

■ 个体理性:每个赢得拍卖的买家或卖家都有非负的效益.也就是对于赢得拍卖的买卖双方来说,买家需要付出的费用不会超过他的出价,卖家得到的报酬不会低于他的报价.因为参与拍卖的请求者和工人,分别会给出自己对任务的出价和报价.对于请求者的出价,是愿意为分配任务出的最大价值,而对于工人的报价,是工人愿意完成任务的最小报酬.

■ 收支平衡:在一个双向拍卖机制中,拍卖商的效益不低于 0,那么这个双向拍卖机制就是收支平衡的.在本文考虑的模型中,平台的效益记为

$$\varphi = \sum_{i \in R} P_i^R - \sum_{j \in W} P_j^W \quad (2)$$

其中, P_i^R 表示对请求者 i 的收费, P_j^W 表示 w_j 获得的报酬.若平台的效益 φ 不低于 0,则保证了收支平衡特性.

2 满足工人多任务需求的双向拍卖机制

首先给出双向拍卖机制的设计细节,主要包括 3 个部分:构建任务分组 G ,选择请求者集 $R, \subseteq R$ 、工人集 W ,以及收费与报酬的定价计算;然后,对所设计的拍卖机制进行理论分析,验证本文的拍卖机制:在满足工人多任务需求的前提下,同时也满足 3 条经济特性.

2.1 构建任务分组

本文满足工人的多任务需求且申请每一任务成功与否互不影响.若考虑一个工人只能申请一个任务的情况,那么对于申请多个感兴趣任务的工人,则可以看作是自身的多个虚拟工人.按照这样的思路,本文把每个工人根据自身感兴趣任务集抽象为多个虚拟工人,并且以工人其他的每个感兴趣任务而区分(如,工人 w_1 的任务计划 $T_1 = \{t_1, t_3, t_5\}$, 则用 3 个虚拟工人 $\{w_1^1, w_1^3, w_1^5\}$ 表示工人 w_1).这样,虚拟工人 w_j^i 表示工人 w_j 申请了任务 t_i , 以下用 W' 表示虚拟工人集.接下来,我们开始构建任务分组 G , 如图 1 所示. 首先将 R 中的请求者按出价非升次序排序, 使得 $r_1 \geq r_2 \geq \dots \geq r_m$; 然后把排序后的每个请求者 i 对应到任务 t_i , 把所有申请了 t_i 的虚拟工人 w_j^i 加入分组 G_i 和 W' 中; 最后得到了 m 个任务分组和 W' , 对不空的 G_i 中所有虚拟工人按报价非降次序排序, 使得 $b_{1,j} \leq b_{2,j} \leq \dots \leq b_{q,j}$, 至此完成对 G 的构建, 具体算法描述见算法 1.

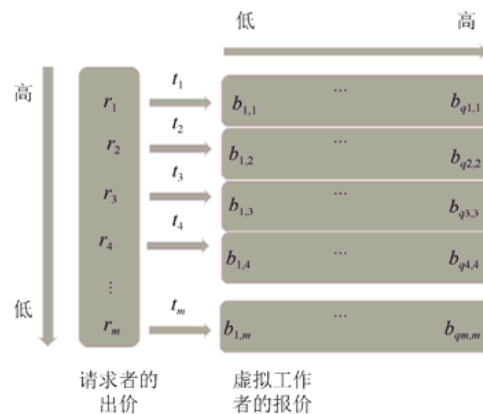


Fig.1 The construction of task groups

图 1 任务分组的构建模型

2.2 分配和定价

G 构建完成后,得到了 m 个分组,请求者 i 对应分组 G_i ,并且在 G_i 中虚拟工人按 $b_{j,i}$ 非降次序排序.接下来,本文将通过分配算法为平台选择可以赢得拍卖的请求者和工人.本文采用高出价低报价优先匹配的原则,为请求者和工人进行匹配,主要包括两个步骤.第 1 步,若在请求者 i 对应的任务分组 G_i 中,满足不大于出价 r_i 和任务规定数 f_i 条件的虚拟工人个数不超过 1,那么请求者 i 请求的任务没有工人申请完成,不能进行任务匹配;反之,

将 i 加入 R_s 中.最后从 R_s 中删除出价最低的请求者 s ,并将 s 的出价 r_s 作为选择虚拟工人的一个约束条件.

第 2 步,在 R_s 中的请求者 i 对应的分组 G_i 中,按照 i 的任务规定数 f_i 以及出价限制 r_s ,若 G_i 中所在序列位置不超过 f_i 且报价不超过 r_s 的虚拟工人个数超过 1,则将最后一个满足条件 w_q^i 之前的虚拟工人加入 W'_s ;否则,从 R_s 中删除请求者 i .

在上面的两个步骤中,通过第 1 步我们确定了请求者 s 和 R_s ,其中我们把 s 称作为关键任务请求者(critical requester);通过第 2 步,我们更新了 R_s ,并得到 W'_s .于是,我们的任务分配就结束了.接下来,对 R_s 、 W'_s 中的用户分别给出收费和报酬的定价计算.

本文采用统一定价技术,对 R_s 中的请求者与 W'_s 中的虚拟工人进行定价计算.由于本文在构建 G 时,把每个 w_j 按任务集 T_j 抽象为多个虚拟工人 w_j^i ,并以 $t_i \in T_j$ 区分抽象得到的多个虚拟工人,那么 w_j 的报酬,就是分配结束后他抽象出的所有 w_j^i 所得报酬 $p_{j,i}$ 的总和.对于分配算法中,本文把每个可以赢得拍卖的 w_j^i 的报酬计为分组中最后一个可以赢得拍卖的 w_q^i 的报价 $b_{q,i}$,再令 w_q^i 输掉拍卖,也就是, w_q^i 对分组所表示的任务 t_i 申请失败.对于没有赢得拍卖的虚拟工人,报酬计为 0.本文对每个赢得拍卖的请求者的收费 r_s ,计为在首轮分配中找到的关键请求者的出价 r_s ,也就是把最后一个赢得拍卖而令其输掉拍卖的请求者的出价作为收费标准.对于没有赢得拍卖的请求者,平台不对他进行收费.于是,得到定价公式(3).

到这里,我们完成了对请求者与工人的分配和定价,具体算法步骤参见算法 2.

$$P_j^W = \sum_{t_i \in T_j} x_{j,i} b_{q,i}, x_{j,i} \in \{0,1\}, p_i^R = r_s \quad (3)$$

算法 1. 构建任务分组 G .

输入:请求者集合 R ,工人集合 W ;

输出:任务分组 G .

- ① $W' = \emptyset, G = \emptyset$;
- ② 对所有请求者按出价非升次序排序,使得 $r_1 \geq r_2 \geq \dots \geq r_m$;
- ③ for each $w_j \in W$
- ④ for each $t_i \in T_j$
- ⑤ $W' = W' \cup w_j^i$;
- ⑥ $G_i = G_i \cup w_j^i$;
- ⑦ end for
- ⑧ end for
- ⑨ for each $G_i \in G, i \in R$
- ⑩ 对 G_i 中 w_j^i 以报价非降次序排序,使得 $b_{1,i} \leq b_{2,i} \leq \dots \leq b_{q,i}$;
- ⑪ end for
- ⑫ return G, W' ;

算法 2. 选择赢得拍卖的用户,并为其定价.

输入:请求者集合 R ,工人集合 W, W' ,任务分组 G ;

输出:赢得拍卖的请求者集 R_s ,赢得拍卖的虚拟工人集 W'_s ,收费集 $\{p_i\}_{i \in R}$ 以及报酬集 $\{p_{j,i}\}_{w_j \in W, t_i \in T_j}$.

- ① $R_s = \emptyset, W'_s = \emptyset$;
- ② for each $G_i \in G$
- ③ if 报价不超出 r_i 的元素个数大于 1
- ④ $R_s = R_s \cup \{i\}, s = i$;
- ⑤ end if
- ⑥ $R_s = R_s \setminus \{s\}$;

```

⑦ end for
⑧ for each  $G_i \in G, i \in R_s$ 
⑨   set  $k=0$ ;
⑩   while  $k < G_i$  的长度 &&  $k < f_i$  &&  $b_k^i \leq r_s$ 
⑪      $k++$ ;
⑫    $k--$ ; /* $k$  表示满足条件的边界用户*/
⑬   if  $k \geq 1$ 
⑭     set  $x_{k,i}=0$ ;
⑮     对前  $k-1$  个用户设置  $x_{j,i}=1$ , 并将其加入  $W'_s$ , 设置  $p_{j,i} = b_k^i, p_i = r_s$ ;
⑯   else  $R_s = R_s \setminus \{i\}$ ;
⑰   end if
⑱ end for
⑲ return  $R_s, W'_s, \{p_i\}_{i \in R_s}, \{p_{j,i}\}_{w_j \in W, i_j \in T_j}$ ;

```

2.3 理论分析

本文提出的双向拍卖机制,满足工人多任务需求性的同时,保证机制满足 3 条经济特性,更符合众包实际应用场景.下面,将给出本文机制的诚实性、个体理性及收支平衡的证明.

本文所设计机制的诚实性主要体现在两个方面:一方面,对于请求者是诚实的;另一方面,对于参与拍卖的工人是诚实的.首先证明本文机制对于请求者的诚实性,我们引入两条定理.

定理 1. 给定包括出价信息的请求者集合以及包括报价信息的工人集合,若请求者 i 以出价 r_i 赢得了拍卖,那么当 i 出价 $r_i' > r_i$ 时,他也一定会赢得拍卖.

由于本文将请求者按出价非升序排列,因此,若 i 提高自己的出价为 r_i' ,不会降低自己在序列中的位置.同时, i 对应的任务分组 G_i 内虚拟工人的报价不变,这样满足出价条件的虚拟工人不会减少,于是当请求者提高了报价后,也一定会赢得拍卖.

定理 2. 给定包括出价信息的请求者集合和包括报价信息的工人集,若请求者 i 以出价 r_i 和 r_i' 赢得了拍卖,那么对请求者 i 的收费相同,或者说, i 不会通过降低自己的出价而减少收费.

由于本文对请求者 i 的定价依赖于首轮分配中的 critical requester s .若 i 以出价 r_i 和 r_i' 均赢得了拍卖,那么平台对 i 的收费相同,均为 r_s .

其次我们需要证明本文提出的机制对于工人是诚实的.也就是说,诚实报价是工人的占优策略,即不诚实报价带来的效益没有诚实报价带来的效益高.而对本文的模型,只需证明对虚拟工人是诚实的.对于参与拍卖的虚拟工人 w_j^i ,我们可以得到如下 4 种关于能否赢得拍卖的报价情况分析.

1) w_j^i 无论诚实或不诚实报价, w_j^i 都会赢得拍卖.由于本文把 w_j^i 所在任务分组 G_i 内最后一个满足条件 (critical requester s 的出价条件及 i 任务规定数条件)且报价最高的虚拟工人的报价 $b_{q,i}$ 作为 w_j^i 的报酬 $p_{j,i}$,所以,即使 w_j^i 诚实或不诚实报价都在赢得拍卖的情况下, w_j^i 的报酬均为 w_q^i 的报价.对于诚实报价而赢得拍卖即真实估价小于报酬,得到的效益大于 0;对于不诚实报价而赢得拍卖,若真实估价大于报酬,效益为负,否则效益仍为报酬减去真实估价.

2) w_j^i 诚实报价赢得拍卖,不诚实报价输掉拍卖.根据我们的分配机制,诚实报价赢得拍卖,说明诚实报价是小于 r_s 的,而得到的报酬一定会大于诚实报价,所以诚实报价可以得到大于 0 的效益;而不诚实报价输掉拍卖,效益为 0.所以这种情况下,诚实报价效益最高.

3) w_j^i 诚实报价输掉拍卖,不诚实报价赢得拍卖.根据我们的分配机制,若诚实报价而输掉拍卖,那么 w_j^i 的诚实报价大于 r_s ,效益为 0;当 w_j^i 降低对任务的估价进行不诚实报价,赢得了拍卖,于是得到序列靠后的 w_q^i 的报

价作为报酬.这种情况下, w_j^i 诚实报价可能刚好作为 w_q^i 而输掉拍卖,那么 w_j^i 的真实估价不小于 w_q^i 的报价即赢得拍卖的报酬.因而不诚实报价,即使赢得拍卖,也会导致 w_j^i 的效益非正.

4) 无论诚实或不诚实报价,均输掉拍卖.在这两种情况下,无论虚拟工人怎样报价都会输掉拍卖,那么得到的效益都为 0.

结合上面 4 点分析,我们可以得到,诚实报价是虚拟工人的占优策略.接下来,我们看一下本文如何体现个体理性以及收支平衡.

根据前文可知,若对请求者的收费不高于其出价而对虚拟工人的报酬不少于其报价,那么本文的机制满足个体理性.由于本文对赢得拍卖的任务请求者的收费为 r_s ,不会超过他们的出价,而对赢得拍卖的虚拟工人的报酬为 $b_{q,i}$ 不会低于其报价,因而,本文机制满足个体理性.最后,由于本文对请求者单任务的收费为 r_s ,而对工人单任务的报酬不高于 r_s ,所以平台得到的效益是不小于 0 的.因而,本文提出的机制保证了收支平衡.

3 仿真模拟与分析

本文提出了一个诚实的双向拍卖机制,并满足工人的多任务需求.在实验仿真部分,实验首先对平台单任务的收益进行分析;其次将从工人满意度与任务完成率(或者成功匹配率)两个角度评定实验效果.为了体现本文算法的优势,本文通过与文献[16]提出的方法(double auction task assignment,简称 DATA)进行对比实验.在我们的实验设置中,请求者的出价均匀分布在区间[50,249],工人的报价均匀分布在区间[1,100].所有的仿真结果均取 1 000 组独立运行结果的算术平均.

3.1 度量标准

本文将工人满意度、任务完成率、平均任务完成率的定义用公式(4)~公式(6)表示.

$$\text{工人满意度} = \frac{\text{有任务可做的人数}}{\text{总工人数}} \quad (4)$$

$$\text{任务完成率} = \frac{\text{分组内赢得拍卖的虚拟工人个数}}{\text{任务规定数}} \quad (5)$$

$$\text{平均任务完成率} = \frac{\text{请求者的任务完成率总和}}{\text{赢得拍卖的请求者个数}} \quad (6)$$

3.2 平台单任务收益分析

在本文提出的双向拍卖机制中,平台作为拍卖商,实现任务交易过程.在这个过程中,平台向成功的请求者收取费用,按单价乘以被做任务数计算;并且平台付给工人报酬.由于每个工人被抽象为多个虚拟工人,因而每个工人的报酬,就是其对应赢得拍卖的虚拟工人获得的报酬总和.如图 2 所示,观察当工人数从 100 变化到 1 000,间隔为 100 时,平台对单任务的平均收费与平均报酬之间的变化关系,进而分析平台的单任务收益.

从图 2 中可以看到,当平台的请求者人数一定时,随着工人数量的增加,平台对请求者单任务平均收费基本处于稳定状态,而对工人单任务的平均报酬数值逐渐下降.这是因为,当请求者数量一定,工人数量增加时,供不应求,工人之间竞争强烈,因而工人的报酬逐渐下降.因此,当工人数量逐渐增大时,平台的单任务收益会越来越高.

3.3 满意度对比分析

如图 3 所示,当请求者数量从 100 变化至 1 000 时,两个算法得到的工人满意度最初随着请求者数量的增加而逐渐上升,而最终趋于平缓;并且本文提出的算法得到的工人满意度较早趋于平缓并高于 DATA 算法的.这是因为随着请求者数目的增加,每个工人被分配到任务的可能性增大,因而,在两个算法中,工人满意度均呈现上升趋势;当请求者增加到可以满足所有工人需求时,工人满意度不再上升.又因为本文考虑工人的多任务需求,而 DATA 仅考虑工人的单任务需求,因而本文工人的满意度高于 DATA.

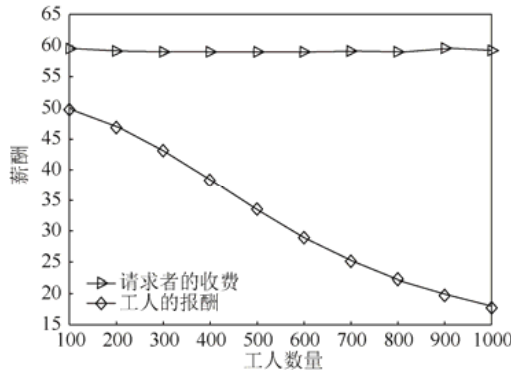


Fig.2 The performance of payment under different number of workers

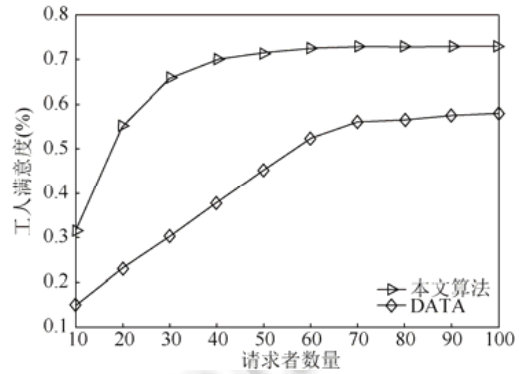


Fig.3 The performance of worker satisfaction ratio under different number of requesters

图 2 不同工人数量对请求者收费与工人报酬的影响 图 3 不同请求者数量对两个算法的工人满意度影响

在图 4 中,观察工人数从 100 变化至 1 000 时两个算法得到的工人满意度的变化趋势.从图 4 中,我们可以看到,随着工人数量的增加,两个算法的工人的满意度均呈下降趋势且最终趋于平缓;并且本文算法产生的工人满意度一直高于 DATA 的.这是因为,当工人数量增加,任务供不应求,工人之间竞争激烈,因而其满意度呈现下降趋势;又因为本文算法考虑到工人多任务需求,尽可能使更多的工人满足任务需求,而 DATA 只考虑了工人的单任务,因而本文算法带来的工人满意度高于 DATA 算法的.

通过满意度对比分析,可以发现,我们的算法更适用于工人数较多的平台,从而避免分配不足的现象.

3.4 任务完成率分析

在图 5 中,我们通过比较在请求者的任务规定数限制在 50 以内和 100 以内的两种情况下,随着工人数量的增加,请求者的平均任务完成率的变化规律,来分析本文机制的实用性能.从图 5 中,我们可以看到,随着工人的数量增加,两种情况下的平均任务完成率均呈现上升趋势,且在初期,规定数为 50 情况下的任务完成率上升更快并高于规定数 100 情况下的;当工人数量饱和时,两种情况下的任务完成率均不再增加而趋于平缓,并且前者比后者更早趋于平缓.这是因为,请求者的任务完成率与满足条件的工人个数、任务规定数有关,当工人数量较少时,任务规定数越大的请求者对满足条件的工人的需求越大,因而任务完成率低于任务规定数较小的请求者,并且随着工人的数量增加均呈现上升趋势;当工人数量充足时,请求者的任务需求都尽可能被满足,并且规定数小的请求者更先于被满足,因而,两种情况下的任务完成率最终趋于平缓并且任务完成率高达 93%.

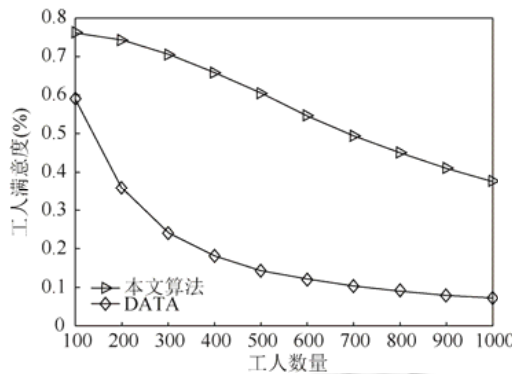


Fig.4 The performance of worker satisfaction ratio under different number of workers

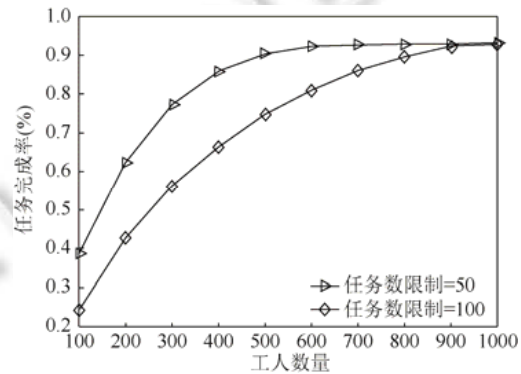


Fig.5 the performance of task completion ratio under different number of workers and task limit

图 4 不同工人数量对两个算法中工人满意度的影响

图 5 不同工人数量以及不同最大任务规定数限制对任务完成率的影响

4 总 结

基于双向拍卖的多任务分配模型,本文提出了一种满足工人多任务需求的双向拍卖机制.该机制在请求者限定任务数的基础上,通过双向拍卖,实现请求者与工人之间的任务交易过程,保证了参与者的诚实性,在激励用户的同时满足用户的高参与度并且保证了高达 93%的任务完成率,具有实用价值.

References:

- [1] Howe J. The rise of crowdsourcing. *Wired*, 2006,14(14):1–5.
- [2] <https://www.mturk.com/>
- [3] Brabham DC. Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence the Int'l Journal of Research Into New Media Technologies*, 2008,14(1):75–90.
- [4] Rai A, Chintalapudi KK, Padmanabhan VN, *et al.* Zee: Zero-Effort crowdsourcing for indoor localization. In: *Proc. of the Mobicom 2012*. 2012. 293–304.
- [5] Ballesteros J, Rahman M, Carbunar B, *et al.* Safe cities. A participatory sensing approach. *Local Computer Networks, IEEE*, 2012. 626–634.
- [6] Rana RK, Chou CT, Kanhere SS, *et al.* Ear-phone: An end-to-end participatory urban noise mapping system. In: *Proc. of the ACM/IEEE Int'l Conf. on Information Processing in Sensor Networks (IPSN)*. 2010. 105–116.
- [7] Deng J, Dong W, Socher R, *et al.* Imagenet: A large-scale hierarchical image database. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. 2009. 248–255.
- [8] Torralba A, Fergus R, Freeman WT. 80 million tiny images: A large data set for nonparametric object and scene recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2008,30(11):1958–1970.
- [9] Demartini G, Difallah DE, Cudré-Mauroux P. ZenCrowd: Leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking. In: *Proc. of the Int'l Conf. on World Wide Web. ACM*, 2012. 469–478.
- [10] <https://www.nasa.gov/>
- [11] Munro R. Crowdsourcing and the crisis-affected community. *Information retrieval*, 2013,16(2):210–266.
- [12] Lebraty JF, Lobre-Lebraty K. *Crowdsourcing: One Step Beyond*. Wiley-ISTE, 2013. 1–119.
- [13] Ye H, Kankanhalli A. Solvers' participation in crowdsourcing platforms: Examining the impacts of trust, and benefit and cost factors. *Journal of Strategic Information Systems*, 2017,26(2).
- [14] Zhang X, Xue G, Yu R, *et al.* Truthful incentive mechanisms for crowdsourcing. In: *Proc. of the Computer Communications (COMCOM). IEEE*, 2015. 2830–2838.
- [15] Yang D, Xue G, Fang X, *et al.* Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing. In: *Proc. of the Int'l Conf. on Mobile Computing and Networking (Mobicom 2012). ACM*, 2012. 173–184.
- [16] Xu W, Huang H, Sun YE, *et al.* DATA: A double auction based task assignment mechanism in crowdsourcing systems. In: *Proc. of the Int'l ICST Conf. on Communications and Networking*. 2014. 172–177.
- [17] Feng Z, Zhu Y, Zhang Q, *et al.* Towards truthful mechanisms for mobile crowdsourcing with dynamic smartphones. In: *Proc. of the IEEE ICDCS*. 2014. 11–20.
- [18] Feng Z, Zhu Y, Zhang Q, *et al.* TRAC: Truthful auction for location-aware collaborative sensing in mobile crowdsourcing. In: *Proc. of the IEEE INFOCOM*. 2014. 1231–1239.
- [19] Wei Y, Zhu Y, Zhu H, *et al.* Truthful online double auctions for dynamic mobile crowdsourcing. In: *Proc. of the IEEE INFOCOM*. 2015. 2074–2082.
- [20] Qin H, Zhang Y, Li B. Truthful mechanism for crowdsourcing task assignment. In: *Proc. of the IEEE Int'l Conf. on Cloud Computing*. 2017. 520–527.
- [21] Wang H, Guo S, Cao J, *et al.* MeLoDy: A long-term dynamic quality-aware incentive mechanism for crowdsourcing. In: *IEEE Trans. on Parallel & Distributed Systems*, 2018,29(4):901–914.
- [22] Cui JM, Sun YE, Huang H, *et al.* TCAM: A truthful combinatorial auction mechanism for crowdsourcing systems. In: *Proc. of the WCNC*. 2018. 1–6.



崔景妹(1993-),女,安徽省蚌埠人,硕士生, CCF 学生会员,主要研究领域为众包.



孙玉娥(1983-),女,博士,副教授,CCF 专业会员,主要研究领域为无线通信网络,博弈论,交通流量统计.



黄河(1983-),男,博士,副教授,CCF 专业会员,主要研究领域为无线通信网络,网络经济学,流量统计.



辛煜(1988-),男,博士,工程师,主要研究领域为遥感大数据信息挖掘.



郭寒松(1990-),男,博士生,主要研究领域为众包,机器学习.



杜扬(1992-),男,博士生,CCF 学生会员,主要研究领域为众包,真值发现,机器学习.

www.jos.org.cn

www.jos.org.cn