的基本策略主要有两个:一是事后控制,即对众包服务的结果进行筛查处理;二是事前控制,即对参与者的能力和诚信度进行评估,选择更可靠更合适的参与者。本文从社会关系角度出发,选择更合适、更可靠的参与者,从而提升众包服务质量,属于事前控制策略。

在质量控制方面:Howe 提出了 golden standard data 模式,即在众包任务中随机插入一些具有标准答案的任务,根据参与者对这些任务的应答,判断参与者对其他任务的完成质量^[1];Peng 等将激励机制与参与者完成众包任务的结果质量相结合,根据众包任务参与者的有效贡献来决定奖励大小,与统一定价方案相比,该方法具有较好的性能^[3];Tong 等采用了 simple majority voting 方法,即根据多个参与者对同一个任务的应答,采用简单的多数投票方法确定最终的结果^[9]。此外,一种称为 historical performance-based methods 的方法被广泛使用,该方法根据参与者的历史任务完成情况和评价,对其执行结果赋予一定的权值。但是,以上这些方法没有考虑任务需求与参与者能力之间的匹配,而且存在冷启动问题,即刚开始的时候无法对该参与者进行评价^[7,10-11]。另外,Zhang 等提出了一个简单有效的称为 MacroWiz 的框架,并利用 MSN 进行众包活动,当收到一个众包任务后,MacroWiz 可以激励在线 MSN 用户贡献他们的知识或建议,协助任务请求者选择可靠答案并作出最终决策^[12]。Eickhoff 等提出了一个反作弊策略,并且得出 2 条结

论:①过滤恶意的参与者可以减少低质量的数据,但是会延长任务的完成时间;②参与者的可靠性无法由其历史任务完成率进行保证^[13]。Qin 等针对图片的易修改性,提出了一个简单有效的轻量级的图片修改检测方法,用于保证基于手机的照片版权^[14]。Shamir 等分析了人工分类和标注的一致性,提出了一种基于监督机器学习的众包质量控制方法^[15]。

在参与者选择方面,考虑到终端的移动性会导致收集数据的路由发生变化,Zhao 等提出了一个无中心化的路由策略,即利用服务节点之间的机会相遇来转发任务、收集数据^[16-18]。众包任务一般可分为时间敏感任务和可容忍延时的任务,基于此,Guo 等提出了一个适用于多任务的移动用户选择和大范围的众包服务场景的 ActiveCrowd 的框架,并利用 greedy enhanced genetic algorithm 为参与者实现总距离或总开销的最优或近似最优^[19]。张君涛等将子道路和任务绑定在一起,提出了一种基于矢量任务地图的渐进式的任务分发方法,并提高了任务地图的分辨率^[20]。Peng 等提出了一种基于泰森多边形的参与者节点选择方法,将传统的基于标准栅格的参与者区域覆盖变成基于任务特征的不规则区域覆盖,提升了参与者的执行效率,降低了平台或任务发布者的总体成本^[21]。安健等利用城市中公交载体的轨迹可预测、活动覆盖范围大、乘客节点自主聚集且交互时间有保证等优势,提出了一种基于公交系统的众包任务差异化分发方法^[22]。张文东等利用社会关系的多维要素,构造多维语义分级树和空间索引编码,引入交互信息熵等相关概念,提出了一种社会属性亲密度量化方法,对社会关系的不对称性进行分析和比较,从而提高了亲密度量化方法的客观性^[23]。

这些方法通过选择更优质的服务节点,在一定程度上提升了众包的服务质量,但是通常需要所有的参与者都事先在平台上注册,而且他们的信息(兴

趣、爱好、专长等)是固定不变的,任务发布者或者平台能够获得他们的所有信息。但是,实践中参与者的位置属性、专长都有可能随着时间而变化,即使是众包服务平台,也不可能获知那些在平台上注册用户的全部信息,更不可能获知那些未在平台上注册的潜在参与者的全部信息。

基于此,本文提出了基于社会关系的众包服务节点选择算法,任务发布者不需要获得全部参与者的信息,算法即会根据任务发布者与其朋友的社会关系、聊天信息等,对朋友能力进行动态、精准地分类与刻画,将任务与朋友的能力值进行匹配,从而选择最合适的参与者参与众包任务。当朋友接受任务后,以同样的方式在自己的朋友圈里发布该任务。如此循环迭代,能够更高效、准确地完成众包任务。本文算法充分利用了社会关系中的天然信任关系,构建了众包任务参与者之间的信任链条,并且可以由社会关系中的六度分隔理论保证任务传输的时间效率^[24-25]。

1 基于社会网络的能力发现与任务传播

1.1 社会网络

社会网络是由许多节点构成的一种社会关系结构,节点通常是指个人或组织,社会网络代表各种社会关系,这些社会关系可以把从偶然相识的泛泛之交到紧密结合的家庭关系的各种人们或组织串连起来^[24]。社会网络内个体成员之间以任何形式进行的联系称为交互。交互能帮助社会网络中的节点之间形成相对稳定的关系体系,进而影响节点的社会行为。

1.2 基本定义

每个节点都有其自身的属性,如兴趣爱好、专业特长等,并且对他的所有朋友都会有能力区分和认识,如认为某个朋友适合干什么、某个朋友对什么感兴趣等。社会网络中的一个节点用 p 表示, p 的朋友数用 m 表示。众包系统约定的能力类别数用 n 表示。任务参与者既包括任务发布者,又包括任务接受者。

定义1:能力。能力是完成一个项目或者任务所需要的基本技能或素质,用 a 表示。

定义2:能力集合。能力集合是完成一个众包任务需要的所有能力的集合。平台能力集合表示为 $A = \{a_1, a_2, \dots, a_n\}$,共 n 种能力。

定义3:能力值。众包任务参与者的能力值是该参与者对应能力集合中各个能力的数值。由于一个节

点的能力值在其不同朋友中会有不同认知,因此即使是同一种能力,也可能会有不同的能力值。能力值的初始值 C_p 是由节点 p 的朋友根据其各自的认知设定的,没有设定的默认为0。 C_p 具体表示为

$$C_p = \{C_p^{a_1}, C_p^{a_2}, \dots, C_p^{a_n}\} \quad (1)$$

定义4:质量因子。质量因子是指众包参与者接受并完成某个众包任务后,获得任务发布者对自己的综合评价,用 Q 表示。设 s_j 是朋友 p_j 委托 p 完成的所有众包任务数量,任务完成后,朋友 p_j 对 p 所完成的各个任务的情况打分,记为 $y_z, z=1, 2, \dots, s_j, y_z \in [0, 1]$ 。此时,质量因子 Q_p^j 可表示为

$$Q_p^j = \sum_{z=1}^{s_j} w_z^j y_z^j / s_j, \quad j = 1, 2, \dots, m \quad (2)$$

式中 w_z^j 表示第 j 个朋友的第 z 个任务的权重,可在 $[0, 1]$ 之间设定。

定义5:诚信指数。诚信指数是指某个众包参与者的所有朋友对该参与者质量因子的加权平均。 p 的所有朋友对 p 总体评价可用诚信指数 h_p 表示,表达式为

$$h_p = \sum_{j=1}^m W_p^j Q_p^j / m \quad (3)$$

式中 W_p^j 为第 j 个朋友对节点 p 的社会关系的权重,一般情况下,可将其全部设为1。

定义6:朋友能力向量。社会网络中的节点对其每个朋友在能力集合中对应的能力大小,构成该节点的一个朋友能力向量。如节点 p 与朋友 p_j 之间的能力向量 e_{p, p_j} 可表示为

$$e_{p, p_j} = [x_{p, p_j}^1, x_{p, p_j}^2, \dots, x_{p, p_j}^n] \quad (4)$$

定义7:朋友能力矩阵。社会网络中的某一节点的所有朋友能力向量,构成了该节点的朋友能力矩阵。节点 p 的朋友能力矩阵 E_p 可表示为

$$E_p = \begin{bmatrix} e_{p, p_1} \\ e_{p, p_2} \\ \vdots \\ e_{p, p_m} \end{bmatrix} \quad (5)$$

2 社会关系中基于内容的能力发现

2.1 参与者社会关系重构

结合式(1)(3)(5),可以将社会网络中的节点 p 重新定义为一个三元组

$$p = \langle h_p, C_p, E_p \rangle \quad (6)$$

2.2 能力值的计算与更新

在能力值计算上,主要统计节点 p 与其朋友之间就某个关键词的出现次数。如果某个关键词出现

在其朋友圈中,则以更大的权重计算。含加权朋友圈的能力值动态更新算法的伪代码如下。

```
1:  $k=1$ ;  
2: Do while  $k \leq n$   
3:    $C_p^{a_k}=0$ ; // 初始化节点  $p$  的  $a_k$  能力的能力值;  
4:    $C_p^{a_k} = C_p^{a_k} + \sum_{j=1}^m C_{p,p_j}^{a_k} + \sum_{j=1}^n w_{p,p_j}^f C_{p,p_j}^{f,a_k}$   
   // 动态更新节点  $p$  的  $a_k$  能力的能力值  
5:    $k=k+1$ ;  
6: End Do;
```

伪码中: $C_{p,p_j}^{a_k}$ 表示 p 与朋友 p_j 在能力 a_k 上的交互次数,每出现能力 a_k 所包括的关键词 1 次, a_k 的交互次数就加 1; C_{p,p_j}^{f,a_k} 表示 p_j 朋友圈中出现能力 a_k 的次数; w_{p,p_j}^f 为该朋友能力的权重,通常情况下,如果朋友圈出现某个能力,会比在单个聊天记录里出现的重要性更大些,因此 w_{p,p_j}^f 的值需要大于 1,一般可取 2。每经过一个采样周期,节点 p 的每个属性能力值都会进行调整。

2.3 含加权朋友圈能力的能力矩阵计算与更新

2.2 节主要对节点 p 的能力值进行计算和更新,更新后可以生成一个 $1 \times n$ 的行向量。在此基础上,还需要对所有的朋友进行能力计算,最后生成一个朋友能力矩阵,算法伪代码如下。

```
1:  $k=1$ ;  $j=1$ ;  
2: Do while  $j \leq m$   
3:   Do while  $k \leq n$   
4:      $x_{p,p_j}^k = h_{p_j}(F_{p,p_j}^{a_k} + w_{p,p_j}^f C_{p,p_j}^{f,a_k})$   
5:      $k=k+1$ ;  
6:   End Do;  
7:    $j=j+1$ ;  
8: End Do;
```

伪代码中: h_{p_j} 为朋友 p_j 的诚信指数; $F_{p,p_j}^{a_k}$ 为 p 与朋友 p_j 交互过程中提到关键词 a_k 的次数。算法结束后将得到一个 $m \times n$ 的朋友能力矩阵。

通过式(2)和能力值与能力矩阵的计算与更新,节点 p 完成了对自己社会关系的重构,可重写式(5)为

$$\langle h_p, \{C_p^{a_1}, C_p^{a_2}, \dots, C_p^{a_n}\},$$

$$p =$$

$$\begin{bmatrix} x_{p,p_1}^1 & x_{p,p_1}^2 & \cdots & x_{p,p_1}^n \\ x_{p,p_2}^1 & x_{p,p_2}^2 & \cdots & x_{p,p_2}^n \\ \vdots & \vdots & & \vdots \\ x_{p,p_m}^1 & x_{p,p_m}^2 & \cdots & x_{p,p_m}^n \end{bmatrix}$$

$$\rangle$$

(7)

3 仿真实验及结果分析

微信是中国最流行的一款移动即时聊天工具,根据最近腾讯公布的 2019 年第一季度财报,微信月活跃用户数量已正式突破了 11 亿,同比增长达 6.9%,2018 年微信日均发送信息 450 亿条。微信为用户构建了一个庞大的社会关系。本文系统主要以微信交互数据为基础,为了避免隐私泄漏,本文系统只允许在使用者授权同意的情况下,分析自己的聊天记录和内容,对自己的朋友按主题类别进行能力区分,重构自己的社会关系。本文实验模拟微信交互并进行数据分析。

3.1 数据准备

猪八戒网^[26]是中国著名的众包服务网站之一,通过对其网站内容进行分析发现,众包项目可以分成 20 个类目,分别是:知识产权、品牌设计、IT/软件、营销传播、工商财税、科技服务、电商服务等。不失一般性,选择其中的 10 个类目进行数据分析。由第 1 节分析可知,每种类目都代表一种能力或主题。10 种能力的编号分别为 1~10,代表了平台能力集合为 $A=\{1,2,\dots,10\}$ 。对每种能力提炼了部分关键词,如表 1 所示。

表 1 能力分类及部分关键词

能力编号	能力	部分关键词
1	运动	球、篮球、足球、羽毛球、乒乓球、跑步、游泳……
2	企业管理	金融、法律、注册、人力资源……
3	兼职	兼职、地推、打字、购物、分享……
4	品牌创意	Logo、包装、广告、VI、漫画……
5	医疗医学	病、卫生、传染、饮食、健康……
6	软件开发	APP、测试、网站、H5、UI……
7	科学研究	问卷、标注、调查……
8	个人生活	翻译、摄影、取名、画像……
9	社会公益	捐款、民意、拥堵、安全……
10	产品制造	硬件、工业、机械……

3.2 能力发现

设定节点 p 有 100 个朋友,编号分别为 1~100,本节重点介绍该节点的朋友能力矩阵 E_p 的生成过程。

设定一次采样周期为 1 个月,随机从 10 个主题中选择若干个交互主题,交互主题的概率密度呈泊松分布,与每个朋友的交互次数在 5~500 之间。

http://zkxb.xjtu.edu.cn

根据表 1,分别统计了 1 次采样周期与 5 次采样周期下,100 个朋友中前 10 个朋友的能力值(交互次数),分别如表 2 和表 3 所示。

表 2 1 次采样下前 10 个朋友 10 种主题的交互次数

朋友 编号	不同主题编号下的交互次数									
	1	2	3	4	5	6	7	8	9	10
1	171	79	150	62	107	155	121	154	141	47
2	146	93	126	130	90	102	27	101	44	71
3	3	6	7	4	9	1	3	4	7	8
4	20	15	25	28	31	12	20	22	14	11
5	88	29	62	38	80	30	41	59	90	92
6	29	17	40	92	47	78	23	84	55	39
7	32	20	23	43	35	27	37	15	41	40
8	45	7	16	46	62	62	9	37	45	17
9	19	43	28	41	24	38	15	38	51	52
10	21	19	77	29	76	46	70	11	51	26

表 3 5 次采样下前 10 个朋友 10 种主题的交互次数

朋友 编号	不同主题编号下的交互次数									
	1	2	3	4	5	6	7	8	9	10
1	429	241	389	286	389	351	181	322	383	374
2	346	180	390	404	361	238	245	359	382	287
3	268	202	254	264	274	175	111	185	291	271
4	242	191	206	243	241	210	210	106	325	269
5	394	323	424	441	365	320	174	351	544	367
6	194	227	254	364	341	355	147	288	383	242
7	120	118	119	171	113	121	167	93	141	90
8	215	110	290	228	264	175	99	211	206	212
8	398	305	434	400	374	393	144	403	395	400
10	147	118	225	191	234	134	291	163	238	131

表 4 则显示了 5 次采样周期下,前 10 个朋友的朋友圈中出现各类关键词的次数,该值表明了该朋友在不同能力类别的朋友圈能力大小。假设朋友圈能力的权重为 2,综合表 3 和表 4,可得到最终的朋友综合能力矩阵,如表 5 所示。

根据表 2 中的初始能力,随机选取了编号为 5、8、69 的朋友进行研究,图 1 和图 2 分别显示了在 1 次采样下和 5 次采样下,节点 p 与所有朋友在不同主题(能力)下的交互次数(能力值)。图 3 为 1 次采样与 5 次采样(标准化后)下节点 p 与所有朋友的总交互次数的对比图。可以看出,两条线的趋势基本保持一致,说明 1 次采样与 5 次采样的结果相对稳定。图 4 为加入加权朋友圈能力后,节点 p 与朋友的交互次数。从图 1、2、4 可以看出,总体而言,交互

表 4 朋友圈 5 次采样下前 10 个朋友 10 种主题的交互次数

朋友 编号	不同主题编号下的交互次数									
	1	2	3	4	5	6	7	8	9	10
1	55	33	62	26	44	20	50	56	58	0
2	55	38	52	54	37	42	11	42	3	29
3	0	2	3	2	4	0	0	2	3	3
4	8	6	10	12	13	0	8	9	0	0
5	36	12	26	16	33	0	17	24	37	38
6	12	0	17	38	19	32	10	35	23	16
7	13	0	10	18	14	11	15	6	17	17
8	19	0	7	19	26	26	0	15	19	7
9	8	18	12	17	10	16	6	16	21	22
10	0	8	32	12	31	19	29	0	21	11

表 5 计算朋友圈能力值后前 10 个朋友的最最终综合能力矩阵

朋友 编号	不同主题编号下的交互次数									
	1	2	3	4	5	6	7	8	9	10
1	539	307	513	338	477	391	281	434	499	374
2	456	256	494	512	435	322	267	443	388	345
3	268	206	260	268	282	175	111	189	297	277
4	258	203	226	267	267	210	226	124	325	269
5	466	347	476	473	431	320	208	399	618	443
6	218	227	288	440	379	419	167	358	429	274
7	146	118	139	207	141	143	197	105	175	124
8	253	110	304	266	316	227	99	241	244	226
9	414	341	458	434	394	425	156	435	437	444
10	147	134	289	215	296	172	349	163	280	153

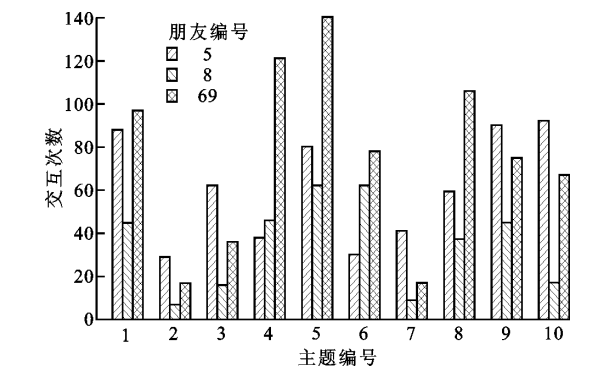


图 1 部分节点 1 次采样下不同主题的交互次数

次数具有相对稳定性,说明节点 p 的能力具有相对稳定性,且与某个朋友在某种能力上的交互次数越多,则该能力值越大,反之亦然。

3.3 候选传播节点选择

当一个任务发布者想发布任务时,需要选择任务的类目或者所需的能力关键词,系统将从他自己

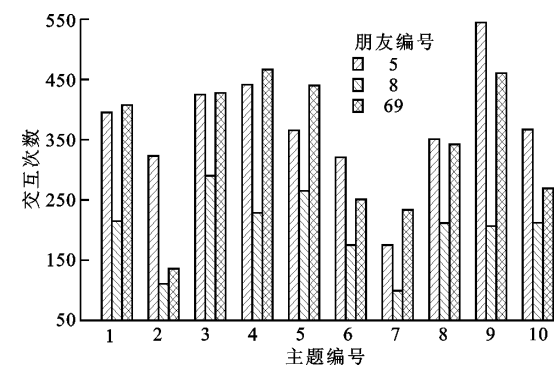


图 2 部分节点 5 次采样下不同主题的交互次数

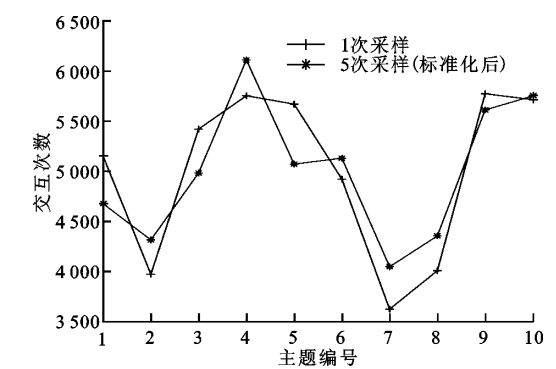


图 3 1 次采样与 5 次采样下不同主题所有朋友的总交互次数

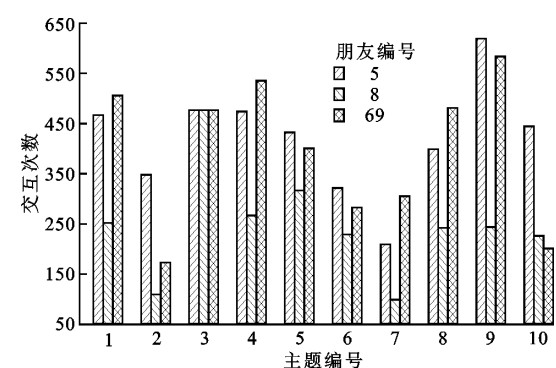


图 4 5 次采样下加入朋友圈权重后不同主题的交互次数

的朋友中选择匹配度最高的若干用户做第 1 次任务分发。表 6 所示为 1 次仿真实验中得到的单能力候选节点排序表,以 5 次采样数据得到的最终综合能力矩阵(表 5)为基础,将候选节点编号按照能力值从大到小进行顺序,只显示能力值最高的前 10 个节点。

根据表 6 中数据,如果节点 p 想要发布能力编号为 2 的众包任务,则应该在其好友列表中选择编号为 57、28、31、50……的朋友,这是因为这些朋友的能力属性最为匹配。如果众包任务需要多种能力,只需要对不同的能力设计相应的权重,即可找到最合适的朋友。实验中,选择了不同的能力要求,并设置了相应的权重,具体选择的结果如表 7 所示。

表 6 1 次实验得到的单能力候选节点排序表
(以 5 次采样数据为基础)

能力编号	候选节点编号									
1	16	92	79	94	1	56	32	69	62	9
2	57	28	31	50	22	88	92	79	16	39
3	34	57	32	56	94	9	50	69	5	28
4	57	52	77	11	69	50	28	94	48	5
5	57	92	60	50	31	77	69	88	35	46
6	31	34	9	16	32	35	57	96	6	39
7	57	72	28	77	92	11	32	94	10	88
8	34	57	9	56	11	39	2	65	77	5
9	5	56	52	57	23	77	50	31	69	28
10	11	50	28	51	16	72	27	96	60	9

表 7 多能力及不同权重下的节点选择结果

能力要求编号	相应权重	前 10 个候选节点编号				
2,9	0.5,0.5	57	5	31	28	50
		56	79	16	92	77
5,10	0.4,0.6	57	50	60	11	28
		92	51	16	31	72
4,6	0.8,0.2	57	77	52	11	94
		16	28	50	69	5
3,7	0.1,0.9	57	28	72	77	92
		11	32	94	10	52
6,9	0.6,0.4	31	57	5	9	16
		34	32	39	96	77
1,8	0.3,0.7	9	56	57	94	34
		16	5	69	11	2

3.4 任务分发的时效性分析

众包系统中任务分发的时间效率是很重要的,特别是对于那些延迟敏感的任务。因此,随机选择了 4 种任务类目:兼职、品牌创意、软件开发、个人生活,即表 1 中编号为 3、4、6、8 的 4 种能力。本文采用了 3 种仿真实验,即随机分发方法、全分发方法、本文方法。假设节点 p 有 2 000 个朋友,并设定任务发布成功的标准是收到 50 个有效的任务转发或执行结果数据。在本文方法和随机分发方法中,都选择 200 个朋友进行,全分发则意味着任务发送给所有朋友。在实验中,还假设当任务能力需求落在朋友能力矩阵 E_p 中各能力主题的前 50 个朋友内时,意味着该朋友将执行任务,任务完成时间是 50 个朋友完成任务的平均时间。不同分发策略下的完

成任务平均时间如图5所示,可以看出:随机分发方法需要花费最长的时间,这是因为其不能精确地发现最合适的参与者;本文方法比随机分发平均节省了约64.5%的时间。进一步地,本文方法和全分发方法的结果基本相似,表明本文方法可以精准地发现合适的工作者,但是本文方法选择的样本数只有全分发方法的1/10,意味着本文方法给无关朋友带来的信息干扰要比全分发策略少得多。此外,大多数任务分配都伴随着一定的激励机制,更少的样本空间意味着能够节约更多的成本。

第3节仿真实验表明:本文系统能顺利地按主题计算各个朋友的能力值大小,并进行采样与更新,同时可以对朋友的诚信指数进行采集与计算更新,形成能力矩阵;对任意一个在能力集合内的众包任务,都能正确选择出最匹配的朋友,据此找到传播候选节点并进行任务分发或数据回传。

任何一个候选节点都可以重复以上步骤,经过若干次迭代,任务能在朋友圈中快速扩展,且朋友圈中的信任关系能被充分利用,在一定程度上增加了众包参与者的可信度,从而提升了众包的数据质量。

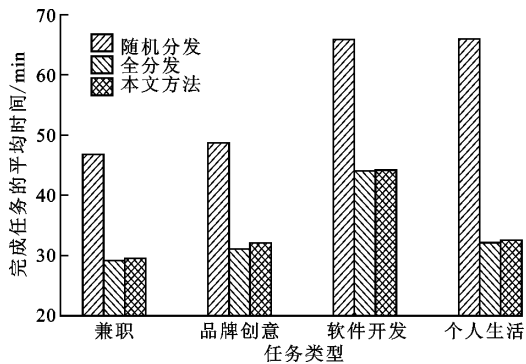


图5 不同分发策略下完成任务的平均时间

4 结束语

众包服务节点的选择应综合考虑其背景知识、社会关系、信誉度等多种属性。本文方法首先需要对参与者进行能力发现,即从参与者日常在社交网络中的交互内容、交互次数、交互频率、朋友圈、加入的群、在群里的活跃度等信息中,对朋友的能力进行提炼,形成能力集合并对能力进行量化分级。通过评估参与者的社会关系和社区归属特性,利用参与者社会关系中隐含的天然信任关系,有针对性地完成任务的分发和后续数据的路由转发。因为人们总是更愿意转发与自己有社会关系的人的信息,所以通过社会关系找到的参与者候选集,比随机选择的参与者具有更高的精准性,仿真实验验证了本文方

法的有效性。未来还需要在以下方面开展研究。

一是客户信息的隐私保护。未来需要开发基于智能终端的轻量级的隐私保护方法,确保感知者的通信录、社交信息、轨迹与位置信息等不被泄漏。

二是交互信息的多样化。本文中仅仅考虑了文本数据。实际上,声音、图片、视频、表情等超文本元素在社交网络中使用非常频繁,这些元素在朋友间的表达感情方面越来越重要。这需要通过AI、语音识别、图像理解等技术,融合社交网络中的全媒体信息,以便更精准地刻画朋友信息。

三是交互的时间维度。本文在分析交互内容时,没有考虑交互发生的时间因素,而时间因素往往具有重要意义。例如同样的交互内容,一个星期内发生的可能比6个月前发生的更准确。如果加入交互内容的时间因素,方法会变得更加复杂,此时可以通过折中方法,给不同的采样阶段设置不同的权重。

参考文献:

- [1] HOWE J. The rise of crowdsourcing [J]. Wired Magazine, 2006, 14(5): 1-4.
- [2] 张志强, 逢居升, 谢晓芹, 等. 众包质量控制策略及评估算法研究 [J]. 计算机学报, 2013, 36(8): 1636-1649.
ZHANG Zhiqiang, PANG Jusheng, XIE Xiaolin, et al. Research on crowdsourcing quality control strategies and evaluation algorithm [J]. Chinese Journal of Computers, 2013, 36(8): 1636-1649.
- [3] PENG D, WU F, CHEN G H. Data quality guided incentive mechanism design for crowdsensing [J]. IEEE Transactions on Mobile Computing, 2018, 17(2): 307-319.
- [4] ZHENG Z Z, WU F, GAO X F, et al. A budget feasible incentive mechanism for weighted coverage maximization in mobile crowdsensing [J]. IEEE Transactions on Mobile Computing, 2017, 16(9): 2392-2407.
- [5] 吴垚, 曾菊儒, 彭辉, 等. 群智感知激励机制研究综述 [J]. 软件学报, 2016, 27(8): 2025-2047.
WU Yao, ZENG Juru, PENG Hui, et al. Survey on incentive mechanisms for crowd sensing [J]. Journal of Software, 2016, 27(8): 2025-2047.
- [6] ASSIS NETO F R, SANTOS C A S. Understanding crowdsourcing projects: a systematic review of tendencies, workflow, and quality management [J]. Information Processing & Management, 2018, 54(4): 490-506.
- [7] WANG J T, WANG Y S, ZHANG D Q, et al. Multi-

- task allocation in mobile crowd sensing with individual task quality assurance [J]. *IEEE Transactions on Mobile Computing*, 2018, 17(9): 2101-2113.
- [8] GAD-ELRAB A A A, ALSHARKAWY A S. Statistical-based data quality model for mobile crowd sensing systems [J]. *Arabian Journal for Science and Engineering*, 2018, 43(12): 8195-8207.
- [9] TONG Y X, CHEN L, ZHOU Z M, et al. SLADE: a smart large-scale task decomposer in crowdsourcing [C]//2019 IEEE 35th International Conference on Data Engineering (ICDE). Piscataway, NJ, USA: IEEE, 2019: 2133-2134.
- [10] MOAYEDIKIA A, YEOH W, ONG K L, et al. Framework and literature analysis for crowdsourcing's answer aggregation [J]. *Journal of Computer Information Systems*, 2017(7): 1-12.
- [11] YUE D J, YU G, SHEN D R, et al. A weighted aggregation rule in crowdsourcing systems for high result accuracy [C]//2014 IEEE 12th International Conference on Dependable, Autonomic and Secure Computing. Piscataway, NJ, USA: IEEE, 2014: 265-270.
- [12] ZHANG X L, SHANGGUAN L F, YUAN Y. A crowd wisdom management framework for crowdsourcing systems [J]. *IEEE Access*, 2016, 4: 9764-9774.
- [13] EICKHOFF C, DE VRIES A. How crowdsourcable is your task [C]//Proceedings of the Workshop on Crowdsourcing for Search and Data Mining (CSDM) at the Fourth ACM International Conference on Web Search and Data Mining (WSDM). New York, NY, USA: ACM, 2011: 11-14.
- [14] QIN T, MA H D, ZHAO D, et al. Crowdsourcing based event reporting system using smart phones with accurate localization and photo tamper detection [M]//Big Data Computing and Communications. Berlin, Germany: Springer International Publishing, 2015: 141-151.
- [15] SHAMIR L, DIAMOND D, WALLIN J. Leveraging pattern recognition consistency estimation for crowdsourcing data analysis [J]. *IEEE Transactions on Human-Machine Systems*, 2016, 46(3): 474-480.
- [16] ZHAO D, MA H D, TANG S J, et al. COUPON: a cooperative framework for building sensing maps in mobile opportunistic networks [J]. *IEEE Transactions on Parallel and Distributed Systems*, 2015, 26(2): 392-402.
- [17] PU L J, CHEN X, XU J D, et al. Crowd foraging: a QoS-oriented self-organized mobile crowdsourcing framework over opportunistic networks [J]. *IEEE Journal on Selected Areas in Communications*, 2017, 35(4): 848-862.
- [18] ZHAN Y F, XIA Y Q, LIU Y, et al. Incentive-aware time-sensitive data collection in mobile opportunistic crowdsensing [J]. *IEEE Transactions on Vehicular Technology*, 2017, 66(9): 7849-7861.
- [19] GUO B, LIU Y, WU W L, et al. ActiveCrowd: a framework for optimized multitask allocation in mobile crowdsensing systems [J]. *IEEE Transactions on Human-Machine Systems*, 2017, 47(3): 392-403.
- [20] 张君涛, 赵智慧, 周四望. 矢量任务地图: 群智感知任务渐进式分发方法 [J]. *计算机学报*, 2017, 40(8): 1946-1960.
- ZHANG Juntao, ZHAO Zhihui, ZHOU Siwang. Vector task map: progressive task allocation in crowdsensing [J]. *Chinese Journal of Computers*, 2017, 40(8): 1946-1960.
- [21] PENG Z L, GUI X L, AN J, et al. A node selection paradigm for crowdsourcing service based on region feature in crowd sensing [J]. *Mathematical Problems in Engineering*, 2018, 2018: 6434083.
- [22] 安健, 彭振龙, 桂小林, 等. 群智感知中基于公交系统的任务分发机制研究 [J]. *计算机学报*, 2019, 42(2): 295-308.
- AN Jian, PENG ZhenLong, GUI XiaoLin, et al. Research on task distribution mechanism based on public transit system in crowd sensing [J]. *Chinese Journal of Computers*, 2019, 42(2): 295-308.
- [23] 张文东, 桂小林, 蔡宁超, 等. 群智感知中采用节点社会属性的亲密度量化方法 [J]. *西安交通大学学报*, 2017, 51(6): 73-78.
- ZHANG Wendong, GUI Xiaolin, CAI Ningchao, et al. Intimacy relationship quantification method among nodes based on social attributes in crowd sensing [J]. *Journal of Xi'an Jiaotong University*, 2017, 51(6): 73-78.
- [24] MARTÍNEZ-NÚÑEZ M, PÉREZ-AGUIAR W S. Efficiency analysis of information technology and online social networks management: an integrated DEA-model assessment [J]. *Information & Management*, 2014, 51(6): 712-725.
- [25] MENG Y, JIANG C X, QUEK T Q S, et al. Social learning based inference for crowdsensing in mobile social networks [J]. *IEEE Transactions on Mobile Computing*, 2018, 17(8): 1966-1979.
- [26] 猪八戒. 猪八戒中国著名众包平台 [EB/OL]. (2013-10-06)[2019-01-03]. <http://www.zbj.com>.

(编辑 陶晴)