

社交网络下的移动群智感知任务扩散^①

孙 驰

(中国科学技术大学 大数据学院, 合肥 230026)
通信作者: 孙 驰, E-mail: sc1997@mail.ustc.edu.cn



摘 要: 移动群智感知是智慧城市数字化建设的核心基础技术之一, 是移动计算领域的热点研究课题. 近年来, 移动群智感知虽然已有许多代表性的研究成果, 但从整体上看距离大规模的普及应用仍有不少距离, 在实际推广应用中仍面临着用户参与度不高的问题. 为此, 引入社交网络 IM (影响最大化) 传播模型, 考虑到现实情况下概率信息的缺失, 通过在线学习的方式, 在进行影响力活动的同时学习影响力概率, 即根据用户反馈不断更新影响力模型信息, 从而提出新的基于该模型的任务扩散方案. 通过使用真实的社交网络数据集进行实验, 结果表明提出的方法在传播范围方面比传统的 IM 方法更有效, 为移动群智感知系统的实际推广应用做出贡献.

关键词: 群智感知; 任务扩散; 在线学习

引用格式: 孙驰. 社交网络下的移动群智感知任务扩散. 计算机系统应用, 2023, 32(10): 166-174. <http://www.c-s-a.org.cn/1003-3254/9255.html>

Task Diffusion of Mobile Crowdsensing in Social Network

SUN Chi

(School of Data Science, University of Science and Technology of China, Hefei 230026, China)

Abstract: Mobile crowdsensing is one of the core basic technologies in the digital construction of smart cities, and it is a hot research topic in the field of mobile computing. In recent years, although there have been many representative research results on mobile crowdsensing, there is still a long way to go before it is widely used on a large scale, and it still faces the problem of low user participation in the actual promotion and application. To this end, the social network influence maximization (IM) transmission model is introduced. It considers the lack of probabilistic information in reality and learns the probability of influence while performing influence activities through online learning, or in other words, the influence model information is constantly updated according to the user feedback, so as to propose a new task diffusion scheme based on the model. Through experiments with real social network data sets, the results show that the proposed method is more effective than the traditional IM method in terms of transmission scope, and it makes a contribution to the practical promotion and application of mobile crowdsensing systems.

Key words: mobile crowdsensing; task diffusion; online learning

近年来, 以智能手机为代表的便携式移动终端设备得到了极大的普及. 在此背景下, 一种基于众包思想和群体智慧的数据收集与信息共享模式即移动群智感知应运而生. 移动群智感知系统主要由位于云端的系统平台、服务请求者、移动终端用户组成, 服务请求

者可以通过系统平台将数据收集任务众包给移动用户群体, 移动用户则移动到相应位置, 利用他们随身携带的移动终端设备来完成相应的数据收集任务, 然后将结果反馈给服务请求者. 通过移动群智感知, 人们可以直接使用便携式移动终端设备收集与日常生活密切相

① 基金项目: 江苏省自然科学基金面上项目 (BK20191194)

收稿时间: 2023-03-19; 修改时间: 2023-04-20; 采用时间: 2023-04-28; csa 在线出版时间: 2023-08-09

CNKI 网络首发时间: 2023-08-10

关的各类数据。正是由于移动群智感知具有普适性、智能性、低成本等诸多优势,因而已发展成为了一种非常灵活有效的大数据收集手段,具有广泛的应用前景,如地图标注、室内定位、城市噪音污染检测、交通流量监控、生活资讯共享、远程医疗等^[1-6]。然而,移动群智感知系统在实际推广应用中仍然存在用户参与度不高的问题,如何提升用户参与度,扩散移动群智感知任务的传播范围,从而使移动群智感知技术发挥最大效用成为亟待解决的重难点问题。

目前,移动群智感知技术已经引起了学术界的广泛研究,研究成果不断涌现,研究工作主要集中在任务分配、激励机制、隐私保护3个方面^[7-14],其中任务分配是移动群智感知的核心问题。然而,移动群智感知系统作为新兴的大数据收集平台,目前尚未实现大规模的普及应用,如何吸引更多移动用户参与执行移动群智感知任务通常被称为冷启动问题,关注到此问题的现有工作较少。一个新的研究视角是从任务扩散的角度来研究移动群智感知系统,由于移动用户经常使用微博、微信、Facebook等各种社交媒体平台相互分享与交流各种信息,所以一个自然的想法是移动群智感知系统中的数据收集任务也可以通过用户之间的各种社交网络关系进行扩散和传播,通过将社交网络下的影响力传播模型应用于移动群智感知系统的任务扩散活动中,移动群智感知任务的传播范围将得到大幅提升。

社交网络中有影响力用户的识别问题被称为影响最大化问题,最早由 Domingos 等^[15]定义为如何寻找 n 个初始节点,使得信息的最终传播范围最广。Kempe 等^[16]第一次将最大化问题归纳为离散最优化问题,并且证明了求最优解是一个 NP 难问题,提出了对后续研究影响极大的两种传播模型:独立级联 (independent cascade, IC) 模型和线性阈值 (linear threshold, LT) 模型,他们提出采用贪心算法求解,即每一步采用当前最优解作为种子。为了降低贪心算法的时间复杂度,后来的研究者又提出了若干改进算法,如 CELF 算法^[17]、PMIA (prefix excluding maximum influence arborescence) 算法^[18]、NewGreedyIC 算法^[19]等。

这些算法虽然效率较高,但必须事先获得网络节点之间相互影响的概率信息。而在现实很多场景,包括在移动群智感知系统中,通过社交网络传播信息的移动用户节点之间互相影响的概率是难以获得的,在此

场景下,这些算法将无法执行。另外,虽然存在一些工作将社交网络中的模型与方法应用到移动群智感知系统中,但仍然侧重于任务分配、激励机制、隐私保护等方面,尚未存在相关工作借助社交网络来解决冷启动问题。

为解决这些问题,本文提出一种新的基于社交网络影响力传播模型的任务扩散方案,在概率缺失的情况下,首先将概率分布初始化为无信息先验分布,接着采用在线学习及多轮次进行任务扩散活动的方式,根据用户反馈不断学习更新初始的概率模型,同时更新种子用户节点,最终实现影响最大化从而最大化扩散任务的传播范围。此外本文将使用真实的社交网络数据集进行实验,并与传统方法进行比较,以验证本文方法的准确性与高效性,从而为移动群智感知系统的实际推广应用贡献可行性方案。

1 任务扩散模型和问题定义

1.1 任务扩散模型

如之前所述,移动群智感知系统一般由服务请求者、执行数据收集任务的移动用户、系统平台这3部分组成。现假设执行数据收集任务的移动用户是潜在的,并且具有各自的社交关系,即存在于一个社交网络当中。当服务请求者发起任务时,为了实现任务扩散的目标,系统平台需要从中选择社交影响力较大的移动用户,向他们支付一定的报酬从而使得他们愿意参与任务的执行并通过自己的社交关系吸引更多潜在的移动用户参与其中。根据上述情形本文构建了一个任务扩散模型,其中各个部分的定义如下所述。

定义1 (服务请求者与任务)。服务请求者向移动群智感知系统平台发起一个长期任务,该长期任务需要一定数量的移动用户通过收集所需数据共同完成,并且需要经历多个时间周期才能完成(如空气质量检测相关数据的收集任务)。任务可以由表达式 $Task \triangleq (N, \Delta)$ 来描述,其中长期任务的完成是一个有时间间隔的过程,整个过程被分为 N 个时间段,并且每个周期 $n \in \{1, 2, \dots, N\}$ 具有相同的持续时间 Δ (例如1天)。

定义2 (系统平台)。作为一个盈利的中间代理,平台会向服务请求者收取一定的金钱报酬来提供服务请求者所需的数据服务,并从中抽取一定比例的报酬作为自己的佣金,以补偿平台的聚合成本。同时,为了激励移动用户尽最大努力完成任务,平台将为招募的移

动用户支付报酬,以弥补其完成任务的成本.另外,在任务扩散过程中,系统平台需要从潜在的移动用户中选择社交影响力较大的移动用户,并向他们支付一定的报酬.

定义3(社交网络).潜在的移动用户共同组成一个社交网络,该社交网络可以用有向加权图 $G=(V, E, P)$ 表示,其中 V 表示社交网络中的移动用户节点集合, E 表示边集, $\forall v \in V$ 表示移动用户节点, $\forall e(u, v) \in E$ 表示一条 u 指向 v 的有向边,代表移动用户 u 能够通过社交网络将任务传递给移动用户 v , P 表示边权值集合,其中边权值 p_{uv} 表示 u 指向 v 的有向边上的权值,代表移动用户 u 将任务传递给移动用户 v 的概率值.

定义4(种子用户).为了达到最优的任务扩散效果,系统平台将在每个任务周期开始之前从社交网络中选择一定数量的社交影响力较大的移动用户作为种子用户,这些种子用户构成了社交网络中的种子节点集 S ,通过他们的社交关系能够使更多的移动用户参与到任务执行当中,以达到执行任务所需的移动用户数量.

定义5(任务扩散过程)基于上述定义与社交网络中的 IC 模型,移动群智感知系统中的任务扩散过程得到定义.初始时所有移动用户节点均处于未激活状态,

当系统平台选择种子用户后,它们共同组成初始的激活节点集合 S_0 ,接着基于 IC 模型,任务在社交网络中不断传播,当传播至第 $t+1$ 轮时,利用在第 t 轮中被激活的用户节点,根据概率边权值 p 尝试激活在其社交网络中相邻的用户节点,即以概率边权值 p 将任务传递给相邻的用户节点,若节点被激活则将其加入集合 S_t 从而扩充为 S_{t+1} ,不断重复此过程直至不再有新的用户节点加入集合,从而数据收集任务通过社交网络得到扩散.总体而言,设 σ 为任务扩散算子,其将选择的初始种子用户节点集合 $S \subseteq V$ 扩充为最终的激活节点集合 $\sigma(S)$,该过程为随机过程,且为与 P 相关的随机变量.

基于以上定义,移动群智感知系统的任务扩散过程如图1所示:(1)数据服务请求者向平台发送数据收集任务与预算;(2)系统平台基于用户反馈信息,利用任务扩散机制在潜在移动用户组成的社交网络中选择种子用户;(3)种子用户选择完成后,任务扩散活动开始,任务信息在社交网络中得到传播;(4)任务扩散活动结束后,移动用户向系统平台反馈数据信息;(5)系统平台向完成数据收集任务的用户与平台选择的种子用户支付报酬.

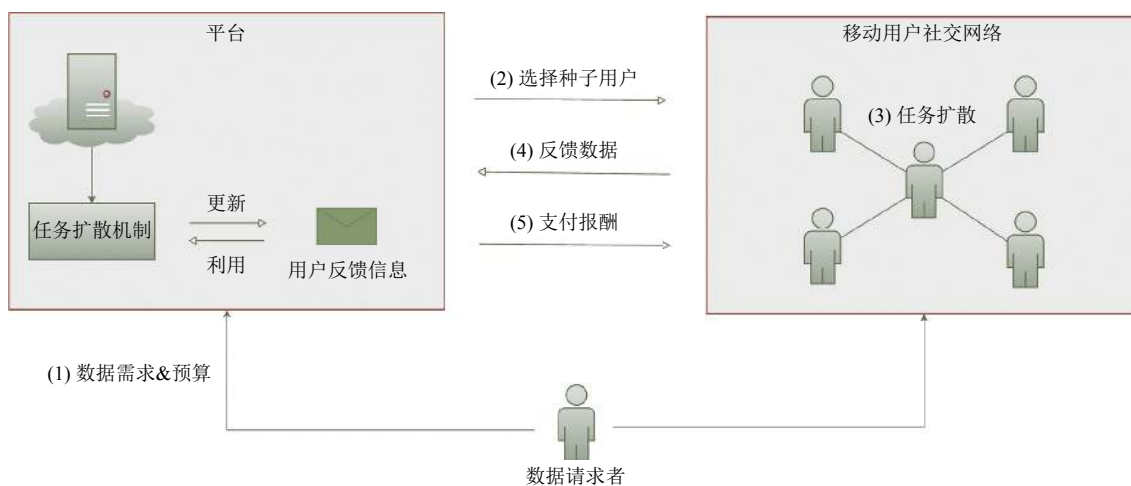


图1 移动群智感知系统的任务扩散过程

在移动群智感知系统中,移动用户传递并完成数据收集任务的概率信息是未知的,因此系统平台需要根据移动用户的反馈信息来学习相关的概率分布参数,系统平台会为每一位用户维护一个反馈信息列表,并且在每一轮任务扩散活动结束后基于用户的反馈信息对相关的概率信息进行评估更新.

1.2 问题定义

基于上述任务扩散模型,本文的研究重点在于系统平台如何在给定成本预算与时延限制下选择一定数量的移动用户作为种子用户,使得任务的扩散范围最广以达到任务对移动用户数量的需求.需要注意的问题是社交网络中移动用户之间传递任务的概率信息往

往难以获得,因此社交网络中的概率边权值 p 是未知的,本文将引入在线学习的方法解决在概率信息缺失情况下基于任务扩散模型的任务扩散最大化问题。

具体而言,一个长期任务需要 N 个时间周期,系统平台也将在任务的每个时间周期开始之前进行任务扩散活动。在每次活动中,系统平台将选择种子用户组成初始的种子节点集合 S_n , 其中 $1 \leq n \leq N$ 。假设在每一轮任务扩散活动中选择种子用户的所需时间不超过一个固定值 T_0 , 且对于任意一个种子用户节点 u , 除执行种子选择之外的所需时间(即从种子节点开始扩散任务直到获得用户反馈的时间)为 t_u , 总的成本预算为 C , 总的时延限制为 T , 节点 u 的激活费用定义为函数 $Cost(u)$, 则任务扩散模型下的任务扩散最大化问题按如下方式定义:

$$\arg \max_{S_n \subset V} \left\{ E \left[\left| \bigcup_{n=1}^N \sigma(S_n) \right| \right] \right\} \quad (1)$$

$$\text{s.t.} \sum_{n=1}^N \sum_{u \in S_n} Cost(u) \leq C \quad (2)$$

$$NT_0 + \sum_{n=1}^N \max\{t_u, u \in S_n\} \leq T \quad (3)$$

其中,式(1)表示任务扩散最大化问题的总体目标,即在每一轮任务扩散活动中从社交网络中选择种子节点使得最终总的任务的期望扩散范围达到最大值。式(2)表示任务扩散最大化问题的成本预算要求,即每一轮任务扩散活动中平台向选择的种子用户支付的报酬总和不超过总的成本预算 C 。式(3)则表示任务扩散最大化问题的时延限制要求,任务扩散活动的所需时间主要包含两个部分,一是选择种子用户的所需时间,该时间已假设为不超过一个固定值 T_0 ,二是对于每个种子用户节点除执行种子选择之外的所需时间 t_u ,在每一轮任务扩散活动中, t_u 代表种子用户节点 u 从该种子节点开始扩散任务直到获得用户反馈的时间,这些时间的最大值则代表了该轮任务扩散活动除执行种子选择之外的所需时间,因此式(3)代表总的任务扩散活动的所需时间不超过总的时延限制 T 。

2 基于在线学习的任务扩散最大化算法

2.1 算法整体框架

在移动群智感知系统中,移动用户通过社交网络传递并完成数据收集任务的概率信息是未知的,一些

传统的影响最大化算法需要知晓所有概率信息,这在现实情况下往往是不可行的。另外,移动群智感知系统的任务扩散过程需要考虑到系统平台选择种子用户、移动用户执行任务、用户反馈等过程存在的时延问题,同时需要考虑到一定的成本因素。因此,本文提出的任务扩散最大化算法将采用在线学习的方式解决概率缺失的问题,同时力求在时延与成本等限制条件下达到最优的任务传播范围。

本文提出的基于在线学习的任务扩散最大化算法为解决任务扩散模型下的任务扩散最大化问题提供了基本的框架。根据定义1,一个长期任务的执行过程分为 N 个时间周期,因此在本文提出的任务扩散最大化算法中也将进行 N 次任务扩散活动,在每次任务扩散活动中算法将选择一定数量的移动用户作为种子用户,在每一轮任务扩散活动完成之后采用在线学习的方式,接受来自移动用户的反馈从而更新初始时建立的概率分布模型,算法的总体框架描述如算法1。

算法1. 任务扩散最大化算法

输入: 任务周期数 N , 预算 C , 时延 T , 社交网络图 G 。

输出: 种子节点集合 $S_n, 1 \leq n \leq N$, 激活节点集合 R 。

- 1) $R \leftarrow \emptyset$, 费用 $\lambda \leftarrow 0$
- 2) for $n=1$ to N do
- 3) $S_n \leftarrow$ 执行种子选择策略
- 4) $R_n, F \leftarrow$ 用户反馈, 其中 R_n 表示每一轮激活的移动用户集合, F 表示用户反馈的信息列表
- 5) if 式(3)成立 and $\lambda + \sum_{u \in S_n} Cost(u) \leq C$ then
- 6) $R \leftarrow R \cup R_n$
- 7) $\lambda \leftarrow \lambda + \sum_{u \in S_n} Cost(u)$
- 8) 执行更新策略,更新 G
- 9) 返回 S_n, R

算法中的第3-8行描述了任务扩散最大化算法的具体实现流程,在每一轮任务传播活动中,系统平台都将进行种子选择与社交网络中概率模型的更新两种操作。算法第3行中种子选择操作即采用传统的IM算法或结合某种策略来选择出种子用户,在每一轮任务传播活动完成后,算法第4行代表系统平台将获得一个被激活的移动用户集合和一个表示用户反馈是否参与执行任务的信息列表 F , 其可定义为元组 (i, j, R_{ij}) 的集合,其中 (i, j) 代表社交网络中的有向边 $e(i, j)$, R_{ij} 代表用户节点 i 将任务传递给用户节点 j 的反馈结果。接着算法第5行检查目前平台需要支付的费用是否超过预算和式(3)中的时延要求是否得到满足,若以上条件满

足, 算法第 6 行将每一轮激活的移动用户集合加入最终激活的移动用户集合 R , 算法第 7 行更新平台所需支付的累计费用, 算法第 8 行中社交网络图的更新操作则基于用户反馈更新社交网络中的概率模型. 本文将在第 2.3 节与第 2.4 节分别介绍种子选择策略与更新策略.

2.2 概率信息建模

在移动群智感知系统中, 用户之间相互影响并传递任务的概率信息是难以获得的, 另外, 数据收集任务一般较为复杂, 需要大量移动用户相互协作, 且对于特定任务只有对应移动用户才能完成. 因此, 本文将社交网络中未知的概率信息先初始化为一个先验分布, 之后通过用户反馈的信息列表不断更新初始时建立的概率模型.

在统计学中, 当初始情况下没有其他有关信息时, 根据同等无知原则, 先验分布的选择一般为无信息先验分布. 但在本文建立的任务扩散模型中, 用户的反馈只有成功与失败两种情况, 当用户接收到可完成的任务并且参与任务的执行时为成功, 其他情况均为失败, 这与掷硬币的过程类似, 因此用户反馈结果的概率分布可视为一次伯努利分布, 其参数 p 即为社交网络中未知的概率边权值. 由于伯努利分布属于二项分布当中的一种特殊情形, 为了方便计算后验分布的相关信息可选取其共轭先验分布即贝塔分布作为参数 p 的先验分布. 同时, 任意两个用户节点之间的概率边权值的影响因素可视为二元随机变量, 贝塔分布恰好包含 α 与 β 两个参数, 因此贝塔分布是一个较为合适的选择.

综上所述, 任务扩散模型中社交网络的各条边的概率边权值均视为服从贝塔分布 $B(\alpha, \beta)$ 的随机变量, 即将贝塔分布作为参数 p 的先验分布, 在每一轮任务扩散活动完成后, 任务扩散最大化算法将根据用户反馈信息计算其后验分布, 从而不断更新最初建立的先验分布概率模型. 根据共轭先验分布的性质, 参数 p 的后验分布将仍然服从贝塔分布, 这将大大简化后验分布的计算过程, 从而提高算法中社交网络图的更新操作的效率.

2.3 种子选择策略

对移动群智感知任务扩散模型中社交网络的概率边权值建立概率模型后, 接下来需要从社交网络中选择种子用户. 由于此时已经建立了概率模型, 一些常见的基于概率信息的 IM 算法均可用于解决该问题, 然而

初始时建立的概率模型与实际情况可能大不相同, 故需要根据用户反馈不断更新原有概率模型.

本文提出一种基于概率分布与社交网络影响最大化算法的种子选择策略. 一般情况下, 根据之前建立的概率模型, 节点 i 与节点 j 之间的影响概率 p_{ij} 服从 Beta 分布, 即 $B(\alpha_{ij}, \beta_{ij})$. p_{ij} 能以期望值 $\alpha_{ij}/(\alpha_{ij} + \beta_{ij})$ 作为其无偏估计, 从而将任务扩散最大化问题转化为概率已知的影响最大化问题, 此时选择任意一种传统的 IM 算法即可选择出种子节点.

然而仅以分布的均值作为影响概率的估计可能会造成误差, 尤其当分布的方差较大时产生误差的可能性更大. 因此, 为了平衡误差, p_{ij} 的估计值可在期望的基础上再加上一个平衡误差项, 如分布的标准差的倍数, 即 $p_{ij} = E + \theta\sigma$, 其中 E 为期望, σ 为标准差. 当 $\theta > 0$ 时, 表明仅以期望作为估计值时低估了影响概率, 需要增大估计值以改善效果. 与此相反, 当 $\theta < 0$ 时, 影响概率被高估, 此时需要降低估计值. θ 的取值则根据用户反馈的信息不断调整.

2.4 更新策略

在选择种子用户之后, 它们将在社交网络中处于激活状态, 开始对任务信息进行传播, 在此过程中可以获得用户的反馈信息, 即具体有哪些用户被激活并能成功完成任务, 利用这些信息可以反过来对模型的参数进行更新, 一般而言, 更新的内容主要体现在以下两个方面.

一是当节点之间尝试激活时, 加权图中对应的有向边的影响概率的概率分布需要根据反馈信息及时更新. 如前所述, 任意两个相邻节点 i 与 j 的影响概率的分布被建模为 $B(\alpha_{ij}, \beta_{ij})$, 由于其为伯努利分布的共轭先验分布, 其后验分布较为容易计算, 由贝叶斯公式易知可按如下规则更新分布的参数信息.

- 1) 若反馈结果 R_{ij} 为成功, 则分布可更新为 $B(\alpha_{ij} + 1, \beta_{ij})$.
- 2) 若反馈结果 R_{ij} 为失败, 则分布可更新为 $B(\alpha_{ij}, \beta_{ij} + 1)$.

因此, 在每一轮任务扩散活动中可统计每一条边上反馈结果的成功与失败次数, 若某条边上成功次数为 s , 失败次数为 f , 则此轮任务扩散活动后该条边的分布更新为 $B(\alpha_{ij} + s, \beta_{ij} + f)$.

二是对之前提出的 θ 值的更新, 首先为 θ 设置一些初始值, 如 $(-1, 0, 1)$, 同时保证 θ 选择每个值的概率相同. 接着当每轮任务扩散活动时, 计算使用当前 θ 值

的收益,即激活节点的数目,若该值的水平较高则增加 θ 值对应的概率权重,即改变初始时 θ 的等概率分布,从而使下一次轮任务扩散活动中选择该 θ 值的可能性更大.实际操作中可采用梯度更新算法^[20]改变 θ 的抽样分布.

3 实验分析

3.1 实验数据

为了验证本文所提方案的有效性,本文实验将在公开的真实社交网络数据集 NetHEPT (collaboration network of high energy physics theory from arXiv.org)^[19]上进行,它代表着来自 arXiv.org 的高能物理理论领域的学术合作网络,其中节点代表作者,边代表合著关系,边权值代表作者之间进行合作的概率真实值.具体而言,NetHEPT 数据集由 3 列数据构成,前两列均为代表作者的节点编号,数据集对每一位作者都进行了唯一的编号,最后一列为概率边权值,其由大量的历史数据信息计算得到.因此对每一行数据而言,两个作者节点编号构成了一条有向边,代表前一个作者向后一个作者发起合作邀请,其概率则为第 3 列对应的概率边权值,其统计信息如表 1 所示.

表 1 数据统计

数据集	节点数	边数	平均度
NetHEPT	15233	32235	4.23

3.2 实验设计

本文实验考虑一个简单情形,即认为所有用户节点的成本费用均相同,且在每一轮任务扩散活动中系统平台的成本预算均相同,因此在总预算确定的情况下,每一轮任务扩散活动中系统平台能够选择的种子用户数量不超过某一正整数 k .

首先按照本文方案编写任务扩散最大化算法程序以基于实验数据集模拟移动群智感知系统中的任务扩散活动过程,该程序提供多个实验参数选项,包括实验数据集文件名、全局先验分布参数 α 与 β 、种子选择策略选项、更新策略选项、模拟的任务扩散时间周期数 N 、每一轮任务扩散活动中选择的种子用户数 k 等.接着通过该算法可以模拟计算出 N 次任务扩散活动后本文方案的任务传播范围,即成功激活的移动用户节点总数.由于概率模型的随机性,实验将重复执行算法程序 20 次,并以其平均值作为结果.另外,实验将对方案中的种子选择策略与更新策略进行评估,通过比较不同种子选择策略与更新策略的组合方案,从而验证

本文方案的有效性与优越性.

默认情况下,实验方案的参数设置如下:首先对图的影响概率设置一个全局先验分布,即 $B(1,20)$,各条边的初始概率分布也设置成全局分布.其次, θ 值的取值集合设置为 $(-1, 0, 1)$.最后,实验中涉及到的影响最大化算法设置为传播效果较好的 TIM 算法^[21].算法的所有程序代码都基于 C++ 语言编程,实验环境为配置 1.19 GHz Intel(R) 四核处理器和 4 GB 内存的 Linux 虚拟机.

3.3 对比算法

由于本文提出的移动群智感知任务扩散机制在概率信息未知的情况下进行,即不使用数据集中的真实概率数据,因此实验将以真实概率下采用传统影响最大化算法 TIM 得到的任务传播范围为基准衡量任务扩散最大化算法的准确性与有效性.另外,本文选取了随机法与最大度数法作为本文方案的比较方案,当种子选择策略设置为随机法与最大度数法时,由于这两种朴素的做法与任务扩散模型中建立的概率模型无关,因此其对应的算法流程将不涉及到更新策略以及概率模型的更新,从而在概率信息未知时可以作为本文的比较方案.

(1) 在概率已知情况下,可将选取任意一种基于概率信息的影响最大化算法得到的结果作为实际参考值,本文选取传播效果较好的 TIM 算法.

(2) 随机法在每一轮任务扩散活动中从所有节点中随机选择 k 个节点作为种子节点,这样每一个节点都有相同的机会被激活.

(3) 最大度数法^[19]是一种基于启发式的影响最大化算法,该算法总是选择当前社交网络图中出度数最大的前 k 个节点作为种子节点,因为从直观上来看如果某一节点拥有更高的出度,则其影响其他节点的机会也越大.

3.4 实验结果分析

图 2 展示了在总预算不变情况下,采用不同的种子选择策略得到的任务传播范围随 k 值的变化情况,其中总预算固定为最多选择 50 个种子节点, k 值代表每一轮任务扩散活动中可选择的最大种子用户数量,实验中记录了 k 分别为 1、5、10、25、50 的任务传播范围,与此顺序相对应,任务扩散活动次数 N 则分别为 50、10、5、2、1.通过比较可以发现,随机算法的传播范围水平较低,且变化幅度不大,原因在于其不依赖于用户反馈,即未采取任何更新策略.最大度数法的效果略低于本文方案,且当 k 值越小时本文方案的传

播效果越好,例如当 $k=5$ 时本文方案的任务传播范围相较于最大度数法提升 12% 左右,而当 $k=1$ 时本文方案的传播范围达到最大值,这表明了本文方案的有效性. 因为 k 值越小则任务扩散活动的轮次越多,通过用户反馈更新模型信息的机会越大,在线学习的方式也能更好地发挥作用,最终的传播效果也越好. 而当 $k=50$ 时,所有预算一次性投入,此时可视为在概率信息未知时不考虑概率模型更新的情况下使用传统 IM 算法解决问题,由于其忽略了概率信息的更新,在选择种子用户节点时极有可能因概率值偏离真实情况而做出错误决策,所达到的效果也是最差的.

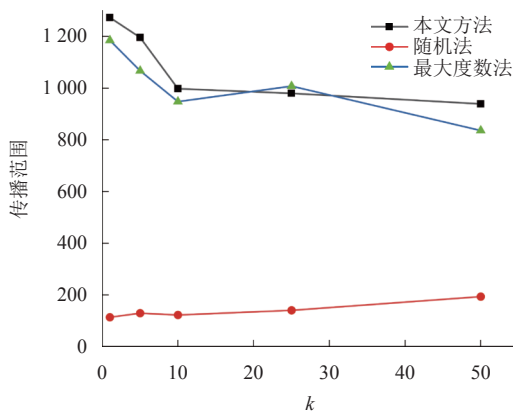


图2 预算不变时各方法传播范围

图3展示了在各个 k 值下不同的种子选择策略的传播范围随任务扩散活动次数的变化. 与之前的结果类似,本文方案总体上优于其他方案,且随着 k 值与任务扩散活动次数的增大,本文方案的传播效果也越好,因为 k 与任务扩散活动次数越大,通过用户反馈更新模型信息的机会越大,这也再一次表明了本文方案的有效性. 需要注意的是,采用随机法、最大度数法与本文方案时数据集中第3列的真实概率值是未使用的,这代表了在现实情况下概率未知的情形. 因此,图中还绘制了在已知真实概率下使用 TIM 算法得到的传播折线,它可以作为其他方案的实际参考值,从图中可以发现,本文方案的表现接近于该实际值,且随着 k 值与任务扩散活动次数的增大,二者的差距逐渐减小,这也表明了本文方案的准确性.

图4展示了在是否使用更新策略下本文方案的传播范围的不同,从图中可以明显看出更新策略对于传播效果的提升,即利用用户反馈更新模型信息的做法是有效的.

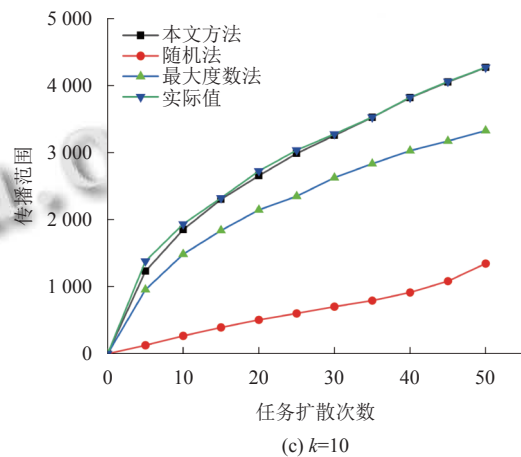
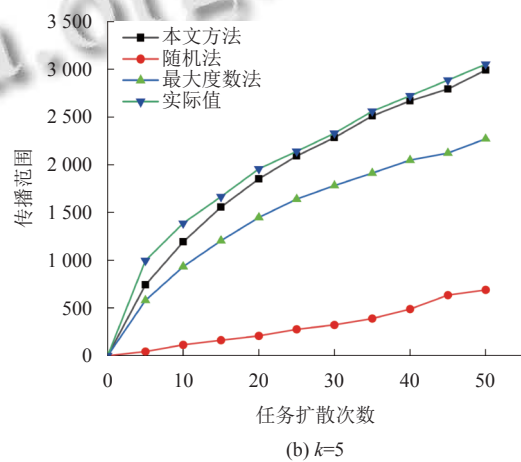
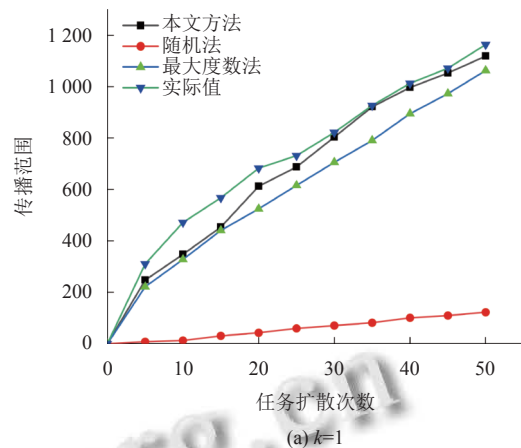


图3 k 不变时各方法传播范围

图5展示了各个方案的运行时间. 其中图5(a)描述了在 $k=1$ 时,各个方案的运行时间随任务扩散活动次数的变化,从图中可以看出,本文方案所需时间远高于随机法与最大度数法,这正是处理用户反馈与更新模型所花费的时间. 图5(b)描述了在预算固定不变时,各个方案的运行时间随 k 值的变化,同样可以发现,随

机法与最大度数法的运行时间十分平稳,不受 k 值影响,而本文方案所需时间则随着 k 值的减小而不断增加,且当 $k=1$ 时远高于其他两种方法。因此,在预算不变时,虽然 k 值越小时方案的传播效果越好,但其时间效率也逐渐变差,故在实际使用时,应当在可容忍的时间范围内选择适当的 k 值。

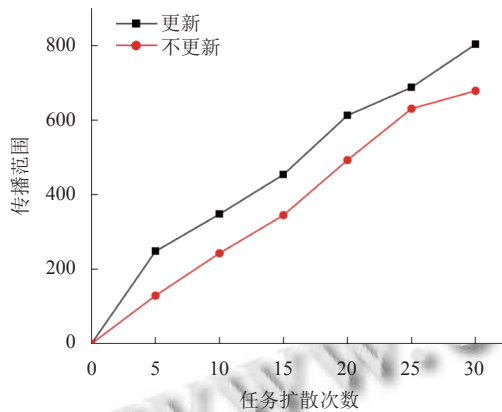
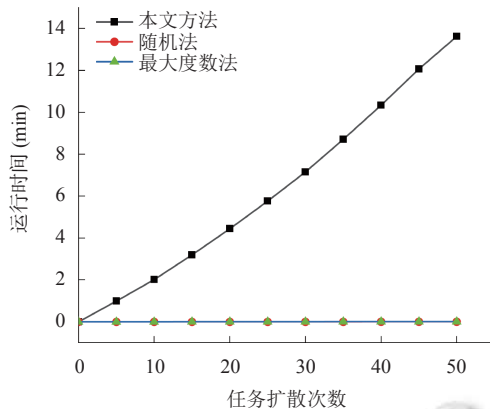
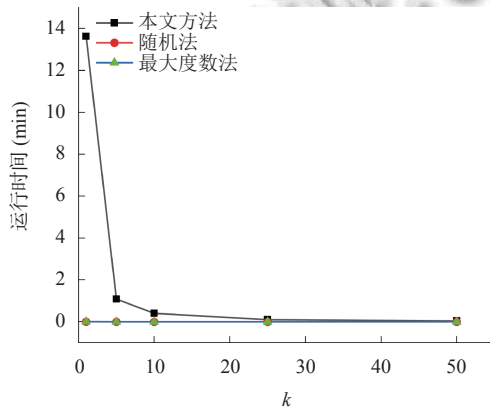


图4 更新策略比较

(a) $k=1$ 

(b) 预算固定为50

图5 时间效率比较

4 结论与展望

本文关注到移动群感知系统中用户参与度不高的问题,提出了基于社交网络的任务扩散模型。由于任务的传播过程较为复杂,用户之间的影响概率难以获得,本文研究了在概率信息缺失的情况下如何实现任务扩散最大化。本文设计了一种新的解决方案,采用在线学习的方法,在进行多轮次任务扩散活动的同时根据用户的反馈不断更新模型信息。实验结果表明即使没有获得概率信息,本文提出的方案与真实情况十分接近,具有较高的准确性,为移动群感知系统的任务扩散贡献了可行性方案,也为移动群感知系统的实际推广应用做出了贡献。

本文也存在一些不足,由于方案仍基于传统IM算法,故在时间效率上仍有较大的提升空间,另外,在现实情况下概率信息可能随时间产生变化,各节点的激活成本也可能不尽相同,因此本文方案仍需改进以适应更多复杂情形。

参考文献

- Ganti RK, Ye F, Lei H. Mobile crowdsensing: Current state and future challenges. *IEEE Communications Magazine*, 2011, 49(11): 32–39. [doi: 10.1109/MCOM.2011.6069707]
- 刘云浩. 群智感知计算. *中国计算机学会通讯*, 2012, 8(10): 38–41.
- 於志文, 於志勇, 周兴社. 社会感知计算: 概念、问题及其研究进展. *计算机学报*, 2012, 35(1): 16–26.
- Ma HD, Zhao D, Yuan PY. Opportunities in mobile crowd sensing. *IEEE Communications Magazine*, 2014, 52(8): 29–35. [doi: 10.1109/MCOM.2014.6871666]
- Liu YT, Kong LH, Chen GH. Data-oriented mobile crowdsensing: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 2019, 21(3): 2849–2885.
- Capponi A, Fiandrino C, Kantarci B, *et al.* A survey on mobile crowdsensing systems: Challenges, solutions, and opportunities. *IEEE Communications Surveys & Tutorials*, 2019, 21(3): 2419–2465.
- Song YW, Jin HM. Minimizing entropy for crowdsourcing with combinatorial multi-armed bandit. *Proceedings of the 2021 IEEE Conference on Computer Communications*. Vancouver: IEEE, 2021. 1–10.
- Dai ZP, Wang H, Liu CH, *et al.* Mobile crowdsensing for data freshness: A deep reinforcement learning approach. *Proceedings of the 2021 IEEE Conference on Computer Communications*. Vancouver: IEEE, 2021. 1–10.

- 9 Ding R, Yang ZX, Wei YF, *et al.* Multi-agent reinforcement learning for urban crowd sensing with for-hire vehicles. Proceedings of the 2021 IEEE Conference on Computer Communications. Vancouver: IEEE, 2021. 1–10.
- 10 Wang XM, Tushar W, Yuen C, *et al.* Promoting users' participation in mobile crowdsourcing: A distributed truthful incentive mechanism (DTIM) approach. IEEE Transactions on Vehicular Technology, 2020, 69(5): 5570–5582.
- 11 Lim WYB, Xiong ZH, Miao CY, *et al.* Hierarchical incentive mechanism design for federated machine learning in mobile networks. IEEE Internet of Things Journal, 2020, 7(10): 9575–9588. [doi: [10.1109/JIOT.2020.2985694](https://doi.org/10.1109/JIOT.2020.2985694)]
- 12 Pandey SR, Tran NH, Bennis M, *et al.* A crowdsourcing framework for on-device federated learning. IEEE Transactions on Wireless Communications, 2020, 19(5): 3241–3256. [doi: [10.1109/TWC.2020.2971981](https://doi.org/10.1109/TWC.2020.2971981)]
- 13 Jin XC, Zhang YC. Privacy-preserving crowdsourced spectrum sensing. IEEE/ACM Transactions on Networking, 2018, 26(3): 1236–1249. [doi: [10.1109/TNET.2018.2823272](https://doi.org/10.1109/TNET.2018.2823272)]
- 14 Wang ZB, Hu JH, Lv RZ, *et al.* Personalized privacy-preserving task allocation for mobile crowdsensing. IEEE Transactions on Mobile Computing, 2019, 18(6): 1330–1341. [doi: [10.1109/TMC.2018.2861393](https://doi.org/10.1109/TMC.2018.2861393)]
- 15 Domingos P, Richardson M. Mining the network value of customers. Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco: ACM, 2001. 57–66.
- 16 Kempe D, Kleinberg J, Tardos É. Maximizing the spread of influence through a social network. Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Washington: ACM, 2003. 137–146.
- 17 Leskovec J, Krause A, Guestrin C, *et al.* Cost-effective outbreak detection in networks. Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Jose: ACM, 2007. 420–429.
- 18 Wang C, Chen W, Wang YJ. Scalable influence maximization for independent cascade model in large-scale social networks. Data Mining and Knowledge Discovery, 2012, 25(3): 545–576. [doi: [10.1007/s10618-012-0262-1](https://doi.org/10.1007/s10618-012-0262-1)]
- 19 Chen W, Wang YJ, Yang SY. Efficient influence maximization in social networks. Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Paris: ACM, 2009. 199–208.
- 20 Cesa-Bianchi N, Lugosi G. Prediction, Learning, and Games. Cambridge: Cambridge University Press, 2006.
- 21 Borgs C, Brautbar M, Chayes J, *et al.* Maximizing social influence in nearly optimal time. Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms. Oregon: SIAM, 2014. 946–957.

(校对责编: 牛欣悦)