

基于组合双向拍卖的物联网搜索任务分配机制

许志凯, 张宏莉, 余翔湛, 周志刚

(哈尔滨工业大学 计算机网络与信息安全技术研究中心, 黑龙江 哈尔滨 150001)

摘 要: 如何合理地分配搜索任务, 进而激励用户加入到搜索中是物联网搜索亟需解决的关键问题。针对物联网中数据实效性强的特点, 结合物联网搜索中用户的高异构性和动态性, 提出一种基于组合双向拍卖的搜索任务分配模型, 从市场供求关系的角度描述了搜索发起者、搜索参与者和搜索引擎之间的关系。首先引入了竞价价值的概念, 提出了一种基于贪心策略的启发式算法确定竞拍成功的用户集合, 然后提出一种基于临界价格的定价算法, 确保用户的竞价反映了其真实估价。理论分析及实验结果证明所提任务分配机制在保证激励相容性、合理性的基础上, 有效提高了物联网搜索引擎的效率。

关键词: 物联网; 信息搜索; 激励机制; 空间众包

中图分类号: TP391.3

文献标识码: A

Combinatorial double auction-based allocation of retrieval tasks in Internet of Things

XU Zhi-kai, ZHANG Hong-li, YU Xiang-zhan, ZHOU Zhi-gang

(Research Center of Computer Network and Information Security Technology, Harbin Institute of Technology, Harbin 150001, China)

Abstract: Task allocation mechanism was greatly important to the success of the search service in Internet of Things (IoT). On basis of analyzing the real time characteristics of the IoT data, and the dynamic characteristics of the users, a combinatorial double auction-based retrieval tasks allocation model was introduced, which described the relationships between the workers, the requesters and the system from the perspective of supply and demand. Firstly, a novel metric to evaluate the value of the users' queries was introduced and a greedy heuristic algorithm to determine the winning requesters and workers was proposed. Then, a critical payment scheme was proposed, which guaranteed that submitted bids of the users reflect their real value. Finally, both the rigid theoretical analysis and simulation result show that the proposed mechanism achieves truthfulness, individual rationality and the efficiency of the service provider is improved.

Key words: Internet of Things; information retrieval; incentive mechanism; spatial crowdsourcing

1 引言

近年来随着 3G、WiFi 等新技术的发展和嵌入了大量传感设备 (如相机、GPS、陀螺仪、噪声传感器等) 的智能终端的普及, 空间众包 (spatial crowdsourcing) 已成为物联网搜索中高效且成本低廉的数据收集方案^[1,2]。如图 1 所示, 基于空间众包的物联网搜索系统中包含 3 个实体: 搜索发起者(发

起查询的用户)、搜索参与者 (执行查询任务的用
户)、搜索引擎。其基本工作流程如下: 1) 搜索发
起者查询与特定地点相关的物联网数据 (如 A 餐厅
现在是否需要等位, B 处现在是否存在噪声污染);
2) 搜索引擎将查询任务分配给愿接受该任务的搜
索参与者; 3) 搜索参与者到达指定地点, 利用随时
携带的智能终端感知并获取相应的数据 (如照片、
音频、视频等), 并通过移动互联网将这些数据发

收稿日期: 2015-05-29; 修回日期: 2015-07-27

基金项目: 国家重点基础研究发展计划 (“973” 计划) 基金资助项目 (2011CB302605, 2013CB329602); 国家自然科学基金资助项目 (61173144, 61073194, 61202457)

Foundation Items: The National Basic Research Program of China (973 Program) (2011CB302605, 2013CB329602); The National Natural Science Foundation of China (61173144, 61073194, 61202457)

送给搜索引擎；4) 搜索引擎将查询结果返回搜索发起者。

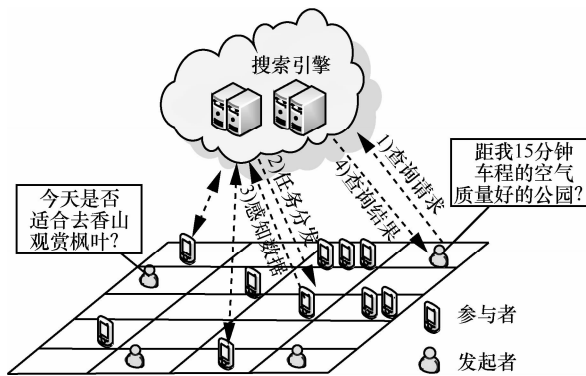


图1 基于空间众包的物联网搜索系统

物联网搜索将搜索空间从面向信息的互联网扩展到人、机、物互联的物联网空间，使用户可以实时地获取自己感兴趣的信息，为人们的生活带来极大的便利^[3]。但物联网中的数据实效性很强，这要求物联网搜索的信息采集工作必须由搜索参与者实时完成，否则将产生大量的噪声数据。若系统中搜索参与者覆盖不足，则直接影响查询结果的准确性和实时性。因此，如何实现搜索任务的合理分配，进而激励用户（搜索发起者、搜索参与者）加入到物联网搜索中成为基于空间众包的物联网搜索系统亟需解决的问题。

当前研究者针对空间众包中的任务分配机制展开了一些研究。但上述研究忽视了众包系统中用户的高异构性和动态性，进而假设搜索参与者的数量远大于搜索发起者（的查询请求）的数量，将搜索引擎与搜索发起者视为同一利益主体，未考虑不同的查询请求之间的竞争关系。而且上述研究均简单地将查询请求假设为原子的，这同样与现实中的一个查询请求可能需要多个搜索参与者协作完成的情况不符。

针对以上研究中存在的问题，本文提出一种基于组合双向拍卖的搜索任务分配模型，从市场供求关系的角度描述了搜索发起者、搜索参与者和搜索引擎之间的关系。首先引入竞价价值的概念，提出了一种基于贪心策略的启发式算法确定竞拍成功的用户集合。然后，提出了一种基于临界价格的定价算法，确保用户的竞价反映了其真实估价。理论分析及实验结果证明本文的任务分配机制在保证激励相容性、合理性的基础上，有效提高了物联网搜索引擎的效率。

2 相关工作

众包（crowdsourcing）是一种分布式的问题解决和生产模式。在众包中，问题以公开招标的方式传播给可能提供解决方案的群体^[4]。随着嵌入了大量传感设备的智能终端的普及，文献[1]将众包模式从互联网引入到物联网中，并提出基于用户所处的地理位置分配搜索任务。

针对空间众包中的任务分配问题，文献[5~8]基于不同的优化目标，提出一系列与之相适应的任务分配算法。根据优化目标的不同，上述方法可分为以下2类。1) 以服务提供商为中心的任务分配机制，即由系统确定搜索参与者应执行的查询任务。为完成尽可能多的查询请求，文献[5]将空间众包中的任务分配视为一个二分匹配问题，进而将该问题转化为最大流问题进行求解。文献[6]在文献[5]的基础上，进一步将包含多个子任务的复杂任务分配问题规约为最大流问题进行求解。文献[7]利用用户的历史轨迹预测用户将来的位置，以实现更高效的搜索任务分配。2) 以用户为中心的任务分配机制，即由搜索参与者主动选择其将要执行的查询任务。其中文献[8]将搜索参与者选择查询任务的过程规约为最大任务调度问题，进而提出一种基于动态规划的启发式算法，可在用户移动距离一定的情况，完成尽可能多的搜索任务。

文献[5~8]均假设搜索参与者无偿加入到空间众包服务中，但搜索参与者在执行查询任务的过程中，往往需消耗其大量的资源（如时间、数据流量、手机电量等），这使物联网中的感知数据具有强烈的所有权特征。因此，文献[5~8]中方法难以保证有足够多的搜索参与者加入到基于空间众包的物联网搜索中。

经济学中的市场机制可充分激励和调动个体的积极性，实现资源的合理配置，由各个高异构性和动态性的用户构成的众包系统与现实经济社会有极大的相似性，因此基于拍卖的任务分配受到研究者的广泛关注。文献[9]将反向拍卖引入到空间众包中，服务提供商依次将竞价最低的参与者加入到竞拍成功的用户集合中，直到满足感知任务的要求。文献[10]指出文献[9]的方法可能导致感知数据在地理上分布不均匀，因此文献[10]在文献[9]的基础上，引入地理位置作为评判标准，以保证感知数据的覆盖范围。文献[11]假定服务提供商知道参与

者执行感知任务时真实估价的分布，并在此基础上提出一种最优化策略，使服务提供商总的支出最小。文献[12]将在线拍卖引入到搜索任务分配中，服务提供商必须实时决定是否接受参与者的竞价请求。但文献[9~12]中将系统及搜索发起者视为同一主体，以执行某一特定的感知任务（如城市噪声监测、空气监测等），上述方法不适用于系统中存在多个查询请求的场景。与本文中工作最接近的是文献[13]，文献[13]假定系统中存在多个查询请求，但参与者只可加入到周围一定范围内的查询请求的拍卖中，且参与者只能接受有限个查询请求。文献[13]以最小化服务商的支出为目标，证明在上述假设下最优化分配问题一个 NP 完全问题，进而将该问题转化为集合覆盖问题进行求解。但文献[13]忽视了空间众包中用户的高异构性和动态性，未考虑不同查询请求之间的竞争关系，且简单地将用户查询请求假设为原子的，与现实中的情况不符。

3 物联网搜索任务分配问题建模

本节首先对物联网搜索任务分配问题进行了描述，并在此基础上提出一种基于组合拍卖的搜索任务分配模型。文中用到的符号列表如表 1 所示。

表 1 文中用到的符号

符号	意义
q_i	搜索发起者 u_i 的查询请求
s_i	搜索参与者
p_i	搜索参与者 s_i 的竞价请求
r_i	搜索资源
R	搜索资源的集合
Q_i^k	完成查询请求 q_i 需要的 k 个资源的集合
$Q(r_i)$	需要搜索资源 r_i 的查询请求的集合
$S(r_i)$	具有搜索资源 r_i 的搜索参与者的集合
b_i	查询请求 q_i 的竞买价格
c_i	竞价请求 p_i 的竞买价格
v_i	查询请求 q_i 的竞买价值
$p(q_i)$	搜索发起者 u_i 需支付的价格
$p(p_i)$	搜索参与者 s_i 应得到的报酬

3.1 问题描述

如图 1 所示，搜索引擎将服务区域（监测的地理区域）分为若干个互不重叠的子区域，即不同的网格。由于物联网中数据时效性很强，搜索参与者仅可接受与其在地理位置上接近的查询任务^[1]，因

此本文假设处于同一网格内的搜索参与者具有相同的搜索资源，即搜索资源与网格之间是一一对应关系。

搜索发起者、搜索参与者及搜索引擎之间的交互关系如图 2 所示。在基于空间众包的物联网搜索系统中，用户的查询请求是与地理位置相关的，且一个用户的查询请求可能需要多个搜索参与者协作完成。如图 1 所示，Alice 发起查询请求“距我 15 分钟车程的空气质量好的公园？”针对这一查询请求，搜索引擎一方面需分配搜索参与者感知周边公园的空气质量；另一方面需分配搜索参与者感知 Alice 附近的交通拥塞状况。搜索引擎只有同时具备上述 2 方面的感知数据，才能完成用户的查询请求。因此，如图 2 所示，搜索发起者（的查询请求）与搜索资源之间是一对多的关系。

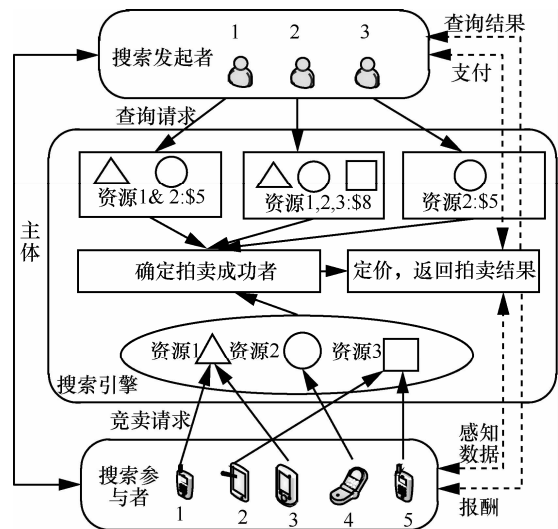


图 2 搜索发起者、搜索参与者及搜索引擎之间的交互关系

此外，由于物联网搜索系统是由高异构性和高动态性的用户构成的，搜索参与者可能随时加入与退出系统，搜索发起者也可能随时发起查询请求。因此搜索引擎无法确保同时满足所有的查询请求。如图 2 所示，不同的查询请求可能会竞买同一资源，而不同的搜索参与者可能会竞买同一搜索资源。

因此，物联网中分配搜索任务问题可视为一个组合双向拍卖问题：搜索发起者对搜索资源的组合进行竞买以完成其查询请求，而搜索参与者对其拥有的搜索资源进行竞买。

3.2 问题建模

本文提出一种基于组合拍卖的物联网搜索任务分配模型，从资源供求关系的角度描述了搜索发

起者、搜索参与者、搜索引擎之间的交互关系。具体流程如图 2 所示。

1) 用户 u_i 将其查询请求 $q_i=(Q_i^k, b_i)$ 提交给搜索引擎。其中 Q_i^k 表示完成查询请求 q_i 所需要的 k 个资源的集合, b_i 表示用户 u_i 对查询请求 q_i 的竞买价格。如图 2 所示, 用户 1 的查询请求可表示为 $q_1=({r_1, r_2}, b_1)$ 。

2) 搜索参与者 s_j 将资源竞卖请求 $p_j=\{r(s_j), c_j\}$ 提交给搜索引擎。其中 $r(s_j)$ 表示搜索参与者 s_j 拥有的搜索资源, c_j 表示 s_j 对其拥有的搜索资源的竞卖价格。如图 2 所示, 参与者 s_1 的竞卖请求可以表示为 $p_1=\{r_1, c_1\}$ 。

3) 搜索引擎确定拍卖成功的搜索发起者及搜索参与者, 并将拍卖结果通知搜索发起者及参与者。

4) 拍卖成功的搜索参与者到达指定地点, 利用智能终端获取相应的数据, 并通过移动互联网将这些感知数据发送给搜索引擎。搜索引擎将感知数据进行融合处理后, 将查询结果返回给搜索发起者。

5) 搜索发起者支付费用 $p(q_i)$, 搜索参与者得到报酬 $p(p_i)$ 。

在上述过程中可以看出, 基于组合双向拍卖的物联网搜索任务分配模型需解决的 2 个子问题是: 1) 确定拍卖成功的用户, 以完成尽可能多的查询请求; 2) 设计合理的定价机制, 激励用户按自己对搜索资源的真实估计进行竞价。下面给出上述 2 个子问题的形式化描述。

定义 1 竞标成功者选择问题 (WBDP, winning bids determination problem)。给定搜索发起者的集合 $U=\{u_1, u_2, \dots, u_m\}$, 搜索参与者的集合 $S=\{s_1, s_2, \dots, s_n\}$, 搜索资源的集合 $R=\{r_1, r_2, \dots, r_l\}$, 竞标成功者选择问题可以定义为

$$\max(\sum_{q_i \in Q} y_i b_i) \quad (1)$$

$$\text{s.t.} \quad x_{ij} \in \{0, 1\} \quad (2)$$

$$y_i = \begin{cases} 1, & \forall r_j \in Q_i^k \ x_{ij} = 1 \\ 0, & \text{其他} \end{cases} \quad (3)$$

$$\sum_{q_i \in Q} x_{ij} \leq |S(r_j)|, \quad \forall r_j \quad (4)$$

其中, 式 (1) 表示搜索引擎的目标是完成尽可能多的查询请求, 以提高搜索引擎效率。 $Q=\{q_i | 0 < i \leq m\}$ 表示所有用户查询请求的集合。 $S(r_j)=\{p_i | r(s_i)=r_j, 0 < i \leq n\}$ 表示拥有搜索资源 r_j 的搜索参与者的集合。 x_{ij} 表示将搜索资源 r_j 分配给查询请求 q_i ,

y_i 表示用户的查询请求 q_i 是否成功。其中约束条件 (4) 限定当搜索发起者竞买成功的搜索资源数小于或等于搜索参与者竞卖成功的搜索资源数。

定义 2 定价算法设计问题。给定用户 u_i 的查询请求 q_i , 令 b_i 表示搜索发起者对查询 q_i 的真实估价, 则查询请求 $q_i=(Q_i^k, b_i)$ 为搜索发起者的真实投标请求, $q_i'=(Q_i^k, b_i')$ 为用户的非真实投标请求, 则本文中定价算法的设计目标为

$$E(q_i) \geq E(q_i') \quad (5)$$

其中, $E(q_i)$ 表示查询请求 q_i 的收益, 可通过式 (6) 计算得出。

$$E(q_i) = y_i b_i - p(q_i) \quad (6)$$

其中, y_i 表示查询 q_i 是否竞买成功, b_i 表示搜索发起者对查询 q_i 的真实估价, $p(q_i)$ 表示搜索发起者需支付的费用。同理可得出搜索参与者 s_i 的收益

$$E(p_i) = p(p_i) - x_i c_i \quad (7)$$

其中, x_i 表示参与者 s_i 是否竞卖成功。本文中的定价算法同样需保证

$$E(p_i) \geq E(p_i') \quad (8)$$

其中, p_i 表示参与者 s_i 的真实投标请求, p_i' 表示参与者的 s_i 非真实投标请求。若本文的定价算法满足式 (5) 和式 (7) 的要求, 则可保证用户按自己对搜索资源的真实估价进行竞价。为激励用户加入到物联网搜索中, 本文中的定价算法还需满足个人理性。

定义 3 个人理性 (individual rationality)。若给定任意一个查询请求 q_i , 其收益 $E(q_i) \geq 0$ 且给定任意一个竞卖请求 p_i , 其收益 $E(p_i) \geq 0$, 则称该分配算法是个人理性的。

基于组合双向拍卖的搜索任务分配模型需同时解决的上述 2 个子问题, 在保证拍卖的激励相容性的基础上, 完成尽可能多的查询请求。针对第一个问题, 本文在第 4 节中提出一种基于贪心策略的竞标成功者选择算法; 针对第二个问题, 本文在第 5 节中提出一种基于临界价格的定价算法。

4 基于贪心策略的竞标成功者选择

如前所述, 尽管 WPDP 问题比较复杂, 但运用整数规划方法可以给予很好的描述, 由于约束条件 (2)、约束条件 (3) 的限制, 式 (1) 可以简化为 0-1 规划问题。但 0-1 规划问题已被证明为一个 NP 完全问题, 因此本节提出一种基于贪心策略的启发

式算法求解 WPDP 问题。

4.1 查询请求的竞买价值

定义 4 查询请求 q_i 的竞买价值 γ_i 。 γ_i 表示查询请求 q_i 对单位搜索资源的竞买价格， γ_i 越高， 查询 q_i 竞买成功的可能性越大。 给定查询请求 q_i 及其对应的搜索资源集合 Q_i^k 及竞价 b_i ， q_i 的竞买价值 γ_i 可定义为

$$\gamma_i = \frac{b_i}{|Q_i^k|} \quad (9)$$

对于任意 2 个查询请求 q_i 和 q_j ， 由于它们对应的搜索资源集合 Q_i^k 和 Q_j^k 不同， 无法直接通过式(9) 计算比较 γ_i 和 γ_j 的大小。 一个简单的计算方法如式 (10) 所示。

$$\gamma_i = \frac{b_i}{|Q_i^k|} \quad (10)$$

但如图 3 所示， 不同的网格内查询请求及搜索参与者的分布是不均匀的， 即不同的搜索资源的供求关系不同。 设查询 q_1 与 q_2 的竞价相同 ($b_1=b_2$)， 且 $Q_1^2=\{R_{11},R_{32}\}$ ， $Q_2^2=\{R_{14},R_{34}\}$ ， 则根据式 (10) 可得出 $\gamma_1=\gamma_2$ 。 但从图 3 中看出， 网格 R_{11} 和 R_{32} 内搜索参与者的数量相对较多， 查询请求的数量相对较少； 但在网格 R_{14} 和 R_{34} 内搜索参与者的数量相对较少， 而查询请求的数量相对较多。 因此可以直观地得出查询请求 q_1 竞拍成功的可能性更大， 即 γ_1 应大于 γ_2 。 因此， 本文进一步考虑了查询请求与搜索参与者的分布情况， 从搜索资源的供求关系出发计算用户的竞买价值， γ_i 的计算如式 (11) 所示。

$$\gamma_i = \frac{b_i \sum_{r_i \in Q_i^k} |S(r_i)| \sum_{r_j \in R} |Q(r_j)|}{|Q_i^k| \sum_{r_i \in R} |S(r_i)| \sum_{r_j \in Q_i^k} |Q(r_j)|} \quad (11)$$

其中， $Q(r_i) = \{q_j | r_i \in Q_j^k, 0 < j \leq m\}$ 表示需要搜索资源 r_i 的查询请求的集合， $S(r_i) = \{p_j | r(s_j) = r_i, 0 < j \leq n\}$ 表示拥有搜索资源 r_i 的搜索参与者的集合。

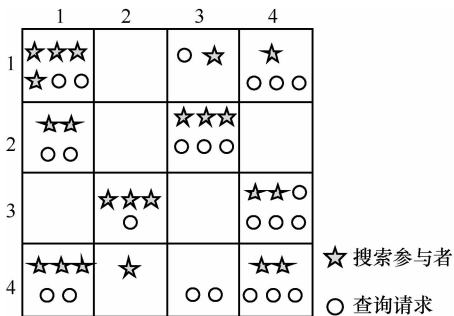


图 3 搜索资源供求关系示意

4.2 基于贪心策略的竞标者成功者选择

为便于描述， 本节中假设搜索发起者及参与者均按其搜索资源的真实估价进行出价， 本文将在 6.1 节中证明用户按其搜索资源的真实估价进行出价是其最优策略。

本文提出一种基于贪心策略的近似算法来求解 WPDP 问题， 令 $Q = \{q_i | 0 < i \leq m\}$ 表示所有查询请求的集合， $S(r_i) = \{p_j | r(s_j) = r_i, 0 < j \leq n\}$ 表示拥有搜索资源 r_i 的搜索参与者的集合表示。 本文的基本思想是将 Q 中竞买价值 γ 最高的查询请求 q_s 取出并加入到竞拍成功的查询请求集合 W_1 中， 直到所有的查询请求的集合 Q 为空。 令 Q_s^k 为完成查询 q_s 所需要的搜索资源的集合。 针对任意的 $r_i \in Q_s^k$ ， 搜索引擎将 $S(r_i)$ 中竞卖价格最低的搜索参与者取出并加入到竞卖成功的搜索参与者集合 W_2 中。 在每一轮迭代中， 若系统中剩余的搜索资源不足以满足查询 q_i 的要求， 则搜索引擎将其从 Q 中删除。 搜索引擎将利用式 (11) 重新计算 Q 中剩余查询请求的竞买价值。 具体流程如算法 1 所示。

算法 1 基于贪心策略的资源分配算法

Input: $Q = \{q_i | 0 < i \leq m\}$, $S(r_i)$, $Q(r_i)$

Output: $W_1 // W_1$: 竞买成功的查询请求的集合
 $W_2 // W_2$: 竞卖成功的搜索参与者的集合

- 1) Initialize $W_1 \leftarrow \emptyset$; $W_2 \leftarrow \emptyset$; int $a[|R|]$; int i, j ;
- // 将具有相同搜索资源的搜索参与者按价格升序排列
- 2) **for** i from 1 to l
- 3) Sort c_j for all $p_j \in S(r_i)$ in the ascending order and the list is denoted $L[i]$;
- 4) 令 $L[i][j]$ 表示 $L[i]$ 中第 j 个元素;
- 5) **end for**
- 6) **while** $Q \neq \emptyset$ **do**
- 7) // 将查询请求按竞买价值降序排列
Sort γ_i for all $q_i \in Q$ in the descending order and the list is denoted by queue L ;
- 8) 令 q_s 是 L 中第一个元素
- 9) $W_1 \leftarrow W_1 \cup \{q_s\}$; // 贪心选择
- 10) delete q_s from Q ;
- 11) **for each** r_i in Q_s^k
- 12) $j = a[i]++$;
- 13) // 选择竞拍成功的搜索参与者
 $W_2 \leftarrow W_2 \cup \{L[i][j]\}$;

```

14) delete  $L[i][j]$  from  $S(r_i)$ ;
15) //删除无法满足其资源要求的查询
请求
if  $|S(r_i)| \leq 0$ 
16) for each  $q_i$  in  $Q(r_i)$ 
17) delete  $q_i$  from  $Q$ ;
18) end for
19) end if
20) end for
21) end while
22) return  $W_1, W_2$ ;

```

分析算法 1 的相对误差, 贪心算法并不能确保在所有场景下得到问题的最优解。若一个最优化问题的最优值为 c^* , 求解该问题的一个近似算法求得的近似最优解相应的目标函数为 c , 则将该近似算法的相对误差定义为 $\frac{c^* - c}{c^*}$ 。不失一般性地, 设在算法 1 中第 i 轮迭代中查询请求 q_s 被加入到竞拍成功的用户集合中, 其竞买价格为 b_s , 竞买价值为 γ_s 。令 q_t 表示该轮迭代中被删除的查询请求中竞买价值最高的查询请求, 则 $c^* - c \leq \max(b_t - b_s, 0)$ 。由于在该轮迭代中 q_s 被选中, 即 $\gamma_s > \gamma_t$, 从竞价价值 γ 的求解步骤中可得出 $c^* - c$ 为搜索参与者所需搜索资源数 k 的一个函数 $\varepsilon(k)$ 。即算法 1 的相对误差 $\frac{c - c^*}{c^*} \leq \varepsilon(k)$ 。

5 支付价格决策机制

搜索引擎需设计合理的定价机制, 激励用户参与到物联网搜索中。为保证用户按自己对搜索资源的真实估价进行出价, 本文引入了文献[14]中提出的临界价格的概念, 每个拍卖成功的用户(参与者)都将支付(得到)其对应的临界价格。

定义 5 临界价格(critical payment)。对于任意一个查询请求 q_i , 若它的竞买价格 $b_i \geq p(b_i)$ 时, q_i 竞买成功; 而当 $b_i < p(b_i)$, q_i 竞买失败; 则称 $p(b_i)$ 为查询请求的临界价格。同理可得出搜索参与者对应的临界价格。

搜索参与者的临界价格计算如下。在物联网搜索资源分配中, 对于任意一个搜索参与者 s_j 在同一时刻只能竞卖一种搜索资源 $r(s_j)$, 即 $p_j = \{r(s_j), c_j\}$ 。不同的搜索参与者竞卖同一搜索资源, 这一过程可视为一次标准的维克瑞拍卖^[14]。设在时刻 t , n 个不

同的搜索参与者竞卖同一搜索资源 r_m , 即 $|S(r_m)| = n$ 。令 $c_1 < c_2 < \dots < c_k < c_{k+1}, \dots, c_n$, 当竞卖成功的搜索参与者数为 k 时, 竞卖请求 p_j 的临界价格, 即拥有搜索资源 r_m 的搜索参与者 s_j 应得的报酬为

$$p(p_j) = \begin{cases} c_{k+1}, & 0 < j \leq k \\ 0, & j > k \end{cases} \quad (12)$$

搜索发起者的临界价格计算如下。如算法 1 所示, 当 $\exists r_i \in Q_j^k, S(r_i) = \emptyset$ 时, 即搜索引擎剩余搜索资源无法满足查询请求 q_j 对搜索资源的需求时, q_j 竞买失败。因此可基于算法 1 中的贪心策略计算搜索参与者的临界价格。令 $Q = \{q_i | 0 < i \leq m\}$ 表示所有查询请求的集合, 首先将查询 q_j 从 Q 中删除。令 $Q_{-(j)} = \{q_i | i \neq j, 0 < i \leq m\}$ 表示剩余查询请求的集合。基本思路是按照查询请求竞价价值的高低, 依次将 $Q_{-(j)}$ 中的查询请求加入到拍卖成功的集合 W_1 中, 直到搜索引擎中剩余的搜索资源无法满足查询请求 q_j 的需求 ($\exists r_i \in Q_i^k, S(r_i) = \emptyset$)。具体流程如算法 2 所示。设当查询请求 q_s 被加入到 W_1 中后, 查询请求 q_j 竞买失败, 则查询请求 q_j 的临界价格, 即搜索发起者 u_j 应支付的价格为

$$p(q_j) = \begin{cases} \max(|Q_j^k| \lambda_s, \sum_{r_i \in Q_j^k} p(r_i)), & q_j \in W_1 \\ 0, & \text{其他} \end{cases} \quad (13)$$

其中, λ_s 表示查询请求 q_s 的竞买价值, $p(r_i)$ 表示拥有搜索资源 r_i 的搜索参与者应得的报酬, 可通过式(12)计算得出。为保证拍卖的合理性 ($U(p_j) \geq 0$), 用户应支付的价格 $p(q_j)$ 应小于或等于其竞买价格 b_j , 若 $p(q_j)$ 大于 b_j , 则将其从竞拍成功的集合 W_1 中删除。具体流程如算法 2 所示。

算法 2 搜索发起者支付价格决策算法

Input: query q_j , other submitted $Q_{-(j)} = \{q_i | i \neq j, 0 < i \leq m\}$

Output: $p(q_j)$ //临界支付价格

- 1) $W_1 \leftarrow \emptyset$ int $flag \leftarrow 0$ int $a[] = \{|S(r_1)|, |S(r_2)|, \dots, |S(r_l)|\}$;
- 2) while $Q_{-(j)} \neq \emptyset$ && $flag == 0$ do
- 3) Sort γ for all $q_i \in Q_{-(j)}$ in the nondecreasing order and the list is denoted by queue L ;
- 4) 令 q_s 是 L 中第一个元素;
- 5) $W_1 \leftarrow W_1 \cup \{q_s\}$;
- 6) delete q_s from Q ;
- 7) for each r_i in Q_s^k

```

8)      if  $a[i] < 1$ 
9)      if  $r_i$  in  $Q_j^k$  //判断  $q_j$  是否
竞买失败
10)     flag=1;
11)     break;
11)     end if
//删除无法满足其资源要求的查询
请求
12)     for each  $q_i$  in  $Q(r_i)$ 
13)         delete  $q_i$  from  $Q_{-i}$ ;
14)     end for
15)     end if
16)     end for
17) end while
18)  $p(q_j) = \max(|Q_j^k| \lambda_s, \sum_{r_i \in Q_j^k} p(r_i))$ ;
19) if  $p(q_j) > b_j$ 
20)      $p(q_j) = 0$ ;
21) end if
22) return  $p(q_j)$ ;

```

6 理论分析

6.1 激励相容与个人理性

本节通过理论分析证明本文的资源分配模型是激励相容的、个人理性的。

定义 6 激励相容性 (incentive compatible or truthfulness)。当拍卖中其他用户的策略固定时,若一个用户按其对商品的真实估计进行出价是其最优策略,则称该拍卖算法是真实的或激励相容的。根据文献[14],当且仅当以下条件被满足时,拍卖算法是激励相容的:1) 其资源分配算法是单调的;2) 每个竞拍成功的用户都支付(得到)了其临界价格。

引理 1 本文提出的资源分配算法是单调的。

证明 设竞拍成功的查询请求集合 $W_1 = \{q_1, q_2, \dots, q_n\}$, 其中 $\gamma_1 > \gamma_2 > \dots > \gamma_n$, 即 W_1 中的元素是被加入到集合 W_1 中的顺序排列的。设 q_i 是第 i 个被加入到 W_1 中的元素, $q_i = \{Q_i^k, b_i\}$ 。设存在一个查询请求 $q_j = \{Q_j^k, b_j\}$, 且 $Q_j^k = Q_i^k$, $b_j > b_i$, 即 q_j 和 q_i 竞买的资源相同, 但 q_j 的竞价价值大于 q_i 的竞价价值。则根据算法 1, q_j 一定会在第 i 轮或更早的轮次竞买成功, 即本文的资源分配算法是单调的。同理可证, 对于搜索参与者, 本文中的资源分配算法同样是单调的。

引理 2 每个竞拍成功的查询请求(搜索参与

者)都支付(得到)了其临界价格。

证明 如算法 2 所示, 每个竞拍成功的查询请求(搜索参与者)都支付(得到)了其临界价格。

定理 1 本文提出的基于组合拍卖的资源分配模型是激励相容的。

证明 根据文献[14], 当一个拍卖算法同时满足引理 1 和引理 2 时, 该拍卖算法是激励相容的。

定理 2 本文提出的基于组合拍卖的资源分配算法是个人理性的。

证明 如 3.2 节所述, $E(q_i)$ 可定义为 $E(q_i) = b_i - p(q_i)$, 其中 b_i 表示查询 q_i 对用户 u_i 的真实价值, $p(q_i)$ 表示用户支付价格。当用户 u_i 的查询请求 q_i 竞拍失败时, $b_i = p(q_i) = 0$, 即 $E(q_i) = 0$, 此时对于查询请求 q_i , 本文提出的基于组合拍卖的资源分配算法是个人理性的。当用户 u_i 的查询请求 q_i 竞拍成功时, 则根据算法 2, $p(q_i) \leq b_i$, 即 $E(q_i) \geq 0$, 此时对于查询请求 q_i , 本文提出的基于组合拍卖的资源分配算法是个人理性的。综上所述, 对于任意一个搜索发起者, 本文中提出的基于组合拍卖的资源分配模型是个人理性的。同理可证, 对于搜索引擎及任意一个搜索参与者, 本文提出的基于组合拍卖的资源分配模型是个人理性的。

6.2 算法的时间复杂度

本节分析了文中提出的任务分配机制的时间复杂度。

引理 3 算法 1 的时间复杂度为多项式时间。

证明 算法 1 中第一个 for 循环(算法第 2~第 5 行)的时间复杂度为 $O(|S| \log |S|)$, 算法第 7 行的排序算法的时间复杂度为 $O(|Q| \log |Q|)$, 算法第 11~第 20 行的循环的时间复杂度为 $O(k|Q|)$ 。因此每执行一次 while 循环(算法第 7~第 20 行)的时间复杂度为 $O(|Q| \log |Q|)$ 。由于 while 循环最多执行 $|Q|$ 次, 因此算法第 6~第 21 行的时间复杂度为 $O(|Q||Q| \log |Q|)$ 。综上所述, 算法 1 的时间复杂度为 $O(|S| \log |S|) + O(|Q| \log |Q| |Q|)$, 即算法 1 的时间复杂度为多项式时间。

引理 4 算法 2 的时间复杂度为多项式时间。

证明 算法 2 第 3 行的排序算法的时间复杂度为 $O(|Q| \log |Q|)$ 。算法第 7~第 11 行的循环的时间复杂度为 $O(k|Q|)$ 。因此每执行一次 while 循环的时间复杂度为 $O(|Q| \log |Q|)$ 。由于 while 循环最多执行 $|Q|$ 次, 因此算法 2 的时间复杂度为 $O(|Q| \log |Q| |Q|)$, 即即算法 2 的时间复杂度为多项式时间。

定理 3 本文中的任务分配机制的时间复杂度

为多项式时间。

证明 基于组合双向拍卖的物联网搜索任务分配模型需解决的 2 个子问题：1)确定竞拍成功的用户；2)计算竞拍成功的用户应支付的价格。因此，由引理 3 和引理 4 可得出，本文提出的任务分配机制的时间复杂度为多项式时间。

7 仿真实验和结果对比

为测试本文提出的基于组合拍卖的资源分配模型，本次实验所处的硬件环境为：Inter(R) Celeron G530, 500 GB 硬盘, 4 GB 内存, Win7 32 位系统。实验数据集为文献[15]公开的 Foursquare 的签到数据集。该数据集包含 18 107 个用户在 2010 年 3 月到 2011 年 1 月间的签到位置、签到时间。实验中从以下几个方面对本文中的资源分配模型进行评估：1) 搜索引擎的效率；2) 搜索引擎的社会成本 (social cost)。

7.1 实验设计

定义 7 用户的签到轨迹。给定用户 j 的签到轨迹可以表示为一个按时间排列的序列 $H(j) = \{(l_1, t_1), (l_2, t_2), \dots, (l_m, t_m)\}$, (l_i, t_i) 是序列 H 中的一个元素, 表示用户 j 在 t_i 时刻在兴趣点 (point of interest), 代表电子地图上某个地标 l_i 签到。

实验中将用户的签到位置 l_i (兴趣点) 视为一种搜索资源 r_i 。将在该位置签到的用户 j 视为拥有该资源的搜索参与者 s_j , 从用户 i 的签到数据轨迹中任意选择 $k(1 \leq k \leq 5)$ 个签到位置作为完成搜索发起者 u_i 的查询要求 q_i 所需要搜索资源的集合 Q_i^k 。如前所述，基于众包的物联网搜索系统是由高异构性和动态性参与者构成的。因此不同的搜索发起者 (参与者) 对其竞买 (竞卖) 的搜索资源价值估价差异较大，实验中随机生成查询请求 q_i 的竞买价格 $b_i(10 \leq b_i \leq 30)$ 与搜索参与者 s_j 的竞卖价格 $c_j(1 \leq c_j \leq 10)$ 。

实验中选取 50 个兴趣点作为搜索资源的集合，为模拟众包系统中用户及参与者的高动态性，本文在不同的资源供需比 ($\frac{\text{查询请求数}}{\text{搜索资源数}}$) 下对基于组合拍卖的资源分配算法进行了测试。

7.2 搜索引擎的效率

定义 8 搜索引擎的效率 μ 。搜索引擎的效率可定义为搜索引擎完成的查询请求数与搜索引擎收到的查询请求数的比值。 μ 的计算如式 (14) 所示。

$$\mu = \frac{\sum_{q \in W_1} b_i}{\sum_{q \in Q} b_i} \tag{14}$$

其中， W_1 表示竞拍成功的查询请求的集合， Q 表示所有查询请求的集合。

实验结果如图 4 和图 5 所示，其中 OPT 表示离线情况下的最优分配策略；CAA 表示本文中提出的基于组合拍卖的分配模型；CAR 表示基于组合拍卖的随机分配模型，即搜索引擎随机选择竞买成功的查询请求；AUC 表示文献[13]中提出的基于反向拍卖的资源分配模型，在该模型中查询请求需分别竞买其所需要的 k 个搜索资源。

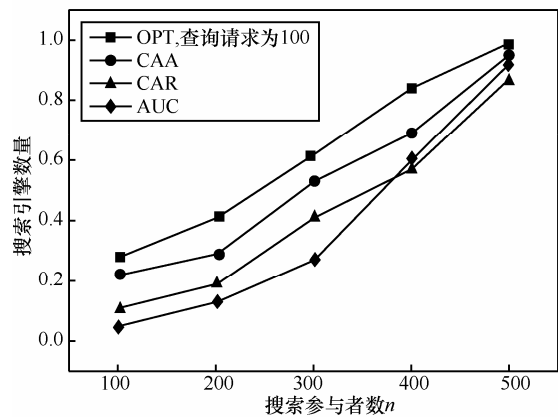


图 4 搜索引擎效率随搜索参与者数变化

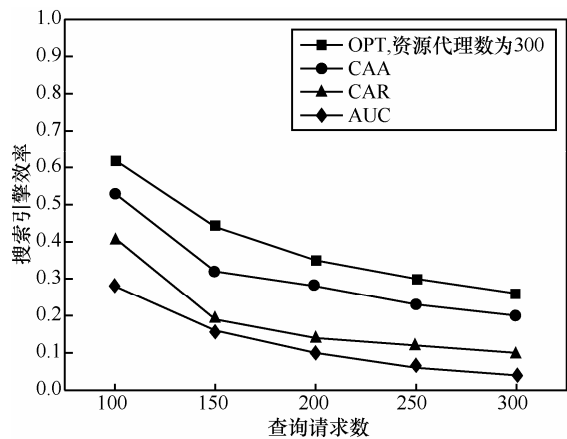


图 5 搜索引擎效率随查询请求数变化

从图 4 和图 5 中可以看出，当系统中查询请求数固定时，搜索引擎的效率与系统中搜索参与者的数量呈正相关。而当搜索参与者数固定时，搜索引擎的效率与查询请求数呈负相关。如图 4 所示，当系统中搜索参与者的数 $n=500$ 时，基于任意一种分配模型的搜索引擎均可达到较高的效率。但随着查

询请求数量的增多，基于 AUC 模型的搜索引擎效率下降速度最为明显，当 $n \leq 300$ 时，基于 AUC 模型的搜索引擎的效率 ($n=300, \mu=0.27$) 甚至低于基于 CAR 模型的搜索引擎 ($n=300, \mu=0.41$)，这是因为只有查询请求的所需的搜索资源集全部竞买成功时，搜索引擎才能完成该查询请求，而基于组合拍卖的分配模型更好地描述了物联网中的搜索资源分配问题。如图 4 和图 5 所示，基于 CAA 的搜索引擎的分配效率 ($n=100, \mu=0.22$) 始终远高于基于 CAR 模型的搜索引擎 ($n=100, \mu=0.11$)，且较接近离线情况下的最优分配效率 ($n=100, \mu=0.28$)。这是因为本文中用户竞价价值 γ 的计算方法较好地反映了用户查询请求对单位搜索资源的竞买价格，从而有效提高了搜索引擎的效率。

7.3 搜索引擎的社会成本

搜索引擎的社会成本是指完成查询请求所需要消耗的搜索资源的价值，可通过统计搜索参与者对其拥有的搜索资源的估价得出。但如图 4、图 5 所示，基于不同分配模型的搜索引擎完成的查询请求数不同，因此本文引入了单位社会成本的概念。单位社会成本 β 的计算如式 (15) 所示。

$$\beta = \frac{C_{\text{total}}}{|W_1|} \quad (15)$$

其中， C_{total} 表示搜索引擎完成 W_1 中查询请求所需要的社会资源，可通过统计竞买成功的搜索参与者对其拥有的搜索资源的估价得出。

$$C_{\text{total}} = \sum_{s_j \in W_2} c_j \quad (16)$$

实验结果如图 6 所示，当查询请求数固定为 100 时，随着搜索参与者数量的减小，基于 4 种模型的搜索引擎的单位成本均明显增加。当 $n=500$ 时，

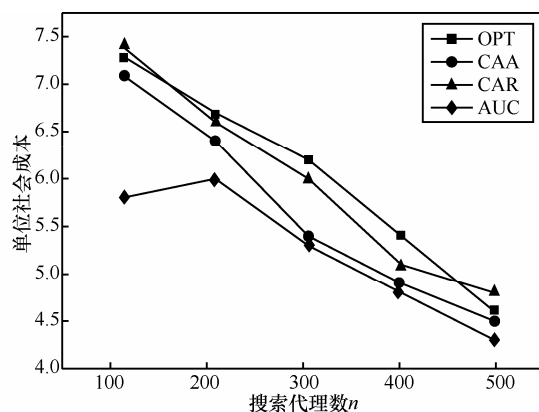


图 6 搜索引擎单位社会成本随搜索参与者数变化

基于 CAR 的搜索引擎的单位社会成本为 4.5，而当 $n=100$ 时，该值为 7.1。实验结果表明搜索引擎的社会成本与搜索参与者的数量呈负相关。搜索引擎需设计合理的资源分配机制，激励用户参与到物联网搜索中。

8 结束语

本文针对物联网中搜索任务分配问题，提出一种基于组合双向拍卖的搜索任务分配模型。该模型将搜索参与者的查询请求视为对搜索资源组合的竞买，将搜索参与者执行搜索任务的请求视为对其拥有的搜索资源的竞卖。本文从搜索资源供求关系的出发，引入用户竞价价值的概念，并提出一种基于贪心策略的启发式算法，确定拍卖成功的用户。最后，本文提出一种基于临界价格的支付算法，确保用户的竞价反映了其对搜索资源的真实估价。理论分析及实验结果表明本文方法在保证拍卖的激励相容性、合理性的基础上，有效地提高了物联网搜索引擎的效率。

对于基于空间众包的物联网搜索，未来仍有若干问题值得研究。搜索参与者的位置隐私保护及模糊位置下的任务分配将是下一阶段工作的重点。

参考文献:

- [1] ALT F, SHIRAZI A S, SCHMIDT A. Location-based crowdsourcing: extending crowdsourcing to the real world [A]. IEEE the 6th Nordic Conference on Human-Computer Interaction[C]. Reykjavik, Iceland, 2010. 13-22.
- [2] KAZEMI L, SHAHABI C. A privacy-aware framework for participatory sensing[J]. ACM SIGKDD Explorations Newsletter, 2011, 13(1): 43-51.
- [3] 方滨兴,等.网络空间大搜索技术白皮书[M].国家自然科学基金委员会信息科学部. 2015. FANG B X, et al. Cyberspace Big Search White Paper [M]. Department of Information Science, National Natural Science Foundation of China.2015.
- [4] AROLAS E, ENRIQUE, GUEVARA F. Towards an integrated crowdsourcing definition[J]. Journal of Information Science, 2012, (38): 189-20.
- [5] KAZEMI L, SHAHABI C. Geocrowd: enabling query answering with spatial crowdsourcing [A]. Proceedings of the 20th International Conference on Advances in Geographic Information Systems[C]. New York, USA, 2012. 189-198.
- [6] DANG H, NGUYEN T, TO H. Maximum complex task assignment:

towards tasks correlation in spatial crowdsourcing [A]. Proceedings of the 15th International Conference on Information Integration and Web-based Applications & Services [C]. Vienna, Austria, 2013. 77.

- [7] HE Z J, CAO J N, LIU X F. High quality participant recruitment in vehicle-based crowdsourcing using predictable mobility [A]. Proceedings of IEEE INFOCOM 2015[C]. Hongkong, China, 2015. 2542-2550.
- [8] DENG D, SHAHABI C, DEMIRYUREK U. Maximizing the number of worker's self-selected tasks in spatial crowdsourcing [A]. Proceedings of ACM GIS[C]. Florida, USA, 2013. 324-333.
- [9] LEE J S, HOH B. Sell your experiences: market mechanism based incentive for participatory sensing [A]. Proceedings of IEEE PERCOM[C]. Mannheim, Germany, 2010. 60-68.
- [10] JAIMES L, VERGARA I, LABRADOR M. A location-based incentive mechanism for participatory sensing systems with budget constraints [A]. Proceedings of IEEE PERCOM[C]. Lugano, Switzerland, 2012. 103-108.
- [11] KOUTSOPOULOS I. Optimal incentive-driven design of participatory sensing systems [A]. Proceedings of IEEE INFOCOM 2013[C]. Turin, Italy, 2013. 1402-1410.
- [12] ZHU Y, ZHANG Q, ZHU H. Towards truthful mechanisms for mobile crowdsourcing with dynamic smartphones[A]. Proceedings of IEEE ICDCS 2014[C].Madrid, Spain, 2014. 11-20.
- [13] FENG Z, ZHU Y, ZHANG Q. TRAC: truthful auction for location-aware collaborative sensing in mobile crowdsourcing[A]. Proceedings of IEEE INFOCOM 2014[C]. Toronto, Canada, 2014. 1231-1239.
- [14] NISAN N, ROUGHGARDEN T, TARDOS E. Algorithmic Game Theory[M]. Cambridge University Press, 2007.
- [15] GAO H, TANG J, LIU H. Exploring social-historical ties on location-based social networks[A]. Proceedings of the 6th International AAAI Conference on Weblogs and Social Media[C]. Dublin, Ireland, 2012.

作者简介:



许志凯 (1988-), 男, 山东潍坊人, 哈尔滨工业大学博士生, 主要研究方向为云计算和信息安全。



张宏莉 (1973-), 女, 吉林榆树人, 哈尔滨工业大学教授、博士生导师, 主要研究方向为网络与信息安全、网络测量与建模、网络计算、并行处理等。



余翔湛 (1973-), 男, 黑龙江哈尔滨人, 哈尔滨工业大学研究员、博士生导师, 主要研究方向为网络容灾和信息安全。



周志刚 (1986-), 男, 山西太原人, 哈尔滨工业大学博士生, 主要研究方向为云计算和云安全。