

以很容易地做出快速推断,也很容易从二项选择推广到多项选择,但也容易导致低质量的众包结果。

(2) 基于统计机器学习的结果推断。主要是通过建立贝叶斯模型,描述工人能力、问题难度、任务真实结果和工人回答等相关参数之间的依赖关系,通过所获得的任务处理数据进行各个变量的联合推断。其中, Dawid & Skene(DS)<sup>[29]</sup> 是一种经典方法,每个工人有一个混淆矩阵,类别有先验,利用最大期望(Expectation Maximization, EM)算法同时评估标注结果、工人的混淆矩阵和类别先验。Snow 等<sup>[30]</sup> 采用同样的模型,但是利用了拉普拉斯滤波的最大似然估计的全监督形式。GLAD<sup>[31]</sup> 只用一个单一参数对工人进行模拟, Rayka 等<sup>[32]</sup> 为每个类别增加了工人特定的先验,在二项标注的场景下,每个工人存在一个敏感参数来描述工人对正类别的倾向,一个特异性参数描述工人对负类别的倾向,并假设这两种参数有贝塔分布先验,进一步利用 EM 算法估计真实标注和模型参数。文献 [33] 对典型的结果汇聚方法进行了概括。

## 众包系统的质量模型与保障方法

众包是一类典型的群智协同计算,其中质量控制问题一直是众包领域的研究热点。本节以众包系统为例介绍我们在理论模型和汇聚方法两方面所开展的群智系统质量保障工作。

### 质量感知的代价复杂性模型

众包系统通过将一个任务分配给多个工人来保障任务的处理质量。从经验上看,任务完成代价(体现为工人数量)和最终结果的质量密切相关,但缺乏理论模型来刻画两者之间的关系。当前,已有的工作主要通过实验的方法来比较不同的算法在代价和质量方面的表现,但在不同的数据集和众包场景下得到的实验结果并不一致,其根本原因是基于有限的数据集难以对不同众包方法的效果进行全面比较。

针对以上问题,我们将众包过程划分为任务分配和结果推断两个子过程,如图 2 所示。任务分配的结

果确定了完成众包任务的工人集合,而结果推断的过程可以抽象为参数学习和结果汇聚两个子过程。参数学习的任务是学习结果汇聚函数中的参数(例如工人能力和问题难度等),从而确定汇聚函数。

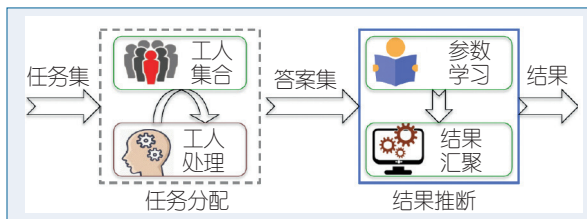


图2 众包任务的处理过程

为了从理论上评估任务分配方法、结果汇聚方法的有效性,我们利用工人能力、质量(用错误率进行度量)、汇聚函数及代价之间的关系来刻画质量感知的代价复杂性模型。不失一般性,假设工人完成单个任务的开销相等,则可利用完成任务所需的人次度量代价。另外,当完成任务的工人集合确定时,则意味着在该集合中工人的能力分布已知,任务分配方法将影响最终参与完成任务的工人能力分布。当确定工人能力分布和结果汇聚函数时,我们将满足给定质量约束下所需最多人次定义为众包系统的代价复杂性。

在此基础上,我们进一步给出该模型的计算方法。利用 Hoeffding 不等式方法对汇聚方法进行分析,并利用计算学习理论对参数学习进行推理,即可得到满足特定质量约束下最多所需代价的计算公式,从而得到了工人能力分布、汇聚函数和代价之间的关系式。

根据代价复杂性模型的计算方法,一方面,可利用工人能力分布的方差和期望来比较任务分配方法的好坏:当工人能力期望越高,并且工人的能力分布的方差越小时,则该任务分配方法较优。另一方面,当工人能力分布确定时,为了判断不同汇聚方法优劣,给定一定质量约束,则可预测出相对的最差需要人次开销,从而比较汇聚方法的好坏。基于实体标注相关数据集的实验结果验证了我们的代价复杂性分析方法的有效性。

文献 [34] 给出了代价复杂性的形式化定义,同

时以定理的形式给出了代价复杂性的计算。所给出的代价复杂性模型建立了代价与质量之间的数学表达。类似计算理论中的时间/空间复杂度的作用,基于众包代价复杂性可以实现对各类结果汇聚及任务分配算法之间的性能进行合理的比较,同时也为设计新的众包任务处理算法提供了理论支撑。

## 知识驱动的众包结果汇聚方法

很多研究工作仅考虑确定性答案个数有限的众包任务,但更多的众包任务的结果是开放型的、不确定的,例如知识获取、问答系统等。考虑下面的场景:在知识获取的众包任务中,要求5个工人对图片中的对象进行标注。给定一幅包含“鸟”的图片,由于每个工人的知识背景不同,导致认知能力不同,所给出的标注可能为{“鸟”,“鸟”,“鸟”,“蜂鸟”,“刺喙蜂鸟”}。按照大多数投票等汇聚方法,汇聚的结果很有可能是“鸟”,但这样的结果所获得的只是通用知识,而我们更希望获得类似“刺喙蜂鸟”的具体标注。



图3 图片标注结果之间的语义关系示例

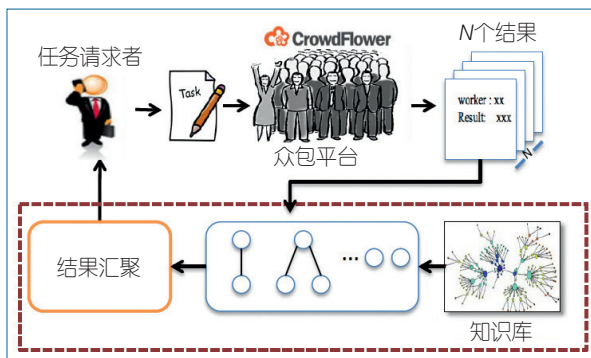


图4 知识驱动的结果汇聚框架

现有的结果汇聚方法存在的问题在于将各个工人提供的答案看作是完全独立的,忽略了答案之间

的关联关系。上例中,实际上“鸟”“蜂鸟”和“刺喙蜂鸟”之间存在着语义关联,如图3所示。如果某个工人给出“刺喙蜂鸟”的答案,那么这也蕴含着他也赞同“鸟”和“蜂鸟”这两个答案。

针对以上问题,我们提出知识驱动的结果汇聚方法(如图4所示)<sup>[35, 36]</sup>,将外部知识库引入到结果汇聚的过程中。该方法不仅考虑工人答案的不可靠性,同时也对答案之间的关联关系进行建模,使得能够通过众包获得专指度更高的知识。

答案之间的关联关系对于众包系统的质量具有重要影响。基于这一认识,我们进一步考虑复杂众包任务中的子任务及相应答案之间的上下文依赖关系,提出了上下文感知的结果汇聚方法<sup>[37]</sup>,并通过手写体识别和语音识别等真实任务对方法进行了实验验证,结果表明,该方法在花费相同的情况下有效地提升了结果的质量。

## 总结

群智系统是在互联网的推动下发展起来的一类重要的应用系统,目前已经在机器学习、问答系统、软件开发等诸多应用领域得到了广泛的应用。随着互联网、大数据、人工智能和物联网等的进一步发展,群智系统将会发挥更加重要的作用。然而,由于群智资源的特殊性以及群智协同计算环境的开放性,质量保障依然是群智系统所面临的重要挑战之一。围绕群智任务的设计、工人处理任务的质量和群智结果的处理等,研究群智系统的质量保障是重中之重。

**协同感知的质量保障。**在已有的众包和人本计算的研究中,通常假定工人之间彼此独立地完成任务。实际上,由于任务的复杂性以及工人对任务完成效率的追求,工人在任务处理过程中往往开展大量的合作。良性的合作有利于提升结果的可靠性,但合作也会使得各个工人提交趋同的处理结果,而降低结果的多样性。因此在结果的多样性和可靠性之间进行有效的均衡,从而建立协同感知的质量保障机制尤为重要。

**人机协作的质量保障。**现有的群智系统中的

任务大多依靠人来完成,然后通过人工或者算法实现结果的汇聚<sup>[38,39]</sup>。虽然在人工智能和数据库等领域,人们已经开始关注人机回路(human-in-the-loop)问题的研究,通过将人的智能引入计算系统实现对更加复杂问题的求解,但是由于人的特殊性和不同任务的差异性,人和计算机的协作机理尚不清晰,研究人机协作的质量保障仍然面临很大挑战。 ■



孙海龙

CCF 高级会员、协同计算专委会常务委员。北京航空航天大学计算机学院特聘副教授。主要研究方向为群体智能、智能化软件方法和分布式系统等。曾获国家技术发明二等奖两项。sunhl@buaa.edu.cn



方毅立

CCF 专业会员。浙江工商大学计算机与信息工程学院讲师。主要研究方向为群智感知计算/众包、机器学习和人机交互等。fangyl@zjgsu.edu.cn



李国良

CCF 杰出会员、CCF 青年科学家奖获得者。清华大学副教授。主要研究方向为群体计算、大数据管理、数据清洗与融合、移动数据管理等。liguoliang@tsinghua.edu.cn

## 参考文献

- [1] Number of internet users worldwide from 2005 to 2017 (in millions). <https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/>.
- [2] 孙海龙, 卢瞰, 李建国, 顾宁. 群智协同计算: 研究进展与发展趋势. CCF2016-2017 中国计算机科学技术发展报告, 机械工业出版社, 2017.
- [3] Doan A, Ramakrishnan R, Alon Y. HALEVY: Crowdsourcing systems on the World-Wide Web[J]. *Commun. ACM*, 2011, 54(4): 86-96.
- [4] Quinn A J, Bederson B B. Human computation: a survey and taxonomy of a growing field[C]// *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2011: 1403-1412.
- [5] [https://en.wikipedia.org/wiki/Collective\\_intelligence](https://en.wikipedia.org/wiki/Collective_intelligence).
- [6] [https://en.wikipedia.org/wiki/Open\\_collaboration](https://en.wikipedia.org/wiki/Open_collaboration).
- [7] Michelucci P, Dickinson J L. The Power of Crowds[J]. *Science*, 2017, 351(6268): 32-33.
- [8] Licklider J C R. Man-Computer Symbiosis[J]. *Ire Transactions on Human Factors in Electronics*, 1960, HFE-1(1):4-11.
- [9] Hendler J, Berners-Lee T. From the Semantic Web to social machines: A research challenge for AI on the World Wide Web[J]. *Artificial Intelligence*, 2009, 174(2): 156-161.
- [10] Pan Y. Heading toward Artificial Intelligence 2.0[J]. *Engineering*, 2016, 2(4): 409-413.
- [11] 国务院关于印发新一代人工智能发展规划的通知[EB/OL]. [2017-07-20]. [http://www.gov.cn/zhengce/content/2017-07/20/content\\_5211996.htm](http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm).
- [12] Wei L I, Wen-Jun W U, Wang H M, et al. Crowd intelligence in AI 2.0 era[J]. *Frontiers of Information Technology & Electronic Engineering*, 2017, 18(1):15-43.
- [13] Kittur A, Smus B, Khamkar S, et al. Crowdforge: Crowdsourcing complex work[C]// *Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, 2011: 43-52.
- [14] Little G, Chilton L B, Goldman M, Miller R C. TurkIt: tools for iterative tasks on mechanical turk[C]// *In KDD workshop on human computation*, Paris, 2009:29-30.
- [15] Dai P, Mausam, Weld D S. Decision—theoretic control of crowd-sourced workflows[C]// *In AAAI*, Atlanta, 2010.
- [16] Lin C H, Mausam, Weld D S. Dynamically switching between synergistic workflows for crowdsourcing[C]// *In AAAI*, Toronto, 2012b.
- [17] Bernstein M S, Little G, Miller R C, Hartmann B, Ackerman M S, Karger D R, Crowell D, Panovich K. Soylent: a word processor with a crowd in-side[C]// *In UIST*, New York, 2010: 313-322.
- [18] Kittur A, Smus B, Kraut R. CrowdForge: crowdsourcing complex work[C]// *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2011:1801-1806.
- [19] Gadiraju U, Yang J, Bozzon A. Clarity is a Worthwhile Quality: On the Role of Task Clarity in Microtask Crowdsourcing[C]// *The ACM Conference*. ACM, 2017:5-14.
- [20] Edith Law, Ming Yin, Joslin Goh, Kevin Chen, Michael A. Terry, Krzysztof Z. Gajos: Curiosity Killed the Cat, but Makes Crowdwork Better. CHI 2016: 4098-4110
- [21] Yao-Xiang Ding, Zhi-Hua Zhou: Crowdsourcing with unsure option. *Machine Learning* 107(4): 749-766