

众包模式在大规模遥感影像信息提取领域的探索

赵江华^{1,2}, 王学志¹, 林青慧¹, 黎建辉¹, 周园春¹

1. 中国科学院计算机网络信息中心, 北京 100190; 2. 中国科学院大学, 北京 100049

摘要

基于地理空间数据云平台,对基于众包的大规模遥感影像信息提取模式进行了研究,提出了一套完整的流程体系,并通过多时期的青藏高原湖泊提取任务对模式的报酬发放机制、任务分配方式、任务划分方法、人才激励等领域进行了探索和完善。实验结果表明提前支付部分报酬并采用小组的方式,对提高数据质量和控制并没有很大影响,而积累人才对获取高质量的数据结果很重要。由于该模式集成了众包和机器计算能力,且对遥感影像处理是通用的,因此可用于更多的需要人工参与的海量遥感影像处理工作中。

关键词

遥感影像;信息提取;众包;地理空间数据云

中图分类号:TP79

文献标识码:A

doi: 10.11959/j.issn.2096-0271.2016066

Exploration of crowdsourcing in information extraction from remote sensing images

ZHAO Jianghua^{1,2}, WANG Xuezhi¹, LIN Qinghui¹, LI Jianhui¹, ZHOU Yuanchun¹

1. Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China

2. University of Chinese Academy of Sciences, Beijing 100049, China

Abstract

Based on geospatial data cloud(GSCloud), the application of crowdsourcing in large scale information extraction from satellite images was studied, and a systematic architecture of this paradigm was proposed. By performing an experiment of extracting lakes on Qinghai-Tibetan plateau from landsat images, various aspects of the paradigm like the incentive mechanism, task assignment method, task division and many others were explored. Results show that paying part of the reward in advance and assigning a task to a team instead of individuals do not help attracting more applicants and improving the quality of results. And the accumulation of talents is of critical importance to obtain high-quality task results. Since this paradigm integrates crowdsourcing and machine computing power, and it is generic, it can be applied in more massive remote sensing image processing work which requires much human intervention.

Key words

remote sensing image, information extraction, crowdsourcing, GSCloud

1 引言

自1957年前苏联发射世界上第一颗人造地球卫星以来,人类利用航天器探索宇宙的研究已有近60年的历史^[1]。随着遥感技术的发展和新一代遥感平台的出现,遥感影像在空间分辨率、时间分辨率、光谱分辨率和辐射分辨率上都有了很大程度的提高。由于遥感具有覆盖面广、及时快速的优势^[2],因此通过遥感影像提取地物信息得到广泛应用^[3-6]。

遥感影像信息提取一般包括分类、识别和特征提取^[7],最早通过目视解译实现,即通过人脑对遥感影像的综合理解和分析,利用判读人员自身的经验,同时从多种非遥感信息资料中获取相关的地学知识,对图像的影像特征和空间特征做出分析,得到判读结果^[8]。由于目视解译能够综合利用地物的色调、形状、大小、阴影、纹理、图案、位置和布局等影像特征知识以及有关地物的专家知识,并结合其他知识进行综合分析和逻辑推理,因而达到较高的信息提取精度。现阶段,目视解译仍然被广泛地应用于精度要求较高的研究和实践中,然而,该方法解译速度慢,需要投入大量的人力和时间。在当今的信息化社会,面对堆积如山的海量高分辨率遥感数据,大范围、高速度的实时解译工作几乎超越了目视解译的极限,因此依靠人工目视解译的方法已远远不能满足数据生产和更新的需要^[9]。

利用计算机进行遥感影像信息提取是以计算机系统为支撑环境,将模式识别技术与人工智能相结合,对遥感图像中地物的影像特征与专家知识库的知识进行比对、分析、推理,从而实现专题信息的提取^[10,11],目前主要包括基于像元的遥感影像信息提

取和面向对象的遥感影像信息提取方法。基于像元的遥感影像信息提取方法主要根据光谱信息,以像元为基础单元进行信息提取,而面向对象的遥感影像信息提取方法首先将遥感影像中空间相邻的像元分割成一个个同质性的对象,然后将这些对象作为最小的分类单元对影像进行分类,完成地物信息的提取^[12,13]。不仅利用了遥感影像的光谱信息,还综合考虑影像对象的纹理、几何和上下文信息,因此得到更广泛的应用。

然而,虽然自动进行遥感影像信息提取的方法和技术一直在提高,但由于遥感影像场景复杂,自动提取方法目前只能用于几何形状和边界有规则的小面积区域,并未在有复杂情况的大面积区域应用。因此,遥感影像信息提取在很大程度上还不能实现完全自动化,多数情况下都需要人工参与^[14,15]。随着遥感数据集的不断增长,各类研究的时间、空间跨度要求越来越大,实现快速准确的大面积长时间序列的遥感影像信息提取,成为迫切要解决的问题之一。

众包作为一种分布式的问题解决和生产模式,通过高效调用分散的人力资源实现海量数据的快速精准分析^[16],为解决人类比较擅长,而计算机难以自动计算的复杂问题的绝佳方案,目前已经在多种领域得到应用,如图像标记、自然语言处理领域等^[17-22]。同时众包平台也得到了很大的发展,其中用户规模最大、知名度最高的众包平台就是亚马逊的土耳其机器人(Mechanical Turk, MTurk)。MTurk充分利用了公众的碎片时间,发布的任务一般比较简单,如图片识别等^[23],这些任务相互独立,通过传统的方法即可完成,需要较少的时间、知识和技能,但报酬通常也非常低。

不同于MTurk上发布的典型任务,遥

感影像信息提取工作更加复杂,往往需要大量的时间和知识精力来完成。Fritz、Dstes等人利用公众协作开发了Geo-Wiki^①和DIYlandcover平台,均通过在网上传布遥感影像,由公众目视解译、勾画土地利用类型,从而实现遥感影像的信息提取^[24-26]。由于其对参与人没有任何专业知识和技能方面的要求,容易出现结果质量参差不齐的情况。See L等人通过Geo-Wiki的人类影响分支进行实验,研究发现,只有在提供大量例子和训练材料或者评价反馈的情况下,大众处理的数据质量才能达到专家的水平^[27]。而且,对于一个独立的遥感影像信息提取任务,无法保证其能够在有限的时间内征集到足够多的公众参加,因此对于有时间限制的大规模遥感影像信息提取任务来说,单纯依靠公众人工绘制无法保证能够在有限时间内完成任务^[28]。

Kaggle是全球最大的数据科学家社区,也是一个通过竞赛形式来解决问题的众包平台。通过提供数据和问题说明,由任务参与者建立模型来解决问题。Kaggle发布的任务多属于机器学习、智能识别以及数据挖掘等领域^[29],参与者可在比赛期间不断完善自己的模型,从而获得更优的结果。对于大规模遥感影像信息提取任务来说,由于目前还没有一种很好的自动进

行质量评价的方法^[30-33],因此,很少通过此类竞赛式众包模式来完成。

近几年,国内也出现了大量的众包平台,如猪八戒网等,但多是围绕软件开发进行的众包。由于遥感影像信息提取不仅需要大量具有相关专业知识和技能公众,还需要提供相关数据资源和基础设施,并有技术团队进行结果的质量控制,因此,目前还没有一个完善的基于众包进行遥感影像信息提取的平台。地理空间数据云(GSCloud)是一个基于云计算技术的海量地学数据资源以及数据处理模型服务的平台,从2007年开始向公众提供服务,目前,已经积累了海量数据资源,形成了完善的基础设施和专业的服务团队,累计注册用户数超过13万人。在这些优势条件的基础上,对基于众包的遥感影像信息提取进行了研究,提出了一套完整的流程框架,并对报酬发放机制、任务分配方式、任务划分方法以及人才激励方案等领域进行了探索。

①
<http://www.geo-wiki.org/>

2 方法

本文提出了基于众包的大规模遥感影像信息提取框架,如图1所示。框架包括任务设计与分配、各子任务的遥感影像信息提取、质量评价以及结果集成与报酬发放

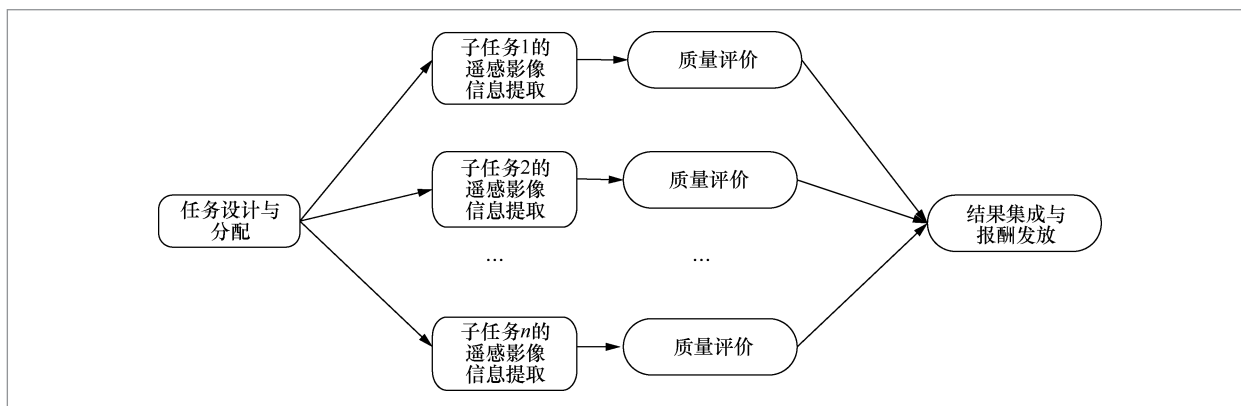


图1 基于众包的数据处理框架

4个过程,即首先清晰地定义任务,并将大任务划分为较小的子任务,然后分配给任务申请人,由任务申请人去分别完成每个子任务,通过对各个子任务进行质量评价,合格的数据结果将被集成到整个任务的结果中,并给各子任务领取人发放报酬。

2.1 任务设计与分配

在开始处理众包任务之前,为方便任务申请人对任务的理解,需对整个任务有清晰的定义和描述,然后将大任务划分为较小的子任务,这样不仅可以降低任务的复杂度,并且处理时间可以大大缩短^[34]。对于大规模遥感影像信息提取任务,通常采用时间和空间相结合的方法

进行任务划分。为保证不同任务领取人提交结果的一致性,应对子任务的处理方法加以规定和说明。同时在对各个子任务进行定价时,需考虑工作量和任务的难易程度。

研究表明,除任务描述外,任务领取人的能力对于任务完成的结果质量影响很大^[35,36],因此,在进行任务分配时,人才选取很关键。本文基于GSCLoud平台,提出基于层次数据库的人才筛选机制,如图2所示。

由于GSCLoud已积累了海量数据资源,且向用户提供基于HTTP和FTP两种数据下载方式,为便于任务的开展和数据的统一提供,平台规定任务领取人首先注册成为GSCLoud用户,然后补充完善相关技能和数据处理经验,成为人才库的一员。对于每项任务,人才库成员在申请时需填写任务解决方案和相关的成果示例等信息。后续,工作人员会综合考虑申请人的技能水平、数据处理经验以及是否能够保证完成任务,对申请人进行筛选。为了保证可以及时获得高质量的数据结果,采用竞争机制,即每个任务会分配给1~2个人。对于任务领取人,GSCLoud会与其签订协议,协议主要规定了数据处理内容、任务报酬、数据处理时间等内容。

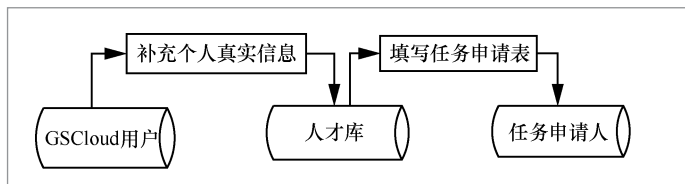


图2 人才筛选机制中的层次数据库

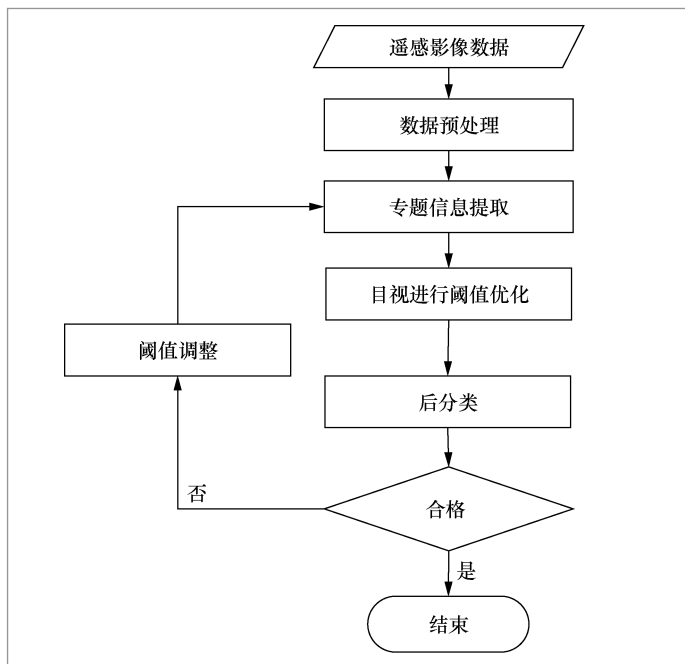


图3 专题信息提取流程

2.2 遥感影像信息提取

为了保证各个子任务的结果的一致性,本文对每个任务的影像信息提取方法进行了规定,包括数据预处理、专题信息提取和后分类3个步骤,如图3所示。通过遥感影像进行信息提取时,常用光谱、纹理等信息进行特征提取,由于涉及多个阈值的设定,因此需要人的参与,无法通过计算机自动实现。

首先对遥感影像进行预处理,包括辐射校正、几何校正、大气校正等;其次,从

影像中提取专题信息。由于每景的成像条件不同,各影像之间存在很大的差异,因此目前没有一个通用的阈值适用于所有的影像信息提取工作。在进行信息提取时,通过目视解译,对每景影像的信息提取阈值进行优化,通过迭代的人工检查,得到最优阈值,并使用该阈值从影像中提取信息。最后,人工检查并修改数据结果,直至数据质量满足要求。

2.3 质量控制

为了在有限的时间内以低成本获得高质量的数据结果,质量控制过程贯穿整个众包任务流程。任务分配阶段,通过将每个任务分配给一到多人,保证及时获得合格的数据结果。同时对每个子任务的数据结果进行精度评价。

遥感影像信息提取工作常见的误差有两类:漏分和错分。由于漏分和错分常见于特殊的情况下,如云层的遮挡、云阴影和山体阴影的影响等,因此本文提出采用定性和定量评价相结合的方式进行数据质量的评价。首先叠加遥感影像作为底图,放大提取的结果数据,对结果进行目视检查;然后根据目视检查结果,对每景影像进行验证点采样。采样方法采用随机采样和目的性采样相结合的方式,即对于容易造成分类误差的区域,如有云和山体阴影覆盖的区域,增加采样点。最后根据采样点数据,分别计算以下3个精度评价。

$$PA = \frac{TP}{TP + OP} \quad (1)$$

$$UA = \frac{TP}{TP + FP} \quad (2)$$

$$OA = \frac{TP}{TP + OP + FP} \quad (3)$$

其中, PA 为生产者精度, UA 为用户精

度, OA 为总体精度, TP 为结果与参考数据同为要提取的地物类型的像素数, OP 为漏分的该地物类型的像素数, FP 为错分为该地物类型的像素数。显然, PA 和 UA 值越大,数据质量越高,理想情况下, PA 和 UA 值应为1。通常视总体精度 OA 大于85%的结果为合格结果。

研究表明,对于遥感影像信息提取任务,对数据质量进行及时的反馈和针对任务的结果检查的反馈都有助于提高结果质量^[37]。因此本文提出迭代回归的质量控制方法,对于数据质量不合格的数据结果,平台专家会根据质量评价结果对任务领取人提出数据修改意见,如果修改次数过多,为保证任务的正常完成,会从任务申请人中重新筛选新的任务领取人来完成整个数据处理任务。具体流程如图4所示。

2.4 结果集成与报酬发放

得到各个子任务的合格数据结果后,对数据结果进行整合与集成。对于质量合格的数据结果,任务领取人会获得全额报酬,对于数据质量不合格的任务领取人,通过评估其任务执行效率,包括数据质量和时间成本,发放部分报酬。

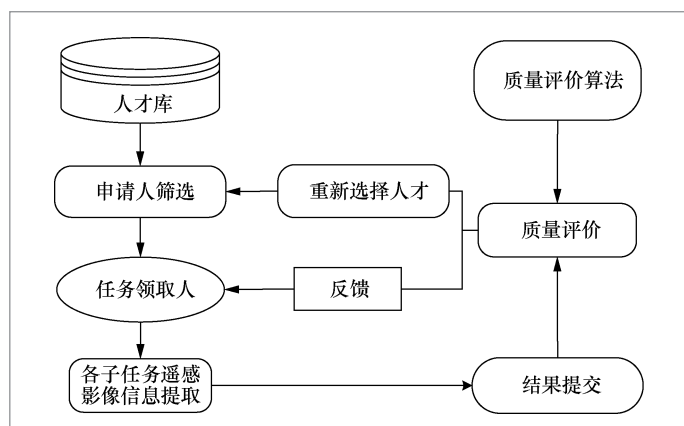


图4 质量控制流程

3 实验与分析

3.1 实验设计

针对本文提出的基于众包的大规模遥感影像信息提取方法及流程,为检测其可用性及有效性,设计并开展了基于Landsat遥感影像的青藏高原湖泊提取任务。青藏高原位于亚洲中部,涉及中国的6个省及自治区,即青海省、四川省、云南省、甘肃省、西藏自治区、新疆维吾尔自治区。由于青藏高原面积广阔,占地约260万平方公里,且海拔高、地形复杂,从遥感影像上提取湖泊的干扰信息多,因此无法使用计算机快速地对遥感影像进行解译来提取湖泊。

本任务需提取中国境内的青藏高原区域在1995年、2000年、2005年、2010年和2015年共5个时期的湖泊信息,所使用的数据为Landsat影像,包括多光谱扫描仪(MSS)、专题成像仪(TM)、增强型主题成像仪(ETM+)和陆地成像仪(OLI)影像。整个研究区需146景Landsat影像才能全部覆盖。由于工作量大,所以,此任务首先按照时间进行划分,每个时期为一个子任务,然后对1995年、2000年、2010年和2015年的子任务再按照空间范围分别划分为3个更小的子任务。由于超过一半的湖泊分布在西部区域,因此在进行空间划分时,整个区域(whole region, WR)被划分为西北(northwest, NW)、西南(southwest, SW)和东部,各子区域面积分别为66.33万平方公里、56.17万平方公里和135.77万平方公里。最后,整个任务共划分为13个子任务,分别为1995-NW、1995-SW、1995-East、2000-NW、2000-SW、2000-East、2005-WR、

2010-NW、2010-SW、2010-East、2015-NW、2015-SW和2015-East。除2005年青藏高原湖泊提取任务报酬为12 000元外,其余12个子任务每个子任务报酬为4 000元。

为了保证任务的易于理解和可执行,对此任务进行了详细的描述和说明,具体如下:

- 为避免季节变化影响,选取每一时期9-11月份,湖泊区域无云或少云(低于5%)的Landsat影像;
- 为保证结果的一致性,湖泊提取方法要一致;
- 有效去除山体阴影、云及云的阴影、积雪、冰川、裸地等非水体地物的影响;
- 当存在覆盖某一区域的两景相邻影像因日期不同造成湖泊边界不同时,在最后结果中,保留较大的湖泊边界;
- 提取精度要求达到30 m及误差控制在一个像元内,并在最终结果中去除小于3 600 m²的湖泊。

3.2 结果与分析

3.2.1 精度分析

在青藏高原湖泊提取工作中,常见的两类误差为小湖泊和由于云、阴影等造成的湖泊漏分以及由于云阴影、积雪、冰川和其他地物影响造成的错分。各时期的青藏高原湖泊提取结果数据精度见表1,其中除2005年任务外,其他时期的数据精度为该时期内3个子任务的精度平均值。

由表1可知,各子任务最终数据结果的整体精度(OA)均达到90%,超过通常所使用的85%精度规定^[30]。

3.2.2 申请人分析

将所有任务在GSCloud网站进行发

表1 提取的湖泊数据的精度数据

子任务	1995年	2000年	2005年	2010年	2015年
生产者精度	92.2%	90.1%	93.4%	91.4%	92.3%
用户精度	97.1%	98.1%	96.5%	97.8%	93.2%
整体精度	96.5%	97.8%	95.7%	96.3%	94.5%

布,为了对影响任务申请人数的因素进行研究,每个子任务在发布时,所规定的申请时间长度不同,且对2000年、2005年和2010年3个时期的子任务,采取提前支付20%的报酬,剩余80%在提交合格结果后支付的方法。1995年和2015年时期的子任务,采取数据处理者提交合格数据结果后,一次性支付所有报酬的方法。各子任务的申请人数如图5所示。

从图5可知,整个青藏高原湖泊提取任务的申请人数共239人。各子任务随着申请时长的增加,申请人数有所增长,但并不成正比。同时,2005年和2010年各子任务申请人数在申请时长和1995年各子任务一致的情况下,申请人数并未增多,因此可知,在任务开始执行前提前支付

20%的报酬,对提高任务申请人数并无多大的影响。但是,同一时期的子任务中,西南区域的子任务申请人数比其他两个子任务多,这是由于西南区域的面积比其他两个区域小,从而直接影响了申请人对工作量的评估。

3.2.3 成本分析

为在有限的时间内获得高质量的数据结果,1995年、2000年、2005年和2010年的每个子任务均由2个申请人或团队来完成。这可能会导致冗余,由于只有满足精度要求的数据才会被采用,并发放报酬,因此整个任务的成本可得到控制。每个子任务的成本信息见表2。

由表2可知,虽然1995年、2000年、

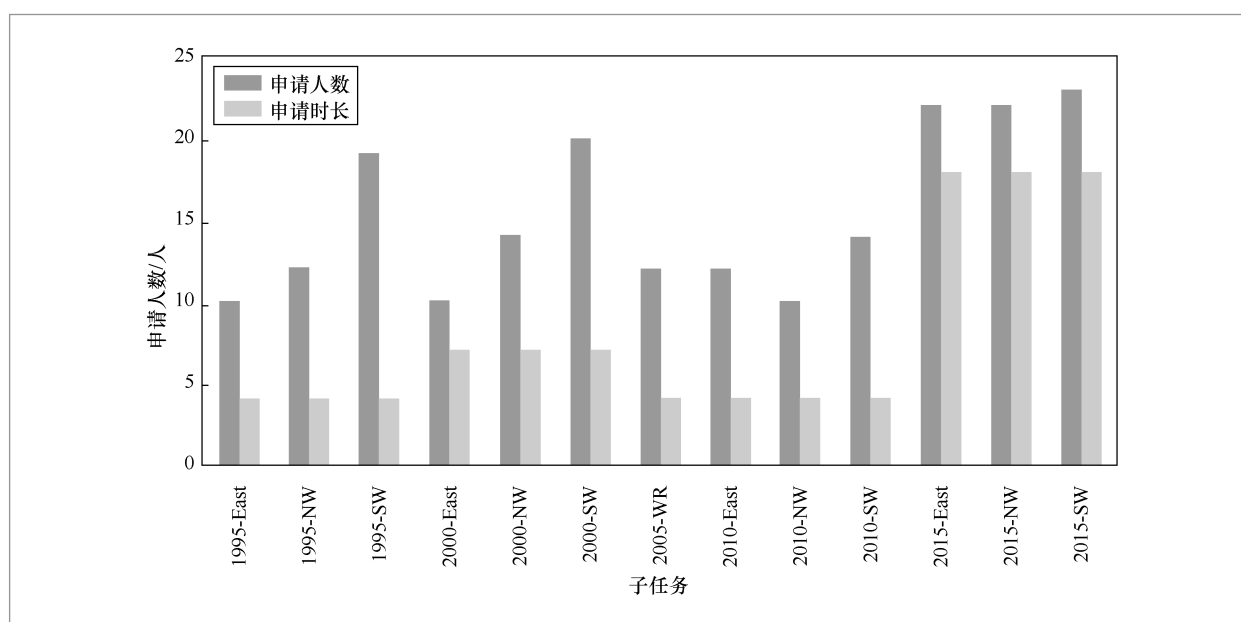


图5 青藏高原湖泊提取任务各子任务的申请人数

表2 每个子任务的报酬和相关统计数据

子任务	处理人数量/个	合格结果数量/个	不合格结果报酬/元	总报酬/元
1995-East	2	1	0	12 000
1995-NW	2	1	0	
1995-SW	2	1	0	
2000-East	2	1	800	16 800
2000-NW	2	2	0	
2000-SW	2	1	0	
2005-WR	2	1	2 400	14 400
2010-East	2	1	800	13 600
2010-NW	2	1	800	
2010-SW	2	1	0	
2015-East	1	1	0	12 000
2015-NW	1	1	0	
2015-SW	1	1	0	

2010年和2005年均选取了两个任务申请人或团队来完成任务,但只有2000-NW子任务最终接收了两份合格的数据结果,其他子任务因不合格的数据结果多支出了共4 800元。由于2015年的各子任务发布时间最晚,而且也已经获得之前任务领取人的能力信息,因此2015年各子任务均有1人完成,并且数据结果合格。由此可知,专业人才的积累能够使得成本逐渐得到控制。

为进一步分析申请人类型对任务处理时间成本的影响,2005年子任务需提取整个区域的湖泊,因此是由团队完成的;其他时期的子任务均由个人来完成。在得到合

格的数据结果前,数据处理人需对数据结果进行修改,直到数据质量合格。数据处理人为个人和团队时,数据修改次数的平均值如图6所示。

由图6可知,团队作为数据处理人,对数据修改次数高达11次,而数据处理人为个人时,平均的修改次数为2次。推测这和团队负责人有较大的关系。但由于在进行数据处理人筛选时,很难获得申请人的领导力信息,且此次研究样本数较少,故申请人类型和时间成本的关系还需要进一步进行研究。

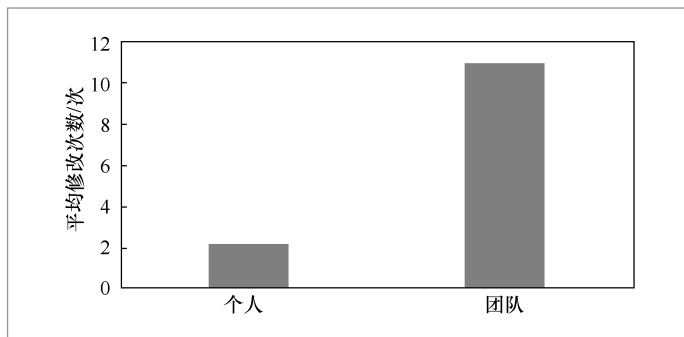


图6 不同申请人类型得到合格结果的平均修改次数

4 结束语

本文提出一种基于众包的大规模遥感影像信息提取模式,并对模式的各个流程进行了研究与设计。采用该模式进行大规模专题进行提取的3个关键方面为人才吸引、质量控制、成本分析。为了在短时间内获得高质量结果,同时控制成本,开展了实验进行相关研究。在青藏高原湖泊提取

实验中,探索了报名时长、提前支付部分报酬、个人和团队领取任务方式、每个任务发给2个人的竞争机制对目标的影响。结果发现,报名时长并不与报名人数成正比;提前支付部分报酬也没有帮助吸引更多的人报名;个人领取的方式比团队的修改次数更少,但由于样本较少,所以还需更多的实例进行进一步的研究;竞争机制有利于在有限时间内保证得到高质量结果,不过需要额外支付一部分钱给不合格的人,作为辛苦劳动的报酬。研究发现,高技能的专业人才的积累,可逐渐消除额外支付的成本,因此如何积累更多的人才,将是下一步要研究的方向。

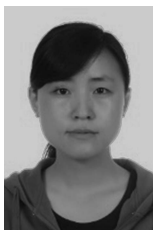
参考文献:

- [1] 黄文虎, 曹登庆, 韩增尧. 航天器动力学与控制的研究进展与展望[J]. 力学进展, 2012, 42(4): 367-394.
HUANG W H, CAO D Q, HAN Z Y. Advances and trends in dynamics and control of spacecrafts[J]. Advances in Mechanics, 2012, 42(4): 367-394.
- [2] BELLO O M, AINA Y A. Satellite remote sensing as a tool in disaster management and sustainable development: towards a synergistic approach[J]. Procedia - Social and Behavioral Sciences, 2014(120): 365-373.
- [3] YAN L, ROY D P. Automated crop field extraction from multi-temporal web enabled landsat data[J]. Remote Sensing of Environment, 2014, 144(144): 42-64.
- [4] MUELLER N, LEWIS A, ROBERTS D, et al. Water observations from space: mapping surface water from 25 years of landsat imagery across Australia[J]. Remote Sensing of Environment, 2016(174): 341-352.
- [5] LESIV M, MOLTCHANOVA E, SCHEPASCHENKO D, et al. Comparison of data fusion methods using crowdsourced data in creating a hybrid forest cover map[J]. Remote Sensing, 2016, 8(3): 261.
- [6] CHU D A, KAUFMAN Y J, ZIBORDI G, et al. Global monitoring of air pollution over land from the earth observing system-terra moderate resolution imaging spectroradiometer (MODIS)[J]. Journal of Geophysical Research, 2003, 108(D21): 1-18.
- [7] 吕超, 吕游. 遥感影像信息提取技术的研究与实现[J]. 黑龙江工程学院学报, 2014, 28(1): 34-37.
LV C, LV Y. Research and realization of remote sensing image information extraction technologies[J]. Journal of Heilongjiang Institute of Technology, 2014, 28(1): 34-37
- [8] 兰泽英. 基于土地利用空间知识挖掘的高分辨率遥感影像分类[D]. 武汉: 武汉大学, 2010.
LAN Z Y. High resolution RS image classification based on spatial knowledge mining of land use[D]. Wuhan: Wuhan University, 2010
- [9] 臧艺. 基于面向对象和集成学习的高分辨率地物要素半自动提取方法与实现[D]. 北京: 中国测绘科学研究院, 2008.
ZANG Y, Semi-automatic approach and implementation of feature extraction for high resolution remote sensing image based on object oriented and ensemble learning[D]. Beijing: Chinese Academy of Surveying and Mapping, 2008.
- [10] POHL C, J V G. Review article multisensor image fusion in remote sensing: concepts, methods and applications[J]. International Journal of Remote Sensing, 1998, 19(5): 823-854.
- [11] YUAN J. Automatic building extraction in aerial scenes using convolutional

- networks[J]. 2016: arXiv: 1602.06564.
- [12] 黄小兵. 高分辨率遥感影像建筑物提取方法研究[D]. 西安: 西安科技大学, 2014.
HUANG X B. A study of building extraction from high resolution remote sensing image[D]. Xi'an: Xi'an Technology University, 2014.
- [13] 崔卫红. 基于图论的面向对象的高分辨率影像分割方法研究[D]. 武汉: 武汉大学, 2010.
CUI W H. Research on graph theory based object oriented high resolution image segmentation[D]. Wuhan: Wuhan University, 2010.
- [14] 吴秀芸. 基于高分辨率遥感影像的建筑物提取及轮廓矢量化研究[D]. 南京: 南京大学, 2011.
WU X Y. Research on extraction and Contour vectorization of buildings from high-resolution remote sensed images[D]. Nanjing: Nanjing University, 2011.
- [15] MA Y, CHEN F, LIU J, et al. An automatic procedure for early disaster change mapping based on optical remote sensing[J]. *Remote Sensing*, 2016, 8(4): 272.
- [16] BARRINGTON L, GHOSH S, GREENE M, et al. Crowdsourcing earthquake damage assessment using remote sensing imagery[J]. *Annals of Geophysics*, 2011, 54(6): 680-687.
- [17] KAZAI G, KAMPS J, MILIC-FRAYLING N. An analysis of human factors and label accuracy in crowdsourcing relevance judgments[J]. *Information Retrieval*, 2013, 16(2): 138-178.
- [18] FIENEN M N, LOWRY C S. Social.water—a crowdsourcing tool for environmental data acquisition[J]. *Computers and Geosciences*, 2012, 49(4): 164-169.
- [19] GAO J, LIU X, OOI B C, et al. An online cost sensitive decision-making method in crowdsourcing systems[C]//The ACM SIGMOD International Conference on Management of Data, June 22-27, 2013, New York, USA. [S.l.:s.n.], 2013: 217-228.
- [20] FRATERALI P, CASTELLETTI A, SONCINI-SESSA R, et al. Putting humans in the loop: social computing for water resources management[J]. *Environmental Modelling and Software*, 2012, 37(17): 68-77.
- [21] AHN L V. Human computation[C]//The 46th Annual Design Automation Conference, July 26-31, 2009, SanFranciso, CA. New York: ACM Press, 2009: 418-419.
- [22] LIU X, LU M, OOI B C, et al. CDAS: a crowdsourcing data analytics system[J]. *Proceedings of the VLDB Endowment*, 2012, 5(10): 1040-1051.
- [23] KITTUR A, SMUS B, KRAUT R. CrowdForge: Crowdsourcing Complex Work[C]//The 24th Annual ACM Symposium on User interface Software and technology, October 16-19, 2011, Santa Barbara, USA. New York: ACM Press, 2011: 43-52.
- [24] FRITZ S, SEE L, MCCALLUM I, et al. Mapping global cropland and field size[J]. *Global Change Biology*, 2015, 21(5): 1980-1992.
- [25] FRITZ S, PERGER C, SEEL, et al. A campaign to collect volunteered geographic information on land cover and human impact information on land cover and human impact[M]. [S.l.:s.n.], 2012: 83-91.
- [26] FRITZ S, MCCALLUM I, SCHILL C, et al. Geo-wiki: an online platform for improving global land cover[J]. *Environmental Modelling and Software*, 2012, 31(7): 110-123.
- [27] SEE L, COMBER A, SALK C, et al. Comparing the quality of crowdsourced data contributed by expert and non-

- experts[J]. Plos One, 2013, 8(7): 670-692.
- [28] OFLI F, MEIER P, IMRAN M, et al. Combining human computing and machine learning to make sense of big (aerial) data for disaster response[J]. Big Data, 2016, 4(1): 47-59.
- [29] 路凯. Kaggle众包平台的隶属网络模型构建与特征分析[D]. 大连: 大连理工大学, 2015.
- LU K. Affiliation network model's constructing and characteristics analysis based on Kaggle crowdsourcing platform[D]. Dalian: Dalian University of Technology, 2015.
- [30] FOODY G M. Status of land cover classification accuracy assessment[J]. Remote Sensing of Environment, 2002, 80(1): 185-201.
- [31] 朱小宇. 支持任务推送的众包系统的研究与实现[D]. 北京: 北京邮电大学, 2014.
- ZHU X N. Research and implementation of a crowdsourcing system supporting task routing[D]. Beijing: Beijing University of Posts and Telecommunications, 2014.
- [32] DIPALANTINO D, VOJNOCIC M. Crowdsourcing and all-pay auctions[C]// The 10th ACM Conference on Electronic Commerce, July 6-10, 2009, Stanford, California. New York: ACM Press, 2009: 119-128.
- [33] 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 Z, PANG J, XIE X, et al. Research on crowdsourcing quality control strategies and evaluation algorithm[J]. Chinese Journal of Computers, 2013, 36(8): 1636-1649.
- [34] IMRAN M, LYKOURENTZOU I, CASTILLO C. Engineering crowdsourced stream processing systems[J]. Computer Science, 2013(1): 1-32.
- [35] KITTUR A, NICKERSON J V, BERNSTEIN M, et al. The future of crowd work[C]// Conference on Computer Supported Cooperative Work, June 27-29, 2013, Whistler, Canada. [S.l.:s.n.], 2013: 1301-1318.
- [36] HO C J, VAUGHAN J W. Online task assignment in crowdsourcing markets[C]//The Twenty-Sixth AAAI Conference on Artificial Intelligence, July 22-26, 2012, Toronto, Canada. [S.l.:s.n.], 2012: 45-51.
- [37] DOW S, KULKARNI A, BUNGE B, et al. Shepherding the crowd: managing and providing feedback to crowd workers[C]// Extended Abstracts on Human Factors in Computing Systems, April 10-15, 2010, Atlanta, USA. New York: ACM Press, 2011: 1669-1674.

作者简介



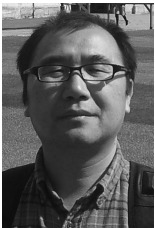
赵江华(1989-), 女, 中国科学院计算机网络信息中心研究实习员, 主要研究方向为大数据分析与管理、数据挖掘。



王学志 (1979-), 男, 中国科学院计算机网络信息中心副研究员, 主要研究方向为海量时空数据处理与分析。



林青慧 (1979-), 女, 中国科学院计算机网络信息中心副研究员, 主要研究方向为大数据资源汇聚、分析、共享及服务。



黎建辉 (1973-), 男, 中国科学院计算机网络信息中心研究员、博士生导师, 主要研究方向为大数据管理、大数据分析处理。



周园春 (1975-), 男, 中国科学院计算机网络信息中心研究员、博士生导师, 主要研究方向为数据挖掘、大数据分析处理。

收稿日期: 2016-09-14

基金项目: 国家重点研发计划基金资助项目“科学大数据管理系统”(No.2016YFB1000600); 国家重点研发计划基金资助项目“协同精密定位技术”(No.2016YFB0501900)

Foundation Items: The National Key Research Program of China “Scientific Big Data Management System” (No.2016YFB1000600), The National Key Research Program of China “Collaborative Precision Positioning Project” (No.2016YFB0501900)