



高可信众包群体构建方法

冯洋¹, 王祎², 房春荣^{1*}, 郭楠楠¹, 陈振宇¹

1. 南京大学计算机软件新技术国家重点实验室, 南京 210093, 中国

2. Department of Software Engineering, Rochester Institute of Technology, Rochester 14623, USA

* 通信作者. E-mail: fangchunrong@nju.edu.cn

收稿日期: 2019-05-31; 修回日期: 2019-08-08; 接受日期: 2019-09-11; 网络出版日期: 2019-11-04

国家重点研发计划 (批准号: 2018YFB1403400) 和国家自然科学基金 (批准号: 61690201, 61772014, 61802171) 资助项目

摘要 构建高可信众包群体是众包活动的一个重要挑战. 目前, 众包机制研究主要集中在激励众包个体, 而忽略了能够反映微服务质量的高诚信众包群体的构建. 本文根据通过融合委托-代理模型和信号博弈理论提出了一种新颖的高可信众包群体构建机制. 该机制实现了众包工人和发包方的信息交换, 并通过随机质量审查来为诚信/失信工人提供动态的经济激励/惩罚. 本文通过大规模的仿真实验来对该机制进行验证. 实验结果表明该机制具有良好的有效性和高效性, 能够通过极小的代价让能力强但失信的工人在任务提交前改正行为做出诚信决策.

关键词 可信众包, 众包机制, 信号博弈理论, 委托代理模型

1 引言

在众包活动中, 确定一个高可信众包群体是非常困难的^[1~3]. 由于众包活动中的微任务成千上万, 发包方¹⁾难以逐个审查提交结果, 只能“假设”匿名的众包工人是一个高可信的群体. 然而, 心理学、经济学, 及管理学的研究都表明众包工人没有“假设”的那么可信^[4~6]. 在实际的众包活动中, 这种超出假设的并且难以约束的众包工人失信行为存在严重风险.

另一方面, 在无法对众包工人自身汇报的任务完成质量进行完全信任的情况下, 从大量完成报告中剔除低质量的标注结果是一件非常耗时耗力的事. 逐个审查所付出的代价基本等同于完全自己标记. 而朴素的随机抽样审查方法在样本量巨大的情况下, 只能估计出所有众包任务整体的完成质量, 而无法剔除所有的低质量结果. 在此情况下, 众包任务发起方 (requester) 很容易陷入一个困境: 是花费大量资源对众包任务质量进行进一步审查, 还是接受包含一定数量的低质量完成结果.

1) 本文中众包任务发起方简称为“发包方”.

引用格式: 冯洋, 王祎, 房春荣, 等. 高可信众包群体构建方法. 中国科学: 信息科学, 2019, 49: 1412–1427, doi: 10.1360/N112018-00303
Feng Y, Wang Y, Fang C R, et al. An approach for developing a highly trustworthy crowd-sourced workforce (in Chinese). Sci Sin Inform, 2019, 49: 1412–1427, doi: 10.1360/N112018-00303

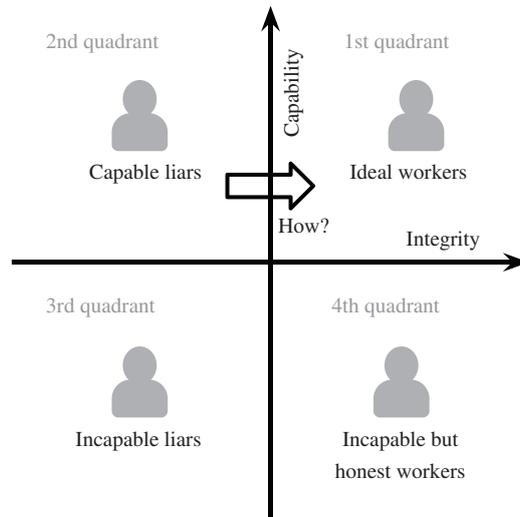


图1 根据诚信度和能力划分的4种工人类型划分

Figure 1 Four types of crowd workers according to their capability and integrity

如果众包工人群体可信, 众包发起方就不会陷入这样的困境. 根据工作能力和诚信度, 我们可以将众包工人划分为如下4类(如图1所示):

- **高能力诚信工人.** 具有较高工作能力, 较大概率能够高质量完成众包任务. 同时, 可以如实反馈众包任务完成质量.

- **高能力失信工人.** 具有较高工作能力, 较大概率能够高质量完成众包任务. 同时, 对于没能完成或完成质量不够高的任务, 可能报告为高质量完成结果.

- **低能力诚信工人.** 不具有较高工作能力, 很小概率能够高质量完成众包任务. 同时, 可以如实反馈众包任务完成质量. 对于没能完成或完成质量不够高的任务, 会报告为低质量完成结果.

- **低能力失信工人.** 不具有较高工作能力, 很小概率能够高质量完成众包任务. 同时, 对于没能完成或完成质量不够高的任务, 可能报告为高质量完成结果.

虽然发包方期望招募高能力诚信众包工人, 但实际上高能力诚信工人只占有众包工人人数的一小部分. 而大部分众包工人是低能力诚信工人、低能力失信工人和高能力失信工人. 一般来说, 考虑到当前绝大多数众包工程是根据任务完成质量对众包工人进行最终奖励的评估, 低能力诚信工人对整个工程的损害是极其有限的. 因为他们会如实地告诉发包方他们所提交众包任务完成结果的质量, 发包方可以根据他们的汇报简单地剔除低质量的完成结果并重新发布该众包任务. 低能力失信工人的报告也很比较容易处理. 由于可以通过多种方法(例如简单地随机抽象)了解到他们的能力. 因此发送特定的验证任务就可以判定其能力, 然后将其从众包工人群体中剔除. 然而, 对于“高能力失信工人”的处理方法则具有较大挑战. 因为他们的能力可以提供良好的众包任务结果, 所以既不能通过简单的众包任务进行验证, 也不能简单地将其从众包群体中移除. 因此, 本文设计了一种机制以激励“高能力失信工人”像理想工人——“高能力诚信工人”——一样完成众包任务.

在当前众包任务的完成过程中, 每个众包工人都需要提交一份对于自己任务完成情况的评估. 同时, 通常情况下众包任务的数量较大, 这导致任务完成反馈报告数量众多, 众包发起者审核这些反馈报告需要耗费巨大的时间精力. 因此, 发包方通常按照某个给定概率来随机抽样审查少量的众包结果, 并根据审核结果评估众包工人的表现, 并给予众包工人报酬. 注意到, 在该过程中, 发包方对于报告的

审核需要消耗一定资源. 可以预见到, 如果随机抽样的概率设定得越高, 那么审核需要消耗的资源也会更高.

一个较为理想的解决方案是降低众包工人群体中“高能力失信工人”的比例, 从而从整体上提升众包任务的完成质量. 为了激励“高能力失信工人”成为理想的工作者, 进而培养出一个“高能力诚信”的众包群体, 我们设计了一种结合委托-代理模型和信号博弈理论结合的高可信众包群体构建机制. 与目前主流的相比, 该机制主要有以下优势: (1) 有效地构建一个高可信众包群体; (2) 通过以“提交微任务完成情况”的信号形式进行信息交换, 可以自动排除大部分低质量的众包结果; (3) 通过不依赖于外部评价体系^[7]的方法提高众包工人的诚信度, 超出了基于绩效的激励机制^[8~10]; (4) 通过容忍众包工人在能力和诚实度方面的差异避免了众包群体的单一性假设.

针对这一机制, 本文实施了一个大规模的仿真实验. 实验结果表明:

(1) 该机制能够促使众包工人的行为进化, 可以激励“高能力失信工人”转变为“高能力可信工人”.

(2) 该机制帮助发包方获取较好的总收益, 同时保持单个众包工人的贡献处于相当高的水平.

(3) 该机制帮助众包工人行为转化成本极小, 可接受.

第 2 节介绍了委托-代理模型、信号博弈等相关的背景工作; 第 3 节具体描述了文本所提机制; 第 4 节介绍了仿真实验设计; 第 5 节报告了仿真实验结果; 第 6 节讨论了相关的问题; 第 7 节介绍了相关工作; 第 8 节总结了本文工作.

2 相关背景

本文所提机制主要基于两个模型: 委托-代理模型和信号博弈理论. 我们将在本节中介绍这两个模型的背景以及我们所做的适配工作.

2.1 众包中的责任-代理模型

众包过程一直被认为是一种经典的基于契约理论的多轮多工人版本委托-代理模型^[8, 11, 12]. 与现有经典的委托-代理模型^[13]不同, 本文基于如下多方面考虑摒弃了经典委托-代理模型中, 众包工人为了完成众包任务在个人“投入”上的不可观察的概念. 首先, 在多轮博弈中, 投入水平具有不可观察性并且难以学习. 其次, 考虑到众包项目可能涉及数千名工人和大量微任务, 仅仅通过优化合同的方式来激励单个工人在微任务上的提升投入水平, 并不能起到很好的效果. 此外, 已有文献^[13]发现, 许多典型的任务可能不是投入响应型, 即付出更多的投入也不一定会提高任务完成质量.

众包工人的诚信度在许多类型的众包微任务中至关重要. 发包方希望众包工人能够诚实地评价其任务完成质量. 即使这些工人的完成质量不高, 即使会把这些结果忽略, 发包方也希望奖励这些诚信工人. 因为对发包方而言, 用以训练的数据的正确性非常重要, 发包方宁可选择付钱给诚信工人将错误结果排除在外. 针对这种场景, 我们采用并扩展了委托-代理问题的“逆向选择”变化. 众包工人的类型由其能力值和诚信值组成, 该值在任务初始阶段是未知的. 在不断地完成微任务过程中, 我们逐渐学习到工人的能力值, 然后通过工人对其自身工作的评价学习到工人的“诚信值”.

因此, 我们的方法不同于契约理论和基于绩效的支付 (PBPs). 基于上述讨论, “诚信值”意味着工人与发包方间存在着信息交换. 我们将在 2.2 小节中使用信号博弈概念化这种类型的信息交换.

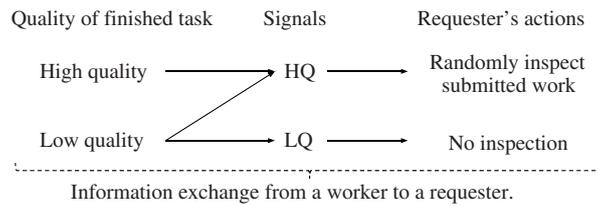


图 2 发包方与众包工人之间的信号博弈

Figure 2 Signaling game between a crowd worker and a requester

2.2 信号博弈理论

标准的委托-代理模型忽略了从代理方到委托人的信息传递,导致交互中的信息传递完全取决于单向信号.因此,我们选择了将委托-代理模型与信号博弈相结合.对于众包工人来说,他们在很大程度上可以感知任务的完成质量,而发包方不可能检查所有已完成的任務.当允许信息从工人流向发包方时,工人发送的信号的真假程度完全取决于工人的诚信值,然后发包方可以根据接收的信号采取相应的动作.

如图 2 所示,可以使用简单的双人信号博弈^[14]来描述这种类型的信息交换.显然,工人所完成的任務可能是高质量的,也可能是低质量的,因此设定两种信号:高质量(HQ)、低质量(LQ).理想情况下,工人应该为高质量完成的任務发送HQ,否则发送LQ.然而,对工作质量的自我感知与评价的准确程度是不确定的,即使他的工作完成的质量不高,他也可能会发送HQ.考虑到实践中,众包工人的目标是最大化他们可以获得的收益,因此当他们以较高质量完成了一个众包任务后,他们不大可能向发包方发送一个LQ信号.因此,当收到HQ信号时,发包方需要对收到的任务完成结果的质量进行审查.这种情况下,发包方可以随机检查提交的工作并根据质量检验的结果动态调整报酬.如果发包方收到LQ信号,则无需检查提交的任务结果,只需要将该结果忽略掉即可.在这种情况下,发包方只需要向工作人员分配正常的报酬.我们将在接下来的章节中详细介绍支付结构.显然,图2中描述的博弈没有分离均衡.因为信号本身携带错误信息,发包方永远不会有关于工人是否发送伪造信号的系统知识.然而,通过强制执行一些激励/惩罚,至少可以使一个失信工人表现得更诚实^[15,16].在我们提出的机制中,这形成了激励“高能力失信工人”表现得像“高能力诚信工人”的基本原理.

3 建模与机制设计

3.1 参与者建模

本文中提出的模型定义了两种类型的参与者:发包方和众包工人,其中,发包方在委托-代理模型中扮演委托方的角色,众包工人扮演代理的角色.假设总共有 m 名工人构成了众包工人集合 $W = \{w_i | 0 \leq i < m\}$.因为代理之间没有交互,所以每次博弈都可以被视为双人博弈(发包方与众包工人).对于任意工人 w_i ,用 n_i 表示该工人的博弈过程包含的总周期数.同时,用 n_{ij} 表示工人 w_i 的第 j 个博弈周期.需要注意的是,本文提出的模型对于微任务的类型和特点并没有特别的限制.唯一的要求是微任务允许质量差异,除此之外没有任何限制.

本文采用一种直接的发包方建模方法.设定概率 p 来描述发包方审查信号HQ的可能性.对于工人的建模略微复杂,一个工人具有两个属性,能力值和诚信值.对于任一工人 w_i ,本文将其的能力值和诚信值分别表示为 cap_i, hon_i .Ones等^[17]开展的关于跨种族以及各年龄段工作申请群体的研究表明,

一般情况下, 人类群体的诚实度在数值化并进行归一化之后, 是符合正态分布的. 同样地, Mayer 等^[18]在研究中发现, 人群的工作能力进行数值化之后, 其数值分布也是符合正态分布的. 这两个研究结果的群体完全符合众包工人的特点, 因此, 假设这两个属性在 m 个工人的总体上服从正态分布. 这两个属性共同决定了众包工人提供高质量提交以及发送失信信号的可能性. 那么, 对于任一工人 w_i , 其为微任务发送假 HQ 信号的概率可以表示为

$$P_i = \varphi(\text{cap}_i, \text{hon}_i). \quad (1)$$

在本文模型中, 假设是 φ 的联合分布的线性函数, 并且假设 (cap, hon) 的分布是独立的. 可以看到, cap 或 hon 的上升将导致 P 的下降. 在对工人进行建模时还需要考虑另外一个重要问题, 即工人的行为变化是由他们被发现低诚信行为后得到的反馈信息的变化来决定. 当然, 可以简单地假设一个工人被发现低诚信后会自动成为一名高诚信的工人, 并得到一些经济损失. 然而, 这在实际中并不可行, 因为诚信被广泛接受为一种基本的人格特质, 在短期内不易改变^[19]. 因此, 采取间接的方式, 根据工人的经济损失调整其在未来轮次中发生低诚信行为的概率^[20].

调整后的假 HQ 信号发送概率表示为

$$P'_i = \varphi(\text{cap}_i, \text{hon}_i) - \Delta(x_i \times \theta), \quad (2)$$

其中, θ 代表惩罚金额数; x_i 表示 w_i 的失信累积次数. 注意到, 在本文提出的机制中, 低诚信行为的概率降低, 是来源于式 (2) 中 $\Delta(x_i \times \theta)$ 的惩罚效应. 这种惩罚效应是通过直接经济损失的反馈方式来实现的. 通过分析可以看到, 在本文机制中, 如果某一众包工人多次产生失信行为, $(x_i \times \theta)$ 所产生的惩罚金额越高. 而 $\Delta(x_i \times \theta)$ 会降低假 HQ 信号的发送概率. 这种设计是基于一个核心假设, 即“众包工人会努力获得较高的收益”. 在该假设之下, 当众包工人接受了任务, 但是发现因为自己的作弊行为使得收益下降的时候, 就会改变自身行为. 这种改变是一个由众包工人自发的行为, 在本文机制中, 假设这种改变是一个线性的过程.

3.2 机制过程

对于特定的 w_i , 图 3 描述了其与发包方交互的整个过程. 在初始阶段, 发包方与工人签订雇佣合同. 然后, 工人开始第 1 个周期中的第 1 个微任务, 并且持续为发包方工作, 直到在第 n_i 个周期的某个事件触发终止这个过程. 在本文讨论的模型中, 有两个事件可以终止这个过程: (1) 没有新的微任务发布或者分派到该工人; (2) 工人被发包方剔除.

在周期 j ($0 < j \leq n_i$) 中, 图 3 右侧部分描述了发包方和工人 w_i 之间的交互.

交互从正常的众包过程开始: 请求者发布任务, 然后, 工人接受并进行处理该任务, 并要求工人 w_i 发送信号来表征微任务的完成质量. 显然, 工人 w_i 有可能不会一直保持诚信. 正如在 2.2 小节中所讨论的那样, 即使提交的完成结果质量很低, 工人也可能将信号发送为“HQ”. 发包方在收到该信号后就会将该低质量结果混入到高质量结果集合中, 从而导致损失. 然而, 本文所提出的机制允许发包方做一些额外的处理. 如果信号是“LQ”, 发包方可以简单地忽略这次提交, 并向工人支付少量报酬 V_L 作为工人 w_i 诚信的奖励. 如果信号是“HQ”, 那么该信号有一定的概率是一个失信行为. 由于不可能详尽地检查所有提交, 发包方可以根据概率 p 决定是否检查工人的提交. 如果发包方决定不进行检查, 工人 w_i 将收到高质量提交 V_H 的报酬. 如果发包方决定检查, 工人的报酬则根据实际检查结果和工人行为而动态变化. 我们把这笔报酬写成 V_D . 相关奖励机制的细节将在 3.3 小节中介绍.

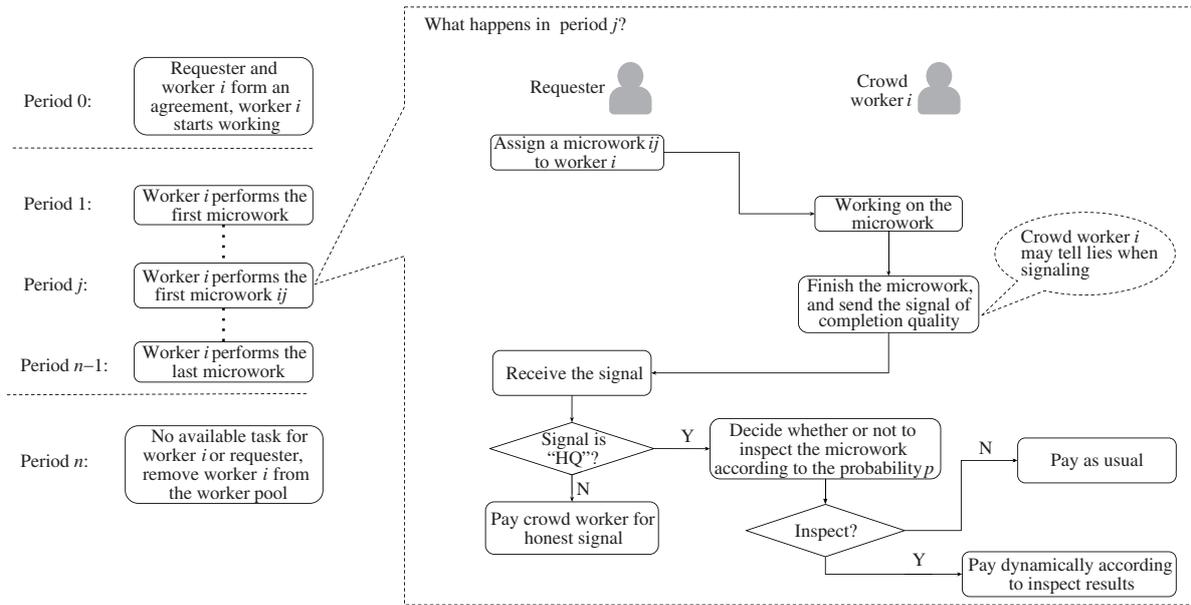


图 3 本文机制过程图 (图的左边是整体描述; 图的右边是每个周期的交互细节)

Figure 3 The process of the proposed mechanism. The left part is the overview of all periods, and the right part is the details of interactions in each period

3.3 奖励机制

由于所提出机制的动态特性, 奖励机制具有较为复杂结构. 正如第 2 节所提, 由于工人自身投入水平的不可观察性, 本文提出的模型并不讨论工人投入水平对于整个过程的影响. 因此, 假设一个微任务所需的劳动代价 e 对所有的工人是一致的. 因此, 工人 w_i 在第 j 轮的报酬可以表示为

$$G_{ij} = g(i, j) - e. \tag{3}$$

注意到, 在式 (3) 中, 用 t_{ij} 表示工人 w_i 在第 j 个周期中完成的微任务. 其中, $g(i, j)$ 表示发包方所支付的酬金. 其中, $g(i, j)$ 的定义如下:

$$g(i, j) = \begin{cases} (V_L), & \text{信号为 "LQ", 不需要审查;} \\ (V_H)_n, & \text{信号为 "HQ", 没有审查;} \\ (V_H)_y, & \text{信号为 "HQ", 审查确认;} \\ (V_F), & \text{信号为 "HQ", 审查拒绝.} \end{cases} \tag{4}$$

对失信者的惩罚不仅影响其当前周期的收益, 而且还可能导致失信者在未来周期的行为改变. 对于发包方来说, 其总收益是取决于从所有高质量的微任务中获得的总体收益, 减去其向工人支付的总体报酬, 以及对提交的微任务的样本进行检查的总体管理费用. 假设发包方会从每个高质量完成的微任务 t 获得一个固定的收益 u , 则整个过程可以表达为

$$u(i, j) = \begin{cases} u, & \text{如果微任务 } t_{ij} \text{ 的完成质量较高;} \\ 0, & \text{如果微任务 } t_{ij} \text{ 的完成质量较低.} \end{cases} \tag{5}$$

由于在实践中, 审查微任务的完成质量, 通常需要消耗一定的资源. 那么我们假设审查一微任务 t_{ij} 的消耗是 $c(i, j)$, 并定义

$$c(i, j) = \begin{cases} c, & \text{如果对微任务 } t_{ij} \text{ 进行审查;} \\ 0, & \text{如果不对微任务 } t_{ij} \text{ 进行审查.} \end{cases} \quad (6)$$

结合以上定义, 发包方的总收益 U_R 可以被表示为

$$U_R = \sum_{i=1}^m \sum_{j=1}^{n_i} [u(i, j) - g(i, j) - c(i, j)]. \quad (7)$$

4 仿真实验设计

本文设计了一个大规模的仿真实验来验证本文所提出的机制对众包工人表现的影响. 首先在本节中描述了基准、参数和实验设置. 然后在第 5 节中讨论了实验结果. 为了确保本文提出的方法的正确性和结果的可重复性, 我们开源了本实验所涉及代码 (用 Java 实现).

4.1 基准方法

本文提出的机制将与以下 3 个基准方法进行比较. 其中, 在本文中使用的术语“报告”表示“已完成的微任务”.

基准 1: Signal-no-tolerance (S-N-T). 第 1 个基准是模拟最朴素的审核机制, 即采用随机策略检测众包工人发送的质量评估信号. 该基准要求众包工人提交微任务的完成结果, 并且伴随一个质量评估信号. 发包方不会审查质量评估信号为 LQ 的报告, 而会对质量评估信号为 HQ 的报告进行随机抽样检测. 在随机检测中, 这种机制对于任何不诚信行为是零容忍. 如果检测到任何不诚信行为者, 该机制会立即剔除失信者, 并且不会为该任务支付费用.

基准 2: Report-no-tolerance (R-N-T). 该基准类似于无容忍信号, 但是在这种机制下, 发包方将从每个工人提交的所有报告中随机抽取一些进行审查, 而不会过滤掉质量评估信号为 LQ 的报告. 如果发现任何有害的不诚信行为, 发包方将立即剔除不诚信者, 并且不会为该任务支付费用.

基准 3: Report-sample-and-pay (R-S-P). 与基准 2 相比, 这种基准机制相对温和. 发包方将要求众包工人提交报告, 并进行相同的抽样和审查. 但发包方不会驱逐任何工人, 相反, 发包方根据人工诊断结果来支付报酬. 注意到, 以上 3 种基准中, 对于报酬的支付以及审核报告所消耗的资源都符合前文中所定义的公式.

4.2 参数设置

4.2.1 众包工人相关的参数设置

正如 3.1 小节讨论的那样, 每个工人具有两个属性: 能力值和诚信值. 根据关于个体认知能力值和诚信值分布的心理学文献^[17, 21], 每个属性都是从正态分布中随机分配的. 对于能力值, 设置均值为 0.5, 标准差为 0.2; 对于诚信度, 均值为 0.7, 标准差为 0.2. 同样, 我们进行了敏感性分析, 以检查这两种分布的变化结果. 通过敏感性分析, 发现当该分布允许这两个属性有足够的方差时, 结果是鲁棒的.

4.2.2 成本相关的参数设置

本小节将详细解释仿真实验中与支付相关的关键参数. 如上所述, 在式 (7) 中定义了净效用增益. 在本实验中, 需要数值化价值变化, 因此需要详细说明审查成本和微任务的支付成本. 首先, 定义 $(V_F) = 0$, 这意味着不会对检测到的不诚信行为支付任何报酬. 此外, 根据式 (5) 使用 u 来表示高质量微任务的效用增益, 式 (6) 使用 c 来表示检查成本, 定义 $(V_H)_n = \alpha u$, $(V_H)_y = \beta u$, $(V_L) = \gamma u$, 并且 $c = \sigma u$.

许多关于雇员与雇主之间收入分配的实证研究一致表明雇员分享一半以上的总利润在现代资本主义中是非常罕见的^[22]. 在监管不那么严格的众包劳动力市场中, 可以合理地假设发包方 (雇主) 在微任务的“利润”中的份额不应低于 50%. 由于最高薪酬分派是 $(V_H)_y$, 并且接收此付款需要检查, $(V_H)_y + c = (\beta + \sigma)u$ 不应超过 50%. 因此在模拟实验中设定 $\beta = 0.45$ 且 $\sigma = 0.05$. 此外, 假设 $\alpha = 0.3$ 且 $\gamma = 0.1$. 我们对这些参数进行了敏感性分析, 发现如果相对数值关系 $(\beta > \alpha \gg \gamma = 0.1 > \sigma)$ 成立, 则结果是一致的.

4.3 实验设计

本文在仿真实验实现了 4 种机制: 4.1 小节中定义的 3 个基准以及本文机制. 每个模拟实例 (映射到每个机制) 包含 50 个周期. 运行每个模拟实例 20 次.

在每个仿真实例中, 假设每个机制都有一个包含 1000 个众包工人的工作池. 有些工人可能会在任一周期被踢出, 剩下的工作者将自动纳入到下一个周期. 在每个周期的开始, 为每个可用的工作者分配一个包含固定数量原子任务的微任务. 为了避免微任务困难程度的偏差, 即对于众包工人来说太困难或太容易, 采用跟工人能力分配相同的分布来随机生成难度值. 在这个实验中, 每个微任务包含 10 个原子任务. 众包工人一旦收到微任务, 就会评估自身成功完成该任务的能力. 对于微任务的每个原子任务, 如果工作能力高于难度, 工人就可以成功完成, 反之亦然. 此外, 如果 80% 的原子任务成功完成, 就定义该微任务的提交是高质量的.

众包工人是理性的, 如果完成了一个高质量的微任务, 则一定会向发包方发送一个 HQ 信号作为自我评估. 正如在第 2.2 小节中所讨论, 工人如果未能达到高质量的表现, 就可以诚信地向发包方发送 LQ 信号, 并收到 LQ 报酬; 或者向发包方发送 HQ 信号 (欺骗), 并希望发包方永远不会检测到该失信行为.

为了模拟失信行为, 我们采用同为工人生成诚信度相同的分布来随机生成一个失信意愿值. 如果这个不诚信意愿值小于工人的诚信度, 工人就选择诚信并发送 LQ 信号, 否则, 工人决定不诚信地发送 HQ. 由于不可能审查所有报告, 发包方使用随机抽样来评估工人的表现. 然后, 基于不同的机制, 发包方应用不同的方法来剔除不诚信的工人. 注意到, 在实践中, 对于报告的审查, 通常会消耗一定的资源, 所以抽样比率通常不会太高. 因此, 所有机制设定的固定抽样率为 0.1. 也就是说, 在所有 4 种机制中, 发包方在初始阶段检查了相同数量的微任务.

5 实验结果

本文机制有两个主要目标: 提高净收益并构建诚信和有能力的高质量众包群体. 基于这两个目标, 将评估不同类型的众包工人的群体构成、发包方收益, 以及实施本文机制所需的审查任务的成本.

表 1 50 个任务周期后的工人组成 (20 次实验结果平均值) (基准 3 由于没有剔除工人, 所以工人组成不变)

Table 1 The composition workforce after 50 periods (averaged over 20 runs), report-sample-and-pay is excluded because it does not kick out any workers

机制	总余额	高能力工人数		诚信工人数		理想工作者		成功转化工人数
		运行前	运行后	运行前	运行后	运行前	运行后	
Signal-no-tolerance	353.10	102.40	98.50	181.10	137.10	16.75	16.55	336.55
Report-no-tolerance	17.60	99.65	17.55	183.15	3.30	17.55	3.30	14.30
Our mechanism	393.15	102.40	102.10	180.40	137.10	18.65	18.60	374.55

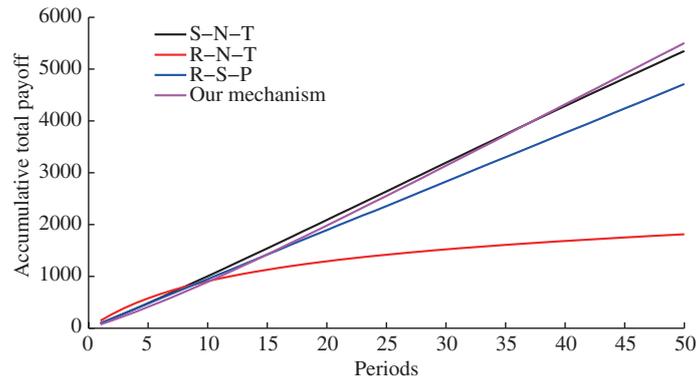


图 4 (网络版彩图) 50 个周期之后发包方支付的报酬总数

Figure 4 (Color online) Requester's accumulative payoff over 50 periods

5.1 不同类型的众包工人群体组成

本研究的主要目标是建立一个能够识别和增加理想众包工人数量的机制, 我们首先考察了不同类型的众包工人的群体动态.

表 1 中的高能力一栏描述了基准 1, 2 以及本文机制²⁾ 中能力值大于 0.75 的工作者数量, 其中“运行前”表示原始工作者群体, “运行后”表示 50 个时期之后的工作者群体. 同样, 诚信工人数一栏中显示诚实度大于 0.85 的工作者数量, “理想工作者”一栏显示诚实度大于 0.85 且能力值大于 0.75 的工作者数量. 最后一栏表明本文机制使得最多的被转化的“有能力的不诚信者”保留在工作者池中.

虽然在实验开始的初期, 这些工作者的诚实并不高, 但在一段时间后, 发包方并没有发现其不诚信行为. 这些具有较低诚信值的工人开始表现得“诚信”, 并像一个理想的工作者在工作. 相比之下, 基准 1 不仅在激励“高能力失信者”的方面效果不佳, 而且更有可能剔除一些理想的工作者. 基准 2 在保留好员工 (和坏员工) 方面效果很差. 在 50 个周期后, 工作者池中只剩下 1.8% 的工人. 随着众包流程的推进, 它几乎将所有员工都赶剔除掉. 显然, 它在保持一定水准的劳动力水平上存在系统性缺陷. 我们总结了在 3 种不同机制下仿真实验前后高能力值和高诚信值的个体数量的变化. 在表 1 中, 很容易得出结论, 基于信号的审核机制 (基准 1 和本文机制) 特别适用于在参与者群体中保留具有较高诚信值的员工. 因此, 如果发包方希望构建高效的众包工人群体, 应该考虑基于信号的机制.

5.2 众包发起者的总支付额

图 4~6 描述了发包方在 50 个周期中的支付动态情况. 其中, 图 4 关注 50 个周期中的累计总支

2) 基准 3 由于没有剔除工人, 所以工人组成不变.

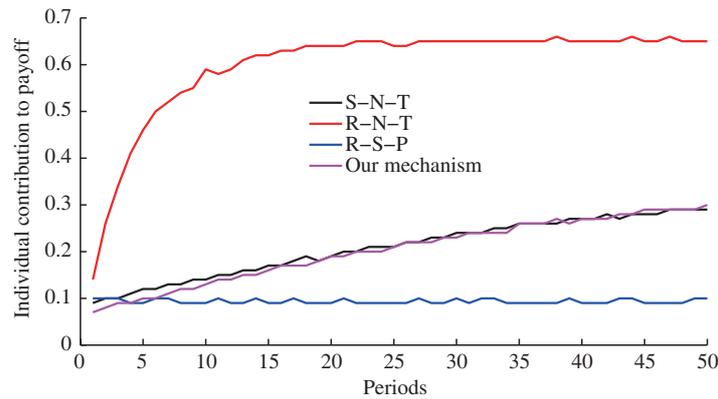


图 5 (网络版彩图) 50 个周期中针对发包方报酬的个人获取量
Figure 5 (Color online) Individual contribution to requester's payoff in each period

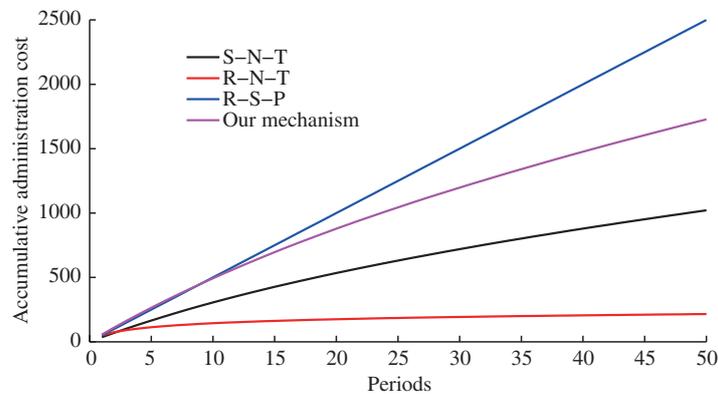


图 6 (网络版彩图) 50 个周期中发包方审核报告的总体消耗
Figure 6 (Color online) Accumulative administration costs

出. 首先, 本文机制优于根据不同行为剔除众包工人的基准 2. 其次, 本文机制比保持 1000 名工人的基准 3 具有更好的效益产出比. 相比于基准 1, 本文所提机制初期性能要差, 但是后续随着不断的任务迭代慢慢胜出.

图 5 描绘了每个周期中每个工人的贡献动态. 本文机制再次优于基准 3. 在后期, 本文机制中的每个工人的贡献是在基准 3 中的 3 倍. 对于基准 2, 本文机制的个人贡献基本持平 (实际上, 后期略高). 虽然基准 2 的确产生了最高的个人贡献, 但这种增长的代价是剔除大多数工人 (见表 1 中的第 2 行), 这最终会损害总体收益. 这表明本文中设计的机制在保留更多“高能力”的员工与避免其失信行为的伤害之间取得了显著平衡. 在本文机制中, 总收益和个体工人的贡献都是稳定的. 这个结果表明了本文机制的另一个优点, 即发包方可以更好地估计预期收益, 这有助于其更好地对众包过程进行管理.

5.3 众包发起者的额外消耗

本文机制的确会产生一些额外费用, 该额外费用来自两个方面: (1) 众包工人的诚实奖励; (2) 任务审查管理费用. 虽然已经证明, 本文机制下总收益明显优于 3 个基准, 但仍然要检查额外支出成本, 从而确保在特定时期内额外成本不会过高, 否则可能会威胁到发包方的支付能力. 图 6 描述了 50 个

周期累计的额外成本. 显然, 本文机制的总成本高于基准 1 和基准 2, 但低于基准 3. 显然, 额外成本不高于不剔除不诚信者的机制. 与基准 1 相比, 总成本增加约 70% (1019.05 单位增加到 1725.50 单位). 此外, 额外成本的增长非常平稳, 因此发包方在任何时期都不会面临突然的支付危机. 我们还计算了本文机制的个性化管理成本. 鉴于本文机制比基准 1 和基准 2 保留了更多的工人, 这种差异就变得微不足道了 (0.06 对 0.05).

6 实验结果讨论

6.1 内在诚信与行为诚信

许多情况下, 内在的诚信必须与社会属性相结合, 以避免在行为层面上欺骗^[23]. 而在目前的众包市场中, 只有内在的诚信才能让众包工人诚信并遵守规则. 心理学中的结论是, 破坏规则和失信行为不仅会产生直接的经济损失, 而且还可能损害个人的内在诚信, 反之亦然^[24, 25]. 这表明, 通过减少行为层面的失信, 可能给众包工人的内在诚信带来积极影响. 如果这种假设成立, 那将有助于在目前缺乏系统性管理的众包市场中建立一个良性价值体系. 因此, 从更广泛的角度来看, 包括我们自己在内的研究人员需要像关注激励众包工人付出更多劳动一样来关注如何激励众包工人的行为诚信.

6.2 投入响应型任务

本文机制的一个假设是, 投入程度不会影响工作质量. 但是, 即使任务是投入响应的, 本文机制仍然有效. 由于我们按照众包工人提交的结果提供激励, 本文机制仍然能够识别和增加那些有意愿贡献更多投入的人. 然而, 在为众包工人建模时, 这需要一个额外的特征来处理愿意付出更多投入的问题. 这种额外特征可能带来副作用, 将会有 $2 \times 2 \times 2 = 8$ 种类型的工作类型, 并且个人的能力和投入之间的相互作用也会非常复杂. 为了验证投入响应任务的有效性, 需要进行进一步的理论和实证分析.

6.3 随机抽样审查与众包工人群体的构建

本文提出利用信号机制构建较好的众包工人群体. 实验中的 3 个基准方法分别模拟了当前使用比较广泛的审查策略. 其中基准 1 模拟的是只对反馈信号为“HQ”的报告, 进行随机抽样审查策略. 该策略的特点是对众包工人的失信行为零容忍, 一旦发现为较低完成质量的报告发送“HQ”信号, 则立即将该众包工人进行剔除. 基准 2 也是一种随机审查策略, 即直接对所有的报告进行抽样审查. 与基准 1 类似, 该策略对众包工人的失信行为是零容忍.

基准 1 与基准 2 的最大区别在于是否利用了信号机制. 不难看到, 如果不利用信号机制, 发包方不可避免需要审查一些诚信工人如实提交的完成质量较低的报告. 如果基于信号机制, 这些报告是可以自动过滤出来的, 并不需要引入额外的人工审查.

我们认为基准 1 与基准 2 具有相似的理论上的缺陷: 他们确实能够很大程度上剔除失信的众包工人, 但这个过程可能同时剔除“高能力失信工人”与“低能力失信工人”. 其中, “低能力失信工人”的剔除对于整个众包工人群体是有益的, 但是“高能力失信工人”在大多数情况下是可以较好地完成任务的. 特别是当众包任务需要较高的领域知识和能力的情况下, 高能力工人可能是非常稀缺的. 同时, 本文实验结果也在一定程度上说明了这一点. 虽然这两种基准具有最低的报告审查成本, 但由于其零容忍的特性, 导致其所获得的高质量完成的报告数量最少. 在实践中, 这会使得整个众包工程任务无法完成, 或不能按时完成. 因此我们认为在实践中这两个基准并不是较为优秀的解决方案.

基准 3 直接采用抽样审查支付的策略, 不对产生“失信行为”的工人进行剔除. 该方法避免了“高能力工人”的流失, 但是不可避免会因为一些“高能力失信工人”的失信行为受到一些损失. 但是从图 6 中可以看到, 这种对于失信行为包含一定容忍度的策略, 会使得发包方需要负担较高的报告审查费用.

与以上 3 种方法相比, 本文中提出的策略展现了一个较为新颖的解决方案.

6.4 众包工人历史信息与诚信群体构建

众包工人的历史行为记录以及背景知识信息对于构建高质量的众包工人群体具有极其重要的意义. 历史行为记录信息可以反映众包工人在过去对于众包任务的完成效率与完成质量等. 当前, 已经有一些众包平台构建了依赖于工人历史行为档案等信息评估众包工人能力以及诚信度的技术^[26, 27]. 然而, 这些技术对于历史行为的记录信息具有较强的依赖. 这种较强的依赖使得它们在实际应用上具有一定的局限性.

首先, 众包工人历史行为的记录信息并不总是可以获取的. 对于一些新加入的众包工人, 他们并没有或只有极其少量的历史行为记录信息. 在这种情况下, 上述的各种技术并不能得到较好的应用. 另一方面, 考虑到实践中众包任务的多样性, 使用众包工人的历史行为记录信息来预测他们对于其他类型众包任务的行为可能并不准确. 因为对于不同类型的众包任务, 众包工人的完成能力并不一样. 例如, 某些众包工人对于图片标注任务非常擅长, 他们可能可以根据要求很好地标注出图片中的目标区域的位置. 如果他们对于文本分析任务就不是那么擅长, 那么他们在图片标注任务中的出色表现, 并不能保证他们对于文本分析任务也能较好完成.

本文中提出的方法并不依赖于众包工人的历史行为记录. 在众包工人的历史行为记录信息较难获得的情况下, 本文所提方法能够帮助众包任务发起者构建兼具诚信和较高能力的众包工人群体. 这种不依赖于众包工人的历史行为记录的群体构建技术, 具有更好的实用性.

7 相关工作

本文涉及两个不同的领域: 劳动经济学中的最优契约设计, 众包支付方案的设计和运营. 下面将概述我们的工作与这两个领域的联系和差异.

7.1 最优契约设计

研究者已经对劳动经济学中关于最优契约的设计进行了深入的研究^[28]. 显然, 我们的工作可以被视为对经典委托-代理模型的延伸和适配. 正如在第 2 节中提到的那样, 不考虑群体工作者的不可观察的投入水平. 因为对于发包方和工作者而言, 工作绩效 (工作质量) 更明确, 可以基于共同认知动态地建立特定的合作关系. 本文机制也适应文献^[29]中的思想, 其中委托人只能观察基于可量化的工作质量产生的总结果.

多臂老虎机 (MAB) 和动态定价也被用于众包机制设计领域. MAB 问题关注于: 代理反复地选择不确定报酬的行为时必须考虑探索 (学习支付的细节) 和利用 (最大化获得的报酬). 经营研究人员和经济学家对此进行了深入的研究. Gittins 等^[30]概述了这类模型. 在众包的最优合同设计中, 有大量的工作运用了 MAB 模型, 如 Biswas 等^[31]、Frazier 等^[32]、Ho 等^[11]、Jain 等^[33]、Tran-Thanh 等^[34], 以及 Singla 和 Krause^[10]. 我们没有应用 MAB 框架, 因为我们更关注于对工人与请求者之间信息交换的影响. 动态定价处理中委托人可以随时为可用的工人提供不同的合同. 在众包中, 它通过工人可以选择接受或拒绝请求者提供的合同来实现^[35].

对于在线市场, Badanidiyuru 等^[36]提出了一种动态定价的策略. 该策略中, 委托人的定价策略主要是基于分析发包方对于微任务的完成意愿. 然而, 大多数在线市场的动态定价研究更关注的是将委托人的预算限制在特定的水平^[37]. 由于本文的研究需要讨论发包方采用不同的机制所需要负担的资源消耗, 因此本文没有执行这样的限制. 同时, 可以从本文实验结果中看到, 我们的方法在合理的资源限制之下, 都是优于基准方法的.

最近, 研究人员也开始运用一些行为经济学理论来设计众包市场中的最优契约, 从而缓解工人经常偏离标准经济理性假设的事实. 例如, Easley 和 Ghosh^[38]建议在众包机制设计中使用行为经济学理论, 特别是前景理论. 我们计划在未来的研究中, 通过设计允许替代工人的信念 - 偏好 - 约束模型^[39]和进行经验行为实验的机制来探索众包工人偏好中的异质性如何影响本文机制的有效性和效率.

7.2 众包支付方案

近年来已经有越来越多的工作关注于理解在线劳动力市场中工人的正确支付方案. Horton 和 Chilton^[2]提出了一个工人模型, 并引入了估算工人适当工资的方法. 文献 [40] 中介绍的 “hagglebot” 能够与 Amazon MTurk 上的工作人员协商图像标签任务的报酬率. 在现实世界的众包市场中, 发包方可能有一定的预算限制. 为了解决这个问题, Sigla 和 Krause^[10]设计的支付方案可以最大限度地提高预算发包方的效用, 同时保持与工人的激励相容.

还有一些实证研究关注于工人的行为如何基于所获经济激励而变化. 例如, Manson 和 Watt^[41]研究了与绩效独立的经济激励对员工行为的影响. 在他们的研究中, 增加经济激励会增加工人完成任务的数量, 但不会增加工作质量. Shaw 等^[42]认为, 为了提高工作质量, 经济激励可能需要与其他非货币机制相结合. 一般而言, 经济激励是否对提交的微任务的质量产生积极影响在很大程度上取决于任务的背景和设置. 例如, Harris^[43]的 MTurk 简历筛选实验表明, 基于绩效支付的工作质量优于统一支付. 另一方面, 近年来的一些研究^[44]建议批量支付 (而不是每个微任务) 可以鼓励工人的参与, 这表明研究众包工人对支付方案的认知或解读应该是一个有前景的方向. 总而言之, 现有文献表明众包市场中的众包工人的确会对经济激励的变化作出反应, 但是其方式存在一定差异. 本文工作与以前的研究先比, 我们试图使用经济激励或惩罚作为干预手段来识别和增加 “理想” 众包工人的数量, 同时提供一种将信息从众包工人传递到发包方的方法.

8 总结

本文设计了一种机制在在线众包市场中构建高可信众包群体. 我们将经典的委托 - 代理模型与信号博弈相结合, 从机制上允许发包方与众包工人的信息交换. 本文进行了大规模的仿真实验. 实验结果表明本文机制的有效性和效率, 同时不会产生过多的额外成本. 本文中设计的机制是可以帮助发包方提升有能力和高诚信众包工人的数量. 对于提高众包的有效性和效率, 本文机制为开发实用且易于实现的技术迈出了重要的一步.

对于未来的工作, 我们计划在流行的众包平台 Amazon MTurk 上进行随机行为实验, 通过实际应用验证本文机制在现实众包市场中的有效性和效率. 同时也可以帮助我们更好地理解所提机制何时、何地, 以及为何有助于构建一个高可信众包群体.

在本研究中, 我们为请求者提出了一个简单的附加效用函数, 未来将扩展到更复杂和更现实的效用函数中, 例如, 边际效用递减. 此外, 尽管我们确实考虑了个体工人在能力和诚信度方面的差异, 但设计考虑工人的不同效用价值及其偏差的机制会更实用.

参考文献

- 1 Bar A J, Maheswaran M. Confidentiality and Integrity in Crowdsourcing Systems. Berlin: Springer, 2013
- 2 Horton J J, Chilton L B. The labor economics of paid crowdsourcing. In: Proceedings of the 11th ACM Conference on Electronic Commerce, 2010. 209–218
- 3 Slivkins A, Vaughan J W. Online decision making in crowdsourcing markets: theoretical challenges. SIGecom Exch, 2014, 12: 4–23
- 4 Becker T E. Integrity in organizations: beyond honesty and conscientiousness. Acad Manage Rev, 1998, 23: 154–161
- 5 Dineen B R, Lewicki R J, Tomlinson E C. Supervisory guidance and behavioral integrity: relationships with employee citizenship and deviant behavior. J Appl Psychol, 2006, 91: 622–635
- 6 Kidder D L. Is it ‘who I am’, ‘what I can get away with’, or ‘what you’ve done to me’? A multi-theory examination of employee misconduct. J Bus Ethics, 2005, 57: 389–398
- 7 Zhang Y, Schaar M. Reputation-based incentive protocols in crowdsourcing applications. In: Proceedings of International Conference on Computer Communications, 2012. 2140–2148
- 8 Ho C J, Slivkins A, Suri S, et al. Incentivizing high quality crowdwork. In: Proceedings of the 24th International Conference on World Wide Web, 2015. 419–429
- 9 Singer Y, Mittal M. Pricing mechanisms for crowdsourcing markets. In: Proceedings of the 22nd International Conference on World Wide Web, 2013. 1157–1166
- 10 Singla A, Krause A. Truthful incentives in crowdsourcing tasks using regret minimization mechanisms. In: Proceedings of the 22nd International Conference on World Wide Web, 2013. 1167–1178
- 11 Ho C J, Slivkins A, Vaughan J W. Adaptive contract design for crowdsourcing markets: bandit algorithms for repeated principal-agent problems. In: Proceedings of the 5th ACM Conference on Economics and Computation, 2014. 359–376
- 12 Laffont J J, Martimort D. The Theory of Incentives: the Principal-agent Model. Princeton: Princeton University Press, 2009
- 13 Grossman S J, Hart O D. An analysis of the principal-agent problem. Econometrica, 1983, 51: 7–45
- 14 Skyrms B. Signals: Evolution, Learning, and Information. Oxford: Oxford University Press, 2010
- 15 Cateeuw D, Han T A, Manderick B. Evolution of honest signaling by social punishment. In: Proceedings of Annual Conference on Genetic and Evolutionary Computation, 2014. 153–160
- 16 Huttegger S, Skyrms B, Tarres P, et al. Some dynamics of signaling games. In: Proceedings of National Academy of Sciences of the United States of America, 2014. 10873–10880
- 17 Ones D S, Viswesvaran C. Gender, age, and race differences on overt integrity tests: results across four large-scale job applicant datasets. J Appl Psychol, 1998, 83: 35–42
- 18 Mayer T. The distribution of ability and earnings. Rev Econ Stat, 1960, 42: 189–195
- 19 Ashton M C, Lee K, Son C. Honesty as the sixth factor of personality: correlations with machiavellianism, primary psychopathy, and social adroitness. Eur J Pers, 2000, 14: 359–368
- 20 Train K E. Discrete Choice Methods with Simulation. Cambridge: Cambridge University Press, 2009
- 21 Eppig C, Fincher C L, Thornhill R. Parasite prevalence and the worldwide distribution of cognitive ability. Proc R Soc B, 2010, 277: 3801–3808
- 22 Cowell F. Measuring Inequality. Oxford: Oxford University Press, 2011
- 23 Gächter S, Schulz J F. Intrinsic honesty and the prevalence of rule violations across societies. Nature, 2016, 531: 496–499
- 24 Mazar N, Amir O, Ariely D. The dishonesty of honest people: a theory of self-concept maintenance. J Market Res, 2008, 45: 633–644
- 25 Shalvi S, Dana J, Handgraaf M J, et al. Justified ethicality: observing desired counterfactuals modifies ethical perceptions and behavior. Organ Behav Hum Decis Process, 2011, 115: 181–190
- 26 Vakharia D, Lease M. Beyond mechanical turk: an analysis of paid crowd work platforms. In: Proceedings of the iConference, 2015
- 27 Whiting M E, Gamage D, Gaikwad S S, et al. Crowd guilds: worker-led reputation and feedback on crowdsourcing platforms. In: Proceedings of Conference on Computer Supported Cooperative Work and Social Computing, 2017. 1902–1913
- 28 Bolton P, Dewatripont M. Contract Theory. Cambridge: MIT Press, 2005

- 29 Babaioff M, Feldman M, Nisan N. Combinatorial agency. In: Proceedings of the 7th ACM Conference on Electronic Commerce, 2006. 18–28
- 30 Gittins J, Glazebrook K, Weber R. Multi-armed Bandit Allocation Indices. Hoboken: John Wiley and Sons, 2011
- 31 Biswas A, Jain S, Mandal D, et al. A truthful budget feasible multi-armed bandit mechanism for crowdsourcing time critical tasks. In: Proceedings of International Conference on Autonomous Agents and Multiagent Systems, 2015. 1101–1109
- 32 Frazier P, Kempe D, Kleinberg J, et al. Incentivizing exploration. In: Proceedings of the 5th ACM Conference on Economics and Computation, 2014. 5–22
- 33 Jain S, Narayanaswamy B, Narahari Y. A multiarmed bandit incentive mechanism for crowdsourcing demand response in smart grids. In: Proceedings of National Conference on Artificial Intelligence, 2014. 721–727
- 34 Tran-Thanh L, Stein S, Rogers A, et al. Efficient crowdsourcing of unknown experts using multi-armed bandits. In: Proceedings of European Conference on Artificial Intelligence, 2012. 768–773
- 35 den Boer A V. Dynamic pricing and learning: historical origins, current research, and new directions. *Surv Oper Res Manage Sci*, 2015, 20: 1–18
- 36 Badanidiyuru A, Kleinberg R, Singer Y. Learning on a budget: posted price mechanisms for online procurement. In: Proceedings of the 13th ACM Conference on Electronic Commerce, 2012. 128–145
- 37 Kleinberg R, Leighton T. The value of knowing a demand curve: bounds on regret for online posted-price auctions. In: Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science, 2003. 594–605
- 38 Easley D, Ghosh A. Behavioral mechanism design: optimal crowdsourcing contracts and prospect theory. In: Proceedings of the 6th ACM Conference on Economics and Computation, 2015. 679–696
- 39 Herbert Gintis. *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*. Princeton: Princeton University Press, 2009
- 40 Horton J J, Zeckhauser R J. Algorithmic wage negotiations: applications to paid crowdsourcing. In: Proceedings of CrowdConf, 2010
- 41 Mason W, Watts D J. Financial incentives and the “performance of crowds”. *SIGKDD Explor Newsl*, 2010, 11: 100–108
- 42 Shaw A D, Horton J J, Chen D L. Designing incentives for inexpert human raters. In: Proceedings of Conference on Computer Supported Cooperative Work, 2011. 275–284
- 43 Harris C. You’re hired! an examination of crowdsourcing incentive models in human resource tasks. In: Proceedings of Workshop on Crowdsourcing for Search and Data Mining, 2011. 15–18
- 44 Ikeda K, Bernstein M S. Pay it backward: per-task payments on crowdsourcing platforms reduce productivity. In: Proceedings of CHI Conference on Human Factors in Computing Systems, 2016. 4111–4121

An approach for developing a highly trustworthy crowd-sourced workforce

Yang FENG¹, Yi WANG², Chunrong FANG^{1*}, Nannan GUO¹ & Zhenyu CHEN¹

1. *State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China;*

2. *Department of Software Engineering, Rochester Institute of Technology, Rochester 14623, USA*

* Corresponding author. E-mail: fangchunrong@nju.edu.cn

Abstract In applying crowd-sourcing techniques, one of the most critical challenges is building a crowd workforce that is both capable and trustworthy. Previous studies proposed numerous strategies, methods, and mechanism to motivate individuals; however, although the results improved the effectiveness and efficiency of finishing crowd-sourcing tasks, few studies focused on improving the honesty of crowd-sourced workers and assisting requesters in obtaining the correct quality report. To address this, based on the principal-agent model and signaling game theory, we design a novel mechanism for building a capable and trustworthy crowd-sourced workforce. This mechanism enables information exchange between crowd-sourced workers and requesters, and leverages a random inspection strategy to assign financial incentives/punishments to honest/dishonest behaviors accordingly. To validate our mechanism, we conduct an extensive simulation. The results show this mechanism is effective and efficient to motivate workers to behave in a trustworthily manner and capable of changing the behavior of dishonest workers with minimal extra cost.

Keywords trustful crowd-sourcing, crowd-sourced mechanisms, signaling game theory, principal-agent model



Yang FENG was born in 1988. He received his M.S. degree in software engineering from Nanjing University in 2013. Currently, he is a Ph.D. candidate at Nanjing University. His research interests include crowd-sourced software engineering, software testing, software understanding, and AI testing.



Yi WANG was born in 1984. He received his Ph.D. degree in information and computer science from the University of California at Irvine in 2015. He is currently an assistant professor at Rochester Institute of Technology. His research interests include human and social aspects of software development.



Chunrong FANG was born in 1986. He received his Ph.D. degree in software engineering from Nanjing University in 2017. He is currently a research assistant at Nanjing University. His research interests include testing automation collective intelligence, software understanding, and AI testing.



Zhenyu CHEN was born in 1978. He received his Ph.D. degree in mathematics from Nanjing University in 2006. He is currently a full professor at Nanjing University. His research interests include testing automation, collective intelligence, crowd-sourced data analytics, and AI testing.