

合作性移动群智感知中具有一般效用和成本的用户招募方法

刘文彬 杨永健 王 恩

(吉林大学计算机科学与技术学院 长春 130012)

摘 要 移动群智感知(Mobile CrowdSensing, MCS)是一种强力而有效的感知数据收集模式,其允许我们在预算有限的情况下招募一群移动用户并利用其随身携带的智能设备来合作地执行感知任务. 现有的工作通常根据具有子模性质的效用函数(如用户集合的任务覆盖概率,存在边际收益递减)和线性的成本函数(如用户集合的累积支付成本)来招募用户. 然而,在实际应用中,我们往往需要进一步考虑用户间合作意愿,导致实际问题比现有工作所假设的场景更加复杂,相应的效用和成本函数可能不再满足子模和线性关系. 举例来说,如果某个用户基于隐私考虑而不愿意与陌生人合作,他可能会拒绝执行一部分敏感的工作,或者要求支付较高的费用,导致效用和成本发生变化. 本文研究了具有一般效用和成本的合作性移动群智感知用户招募问题. 为了解决非子模的效用和成本,我们提出了一种广义离线贪婪算法,离线地从已知用户池中选择用户,并通过引入子模比率和曲率来证明该算法的理论近似比. 进一步考虑在线场景,我们提出了一种分段的在线招募策略,首先将用户序列划分为数个片段,然后在每个片段中分别进行在线招募,并同样给出该策略的理论近似比. 我们在四个真实数据集上进行了广泛的评估,其结果验证了所提出方法的有效性.

关键词 移动群智感知;用户招募;在线优化;非子模函数;秘书问题

中图法分类号 TP393 **DOI号** 10.11897/SP.J.1016.2022.02576

User Recruitment with General Utility and Cost for Cooperative Mobile CrowdSensing

LIU Wen-Bin YANG Yong-Jian WANG En

(College of Computer Science and Technology, Jilin University, Changchun 130012)

Abstract Mobile CrowdSensing (MCS) is a powerful sensing paradigm, which recruits a crowd of users within a limited budget to cooperatively perform sensing tasks. The existing works usually recruit users based on their submodular utility and linear cost functions, e. g., the users' probabilistic coverage of tasks and total consumed costs. However, we argue that these works do not consider the cooperative willingness of users, which may lead to non-submodular utility and cost. For example, if a user is unwilling to cooperate with strangers, he may refuse to perform some sensitive tasks and/or ask for a higher payment due to the privacy concern. In this paper, we study the user recruitment problem with general utility and cost for cooperative MCS. To deal with the non-submodular utility and cost, we present a generalized online greedy algorithm to select users from a determined user pool, which achieves an approximation ratio by introducing submodularity ratio and curvature. Taking online recruiting into further consideration, we propose

收稿日期:2022-01-12;在线发布日期:2022-09-20. 本课题得到科技创新 2030——“新一代人工智能”重大项目(2021ZD0112501, 2021ZD0112502)、国家自然科学基金(62102161, 61972450, 62072209)、中国博士后基金(2021T140261, 2021M701389)、CCF-百度松果基金(2021PP15002000)资助. 刘文彬, 博士, 助理研究员, 中国计算机学会(CCF)会员, 主要研究方向为移动群智感知、普适计算. E-mail: liuwenbin@jlu.edu.cn. 杨永健, 博士, 教授, 中国计算机学会(CCF)会员, 主要研究领域为无线感知网络. 王 恩(通信作者), 博士, 教授, 中国计算机学会(CCF)会员, 主要研究领域为移动群智感知、移动计算. E-mail: wangen@jlu.edu.cn.

a segmented online strategy, which first divides the user sequence into segments and then recruits them with a theoretical performance approximation. Extensive evaluations have been conducted over four real-world datasets, which verify the effectiveness of the proposals.

Keywords Mobile CrowdSensing; user recruitment; online optimization; non-submodular function; secretary problem

1 引言

近年来,随着以 5G 为代表的无线网络的快速发展,和以手机为代表的移动设备的爆炸式普及,移动群智感知(Mobile CrowdSensing, MCS)^[1-2]作为一种新兴而高效的数据采集模式逐渐展现出了强大的感知与计算能力,在智慧城市、智能交通、环境监测等领域有着广阔的应用前景.典型的移动群智感知系统以大量移动用户及其携带的智能设备作为基础感知单元,利用其广泛存在性、灵活移动性和机会连接性来完成大规模且细粒度的感知任务,同时融合显性或隐性的群体智能来进一步优化执行效率并提高感知质量^[3].典型应用包括利用手机加速度等信息识别“桥梁”、“隧道”等地图语义^[4];根据车辆 GPS 信号实时感知城市道路车速与流量^[5];使用麦克风、声音感知器等识别噪音污染^[6];也能够通过用户主动反馈信息,如健身房排队人数^[7]、空闲停车位^[8]、拍照追踪车辆^[9]和室内外定位^[10]等,为人们日常生活和工作提供便利.

为了获得高质量的感知服务,传统的移动群智感知系统通常需要招募尽可能多的用户来执行感知任务.然而,考虑到实际应用往往存在有限的预算限制,我们必须选择一部分更为有效的用户来执行任务并给予奖励,这就是移动群智感知中最为基础的用户招募问题^[11-13].一般来说,用户招募问题的目标是选择一组能够完成最多感知任务的用户,而他们的总成本不超过预定义的总预算.大多数现有工作都将用户集合的效用和成本分别表示为子模函数和线性函数^[11,14-15],以反映边际收益递减和累积支付成本.然而,在合作性移动群智感知中,我们认为被招募用户的合作意愿将严重影响任务的完成情况和对奖励的需求^[16-18],甚至导致用户集合的效用和成本函数不再具有子模或线性关系.例如,在车辆追踪任务中,尤其考虑在摄像头不足的区域,我们需要招募一组用户通过拍照的方式来合作地追踪移动车辆.需要注意的是,实时位置信息和历史轨迹对于持

续的目标追踪非常重要.因此,该追踪任务的参与用户必须将其敏感的时空数据和照片暴露给其他合作用户.显然,如果一个用户不熟悉该组中的其他人,他可能因为担心隐私泄露^[19]而不愿意执行一些敏感的任务并要求更高的奖励.相反地,如果被招募的用户们都是朋友,他们就不会关心这些隐私问题,因此会以较少的奖励完成任务.此时,在考虑合作意愿的情况下,用户的效用和成本在不同用户集合下产生变化,导致无法利用它们的子模(边际收益递减)和线性(累积支付成本)关系来准确衡量用户贡献,也因此难以挑选适合的用户来执行感知任务.

需要注意的是,合作性移动群智感知中的用户意愿往往是非常复杂的,可能受许多因素影响.本文主要关注用户招募问题中由用户意愿所引起的非子模效用和成本^①.图 1(上半部分)提供了一个直观的例子:我们简单地将完成任务数量看作效用,当考虑用户的合作意愿时,由于第三个用户在不同用户组下有不同的合作意愿,所以在不同的用户组下任务的完成情况也不同.因此,他的效用可能是不规律的,存在非子模性质.此外,在图 1 的例子中,在不同的用户组中的成本可能也不能再简单地用线性加法公式计算.这是因为用户的行程、电池消耗、隐私风险和许多其他因素都会受到合作意愿的影响^②.因此,在这样一个合作性移动群智感知中,用户的效用函数和成本函数都可能是非子模的,这就放松了以往

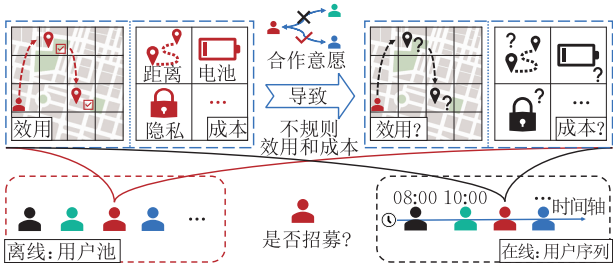


图 1 合作性移动群智感知中具有一般效用和成本的离线和在线用户招募

① 实际上,非子模特性在实际问题和应用中较为普遍,本文工作经过简单修改,就可以为相关工作提供具有理论性能保证的近似解决方案
② 以隐私风险为例,用户会要求与陌生人的合作支付更多的费用,以弥补风险.

工作中的子模边际效用和线性累积成本的假设,使得用户招募问题极具挑战.

同时,除非子模的效用和成本以外,不同的用户参与方式也直接影响用户招募问题.图 1(左下部分)展示了一个相对简单的用户离线参与示例,即一些用户在移动群智感知任务开启前申请参与,我们需要从该用户池中选择最佳用户集合来执行任务^[11,14-15].在该离线场景下,最简单和直接的解决方案是遍历所有可能的用户集合以找到其中最好的一个.显然,这个问题是 NP 难的.因此,之前的工作主要利用用户效用函数的子模性质来迭代地挑选边际效用与成本的比率最大的用户,直至预算耗尽.然而,在合作性移动群智感知下,效用函数和成本函数均可能是非子模的,其不仅打破了边际收益递减的规律,同时非子模成本也使得以前的效用与成本比率优化不再有效.因此,如何解决离线用户招募中的非子模效用和成本问题是本文面临的首要挑战.

离线用户招募约束了所有用户必须提前确认参与,然而在实际应用中,用户往往实时加入,且不会等待较长时间.为此,我们将离线场景中的用户招募问题进一步扩展至一个更加实际的在线场景,称为在线用户招募.如图 1 所示(右下部分),在在线场景下,用户将实时地申请参与移动群智感知活动,我们必须在用户实时申请时立即决定是否招募^[15,20].需要注意的是,在实际应用中,用户可能不会持续参与或等待较长时间.因此,我们必须在极短的时间内给出是否招募的决定^[21-23].此时,用户的效用和成本在其参与之前是不可见的,而我们又必须立即判断是否招募当前用户或等待下一个用户.此外,在合作性移动群智感知中,非子模的效用和成本函数也使得我们难以准确预测未来用户的效用和成本,进而加剧了在线用户招募问题的求解难度.因此,如何处理非子模效用和成本下的在线招募是本文面临的重要挑战.

为了解决上述挑战,本文主要研究了合作性移动群智感知下基于一般效用和成本的用户招募问题.我们首先采用一个关系矩阵来记录用户的成对合作意愿,然后利用这个矩阵建立了一个简单的合作性移动群智感知模型,其中的效用和成本函数可以是非子模的.为了处理非子模效用和成本问题,我们提出了一种广义的离线贪婪用户招募算法,该算法在多项式时间内迭代选择边际效用与边际成本的比率最大的用户.借助于子模比率 α (即集合函数与子模函数的相近程度)和曲率 γ (即子模函数与模函数的

相近程度),我们将成本函数近似为模函数以满足预算限制.同样地,我们将效用函数以 $\frac{\alpha}{2}(1-e^{-\alpha})$ 的近似比逼近于子模函数.针对在线招募问题,我们将其表示为一种基于非子模函数的多选择秘书问题^[20,24-25],并提出了一种分段的在线用户招募策略,该策略首先将参与的用户序列划分为若干相等大小的分段,然后招募每个分段中的最优用户.此外,为了充分利用新获得的信息,我们在招募到一个新用户后,动态地重新调整上述分段策略.最后,我们也证明了该分段优化实际上是一个近似全局最优解,其近似比为 $\frac{\alpha^3(1-e^{-1})(1-e^{-\alpha})}{14}$.

综上所述,本文的主要贡献如下:

(1) 合作性移动群智感知.考虑到合作意愿,我们研究了具有非子模效用函数和成本函数的合作性移动群智感知用户招募问题.

(2) 用户招募问题.为了解决非子模的效用和成本问题,提出了一种离线场景下的广义贪婪用户招募算法,并给出近似比 $\frac{\alpha}{2}(1-e^{-\alpha})$.进一步考虑在线招募,我们提出了一种分段的在线用户招募策略,该策略首先对用户序列进行划分,然后分段地招募用户,其近似比为 $\frac{\alpha^3(1-e^{-1})(1-e^{-\alpha})}{14}$.

(3) 广泛的实验与评估.我们基于四个真实数据集对所提出的算法和策略进行了广泛的实验与评估,结果验证了我们的方法在离线和在线情况下处理非子模函数的有效性.

本文第 2 节回顾相关工作;第 3 节介绍系统模型和研究问题;第 4 节中提出离线算法;第 5 节中提出在线策略;第 6 节评估所提出方法;第 7 节讨论合作意愿和招募策略;第 8 节总结全文.

2 相关工作

2.1 用户招募

移动群智感知允许我们招募用户来合作地执行各种感知任务. Karaliopoulos 等人^[11]从机会网络视角研究了一种基于用户的概率性移动轨迹的离线招募策略. Wang 等人^[14]进一步考虑了用户的两种不同上传方式:月付用户和随用随付用户,并依据其效用和成本进行离线用户招募. Song 等人^[26]则主要关注多任务场景下的任务完成质量,并以此为依据进行用户招募. Bhatti 等人^[27]考虑环境可能发生变化,提出了一种有界和异构的任务分配算法,以在多约束条件下最大化工人的报酬.然而上述方法仅适

用于离线用户招募场景,无法处理在线场景下动态到来的用户和实时发布的任务。

近年来,也有部分研究者关注更为符合实际的在线用户招募场景。Pu 等人^[28]构建了一种自组织的移动众包网络,提出使用动态规划来进行用户招募。Yang 等人^[15]提出了一种基于移动性预测的用户招募框架,并使用秘书问题来处理在线用户招募。纪圣堦等人^[29]使用前向搜索估计未来用户信息,进而通过投票来决定是否招募用户。Liu 等人^[20,30]利用背包问题来离线地处理时间、预算等约束,再通过秘书问题进行在线用户招募。此外,也有部分研究者从类似的激励机制角度研究在线场景。Zhao 等人^[31]主要关注预算分配问题,提出了一种分段式的在线激励机制,从而动态地决定参与用户的报酬。Gao 等人^[32]进一步将数据质量融入激励机制中,通过在线调整的方式确定用户报酬。Wang 等人^[33]则设计了一种在线激励机制,包括在线优胜者选择和报酬确定策略,从而动态选择用户并决定合适的报酬。然而,上述工作都是基于用户子模效用函数和线性成本函数来制定招募决策,不能直接用于处理本文中合作性移动群智感知下非子模效用和成本的问题。

此外,也有一些研究者关注与用户招募非常相近的在线任务分配问题。向朝参等人^[34]提出了一种基于深度强化学习的感知任务分配方法。Han 等人^[35]为移动群智感知离线和在线任务构建了一种树形数据结构,并在其上进行任务和用户的快速匹配。Wang 等人^[36]构建了一个概率图模型并使用无监督学习估计任务和用户信息,从而进行“任务-用户”对的快速匹配。Xiang 等人^[37]则利用历史数据预测不同位置的可能收益,进而基于贪心局部搜索设计了一个任务分配的算法。然而,以上工作均是根据当前已知或预测获得的任务和用户信息来进行快速匹配,其本质上更加关注已有任务和用户等资源的组合与优化,无法处理动态变化且难以准确预测的任务和用户,因此无法完成实时招募决策。

2.2 子模优化

从方法角度来说,现有的基于子模效用和线性成本的用户招募问题实际上是经典子模优化的一个变体^[38]。Krause 等人^[39]提出了一种关注“成本-效益”的算法,可以贪婪地从已知元素池中选择部分元素,并给出 $(1-e^{-1})/2$ 的近似比保证^[39]。进一步结合部分枚举方法,这一经典算法可改进并获得更好的近似比: $(1-e^{-1/2})$ ^[40] 和 $(1-e^{-1})$ ^[41]。Zhang 等

人^[42]将线性成本放宽为一般成本约束,证明了一般近似比为 $(1-e^{-1})/2$ 。Qian 等人^[43]进一步考虑了非子模效用和成本,提出了一种 $\alpha(1-e^{-\alpha})/2$ -近似保证的广义贪婪算法。以上这些工作为处理非子模函数提供了理论基础。然而,本文不仅研究元素效用和成本以进行非子模优化,还需要考虑用户的参与方式,特别是在线场景下的用户招募问题。

3 系统模型和问题描述

3.1 系统模型

本文主要研究较为普遍的移动群智感知场景。大量的用户在目标区域间移动,以执行各种感知任务。我们将用户表示为 $U=\{u_1,u_2,\cdots,u_m\}$,用户总数为 m ;将任务表示为 $S=\{s_1,s_2,\cdots,s_n\}$,任务总数为 n 。根据用户的能力、移动性和意愿等,每个用户可以执行多个任务并获得相应奖励,表示为效用函数 $f(u_i)$ 和感知成本 $c(u_i)$ 。在有限的预算 B 下,我们希望从用户集合 U 中招募 k 个用户组成已招募用户集合 μ ,以最大化他们的效用 $f(\mu)$,而他们的总成本 $c(\mu)$ 不能超过预算 B 。请注意,本文为了简化表达,将集合 $\{u_i\}$ 表示为 u_i ,并将 $\mu\backslash\{u_i\}$ 简化为 $\mu\backslash u_i$ 。本文中使用的符号列于表 1。

表 1 主要的符号

符号	含义
u,s,B	用户,任务和预算。
U,S,μ	用户集合,任务集合和已招募用户集合。
m,n,k	用户总数,任务总数和招募用户数。
$f_o(\cdot),c_o(\cdot)$	不考虑合作意愿的效用和成本。
$f(\cdot),c(\cdot)$	考虑合作意愿的效用和成本。
$\nabla f(\mu,u)$	将用户 u 添加到已招募用户集合 μ 的
$\nabla c(\mu,u)$	边际效用和成本。
W,w	关系矩阵和合作意愿。
α,γ	子模比率 and 曲率。

考虑到用户参与的不同方式,我们分别在离线和在线的场景下进行用户招募。对于离线场景,我们根据已知的效用和成本,从预先确定的用户集合 U 选取并招募部分用户。显然,该离线招募无法处理动态到来的用户与实时发布的任务。因此,我们进一步考虑了在线场景,即用户实时参与并且不会长期持续工作或长时间等待决策。所以,我们必须在不知道未来信息的情况下立即决定是否招募当前到来的用户。

3.2 效用和成本

一般来说,在移动群智感知中衡量一个用户的效用需要考虑很多因素。现有的工作通常使用用户

执行任务的客观能力来衡量其效用. 例如, 使用用户 u_i 能够完成任务 s_j 的概率 $P(u_i, s_j)$, 来估计 u_i 的客观效用 $f_o(u_i) = \sum_{s_j \in S} P(u_i, s_j)$, 即完成任务的数量的期望值. 除客观能力外, 我们认为用户的主观合作意愿也将严重影响其效用. 需要注意的是, 合作意愿是非常复杂且难以衡量的, 它可能由许多因素决定. 在本文中, 我们主要关注的是由合作意愿引起的非子模效用和成本. 为了不失一般性, 我们基于用户的合作意愿引入了两个非子模效用/成本函数, 他们分别反映出较强的合作意愿有助于提高效用, 以及较弱的合作意愿可能导致成本增加.

具体来说, 我们首先通过一个关系矩阵 $\mathbf{W}_{m \times m}$ 来记录用户 u_i 和 u_j 之间的成对合作意愿 $w_{ij} \in [0, 1]$. 为了进一步降低问题和建模难度, 我们直接使用平均合作意愿来更新用户 u_i 加入已招募用户集合 μ 的边际效用 $\nabla f(\mu, u_i)$, 如下所示:

$$\nabla f(\mu, u_i) = f_o(u_i) \cdot \frac{\sum_{u_j \in \mu \setminus u_i} w_{ij}}{|\mu| - 1} \quad (1)$$

同样地, 较弱的合作意愿导致需要更多的奖励来鼓励合作和补偿隐私风险. 基于客观成本 $c_o(u_i)$, 我们也使用平均合作意愿来更新用户 u_i 加入已招募用户集合 μ 的边际成本 $\nabla c(\mu, u_i)$, 而不是使用简单的线性累积, 如下所示:

$$\nabla c(\mu, u_i) = c_o(u_i) + c_o(u_i) \cdot \left[1 - \frac{\sum_{u_j \in \mu \setminus u_i} w_{ij}}{|\mu| - 1} \right] \quad (2)$$

上述式(1)和(2)中的边际函数可以看作 $\nabla f(\mu, u) = f(\mu \cup u) - f(\mu)$ 和 $\nabla c(\mu, u) = c(\mu \cup u) - c(\mu)$. 显然, 因为客观效用和成本是非负的, 所以效用函数 $f(\cdot)$ 和成本函数 $c(\cdot)$ 是非递减的. 此外, 如图 2 中的两个示例所示, $f(\cdot)$ 和 $c(\cdot)$ 都可以是非子模的. 在示例 1 中, 我们首先想要将 u_4 添加到 μ 中, 但是 u_4 不想与其他用户合作. 因此, 边际效用 $\nabla f(\{u_2, u_3\}, u_4) = 0.1 \cdot f_o(u_4)$ 较小. 反之, 在我们招募了 u_4 的朋友 u_1 后, 招募的用户集合中包含了更多的用户, 但

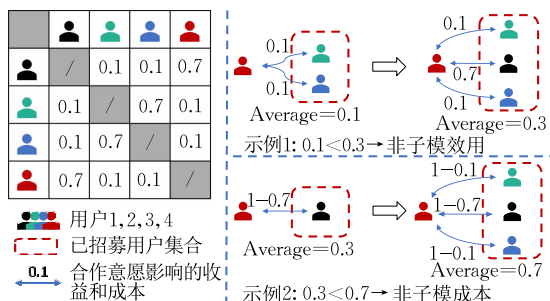


图 2 用户间合作意愿导致的非子模效用和成本

其平均合作意愿有所增加, 边际效用 $\nabla f(\{u_1, u_2, u_3\}, u_4) = 0.3 \cdot f_o(u_4)$ 也因此变大, 其打破了边际收益递减的性质, 体现了效用函数的非子模特性. 类似地, 我们有 $\nabla c(\{u_1\}, u_4) = 0.3 \cdot c_o(u_4) < \nabla c(\{u_1, u_2, u_3\}, u_4) = 0.7 \cdot c_o(u_4)$, 也体现了成本函数的非子模特性.

此外, 我们假设在离线和在线场景中, 关系矩阵都可以从历史数据或外源数据^[16]中获得. 本文仅提供一个具有非子模效用函数和成本函数的通用模型, 其他可能形式可以很容易地通过修改通用模型来应用于特定的用户招募问题. 同时, 我们提出的算法还可以应用于其他涉及非子模函数的应用, 如影响最大化^[44]、稀疏回归和字典选择^[45]等.

3.3 问题描述

问题(合作性移动群智感知中的用户招募): 给定一组移动群智感知任务 S 和一组用户关系矩阵 $\mathbf{W}$, 在有限制的预算 B 下, 从预先确定的用户池(离线场景)或动态参与序列 U (在线场景)中招募一组用户 μ , 目的是最大化其效用 $f(\mu)$:

$$\text{Maximize } f(\mu) \quad (3)$$

$$\text{Subject to } \mu \subseteq U, c(\mu) \leq B \quad (4)$$

我们在图 3 中提供了一个运行示例, 其中, 有若干用户四处移动以合作地执行五个感知任务. 在离线情况下(图 3 上半部分), 用户池是预先确定的, 我们尝试在考虑他们的合作意愿的情况下从中选择一部分用户, 以在有限预算下最大化所选用户集合的效用. 在在线情况下(图 3 下半部分), 用户按时间顺序申请加入. 因此, 在他们参与任务之前, 效用和成本是不可见的. 当第一个用户到来时, 虽然它可以执行一个任务, 但因其不愿同其他用户合作, 导致它的效用不高, 我们决定不招募该用户. 当第二和第三个

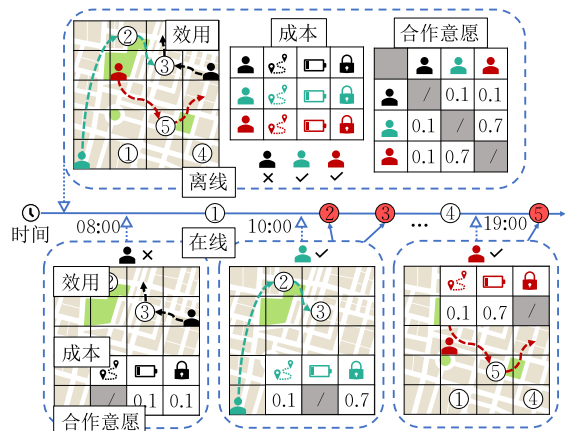


图 3 合作性移动群智感知中具有非子模效用和成本的离线与在线用户招募运行示例

用户参与时,我们发现他们不仅能执行较多任务,也愿意同他人合作,因此效用较高,我们决定招募他们以合作地完成任任务。

4 具有一般效用和成本的离线用户招募

4.1 广义贪婪算法

前人的工作已经证明了具有子模效用和线性成本的用户招募问题是 NP 难的^[11,14-15]。进一步考虑合作意愿,此时的效用函数和成本函数均可能存在更为复杂的非子模特性。为了解决这一问题,我们提出了一个广义的贪婪用户招募算法,其基本思想是迭代而贪婪地选取单位边际成本下边际效用最大的用户。详细的算法在算法 1 中进行了总结。我们首先记录效用与成本比率 $f(u')/c(u')$ 最大的最佳单用户 u' (该数值将应用于第 4.2 节中的理论竞争比分析)。接下来是本算法的核心部分,即迭代地计算每个用户相对于当前已招募用户集合的边际效用和边际成本的比率 $\delta_{u_i} = \nabla f(\mu', u_i) / \nabla c(\mu', u_i)$ (第 4 行),找到拥有最大比率 δ_{u_i} 的最佳用户 u_i (第 5 行),并将它添加到有限预算下的已招募用户集合 μ 中 (第 6 行)。最后,广义贪婪算法返回选定招募用户集合 μ 和最佳单用户 u' 中的较好结果。

算法 1. 广义贪婪算法。

输入: U, S, B

输出: μ

Let $u' = \arg \max_{u_i \in U} f(u_i) / c(u_i), c(u_i) \leq B$;

Initialize $\mu' = \emptyset$ and $U' = U$;

WHILE $U' \neq \emptyset$ DO

 Calculate $\delta_{u_i} = \nabla f(\mu', u_i) / \nabla c(\mu', u_i), \forall u_i \in U'$;

 Find $u^* = \arg \max_{u_i \in U'} \delta_{u_i}$;

 IF $c(\mu' \cup u_i) \leq B$ THEN

$\mu' \leftarrow \mu' \cup u_i$;

 END IF

$U' \leftarrow U' \setminus u^*$;

END WHILE

$\mu = \arg \max_{\mu \in \{\mu', u'\}} f(\mu)$.

此外,在离线场景中,我们进一步考虑用户出价真实性,在算法 1 中嵌入第二价格拍卖模型来确定每个招募用户的最终定价。具体地,在算法 1 的每一轮迭代中,额外记录该轮中次优用户的边际效用和边际成本的比率,即 $\delta_{u'} = \nabla f(\mu, u') / \nabla c(\mu, u')$,其中 $u' = \arg \max_{u_i \in U' \setminus u^*} \delta_{u_i}$ 。然后,使用次优用户的期望报酬

来确定最终招募用户的报酬,即 $p(\mu, u^*) = \nabla f(\mu, u^*) / \delta_{u'}$ 。此时,如用户 u^* 报价过高,它可能失去被招募的机会;如果报价过低,则产生亏损。因此,该策略可以确保用户根据其真实成本出价^[46]。

4.2 理论分析

理论分析部分的详细证明过程请参见附录。我们首先介绍子模比率和曲率的定义,如下:

定义 1. 子模比率. 一个非负函数 $f(\cdot)$ 的子模比率 α_f 为 $\alpha_f = \min \nabla f(\mu_1, u) / \nabla f(\mu_2, u)$, 其中 $\forall \mu_1 \subseteq \mu_2, u \notin \mu_2$ 。

定义 2. 曲率. 子模函数 $g(\cdot)$ 关于集合 U 的曲率 γ_g 为 $1 - \gamma_g = \min \nabla g(U \setminus u, u) / g(u)$, 其中 $\forall u \in U$ 。

定义 1 中的子模比率和定义 2 中的曲率分别刻画了非子模函数 $f(\cdot)$ 与子模函数的相似程度,以及子模函数 $g(\cdot)$ 与模函数的相似程度。需要注意的是,在实际应用中,根据子模比率和曲率的具体情况,本文所证得的近似比可能非常大。一方面,较差的子模比率和曲率往往意味着极为复杂的效用和成本函数,而现有的方法也只能尽力而为地争取获得最大收益,因此效果较差。另一方面,我们的理论近似比是预期结果的下界,即最差情况。本文实验部分也验证了我们的策略实际获得的结果远超所证得的理论近似比。

接下来,我们利用子模比率和曲率来尝试证明我们的广义贪婪算法 1 的有效性和理论近似保证。需要注意的是,由于效用函数和成本函数均为非子模函数,我们实际证得的是广义贪婪算法的一个双准则近似保证,即在一个稍微放松的预算约束下,与最优解相比,我们算法所能保证的最小近似比。基本证明思想是利用子模比率将非子模效用函数和成本函数近似为子模函数,用于估计边际收益。然后,进一步利用曲率把(非)子模成本函数近似为模函数以处理预算限制。

我们首先提出了一个基于子模比率的简单性质,该性质将应用于后续理论分析。

引理 1. 给定 $f(\cdot)$ 的 α_f , 我们有 $\alpha_f \sum_{u \in \mu_2 \setminus \mu_1} \nabla f(\mu_2 \setminus u, u) \leq f(\mu_2) - f(\mu_1) \leq \alpha_f^{-1} \sum_{u \in \mu_2 \setminus \mu_1} \nabla f(\mu_1, u)$, 其中 $\forall \mu_1 \subseteq \mu_2 \subseteq U$ 。

接下来,我们关注非子模成本函数,首先衡量总成本与各用户成本线性累积总和之间的关系。

引理 2. 给定 $c(\cdot)$ 的 α_c 和 $\gamma_c(\mu)$, 集合 μ 的总成本可以达到 $c(\mu) \geq \frac{1 + \alpha_c^2(|\mu| - 1)(1 - \gamma_c)}{|\mu|} \cdot \sum_{u \in \mu} c(u)$ 。

将 K 定义为可行解中可招募用户数的最大值, 其中 $\forall \mu \subseteq U, c(\mu) \leq B$. 设 μ_i 为算法 1 第 i 次迭代后招募的用户组. 然后, 我们证明了每个被招募的用户至少可以获得如下的边际增益.

引理 3. 对于每个已招募用户集合 μ_i , 我们均可获得一个边际增益

$$f(\mu_i) - f(\mu_{i-1}) \geq \alpha_f \frac{c(\mu_i) - c(\mu_{i-1})}{B} (f(\hat{\mu}) - f(\mu_{i-1})),$$

其中 $\hat{\mu}$ 是 $\hat{B} = \alpha_c \frac{1 + \alpha_c^2 (K-1)(1-\gamma_c)}{K} B$ 下的最优集.

最后, 我们计算了总效用以证明我们的算法能达到一个双准则近似保证:

定理 1. 具有非子模效用和成本函数的广义贪婪算法可以实现 $\frac{\alpha_f}{2}(1 - e^{-\alpha_f})$ 的双准则近似比.

5 具有一般效用和成本的在线用户招募

5.1 动态分段策略

与离线场景相比, 在线场景中最大的区别和困难在于未知的未来信息, 即在用户参与之前, 其效用和成本是不可见的. 考虑到合作意愿导致的非子模效用和成本, 在线用户招募变得更具挑战性. 为了降低问题求解难度, 我们首先在在线招募过程之前做出一个满足预算约束的假设:

假设 1. 参与用户数量具有周期性, 其感知成本分布是独立的.

许多现有的应用程序和数据集都展示了类似的统计结果, 例如用于推荐应用的周期性签到记录^[47]等. 同时, 除了公开数据集, 我们还在校园环境构建了一个名为 Campus 的真实移动群智感知应用, 通过招募学生执行校园 PoI 人流监测, 并利用相关数据来验证假设 1 的合理性和正确性. 如图 4 所示, 我们首先在子图(a)中展示了不同时刻的周期性签到记录数据, 然后在子图(b)中展示了参与 Campus 应用的不同学院的用户比例, 这可以简单视作用户成本的分布. 值得注意的是, 各学院的总体比例大致相同, 而每个学院也有其独立分布. 在假设 1 的帮助

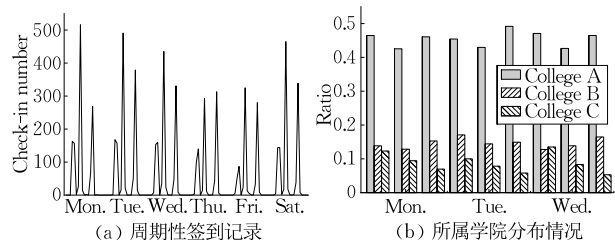


图 4 Campus 数据集中的周期性签到记录与分布情况

下, 我们利用周期性由历史数据建立一些虚拟用户集来模拟未来的用户序列, 然后利用广义贪婪算法对其进行在线招募前的期望招募用户数量估计.

基于上述所估计的预期招募用户数量 k , 在线用户招募问题可以简化为经典的多选择秘书问题^[24], 其目标是选择 k 个秘书以最大化他们的总效用. 考虑子模效用函数, Bateni 等人^[25]提出了经典多秘书问题的一个变种, 即具有子模效用函数的子模 k -秘书问题. 在本文中, 我们进一步考虑非子模的效用和成本函数, 将子模秘书问题进一步拓展为双-非子模多秘书问题, 并提出了一种分段在线招募策略. 如图 5 和算法 2 所示, 我们所提出的分段策略的基本思想是使用估计的预期招募人数 k 来满足非子模成本限制, 再将实时用户序列划分为 k 个等大小的分段, 从每个分段中选出一个当前边际效用最大的最优用户. 具体来讲, 我们首先根据历史数据估计预期参与用户总数 m' 和被招募用户数 k' (第 1 行). 对于特殊情况, 如 $k' = 0$ 和 $m > m'$, 我们直接根据当前剩余预算进行招募 (第 2 行和第 6 行). 然后, 作为该算法的核心, 我们在每个分段中观察前 $\lfloor l/e \rfloor$ 个用户, 并在分段后续到来的用户中, 选择一个边际效用大于所有观察用户的实时用户进行招募 (第 8 行至第 14 行). 更进一步地, 为了纠正估计误差并充分利用新获得的信息, 我们还对算法 2 进行了简单而有效的调整, 具体内容展示在算法 3 中, 其基本思想是在招募新用户后重新估计后续的预期招募用户数量 k 并重新划分后续分段. 最后, 我们从真实的用户序列中分段地招募到了用户集合 μ .

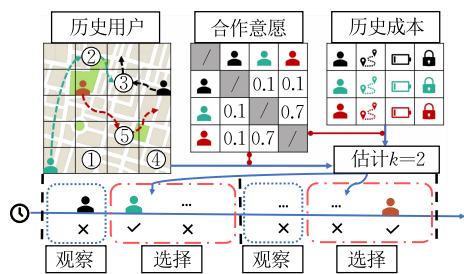


图 5 分段在线策略的一个示例

算法 2. 分段在线策略.

输入: $U = \{u_1, u_2, \dots, u_m\}, S, B, U_h, T$

输出: μ

Estimate m' and k' from historical data U_h according to the time T and the periodicity; {If $k' = 0$, return the first u_i , where $c(u_i) \leq B$ }

Initialize $l = \lfloor m'/k' \rfloor$, $l_{ob} = \lfloor l/e \rfloor$, and $\mu = \emptyset$;

WHILE $U \neq \emptyset$ and $B > 0$ DO

```

Wait until the next user  $u_i$  participate in MCS;
IF  $i > m'$  and  $c(\mu \cup u_i) \leq B$  THEN
     $\mu \leftarrow \mu \cup u_i$ ;
END IF
IF  $i = \lfloor i/l \rfloor \cdot l + 1$  THEN
     $\epsilon_{u_i} = \epsilon_{u_c} = 0$ ;
END IF
Calculate  $\delta_{u_i} = \nabla f(\mu, u_i)$ ,  $\delta_{u_c} = \nabla f(\mu, u_i) / \nabla c(\mu, u_i)$ ;
 $\epsilon = \delta_{u_i}$ ;
IF  $i \leq \lfloor i/l \rfloor \cdot l + l_{ob}$  THEN
    IF  $\delta_{u_i} > \epsilon_{u_i}$  THEN
         $\epsilon_{u_i} = \delta_{u_i}$  and  $\epsilon_{u_c} = \delta_{u_c}$ ;
    ELSE IF  $\delta_{u_i} = \epsilon_{u_i}$  THEN
         $\epsilon_{u_c} = \max\{\epsilon_{u_c}, \delta_{u_c}\}$ 
    END IF
ELSE IF  $i > \lfloor i/l \rfloor \cdot l + l_{ob}$  THEN
    IF  $\delta_{u_i} \geq \epsilon_{u_i}$  and  $\delta_{u_c} \geq \epsilon_{u_c}$  and  $c(\mu, u_i) \leq B$  THEN
         $\mu \leftarrow \mu \cup u_i$ ; {Break at Alg. 3}
        Skip to  $i < \lfloor i/l \rfloor \cdot l + 1$ ;
    END IF
END IF
END WHILE

```

算法 3. 动态分段策略.

输入: $U = \{u_1, u_2, \dots, u_m\}$, S, B, U_h, T

输出: μ

Initialize $\mu = \emptyset$, $U' = U$, $S' = S$, $B' = B$, $T' = T$;

WHILE $U \neq \emptyset$ and $B' > 0$ DO

$\mu' = \text{Segmented}(U', S', B', U_h, T')$

$\mu \leftarrow \mu \cup \mu'$;

Update U', S', B', T' ;

END WHILE

此外,在在线场景中,我们同样提供了一个可轻量化嵌入算法 2 和算法 3 的真实定价机制. 该机制同离线场景下的定价机制类似,其基本思想也是基于第二价格拍卖模型,即为每一位招募用户寻找一个次优解作为最终定价依据. 在在线场景中,虽然我们无法遍历每轮迭代中的所有候选用户,但在算法 2 和算法 3 中,每轮招募需要先观察并记录前 $\lfloor l/e \rfloor$ 个用户的最大边际效用和成本比率作为阈值 ϵ_{u_c} ,其天然地成为该轮招募的近似次优解. 因此,我们可利用该阈值,结合第二价格拍卖模型,给出相应的真实定价 $p(\mu, u^*) = \delta_{u^*} / \epsilon_{u_c}$. 同离线机制类似,此时,如用户 u^* 报价过高,它可能不会被选中;如果报价过低,则产生亏损. 因此,我们的在线招募策略可以保证价格的真实性.

5.2 理论分析

理论分析部分的详细证明过程请参见附录. 与离线策略类似,我们的在线策略分析同样利用子模比率使非子模函数近似于子模函数,从而获得我们的在线策略相比最优策略的近似比. 首先,在给定估计的预期招募用户数量 k 的情况下,我们将整个用户集 U 分成大小相等的 k 段,记为 $U_1, \dots, U_k$. 假设 μ^* 是最优用户组,其基数 $k = |\mu^*|$. 如果 $U_i \cap \mu^* \neq \emptyset$,则将 v_i 设为 $U_i \cap \mu^*$ 中的一个随机用户,我们删除 $v_i = \emptyset$ 并将其重新排序为 $v = \{v_1, \dots, v_{k_v}\}$.

接下来,我们首先分析并展示我们的分段策略的有效性,即从每个用户分段中选择一个分段最优用户实际上意味着我们已经选择到了最优用户集中的很多用户.

引理 4. 给定预期招募用户数量 k ,我们预期有 $|v| \geq (1 - 1/e)k$.

接下来,我们来证明了部分最优用户 μ' 与最优集合 μ^* 相比,已经能够得到一个较好的近似比.

引理 5. 对于 μ^* 的随机子集 μ' ,我们预期有 $f(\mu') \geq \alpha_f \cdot \frac{|\mu'|}{k^*} \cdot f(\mu^*)$.

接下来是在线策略理论分析的核心,我们证明了在非子模效用的情况下,对每个分段进行局部优化就可以得到一个全局近似.

引理 6. 给定 k 和 $|v|$,我们的分段在线策略可以实现一个近似比,即 $E[f(\mu_k)] \geq \alpha_f^2 \cdot \frac{|v|}{7k} \cdot f(\mu_k^*)$.

然后,我们将引理 4 中 $|v|$ 的期望值引入引理 6,得到了在确定的预期招募用户数量 k 下的近似保证.

引理 7. 给定 k ,我们的分段在线策略可以达到一个近似 $E[f(\mu_k)] \geq \frac{\alpha_f^2(1-1/e)}{7} f(\mu_k^*)$.

最后,利用引理 7、假设 1 和定理 1,我们进一步去掉 k 以证明分段在线策略在定理 2 中的近似比.

定理 2. 分段在线策略可以达到的预期逼近比为 $\frac{\alpha_f^3(1-e^{-1})(1-e^{-\alpha_f})}{14}$.

为了修正估计误差(尤其是预期招募用户数量 k)并充分利用新信息,我们在每次招募新用户后,对预期招募用户数量 k 进行重新估计并重置分段策略,成为动态策略. 显然地,动态策略在预期上优于分段策略. 此外,本文的目标是在有限的预算下实现效用最大化,实际上,通过修改效用和成本函数,它可以很容易地转化为“效用-成本”算法,即在保证良好效用的同时,使总成本最小化.

6 性能评估

6.1 实验设置

数据集. 我们从三个城市和一个校园收集了四组真实数据集, 分别是 Feeder^[48]、GeoLife^[49]、City 和 Campus. 我们在这些数据集上对本文所提出的离线和在线用户招募方法进行了评估与验证. 前三个数据集主要包括从三个不同城市收集的出租车、手机和私家车的 GPS 数据. 如图 6 所示, 我们将城市区域划分为 10×15 个网格, 在网格中生成任务, 并利用用户的轨迹来衡量其是否能按时经过, 以衡量用户效用. 对于合作意愿, 我们在基于 GeoLife 数据集的实验中, 随机生成用户关系矩阵. 该矩阵是对称矩阵, 其每个元素代表该用户对的合作意愿, 取值范围是 $[0, 1]$. 在 Feeder 和 City 两个数据集中, 除了用户轨迹数据外, 我们还拥有对应用户的通话记录. 因此, 我们利用用户间相互呼叫记录来建立关系矩阵. 具体地, 我们简单地将平均每天通话次数作为用户对的合作意愿, 并将通话次数大于 1 的用户对置为 1, 将无通话记录的用户对置为 0. 我们还使用用户的行程距离作为客观成本. 另外, 我们还在校园环境构建了一个名为 Campus 的真实校园移动群智感知系统, 在这个系统中, 学生们被招募并在校园里执行各个 PoI(如食堂、图书馆等)的人流量监测任务. 在这个数据集的实验中, 我们认为来自同一学院、年级、专业和宿舍楼的用户有更强的合作意愿, 进而以此构建用户关系矩阵, 即将合作意愿划分为五档, 根据是否来自相同学院等将用户对的合作意愿置为 $0/0.25/0.5/0.75/1$.

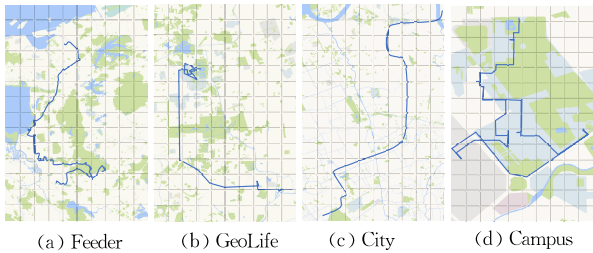


图 6 用户轨迹和网格划分的示例

对比方法. 由于具有非子模效用和非子模成本的用户招募与现有的工作有很大的不同, 我们首先建立了一个在预算约束下随机选择用户作为基线的“RAN”算法. 然后对现有的在线策略^[31-32]进行了修正, 并实现了“OMZ”策略. 该策略首先将总时间 T 和预算分为 $\lfloor \log_2 T \rfloor + 1$ 个阶段, 然后根据前面的阶段计算阈值, 并根据预算的相应比例限制, 在各个阶段中分别招募用户. 此外, 由于在离线和在线场景中, 除了决策时间之外具有相似的设置, 我们也同时评估了广义离线贪婪算法(称作“OFF”), 不重新调整的分段在线策略(称作“SEG”)和动态在线策略(称作“ON”)以及其他方法.

6.2 评估结果

6.2.1 完成的任务

如图 7~图 10 和表 2 所示, 我们首先评估主要指标, 即完成任务的数量. 我们分别改变用户和任务的数量, 以及客观效用和成本的平均值, 同时保持其他参数固定不变, 并将预算设置为 200. 需要注意的是, 我们为每个用户添加了一个活跃时间, 这意味着他们只能在活跃时间内工作, 以便度量平均的客观效用. 四个数据集的结果取得了相似的趋势, 表明我们提出的离线和在线方法可以处理非子模效用问题, 并且具有良好的性能.

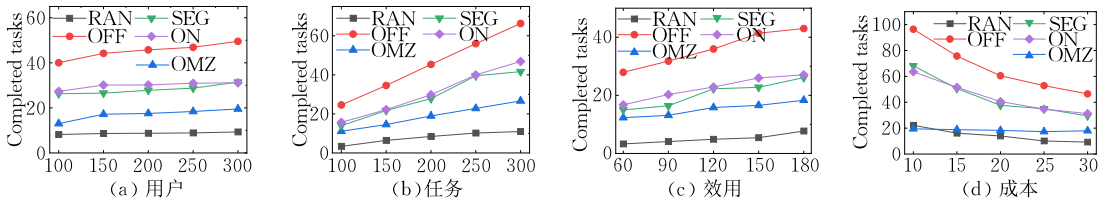


图 7 Feeder 的主要结果

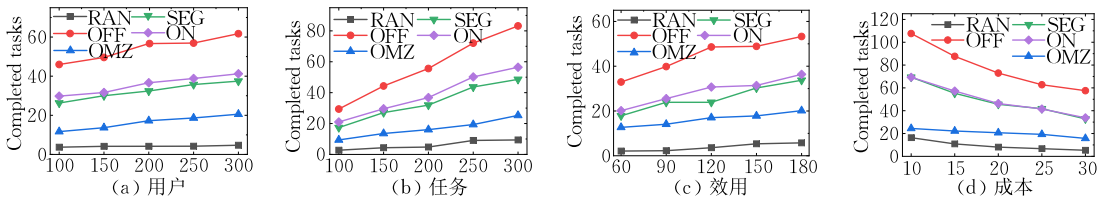


图 8 GeoLife 的主要结果

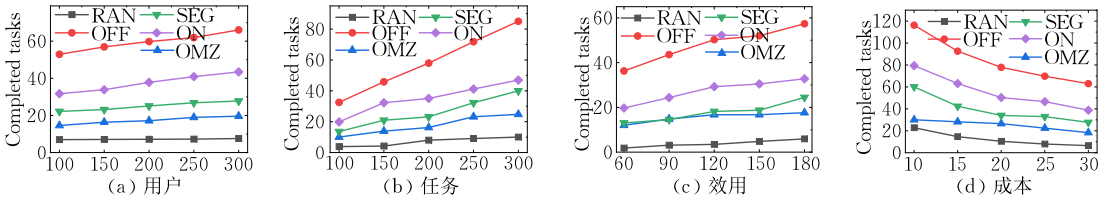


图 9 City 的主要结果

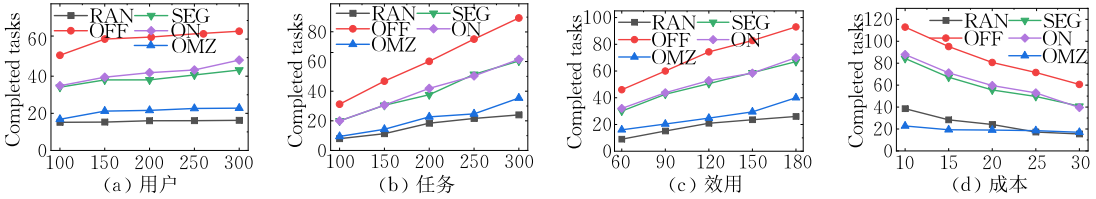


图 10 Campus 的主要结果

表 2 完成任务数量和近似比率

	RAN	OMZ	SEG	ON/OFF
Feeder	11.02	20.62	40.00	43.16/62.34(69%)
GeoLife	8.52	19.96	45.14	45.22/68.68(66%)
City	12.18	21.86	53.36	53.46/76.98(69%)
Campus	22.32	22.20	53.20	55.06/80.32(69%)

具体来说,我们的 OFF、SEG 和 ON 总是比其他方法好,这表明了我们的离线和在线方法的有效性. 注意,ON 通常能比 SEG 完成更多的任务,因为 ON 可以纠正估计误差并利用新信息. 此外,比较图 7~图 10 的子图(a)和其他子图,我们发现子图(a)中用户的增长率低于其他子图. 原因是当我们已经有足够多的候选用户时,再增加用户也不能显著提高所招募用户集合的性能了. 我们还在表 2 中展示了 ON 和 OFF 的已完成任务数量的近似比,结果说明我们的 ON 可以达到 OFF 的高竞争比($\sim 70\%$). 此外,图中效益和成本曲线有一些波动,原因可能是合作意愿对相关结果有一些影响.

6.2.2 预算限制

我们接下来进一步比较并分析方法的性能,首先评估了所消耗的预算和相应的完成任务的数量,结果展示在图 11 中. 我们发现离线算法,即 RUN 和 OFF,可以更好地利用有限的预算,或者说使用了更多的预算. 然而,因为 RUN 忽略了效用,所有性能最差. 需要注意的是 OMZ 消耗的预算总是最少的,原因是因为成比例分段的预算很难处理非子模的成本函数. 同时,ON 可以在招募新用户后及时调整,从而可以比 SEG 表现得更好.

此外,为了证明我们的在线策略可以处理非子模成本函数,我们用之前时间段所消耗的预算和虚拟用户集的平均边际成本来说明估计的 k . 如图 12 所示,我们可以看到大部分估计的 k 值是相同的,但也存在较小的波动. 同时,预算有较大的波动,而平均边际成本非常接近,其原因可能是在线场景中存在很多不确定性,而招募的用户在边际效用和成本方面大致相同.

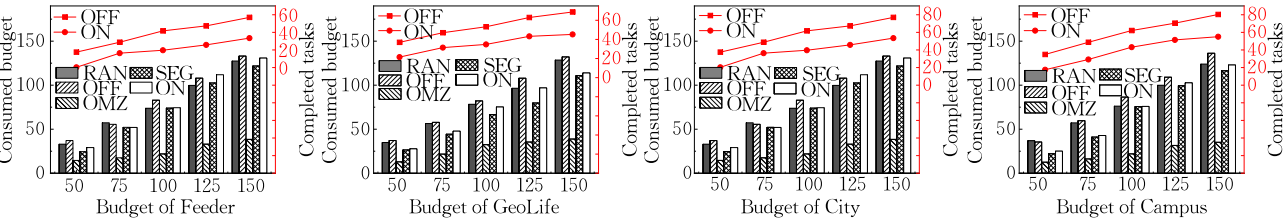


图 11 不同预算限制下预算消耗和完成任务数量的变化

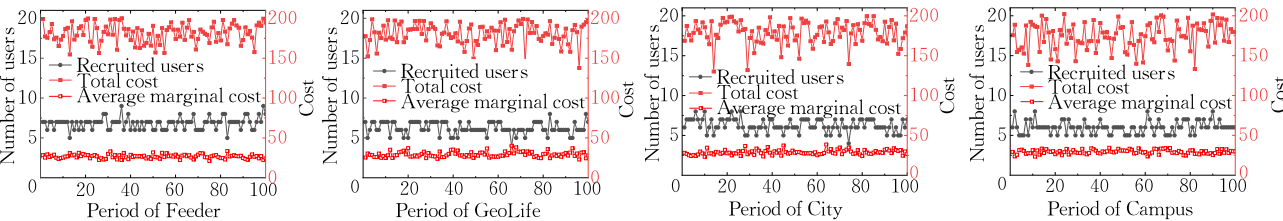


图 12 不同时间段下用户数量和成本的变化

6.2.3 真实定价

为了验证基于第二价格模型的真实定价,我们提供了 Campus 中的两个例子.如图 13 所示,我们首先选择一个真实价格为 23.2 的用户,然后把它的出价从 20 改为 26.显而易见,如果投标价格高于真实价格,用户将失去所有的回报.此外,我们还将所有用户的回报与其成本进行了比较,其中真实的回报总是高于用户的成本,原因在于我们需要支付更多报酬来确保真实出价和个人理性.此外,图 13 还显示,我们实际上选择了成本较小的用户,而且支付的价格也不会超过成本太多.

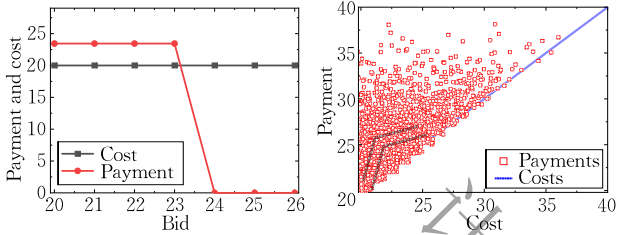


图 13 真实定价和个人理性

6.2.4 运行时间

最后,我们在表 3 中展示了所提出算法的运行时间,实验平台配置为 Intel(R) Xeon(R) CPU E5-2630 v4@2.20 GHz 和 32 GB RAM.我们可以看到,离线方法大约消耗 0.7~3.5 ms,而在线策略需要 2.5~9.2 ms.注意,OMZ 花费的时间最多,因为它的阈值需要反复计算.此外,Campus 成本低于其他三个数据集,因为其数据量最少.同时,在线策略中 k 的估计可以在线下进行以便实际应用.

	表 3 运行时间 (单位:ms)				
	RAN	OFF	SEG	ON	OMZ
Feeder	1.264	2.584	3.530	6.574	25.241
GeoLife	1.268	3.580	4.310	7.346	25.320
City	1.399	3.597	4.528	9.242	117.260
Campus	0.726	2.220	2.531	5.383	20.316

7 讨 论

7.1 合作意愿

在合作性移动群智感知中,用户间合作意愿对用户执行任务的效用和成本有着非常重要的影响.为了不失一般性,本文主要关注由合作意愿所致的非子模效用和成本函数.对于合作意愿本身,仅采用简单的对称关系矩阵来记录用户间成对的合作意愿,这种方法可能会带来性能瓶颈.一方面,在面对海量用户的大规模移动群智感知系统时,大型关系

矩阵的存储空间和计算时间成本极高.我们可以利用矩阵的稀疏性、低秩属性,以及将用户分类、聚簇等多种方式将大矩阵针对性地分解或划分为多个小矩阵,并利用并行计算等方式加速计算.另一方面,用户之间往往缺乏关联,且多用户间合作意愿可能更为复杂,导致关系矩阵构建较为困难.我们需要应用如模糊理论、知识图谱、矩阵补全等新兴技术来帮助构建和完善关系矩阵.

7.2 招募策略

在分段在线用户招募过程中,对于每个用户分段,我们需要观察并抛弃前 $1/e \cdot 100\%$ 的用户,然后在后续到来的用户中进行在线招募.虽然我们在第 5.2 节的理论分析中证明了这种分段遴选能够得到全局最优解的近似保证,但对于观察阶段的用户仍旧是不公平的,且存在最优用户被抛弃、策略性用户拖延至后半段再加入等问题.为了解决该问题,我们需要利用前向搜索、先知不等式等理论和方法,从历史数据中采样并构建一个近似的观察用户集合.此时,我们不再需要抛弃前 $1/e \cdot 100\%$ 的用户,仅需利用近似的观察用户集合来获取阈值,进而在完整的用户分段中进行在线招募.

8 总 结

本文研究了具有一般效用和成本下的合作性移动群智感知用户招募问题.为了处理非子模函数,我们首先提出了一个广义的离线贪婪算法,即从一个预先确定的用户池中挑选合适用户.通过引入子模比率和曲率,该离线算法得到了 $\frac{\alpha_f}{2}(1-e^{-\alpha_f})$ 的双准则近似比.进一步将其扩展到在线场景,我们提出了一种分段在线策略,首先对用户序列进行划分,然后再分段地对用户进行招募.在周期性假设下,该在线策略可以实现的一个期望的逼近比 $\frac{\alpha_f^3(1-e^{-1})(1-e^{-\alpha_f})}{14}$.最后,基于四个真实数据集,我们对所提出方法进行了广泛评估并验证了其有效性.

参 考 文 献

[1] Ganti R K, Ye F, Lei H. Mobile crowdsensing: Current state and future challenges. IEEE Communications Magazine, 2011, 49(11): 32-39

[2] Liu Yun-Hao. Crowdsourcing computing. Communications of the China Computer Federation, 2012, 8(10): 38-41 (in Chinese)

- (刘云浩. 群智感知计算. 中国计算机学会通讯, 2012, 8(10): 38-41)
- [3] Zhang D, Wang L, Xiong H, et al. 4W1H in mobile crowd sensing. *IEEE Communications Magazine*, 2014, 52(8): 42-48
 - [4] Aly H, Basalamah A, Youssef M. Automatic rich map semantics identification through smartphone-based crowd-sensing. *IEEE Transactions on Mobile Computing*, 2017, 16(10): 2712-2725
 - [5] Hu S, Su L, Liu H, et al. SmartRoad: Smartphone-based crowd sensing for traffic regulator detection and identification. *ACM Transactions on Sensor Networks(TOSN)*, 2015, 11(4): 1-27
 - [6] Rana R K, Chou C T, Kanhere S S, et al. Ear-phone: An end-to-end participatory urban noise mapping system//Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks. Stockholm, Sweden, 2010: 105-116
 - [7] Ouyang R W, Kaplan L M, Tonfeto A, et al. Aggregating crowdsourced quantitative claims: Additive and multiplicative models. *IEEE Transactions on Knowledge and Data Engineering*, 2016, 28(7): 1621-1634
 - [8] Shi F, Wu D, Arkhipov D I, et al. ParkCrowd: Reliable crowdsensing for aggregation and dissemination of parking space information. *IEEE Transactions on Intelligent Transportation Systems*, 2018, 20(11): 4032-4044
 - [9] Chen H, Guo B, Yu Z, et al. CrowdTracking: Real-time vehicle tracking through mobile crowdsensing. *IEEE Internet of Things Journal*, 2019, 6(5): 7570-7583
 - [10] Kulshrestha T, Saxena D, Niyogi R, et al. Real-time crowd monitoring using seamless indoor-outdoor localization. *IEEE Transactions on Mobile Computing*, 2020, 19(3): 664-679
 - [11] Karaliopoulos M, Telelis O, Koutsopoulos I. User recruitment for mobile crowdsensing over opportunistic//Proceedings of the IEEE Conference on Computer Communications (IEEE INFOCOM 2015). Hong Kong, China, 2015: 2254-2262
 - [12] Liu Yan, Guo Bin, Wu Wen-Le, et al. Multitask-oriented participant selection in mobile crowd sensing. *Chinese Journal of Computers*, 2017, 40(8): 1872-1887(in Chinese)
(刘琰, 郭斌, 吴文乐等. 移动群智感知多任务参与者优选方法研究. 计算机学报, 2017, 40(8): 1872-1887)
 - [13] Yang Shuo, Wu Fan, Chen Gui-Hai. On designing most informative user selection methods for mobile crowdsensing. *Chinese Journal of Computers*, 2020, 43(3): 409-422(in Chinese)
(杨朔, 吴帆, 陈贵海. 移动群智感知网络中信息量最大化的用户选择方法研究. 计算机学报, 2020, 43(3): 409-422)
 - [14] Wang E, Yang Y, Wu J, et al. An efficient prediction-based user recruitment for mobile crowdsensing. *IEEE Transactions on Mobile Computing*, 2018, 17(1): 16-28
 - [15] Yang Y, Liu W, Wang E, et al. A prediction-based user selection framework for heterogeneous mobile crowdsensing. *IEEE Transactions on Mobile Computing*, 2019, 18(11): 2460-2473
 - [16] Yang G, He S, Shi Z, et al. Promoting cooperation by the social incentive mechanism in mobile crowdsensing. *IEEE Communications Magazine*, 2017, 55(3): 86-92
 - [17] Gao L, Tang M, Pang H, et al. Multi-user cooperative mobile video streaming: Performance analysis and online mechanism design. *IEEE Transactions on Mobile Computing*, 2019, 18(2): 376-389
 - [18] Li Y, Li F, Yang S, et al. MP-Coopetition: Competitive and cooperative mechanism for multiple platforms in mobile crowd sensing. *IEEE Transactions on Services Computing*, 2019, 14(6): 1864-1876
 - [19] Zhang Y, Li M, Yang D, et al. Tradeoff between location quality and privacy in crowdsensing: An optimization perspective. *IEEE Internet of Things Journal*, 2020, 7(4): 3535-3544
 - [20] Liu W, Yang Y, Wang E, et al. Dynamic user recruitment with truthful pricing for mobile crowdsensing//Proceedings of the IEEE Conference on Computer Communications (IEEE INFOCOM 2020). Virtual, 2020: 1-10
 - [21] Jin H, Guo H, Su L, et al. Dynamic task pricing in multi-requester mobile crowd sensing with Markov correlated equilibrium//Proceedings of the IEEE Conference on Computer Communications (IEEE INFOCOM 2019). Paris, France, 2019: 1063-1071
 - [22] Tian X, Zhang W, Yang Y, et al. Toward a quality-aware online pricing mechanism for crowdsensed wireless fingerprints. *IEEE Transactions on Vehicular Technology*, 2018, 67(7): 5958-5964
 - [23] Xiao M, Wu J, Huang L, et al. Online task assignment for crowdsensing in predictable mobile social networks. *IEEE Transactions on Mobile Computing*, 2017, 16(8): 2306-2320
 - [24] Preetar J. On multiple choice secretary problems. *Mathematics of Operations Research*, 1994, 19(3): 597-602
 - [25] Bateni M, Hajiaghayi M T, Zadimoghaddam M. Submodular secretary problem and extensions. *ACM Transactions on Algorithms*, 2013, 9(4): 32:1-32:23
 - [26] Song Z, Liu C H, Wu J, et al. QoI-aware multitask-oriented dynamic participant selection with budget constraints. *IEEE Transactions on Vehicular Technology*, 2014, 63(9): 4618-4632
 - [27] Bhatti S S, Fan J, Wang K, et al. An approximation algorithm for bounded task assignment problem in spatial crowdsourcing. *IEEE Transactions on Mobile Computing*, 2020: 1-1
 - [28] Pu L, Chen X, Xu J, et al. Crowdlet: Optimal worker recruitment for self-organized mobile crowdsourcing//Proceedings of the IEEE Conference on Computer Communications (IEEE INFOCOM 2016). San Francisco, USA, 2016: 1-9
 - [29] Ji Sheng-Gong, Zheng Yu, Wang Zhao-Yuan, Li Tian-Rui. Look-ahead search and voting-based dynamic participant recruitment in mobile crowd sensing. *Chinese Journal of Computers*,

- 2021, 44(10): 1998-2015(in Chinese)
- (纪圣杰, 郑宇, 王诏远, 李天瑞. 基于前向搜索和投票的移动群智感知动态用户招募方法. 计算机学报, 2021, 44(10): 1998-2015)
- [30] Liu W, Yang Y, Wang E, et al. Dynamic online user recruitment with (non-) submodular utility in mobile crowdsensing. *IEEE/ACM Transactions on Networking*, 2021, 29(5): 2156-2169
- [31] Zhao D, Li X, Ma H. Budget-feasible online incentive mechanisms for crowdsourcing tasks truthfully. *IEEE/ACM Transactions on Networking*, 2016, 24(2): 647-661
- [32] Gao H, Liu C H, Tang J, et al. Online quality-aware incentive mechanism for mobile crowd sensing with extra bonus. *IEEE Transactions on Mobile Computing*, 2019, 18(11): 2589-2603
- [33] Wang E, Wang H, Yang Y, et al. Truthful incentive mechanism for budget-constrained online user selection in mobile crowdsensing. *IEEE Transactions on Mobile Computing*, 2021: 1-1
- [34] Xiang Chao-Can, Li Yao-Yu, Feng Liang, et al. Near-optimal vehicular crowdsensing task allocation empowered by deep reinforcement learning. *Chinese Journal of Computers*, 2022, 45(5): 918-934(in Chinese)
- (向朝参, 李耀宇, 冯亮等. 基于深度强化学习的车联网汽车感知任务分配. 计算机学报, 2022, 45(5): 918-934)
- [35] Han L, Yu Z, Yu Z, et al. Online organizing large-scale heterogeneous tasks and multi-skilled participants in mobile crowdsensing. *IEEE Transactions on Mobile Computing*, 2021: 1-1
- [36] Wang X, Jia R, Fu L, et al. Online spatial crowdsensing with expertise-aware truth inference and task allocation. *IEEE Journal on Selected Areas in Communications*, 2022, 40(1): 412-427
- [37] Xiang C, Li Y, Zhou Y, et al. A comparative approach to resurrecting the market of mod vehicular crowdsensing// *Proceedings of the IEEE Conference on Computer Communications (IEEE INFOCOM 2022)*. Virtual, 2022: 1-10
- [38] Fujishige S. *Submodular Functions and Optimization*. Amsterdam: Elsevier, 2005
- [39] Krause A, Guestrin C. A note on the budgeted maximization of submodular functions. Carnegie Mellon University, Pittsburgh, 2005
- [40] Lin H, Bilmes J. Multi-document summarization via budgeted maximization of submodular functions// *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Los Angeles, USA, 2010: 912-920
- [41] Sviridenko M. A note on maximizing a submodular set function subject to a knapsack constraint. *Operations Research Letters*, 2004, 32(1): 41-43
- [42] Zhang H, Vorobeychik Y. Submodular optimization with routing constraints// *Proceedings of the 30th AAAI Conference on Artificial Intelligence*. Phoenix, USA, 2016: 819-826
- [43] Qian C, Shi J C, Yu Y, et al. On subset selection with general cost constraints// *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. Melbourne, Australia, 2017: 2613-2619
- [44] Wu J, Wang N. Approximating special social influence maximization problems. *Tsinghua Science and Technology*, 2020, 25(6): 703-711
- [45] Das A, Kempe D. Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection// *Proceedings of the 28th International Conference on Machine Learning*. Bellevue, Washington, USA, 2011: 1057-1064
- [46] Yang D, Xue G, Fang X, et al. Incentive mechanisms for crowdsensing: Crowdsourcing with smartphones. *IEEE/ACM Transactions on Networking*, 2016, 24(3): 1732-1744
- [47] Wang H, Terrovitis M, Mamoulis N. Location recommendation in location-based social networks using user check-in data// *Proceedings of the 21st SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL 2013)*. Orlando, USA, 2013: 364-373
- [48] Zhang D, Zhao J, Zhang F, et al. Feeder: Supporting last-mile transit with extreme-scale urban infrastructure data// *Proceedings of the 14th International Conference on Information Processing in Sensor Networks (IPSN'15)*. New York, USA, 2015: 226-237
- [49] Zheng Y, Li Q, Chen Y, et al. Understanding mobility based on GPS data// *Proceedings of the 10th International Conference on Ubiquitous Computing (UbiComp'08)*. New York, USA, 2008: 312-321

附录 A. 具有一般效用和成本的离线用户招募.

我们首先提出了一个基于子模比率的简单性质, 该性质将应用于后续理论分析.

引理 1. 给定 $f(\cdot)$ 的 α_f , 我们有 $\alpha_f \sum_{u \in \mu_2 \setminus \mu_1} \nabla f(\mu_2 \setminus u, u) \leq$

$$f(\mu_2) - f(\mu_1) \leq \alpha_f^{-1} \sum_{u \in \mu_2 \setminus \mu_1} \nabla f(\mu_1, u), \text{ 其中 } \forall \mu_1 \subseteq \mu_2 \subseteq U.$$

证明. 设 $k_\mu = |\mu_2 \setminus \mu_1|$, 将 $\{A_i\}_{i=0}^{k_\mu}$ 定义为 $A_0 = \emptyset, A_{k_\mu} = \mu_2 \setminus \mu_1$, 并且 $u_i = A_i \setminus A_{i-1}$. 给定 $f(\cdot)$ 的 α_f , 我们有

$$\begin{aligned}
f(\mu_2) - f(\mu_1) &= \sum_{i=1}^{k_\mu} [f(\mu_1 \cup A_i) - f(\mu_1 \cup A_{i-1})] \\
&= \sum_{i=1}^{k_\mu} \nabla f(\mu_1 \cup A_i \setminus u_i, u_i) \\
&\geq \alpha_f \sum_{i=1}^{k_\mu} \nabla f(\mu_2 \setminus u_i, u_i) \quad (5)
\end{aligned}$$

$$\begin{aligned}
f(\mu_2) - f(\mu_1) &= \sum_{i=1}^{k_\mu} \nabla f(\mu_1 \cup A_{i-1}, u_i) \\
&\leq \alpha_f^{-1} \sum_{i=1}^{k_\mu} \nabla f(\mu_1, u_i) \quad (6)
\end{aligned}$$

由此,我们根据定义 1 得到了式(5)和(6).

证毕.

接下来,我们关注非子模成本函数,首先衡量总成本与各用户成本线性累积总和之间的关系:

引理 2. 给定 $c(\cdot)$ 的 α_c 和 $\gamma_c(\mu)$, 集合 μ 的总成本可以

$$\text{达到 } c(\mu) \geq \frac{1 + \alpha_c^2(|\mu| - 1)(1 - \gamma_c)}{|\mu|} \cdot \sum_{u \in \mu} c(u).$$

证明. 根据引理 1, 我们有

$$c(\mu) - c(u) \geq \alpha_c \sum_{v \in \mu \setminus u} \nabla c(\mu \setminus v, v) \quad (7)$$

其中 $\forall u \in \mu$. 然后,把所有 $u \in \mu$ 累加,得到

$$\begin{aligned}
|\mu|c(\mu) - \sum_{u \in \mu} c(u) &\geq \alpha_c \sum_{u \in \mu} \sum_{v \in \mu \setminus u} \nabla c(\mu \setminus v, v) \\
&= \alpha_c(|\mu| - 1) \sum_{u \in \mu} \nabla c(\mu \setminus u, u) \quad (8)
\end{aligned}$$

令 $\gamma_c(\mu)$ 表示 $c(\cdot)$ 相对于集合 $\mu \subseteq U$ 的曲率. 考虑到成本函数 $c(\cdot)$ 是单调的, 结合定义 1, 我们可以很容易地得到 $1 - \gamma_c(\mu) \geq \alpha_c(1 - \gamma_c) \geq 0$. 进一步结合式(8)和定义 2, 我们可获得

$$\begin{aligned}
|\mu|c(\mu) - \sum_{u \in \mu} c(u) &\geq \alpha_c(|\mu| - 1)(1 - \gamma_c(\mu)) \sum_{u \in \mu} c(u) \\
&\geq \alpha_c^2(|\mu| - 1)(1 - \gamma_c) \sum_{u \in \mu} c(u) \quad (9)
\end{aligned}$$

因此,我们证得 $\sum_{u \in \mu} c(u) \leq c(\mu) \cdot \frac{|\mu|}{1 + \alpha_c^2(|\mu| - 1)(1 - \gamma_c)}$. 证毕.

将 K 定义为可行解中可招募用户数的最大值, 其中 $\forall \mu \subseteq U, c(\mu) \leq B$. 设 μ_i 为算法 1 第 i 次迭代后招募的用户组. 然后,我们证明了每个被招募的用户至少可以获得如下的边际增益:

引理 3. 对于每个已招募用户集合 μ_i , 我们均可获得

$$\begin{aligned}
\text{一个边际增益 } f(\mu_i) - f(\mu_{i-1}) &\geq \alpha_f \frac{c(\mu_i) - c(\mu_{i-1})}{B} (f(\hat{\mu}) - f(\mu_{i-1})), \text{ 其中 } \hat{\mu} \text{ 是 } \hat{B} = \alpha_c \frac{1 + \alpha_c^2(K-1)(1 - \gamma_c)}{K} B \text{ 下的最优集.}
\end{aligned}$$

证明. 根据定义 1 和定义 2, 我们有 $1 - \gamma_c \leq \frac{\nabla c(\mu \setminus u, u)}{c(u)}$

$$\leq \frac{1}{\alpha_c} \cdot \frac{\nabla c(u \setminus u, u)}{c(u)} = \frac{1}{\alpha_c}, 0 \leq \alpha_c \leq 1, \text{ 因此}$$

$$\alpha_c \frac{1 + \alpha_c^2(K-1)(1 - \gamma_c)}{K} \leq \alpha_c \frac{1 + \alpha_c(K-1)}{K} \leq 1 \quad (10)$$

此时, $\hat{\mu}$ 是在略微放松的预算 $\hat{B}$ 限制下的招募到的最优用户组, 该结果也受 α_c 影响.

设 $k' = |\hat{\mu} \setminus \mu_{i-1}|$, 将 $\{A_j\}_{j=0}^{k'}$ 定义为 $A_0 = \emptyset, A_{k'} = \hat{\mu} \setminus \mu_{i-1}$, 并且 $u_j = A_j \setminus A_{j-1}$. 考虑函数单调性, 我们可获得

$$\begin{aligned}
f(\hat{\mu}) - f(\mu_{i-1}) &\leq f(\hat{\mu} \cup \mu_{i-1}) - f(\mu_{i-1}) \\
&= f(\hat{\mu} \setminus \mu_{i-1} \cup \mu_{i-1}) - f(\mu_{i-1}) \\
&= \sum_{j=1}^{k'} \nabla f(\mu_{i-1} \cup A_{j-1}, u_j) \quad (11)
\end{aligned}$$

根据引理 1, 我们有

$$\begin{aligned}
\frac{\nabla f(\mu_{i-1} \cup A_{j-1}, u_j)}{\nabla c(\mu_{i-1}, u_j)} &\leq \frac{1}{\alpha_f} \cdot \frac{\nabla f(\mu_{i-1}, u_j)}{\nabla c(\mu_{i-1}, u_j)} \\
&\leq \frac{1}{\alpha_f} \cdot \frac{f(\mu_i) - f(\mu_{i-1})}{c(\mu_i) - c(\mu_{i-1})} \quad (12)
\end{aligned}$$

进一步结合式(11)和(12), 我们得到

$$\begin{aligned}
f(\hat{\mu}) - f(\mu_{i-1}) &\leq \sum_{j=1}^{k_\mu} \nabla f(\mu_{i-1} \cup A_{j-1}, u_j) \\
&\leq \frac{1}{\alpha_f} \cdot \sum_{j=1}^{k_\mu} [\nabla c(\mu_{i-1}, u_j)] \cdot \frac{f(\mu_i) - f(\mu_{i-1})}{c(\mu_i) - c(\mu_{i-1})} \\
&\leq \frac{1}{\alpha_f} \cdot \frac{1}{\alpha_c} \cdot \sum_{u \in \hat{\mu}} c(u) \cdot \frac{f(\mu_i) - f(\mu_{i-1})}{c(\mu_i) - c(\mu_{i-1})} \quad (13)
\end{aligned}$$

根据引理 2, 可知 $\sum_{u \in \hat{\mu}} c(u) \leq c(\hat{\mu}) \frac{|\hat{\mu}|}{1 + \alpha_c^2(|\hat{\mu}| - 1)(1 - \gamma_c)}$, 其中

$\frac{|\hat{\mu}|}{1 + \alpha_c^2(|\hat{\mu}| - 1)(1 - \gamma_c)}$ 是非递减的, $K \geq |\hat{\mu}|, c(\hat{\mu}) \leq \hat{B}$. 因此, 式(13)变换为

$$\begin{aligned}
f(\hat{\mu}) - f(\mu_{i-1}) &\leq \frac{1}{\alpha_f \alpha_c} \cdot \frac{\hat{B}K}{1 + \alpha_c^2(K-1)(1 - \gamma_c)} \cdot \frac{f(\mu_i) - f(\mu_{i-1})}{c(\mu_i) - c(\mu_{i-1})} \\
&= \frac{1}{\alpha_f} \cdot B \cdot \frac{f(\mu_i) - f(\mu_{i-1})}{c(\mu_i) - c(\mu_{i-1})} \quad (14)
\end{aligned}$$

因此, 引理 3 成立.

证毕.

最后, 我们计算了总效用以证明我们的算法能达到一个双准则近似保证:

定理 1. 具有非子模效用和成本函数的广义贪婪算法

可以实现 $\frac{\alpha_f}{2}(1 - e^{-\alpha_f})$ 的双准则近似比.

证明. 根据引理 3, 我们有

$$\begin{aligned}
f(\mu_{i+1}) - f(\mu_i) &= f(\mu_{i+1}) - f(\mu_i) + (f(\mu_{i+1}) - f(\mu_i)) \\
&\geq \left(1 - \alpha_f \frac{c(\mu_{i+1}) - c(\mu_i)}{B}\right) f(\mu_i) + \\
&\quad \alpha_f \frac{c(\mu_{i+1}) - c(\mu_i)}{B} f(\hat{\mu}) \\
&\geq \left[1 - \prod_{i=1}^{i+1} \left(1 - \alpha_f \frac{c(\mu_{i+1}) - c(\mu_i)}{B}\right)\right] f(\hat{\mu}) \quad (15)
\end{aligned}$$

需要注意的是, 当 $A = \sum_{i=1}^n a_i$ 且 $a_1 = \dots = a_n = A/n$ 时, $1 - \prod_{i=1}^n \frac{a_i}{A}$ 可以取到最小值, 因此

$$\begin{aligned}
f(\mu_{k+1}) &\geq \left[1 - \prod_{i=1}^{k+1} \left(1 - \alpha_f \cdot \frac{c(\mu_i) - c(\mu_{i-1})}{B}\right)\right] f(\hat{\mu}) \\
&\geq \left[1 - \prod_{i=1}^{k+1} \left(1 - \alpha_f \cdot \frac{c(\mu_i) - c(\mu_{i-1})}{c(\mu_{k+1})}\right)\right] f(\hat{\mu}) \\
&\geq \left(1 - \left(1 - \alpha_f \cdot \frac{1}{k+1}\right)^{k+1}\right) f(\hat{\mu}) \\
&\geq (1 - e^{-\alpha_f}) f(\hat{\mu}) \quad (16)
\end{aligned}$$

根据定义 1 和 $f(\cdot)$ 的单调性, 我们有

$$f(\mu_{k+1}) - f(\mu_k) \leq \frac{1}{\alpha_f} f(u') \quad (17)$$

其中, $\mu' = \arg \max_{u \in U} f(u)$. 因此, 可获知

$$\begin{aligned} f(\mu_k) + f(u') &\geq \alpha_f f(\mu_{k+1}) \geq \alpha_f (1 - e^{-\alpha_f}) f(\hat{\mu}), \\ \max\{f(\mu_k), f(u')\} &\geq \frac{\alpha_f}{2} (1 - e^{-\alpha_f}) f(\hat{\mu}) \end{aligned} \quad (18)$$

附录 B. 具有一般效用和成本的在线用户招募.

与离线策略类似,我们的在线策略分析同样利用子模比率使非子模函数近似于子模函数,从而获得我们的在线策略相比最优策略的近似比. 首先,在给定估计的预期招募用户数量 k 的情况下,我们将整个用户集 U 分成大小相等的 k 段,记为 $U_1, \dots, U_k$. 假设 μ^* 是最优用户组. 其基数 $k = |\mu^*|$. 如果 $U_i \cap \mu^* \neq \emptyset$, 则将 v_i 设为 $U_i \cap \mu^*$ 中的一个随机用户, 我们删除 $v_i = \emptyset$ 并将其重新排序为 $v = \{v_1, \dots, v_{k_v}\}$.

接下来,我们首先分析并展示我们的分段策略的有效性,即从每个用户分段中选择一个分段最优用户实际上意味着我们已经选择到了最优用户集中的很多用户.

引理 4. 给定预期招募用户数量 k , 我们预期有 $|v| \geq (1 - 1/e)k$.

证明. 如上文中的定义, $l = |U_i|$, 且 m 是用户总数. 定义 $\mu^* = \{u_1, \dots, u_k\}$, 并使用 ϵ_i 来表示 u_i 不在 U_j 中的事件. 然后, 我们得到 $\mu^* \cap U_j = \emptyset$ 的概率是

$$\begin{aligned} \Pr(\bigcap_{i=1}^k \mathcal{E}_i) &= \Pr(\mathcal{E}_1) \cdot \Pr(\mathcal{E}_2 | \mathcal{E}_1) \cdots \Pr(\mathcal{E}_k | \bigcap_{j=1}^{k-1} \mathcal{E}_j) \\ &\leq \left(1 - \frac{1}{k}\right)^k \leq \frac{1}{e} \end{aligned} \quad (19)$$

因此, $|v|$ 的期望至少为 $k - k/e$. 因此, 对于所有分段来说, 不存在最优用户集合 μ^* 中的用户分段期望数为 k/e . 证毕.

接下来, 我们证明: 与最优集合 μ^* 相比, 招募部分最优用户 μ' 的方案已经能够得到一个较好的近似比.

引理 5. 对于 μ^* 的随机子集 μ' , 我们预期有 $f(\mu') \geq \alpha_f \cdot \frac{|\mu'|}{k^*} \cdot f(\mu^*)$.

证明. 设 $\mu^* = \{u_1, u_2, \dots, u_k\}$, $F_i = f(\{u_1, u_2, \dots, u_i\})$, $F_0 = 0$, $D_i = F_i - F_{i-1}$. 将 $(u'_1, \dots, u'_i)$ 表示为 $(u_1, \dots, u_i)$ 的循环置换, 也就是说, $u'_1 = u_i, u'_2 = u_1, \dots, u'_i = u_{i-1}$, $F_i = f(\{u'_2, u'_3, \dots, u'_{i+1}\})$. 此外, 由于用户具有相同的分布, 我们可以得到 $F_i = f(\{u'_1, \dots, u'_i\})$. 因此, $D_i = f(\{u'_1, \dots, u'_i\}) - f(\{u'_2, \dots, u'_i\})$. 同样地, 我们可得 $D_{i+1} = f(\{u'_1, \dots, u'_{i+1}\}) - f(\{u'_2, \dots, u'_{i+1}\})$. 进一步地, 根据定义 1, 我们可以得到 $D_i \geq \alpha_f \cdot D_j, \forall j > i$. 因此, 我们有 $f(\mu^*) = F_k = D_1 + \dots + D_k$ 和 $f(\mu') = F_{|\mu'|} = D_1 + \dots + D_{|\mu'|}$,

$$\begin{aligned} k \cdot f(\mu') &= k(D_1 + D_2 + \dots + D_{|\mu'|}) \\ &\geq \alpha_f (D_1 + D_2 + \dots + D_{|\mu'|}) + \dots + \\ &\quad \alpha_f (D_{k-|\mu'|+1} + \dots + D_k) + \dots + \\ &\quad \alpha_f (D_k + D_1 + \dots + D_{|\mu'|-1}) \\ &= \alpha_f \cdot |\mu'| \cdot (D_1 + D_2 + \dots + D_k) \\ &= \alpha_f \cdot |\mu'| \cdot f(\mu^*) \end{aligned} \quad (20)$$

因此, 我们能够得到 $f(\mu) \geq \alpha_f \cdot \frac{|\mu'|}{k^*} \cdot f(\mu^*)$. 证毕.

接下来是在线策略理论分析的核心, 我们证明了在非子模效用的情况下, 对每个分段进行局部优化就可以得到一个全局近似.

因此, 我们的广义贪心算法可以实现 $\frac{\alpha_f}{2} (1 - e^{-\alpha_f})$ 的双准则近似比. 证毕.

引理 6. 给定 k 和 $|v|$, 我们的分段在线策略可以实现一个近似比, 即 $E[f(\mu_k)] \geq \alpha_f^2 \cdot \frac{|v|}{7k} \cdot f(\mu^*)$.

证明. 如上文所述, $v = \{v_1, v_2, \dots, v_{k_v}\}$ 和 $v_i \in U_{h_i} \cap \mu^*$. 首先, 定义 $\nabla f(j) = f(\mu_j) - f(\mu_{j-1})$. 此时, 我们可以以概率 $1/e$ 在第 j 部分选择到最优的 v_j , 因此

$$E[\nabla f(h_i)] \geq E[\nabla f(\mu_{h_{i-1}}, v_i)]/e \quad (21)$$

我们使用反证法来证明引理 6. 首先, 假定 $E[f(\mu_k)] < \alpha_f^2 \cdot \frac{|v|}{7k} \cdot f(\mu^*)$. 定义 $\psi = \{v_i, v_{i+1}, \dots, v_{k_v}\}$. 利用引理 1 和 $f(\cdot)$ 的单调性, 可得

$$\begin{aligned} E[f(\psi)] &\leq E[f(\mu_{h_{i-1}})] + \frac{1}{\alpha_f} \sum_{j=i}^{k_v} E[\nabla f(\mu_{h_{i-1}}, v_j)] \\ &= E[f(\mu_{h_{i-1}})] + \frac{k_v - i + 1}{\alpha_f} E[\nabla f(\mu_{h_{i-1}}, v_j)] \end{aligned} \quad (22)$$

因为 $\forall v_i \in \psi$ 的期望是相同的. 结合式 (21)、(22)、引理 5 和反证法的假定, 我们有

$$\begin{aligned} E[\nabla f(h_i)] &\geq \alpha_f \frac{E[f(\psi)] - E[f(\mu_{h_{i-1}})]}{e(k_v - i + 1)} \\ &\geq \frac{\alpha_f^2}{ek} f(\mu^*) - \frac{k_v \alpha_f^3}{7ek(k_v - i + 1)} f(\mu^*) \end{aligned} \quad (23)$$

为了去掉参数 i , 我们把式 (23) 对 $1 \leq i \leq \lfloor k_v/2 \rfloor$ 求和, 得到

$$\begin{aligned} \sum_{i=1}^{\lfloor k_v/2 \rfloor} E[\nabla f(h_i)] &\geq \left(\left\lfloor \frac{k_v}{2} \right\rfloor \frac{\alpha_f^2}{ek} - \frac{k_v \alpha_f^3}{7ek} \sum_{i=1}^{\lfloor k_v/2 \rfloor} \frac{1}{k_v - i + 1} \right) f(\mu^*) \\ &\geq \left(\left\lfloor \frac{k_v}{2} \right\rfloor \frac{\alpha_f^2}{ek} - \frac{k_v \alpha_f^3}{7ek} \ln \frac{k_v}{\lfloor k_v/2 \rfloor} \right) f(\mu^*) \end{aligned} \quad (24)$$

此时, $\sum_{j=a}^b 1/j \leq \ln b / (a - 1), 1 < a \leq b$. 同样的, 对于 $1 \leq i \leq \lfloor k_v/2 \rfloor$, 我们有

$$\sum_{i=1}^{\lfloor k_v/2 \rfloor} E[\nabla f(h_i)] \geq \left(\left\lfloor \frac{k_v}{2} \right\rfloor \frac{\alpha_f^2}{ek} - \frac{k_v \alpha_f^3}{7ek} \ln \frac{k_v}{\lfloor k_v/2 \rfloor} \right) f(\mu^*) \quad (25)$$

把式 (24) 和 (25) 相加, 因为 $\frac{k_v^2}{\lfloor k_v/2 \rfloor \lfloor k_v/2 \rfloor} < 4.5$, 我们有

$$\begin{aligned} 2E[f(\mu_k)] &\geq \sum_{i=1}^{\lfloor k_v/2 \rfloor} E[\nabla f(h_i)] + \sum_{i=1}^{\lfloor k_v/2 \rfloor} E[\nabla f(h_i)] \\ &\geq \frac{k_v \alpha_f^2}{k} \left(\frac{1}{e} - \frac{\ln 4.5}{7e} \alpha_f \right) f(\mu^*) \geq \frac{k_v \alpha_f^2}{k} \cdot \frac{2}{7} f(\mu^*) \end{aligned} \quad (26)$$

这与 $E[f(\mu_k)] < \alpha_f^2 \cdot \frac{k_v}{7k} \cdot f(\mu^*)$ 相悖, 因此我们有 $E[f(\mu_k)] \geq$

$\alpha_f^2 \cdot \frac{k_v}{7k} \cdot f(\mu^*)$. 证毕.

然后, 我们将引理 4 中 $|v|$ 的期望值引入引理 6, 得到了在确定的预期招募用户数量 k 下的近似保证.

引理 7. 给定 k , 我们的分段在线策略可以达到一个近似 $E[f(\mu_k)] \geq \frac{\alpha_f^2 (1 - 1/e)}{7} f(\mu^*)$.

证明. 根据引理 4 和引理 6, 我们可以得到

$$\begin{aligned}
E[f(\mu_k)] &= \sum_{j=1}^k E[f(\mu_k) \mid |\nu| = j] Pr[|\nu| = j] \\
&\geq \sum_{j=1}^k \alpha_f^2 \cdot \frac{j}{7k} \cdot f(\mu^*) \cdot Pr[|\nu| = j] \\
&= \frac{\alpha_f^2}{7k} f(\mu^*) \cdot E[|\nu|] \geq \frac{\alpha_f^2(1-1/e)}{7} f(\mu^*) \quad (27)
\end{aligned}$$

实际上,当效用函数变成子模函数时,例如 $\alpha_f = 1$, 我们的竞争比变成 $\frac{1-1/e}{7}$, 这正是子模块秘书问题^[25]的经典结果。证毕。

最后,利用引理 7、假设 1 和定理 1, 我们进一步去掉 k 以证明分段在线策略在定理 2 中的近似比。

定理 2. 分段在线策略可以达到的预期逼近比为

$$\frac{\alpha_f^3(1-e^{-1})(1-e^{-\alpha_f})}{14}.$$

证明. 如上文所述, μ_k^* 是在已知基数 k 的最佳用户组, 将 μ_k 定义为离线贪婪算法使用效用 $g(\mu_k)$ 从相同的用户序列 U 中选择的用户组。如算法 2(第 1 行), 我们的分段在线策略构建一些虚拟用户集 U' 来模型真实的用户序列 U , 并且贪婪的从每个 U' 中选择 μ_k' 来估计 $|\mu_k|$ 。根据假设 1, 可得期望 $E[|\mu_k'| \mid |\mu_k|]$, 因此 $|\mu_k'| = k$ 且 $f(\mu_k^*) \geq g(\mu_k)$ 。因此, 结合引理 7 和定理 1, 我们有

$$\begin{aligned}
E[f(\mu_k)] &\geq \frac{\alpha_f^2(1-e^{-1})}{7} f(\mu_k^*) \geq \frac{\alpha_f^2(1-e^{-1})}{7} g(\mu_k) \\
&\geq \frac{\alpha_f^3(1-e^{-1})(1-e^{-\alpha_f})}{14} f(\mu) \quad (28)
\end{aligned}$$

因此定理 2 成立。

证毕。



LIU Wen-Bin, Ph.D., assistant research fellow. His research interests include mobile crowdsensing and ubiquitous computing.

YANG Yong-Jian, Ph.D., professor. His research interest is wireless sensor networks.

WANG En, Ph.D., professor. His research interests include mobile crowdsensing and mobile computing.

Background

With the explosive popularity of portable smart devices and the rapid development of wireless networks and communication technologies, Mobile CrowdSensing (MCS) has recently become an interesting topic to the research community, one that has received much attention from practitioners, scholars, and the general public. In general, MCS allows us to recruit a crowd of users within a limited budget to cooperatively perform various sensing tasks, such as environmental monitoring, traffic flow measuring, and target tracking, by using their carried mobile devices equipped with powerful sensors. Compared to the traditional fixed-deployment wireless sensor networks, MCS not only greatly improves the ability of sensing, but also pushes the boundaries of scalability, cost, coverage, and usability.

Benefiting from these above advantages, a number of MCS-based applications have emerged in recent years. For example, a map application utilizes MCS to collect the smartphones' acceleration data to identify map semantics such as bridge and tunnel; a navigation application senses the real-time traffic information according to the GPS signals collected by MCS; a noise monitoring system utilizes MCS to conduct a fine-grained noise map by using the microphones of smartphones. Moreover, MCS can also make use of the information fed back actively by users, such as the number of queu-

ing in restaurants, the number of free parking spaces, the location of shops in shopping malls, etc.

Predictably, MCS can provide great convenience for the masses, while it would provide great services if and only if there are sufficient users and the users perform their sensing tasks. Obviously, user selection is the foundation of MCS and there has been so much research on it. The existing works usually recruit users based on their submodular utility and linear cost functions, e. g., the users' probabilistic coverage of tasks and total consumed costs. However, we argue that these works do not consider the cooperative willingness of users, which may lead to non-submodular utility and cost. In this paper, we study the user recruitment problem with general utility and cost, and propose both the offline and online user recruitment strategies.

This work is supported by the National Key R&D Program of China under Grant Nos. 2021ZD0112501, 2021ZD0112502, the National Natural Science Foundations of China under Grant Nos. 62102161, 61972450, and 62072209, the CCF-Baidu Open Fund (No. 2021PP15002000).

From 2014, our research group has already published more than 60 high-quality papers in the field of user recruitment, incentive mechanism, and data inference in MCS.