



引用格式:孙端. 国内众包模式研究的空间分布、热点与趋势——基于中国知网数据库2006—2019年数据分析[J]. 郑州轻工业学院学报(社会科学版),2020,21(1):58-66.

中图分类号:F210 文献标识码:A

DOI:10.12186/2020.01.008

文章编号:1009-3729(2020)01-0058-09

国内众包模式研究的空间分布、热点与趋势

——基于中国知网数据库2006—2019年数据分析

Spatial distribution, hotspots and trends of crowdsourcing study in China

—Analyse based on CNKI database from 2006 to 2019

孙端^{1,2}

SUN Duan

- 1. 长沙师范学院 经济管理学院,湖南 长沙 410100;
- 2. 湘潭大学 公共管理学院,湖南 湘潭 411105

关键词:
众包模式;
科学知识;
图谱;
可视化分析;
共现网络图谱;
聚类知识图谱

摘要:为全面揭示和分析国内众包模式研究的空间分布、研究热点和研究趋势,选择中国知网数据库收录的669篇期刊论文作为样本,采用CiteSpace,Excel等工具对国内众包研究的时间分布、空间分布、关键词聚类、名词术语词频进行分析,发现:无论在时间分布上还是在空间分布上,众包模式研究在我国都得到迅速发展,研究热点主要集中在高校图书馆、众包模式、隐私保护和公众参与等领域,未来众包任务分配研究领域将会得到学者们更多的关注和持续深入的研究。

[收稿日期]2019-10-17

[基金项目]湖南省社会科学成果评审委员会基金项目(XSP18YBC220);湖南省哲学社会科学基金项目(17YBA025);长沙师范学院校级重点建设学科工商管理学阶段性成果

[作者简介]孙端(1982—),男,河南省周口市人,长沙师范学院讲师,湘潭大学博士研究生,主要研究方向:公共管理、商务管理。

“众包”始于西方企业的创新活动,2006年才进入学术研究领域。J. Howe^[1]首次提出“众包”概念,认为众包是一个公司或机构将过去由员工执行的工作任务以自由自愿的形式外包给非特定的大众网络的做法。众包2007年被引入中国后,其相关研究和实践得到了快速发展。刘文华等^[2]认为众包借助网络平台,能够充分挖掘广大消费者的力量和潜力来参与企业的研究开发和创新活动,这将大大降低企业产品上市的风险。李克强总理在2016年政府工作报告中提出“打造众创、众包、众扶、众筹平台,构建大中小企业、高校、科研机构、创客多方协同的新型创业创新机制”。《湖南国民经济和社会发展“十三五”规划纲要》也明确提出,“推广研发创意、制造运维、知识内容和生活服务众包,推动大众参与线上生产流通分工”。可见,众包作为一种创新模式,已经改变了传统的企业生产模式,并不断得到政府、企业和学者的重视。自2007年至今,众包被引入中国已经12年了,这些年来国内众包模式取得了哪些有代表性的成果?有哪些研究热点?未来发展趋势如何?这些问题是许多对众包感兴趣的相关人士希望深入了解的。鉴于此,本文拟采用CiteSpace, Excel等科研工具,对2007年以来的研究文献进行梳理和总结,选取有代表性的文献样本进行计量分析,概括总结我国众包模式研究现状,以期为深化众包模式研究提供参考。

一、研究设计

本研究采用CiteSpace, Excel等科研计量工具对文献样本数据进行可视化分析处理,重点围绕国内众包模式研究的时空分布、热点与趋势进行分析和总结,揭示我国众包模式发展的内在规律。找到有代表性的权威文献数据,这是实现研究目标的基本条件。为此,我们选取我国最大的实时动态数据库——中国知网

(CNKI)数据库为文献数据采集来源库,依据CNKI数据库的期刊论文检索规则,选取具有代表性的期刊论文作为样本数据。

在中国知网输入关键词“众包”,选择“2006—2019”年与主题词“中文核心期刊+CSSCI+CSCD”,获得有效文献样本669篇,该样本数据的采集时间为2019年9月13日。

本文的研究路径为:第一步,登录中国知网数据库,按照目标设计要求检索文献资料,并对文献数据进行清洗和整理,取得有效数据样本;第二步,打开CiteSpace计量分析软件,对数据进行转换,使之能够适用于软件分析,并建立相应的文件夹,储存不同的数据,原始样本数据储存于input文件夹,转换后的文件自动保存于output文件夹;第三步,建立data和project文件夹,将转化后的样本数据拷贝到data文件夹;第四步,在CiteSpace首页建立工程项目,取名为“众包模式”,根据分析需要对页面栏目进行设置,然后进行可视化处理,依据显示需要对控制面板栏目进行相应设置,从而使得可视化图片更加符合目标分析需要。

二、研究过程与结果分析

1. 国内众包模式研究的时空分布及其规律

国内众包模式研究的时空分布包括时间分布和空间分布。时间分布用样本数据中各个发文年度的论文篇数来表示,从其时间分布可以看出国内众包模式研究的发展趋势;空间分布用样本数据各个发文机构总体发文数量来表示,从其空间分布可以发现国内相关发文机构在众包模式研究方面的学术影响力。

(1) 国内众包模式研究的时间分布及其规律

国内众包模式研究的时间分布体现在我国学者在中文核心期刊、CSSCI和CSCD 3个作为数据库来源的学术期刊上发表众包模式研究的论文数量,这些数据库来源期刊是我国人文社

科期刊中的代表性期刊,有一定权威性,具有较强的影响力。根据样本数据,采用 Excel 表格制作的“众包”理论研究的时间分布见图 1,其显示了自“众包”进入学术研究领域以来众包模式研究文献的时间分布状况。由图 1 可知,2006 年被视为众包模式研究元年,国内没有相关研究文献;2007 年 2 篇,2008 年 1 篇,2009 年 2 篇,2010 年 10 篇,2011 年 8 篇,这 5 年众包研究热度不足,发文数量较少;自 2012 年 19 篇开始,国内众包模式研究论文数量开始飙升,2016 年达到峰值 131 篇,2017 年、2018 年略有下降,但是整体保持 120 篇/年以上;截至统计数据当日,2019 年已经发文 69 篇。总体来看,国内众包模式研究稍晚于西方国家,随着 2007—2011 年少量论文的发表,越来越多的学者开始关注众包研究,并且发文数量飞速上升,整体趋势不断增长。

(2) 国内众包模式研究的时空分布及其规律

国内众包模式研究的时空分布主要体现在国内中文核心期刊、CSSCI 和 CSCD 期刊论文中相关学术机构发表论文的数量分布上,反映了学术机构在众包研究方面的学术影响力。

利用 CiteSpace 软件制作的国内众包模式研究文献的时空分布状况见图 2,该图显示了在众包模式研究方面具有重要影响力的学术机构名称。圆形节点代表学术机构;节点大小代表发文数量的多少,节点越大意味着发文数量

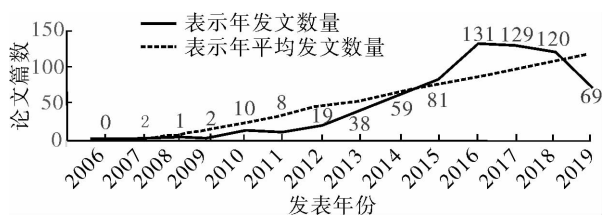


图 1 CNKI 数据库“众包”理论研究的时间分布(2006—2019 年)

越多;颜色代表年份,颜色越深代表年份相对久远。由图 2 可知,发文数量最多的机构是中国科学院大学,紧随其后的是南京理工大学、南京大学、天津大学、上海交通大学、南开大学、中国人民大学、东华大学、湖北大学(见表 1),这些学术机构在众包模式研究方面具有较强的影响力,开展的众包研究领域主要集中在经济管理、工商管理、新闻与传播、计算机与信息工程等学科领域。由表 1 可知,排名靠前的 30 个学术机构发文总计 198 篇,约占总体发文数量的 29.60%,平均发文数量 6.6 篇,这些学术机构多为国内重点院校,科研实力较强;也有一些地方院校,如湖北大学、武汉纺织大学、上海大学、南昌大学等,在众包研究领域也取得了不错的成绩。

2. 国内众包模式的研究热点及其规律

关键词是学术研究与学术评价中具有代表性意义的重要词汇,集中体现了学术研究的重点。在 CiteSpace 中进行可视化,得到“众包”主题关键词共现网络图谱(见图 3)。图 3 中,1 个节点代表 1 个关键词,共有节点 141 个,连线 243 条,网络密度 0.0246。在 CiteSpace 中,节点代表关键词的词频,节点越大,词频越高,说明该词越热;紫色圆环代表关键词的中介中心性,年轮越宽,意味着该词越具有影响力。由图 3 可知,“众包”“众包模式”“大数据”“科研众包”“众包竞赛”“开放式创新”“众筹”“质量控制”“众包平台”等词汇属于热点关键词,具有较强的影响力。

在图 3 的基础上,对样本数据进行聚类,适当调整控制面板参数,可以得到“众包”理论研究的聚类知识图谱(见图 4)。由图 4 可知,众包模式研究包含高校图书馆、众包模式、隐私保护、公众参与、众包竞赛、众包新闻、众包社区、软件众包、激励机制、扎根理论和数据新闻 11 个研究领域。其中,排名靠前、最具有影响力的

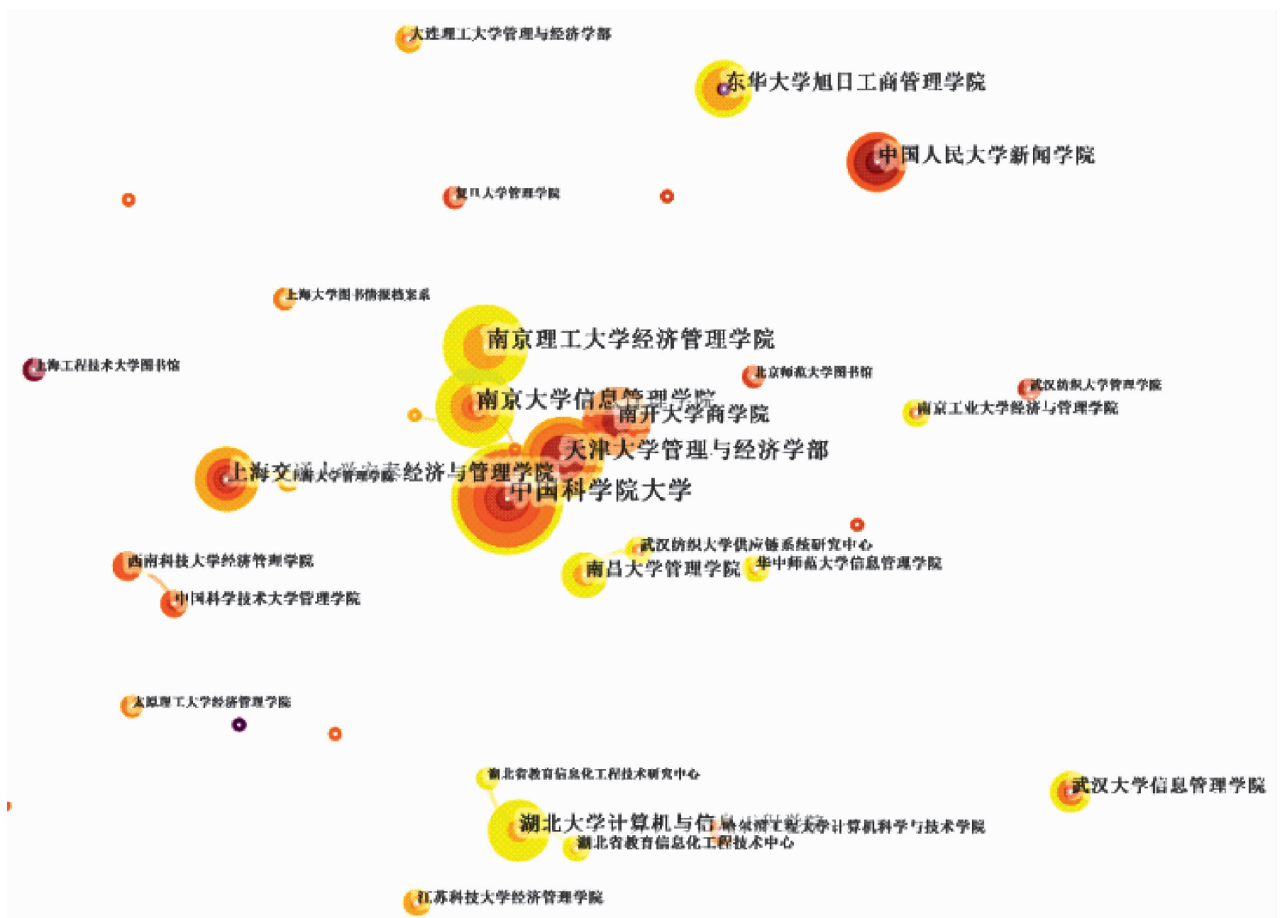


图 2 CNKI 数据库“众包”理论研究的空间分布(2006—2019 年)

表 1 CNKI 数据库“众包”研究发文数量前 30 位学术机构列表

| 排名 | 发文篇数 | 机构名称 | 排名 | 发文篇数 | 机构名称 |
|----|------|----------------|----|------|----------|
| 1 | 19 | 中国科学院大学 | 12 | 6 | 大连理工大学 |
| 2 | 13 | 南京理工大学 | 17 | 5 | 复旦大学 |
| 2 | 13 | 南京大学 | 17 | 5 | 北京师范大学 |
| 4 | 11 | 天津大学 | 19 | 4 | 重庆大学 |
| 5 | 9 | 上海交通大学 | 19 | 4 | 中国科学技术大学 |
| 5 | 9 | 南开大学 | 19 | 4 | 西南科技大学 |
| 7 | 8 | 中国人民大学 | 19 | 4 | 南京邮电大学 |
| 7 | 8 | 东华大学 | 19 | 4 | 南京工业大学 |
| 7 | 8 | 湖北大学 | 19 | 4 | 江苏科技大学 |
| 10 | 7 | 武汉纺织大学 | 19 | 4 | 华中师范大学 |
| 10 | 7 | 湖北省教育信息化工程技术中心 | 19 | 4 | 哈尔滨工程大学 |
| 12 | 6 | 西北工业大学 | 19 | 4 | 北京理工大学 |
| 12 | 6 | 武汉大学 | 19 | 4 | 北京航空航天大学 |
| 12 | 6 | 上海大学 | 29 | 3 | 太原理工大学 |
| 12 | 6 | 南昌大学 | 29 | 3 | 上海工程技术大学 |

资料来源:根据 CNKI 数据库样本数据统计(2006—2019 年)

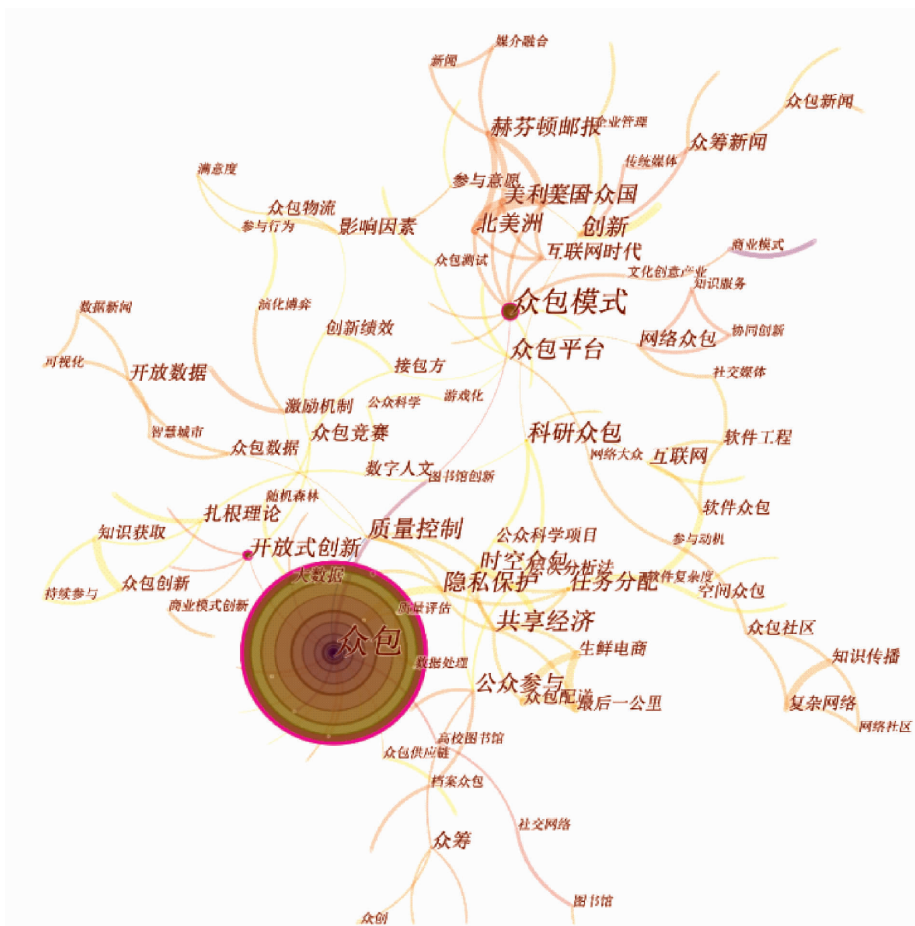


图3 CNKI数据库“众包”关键词共现网络图谱(2006—2019年)

4个类别分别是高校图书馆、众包模式、隐私保护、公众参与,其相似值分别为0.94,1,0.946,1。下面仅就最具影响力的4个研究领域的热点及其规律进行梳理,可得出如下结论。

其一,关于高校图书馆众包研究,相关的关键词有“众包”“图书馆”“高校图书馆”“图书馆创新”“数字图书馆”等(见表2)。研究的热点问题主要集中于众包在高校图书馆建设中的应用,如盛芳等^[3]分析了高校图书馆实施众包的可行性,论述了高校图书馆在图书馆宣传、资源采购、资源揭示和参考咨询服务4个方面引入众包的具体策略;胡永生^[4]分析了图书馆阅读推广服务众包的必要性和可行性,提出了基于众包的阅读推广服务策略;乔亚铭等^[5]指出,众包在图书馆竞争情报领域的应用是深化图书

馆为学科建设和科学研究提供情报服务的有效途径。可见,高校图书馆众包研究重在分析论证引用众包的可行性与运用策略,其研究的广度和深度还有待于进一步拓展和深化。

其二,关于众包模式研究,相关的关键词有“众包模式”“创新模式”“商业模式”“模式创新”等(见表2)。研究的热点问题主要集中于众包模式的概念、特征、意义、创新方式和实现路径等。魏拴成等^[6]总结了众包的概念,梳理了众包与外包的关系,分析了众包的产生、发展、特征,以及对传统企业模式的影响;张利斌等^[7]在前人研究成果的基础上,从众包起因、大众参与动机、众包意义等方面进行了综述;叶伟巍等^[8]从网络众包案例分析入手,分析了合作型和竞争型两类网络众包创新模式的主要特

众包理念催生了自下而上的海量公共服务个性化需求与新型的公众参与,给规划管理带来了新的理念、机遇与挑战;郭捷等^[14]提出了众包物流大众参与行为影响因素研究模型,并通过实证研究得出其影响因素为便利条件、参与意愿、获益期望、社会影响和感知风险;韩普等^[15]

在前人研究基础上,采用层次分析法提炼了公众参与智慧城市管理众包的关键影响因素。可见,众包公众参与研究主要集中在公众参与众包的影响因素上,学者们结合不同的众包领域对其进行了分析和验证。

3. 国内众包模式研究的前沿趋势及其规律

CiteSpace 嵌入了突发性探测算法功能,可以用来探测一篇文献是否切中该领域中的要害问题。因此,我们可以利用该功能对“众包”样本数据进行探测。利用 CiteSpace 导出的“众包”样本数据库,可得出最强的7个突发关键词(见图5)。图5清晰地展现了每个突发关键词的突发强度和起始年份,红色线代表突发关键词持续时间。“众包”关键词的突发强度为16.0396,持续时间为2006—2013年,该词突发强度最高,持续时间最长,说明“众包”一词在该研究领域最具有影响力;其次是“质量控制”,突发强度为6.8247,持续时间为2006—2012年;“图书馆创新”“大数据”“高校图书馆”“众筹”4个关键词的突发强度分别为2.9908,3.8521,3.638,3.9481,持续时间分别为2011—2013年、2013—2014年、2013—2014年、2014—2016年。上述关键词在其持续年份内为该领域的研究前沿,代表着研究方向。“任务分配”突发于2016年,其突发强度为3.1889,并一直持续到2019年,说明该词可能成为众包研究领域的热点词汇,并成为该领域的前沿学术词汇。

由上述分析可知,“任务分配”将成为众包研究新趋势,与其相关的关键词主要有“质量

控制”“共享经济”“时空众包”“空间众包”“隐私保护”等(见图6)。这些关键词从不同侧面反映了众包任务分配问题。目前,国内学者主

表2 CNKI 数据库“众包”频次排名前50位的
热点关键词列表

| 序号 | 频次 | 中介中心性 | 关键词 | 序号 | 频次 | 中介中心性 | 关键词 |
|----|-----|-------|-------|----|----|-------|-------|
| 1 | 318 | 0.9 | 众包 | 26 | 6 | 0.03 | 众包新闻 |
| 2 | 37 | 0.31 | 众包模式 | 27 | 6 | 0.01 | 隐私保护 |
| 3 | 33 | 0.22 | 质量控制 | 28 | 6 | 0 | 新产品开发 |
| 4 | 28 | 0.16 | 大数据 | 29 | 6 | 0.07 | 网络众包 |
| 5 | 20 | 0.13 | 开放式创新 | 30 | 6 | 0 | 社交媒体 |
| 6 | 17 | 0.08 | 科研众包 | 31 | 6 | 0.03 | 商业模式 |
| 7 | 16 | 0.01 | 众包竞赛 | 32 | 6 | 0.03 | 空间众包 |
| 8 | 14 | 0.15 | 众筹 | 33 | 6 | 0.07 | 共享经济 |
| 9 | 13 | 0.16 | 众包平台 | 34 | 6 | 0.02 | 公众科学 |
| 10 | 12 | 0.06 | 创新 | 35 | 6 | 0.07 | 公众参与 |
| 11 | 10 | 0 | 众包翻译 | 36 | 5 | 0.04 | 众筹新闻 |
| 12 | 10 | 0.02 | 互联网+ | 37 | 5 | 0.02 | 众包测试 |
| 13 | 9 | 0.02 | 图书馆 | 38 | 5 | 0.01 | 图书馆创新 |
| 14 | 9 | 0.02 | 软件众包 | 39 | 5 | 0 | 数字图书馆 |
| 15 | 9 | 0.04 | 任务分配 | 40 | 4 | 0.02 | 众创 |
| 16 | 9 | 0.03 | 激励机制 | 41 | 4 | 0.07 | 众包社区 |
| 17 | 9 | 0.01 | 高校图书馆 | 42 | 4 | 0 | 众包创新 |
| 18 | 9 | 0.03 | 参与意愿 | 43 | 4 | 0.01 | 扎根理论 |
| 19 | 8 | 0.02 | 众包物流 | 44 | 4 | 0.01 | 演化博弈 |
| 20 | 8 | 0.02 | 众包供应链 | 45 | 4 | 0.03 | 新闻 |
| 21 | 8 | 0.02 | 时空众包 | 46 | 4 | 0.04 | 数字人文 |
| 22 | 8 | 0 | 社会网络 | 47 | 4 | 0 | 数字出版 |
| 23 | 7 | 0.06 | 影响因素 | 48 | 4 | 0 | 企业 |
| 24 | 7 | 0 | 任务推荐 | 49 | 4 | 0.09 | 互联网 |
| 25 | 7 | 0.01 | 机器学习 | 50 | 4 | 0 | 分享经济 |

资料来源:根据 CNKI 数据库样本数据统计(2006—2019年)

Top 7 Keywords with the Strongest Citation Bursts

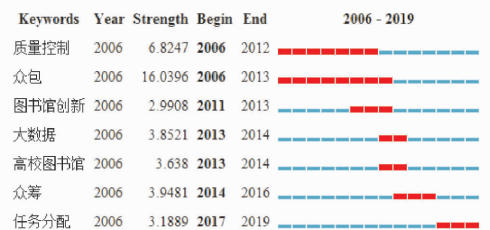


图5 CNKI 数据库“众包”最强突发
关键词图示(2006—2019年)



图6 CNKI数据库“任务分配”关键词节点网络图(2006—2019年)

要从以下几个方面来探讨众包的任务分配问题。

一是用户任务分配研究。例如,张晓航等^[16]在分析用户主题感知机制基础上,通过对用户主题和准确率建模,提出了一种基于用户主题感知的迭代式任务分配算法;施战等^[17]根据每个用户针对不同类型任务具有不同的可靠性,设计了一种基于用户可靠性的众包系统任务分配机制。从目前的研究成果来看,用户任务分配研究内容不多、层次深度不够,特别是针对网络用户任务分配的相关内容还未出现,这将成为未来的研究趋势之一。

二是在线任务分配研究。例如,宋天舒等^[18]基于空间众包环境,提出了一种基于众包任务、众包工人和众包工作地点的新型动态在线任务分配算法,并通过数据集进行大量实验,验证了该算法的效果与性能;余敦辉等^[19]利用CT-KNN算法选出合适的众包工人,并动态选择不同的阈值进行分配,从而提出一种基于动态效用的在线任务分配方法。在线任务分配研究内容也不多,主要是众包任务分配的线上延伸,主要集中在任务分配的算法上,而对网络用户主观性因素的研究不多,未来该领域将成为研究趋势之一。

三是最优任务分配研究。例如,李洋等^[20]利用树分解技术将工人分割成独立的集合,并提出一种启发式的深度优先搜索算法,实验证

明,该算法可以很好地解决最优任务分配问题;赵焜松等^[21]提出一种基于任务发布者权重与任务权重的软件众包任务发布优先级(TRP)计算方法,经验证该方法具有较高的有效性和合理性。目前任务分配研究主要集中在算法上,而对任务分配的要素尚未形成体系,因而在算法上偏向某个特定领域,研究成果的适用性和推广性有待考验,未来在任务体系构建上会产生更多算法,其兼容性、应用性和适用性会得到优化。

可见,国内众包模式经历了引进、消化、吸收和创新的过程,在短短的十多年中已经取得了丰富的研究成果。随着网络经济的发展和社会主义市场经济的不断成熟,国内众包有望成为国内创新领域的一项重要内容,其发展趋势引人注目。

参考文献:

- [1] HOWE J. The rise of crowdsourcing[J]. Wired Magazine, 2006(6):1.
- [2] 刘文华,阮值华. 众包:让消费者参与创新[J]. 企业管理,2009(7):93.
- [3] 盛芳,耿艾莉. 网络环境下高校图书馆四项工作的众包策略[J]. 图书馆论坛,2012(1):15.
- [4] 胡永生. 引入众包模式的阅读推广服务路径、策略及思考[J]. 新世纪图书馆,2016(9):52.
- [5] 乔亚铭,冯永财. 高校图书馆科研竞争情报服务众包探究[J]. 图书馆工作与研究,2018(12):50.
- [6] 魏拴成,邬适融. 众包的产生、发展以及构建众包商业模式应遵循的路径[J]. 上海管理科学,2010(1):55.
- [7] 张利斌,钟复平,涂慧. 众包问题研究综述[J]. 科技进步与对策,2012(6):154.
- [8] 叶伟巍,朱凌. 面向创新的网络众包模式特征及实现路径研究[J]. 科学学研究,2012(1):145.

- [9] 燕彩蓉,张洋舜,徐光伟.支持隐私保护的众包实体解析[J].计算机科学与探索,2014(7):802.
- [10] 童咏昕,袁野,成雨蓉,等.时空众包数据管理技术研究综述[J].软件学报,2017(1):35.
- [11] 张贤贤,王浩宇,郭耀,等.基于众包和机器学习的移动应用隐私评级研究[J].计算机科学与探索,2018(8):1238.
- [12] 孟韬,张媛,董大海.基于威客模式的众包参与行为影响因素研究[J].中国软科学,2014(12):112.
- [13] 周素红.规划管理必须应对众包、众筹、众创的共享理念[J].城市规划,2015(12):96.
- [14] 郭捷,王嘉伟.基于UTAUT视角的众包物流大众参与行为影响因素研究[J].运筹与管理,2017(11):1.
- [15] 韩普.社会化媒体环境下公众参与智慧城市管理的众包概念模型研究[J].现代情报,2018(7):19.
- [16] 张晓航,李国良,冯建华.大数据群体计算中用户主题感知的任务分配[J].计算机研究与发展,2015(2):309.
- [17] 施战,辛煜,孙玉娥,等.基于用户可靠性的众包系统任务分配机制[J].计算机应用,2017(9):2449.
- [18] 宋天舒,童咏昕,王立斌,等.空间众包环境下的3类对象在线任务分配[J].软件学报,2017(3):611.
- [19] 余敦辉,张灵莉,付聪.基于动态效用的时空众包在线任务分配[J].电子与信息学报,2018(7):1699.
- [20] 李洋,贾梦迪,杨文彦,等.基于树分解的空间众包最优任务分配算法[J].软件学报,2018(3):824.
- [21] 赵焜松,余敦辉,张万山.软件众包任务发布优先级计算方法[J].计算机应用,2018(7):2032.