

地理社交网络中重叠种子的广告博弈决策机制

于亚新 王 磊

(东北大学计算机科学与工程学院 沈阳 110169)
(yuyaxin@ise.neu.edu.cn)

An Advertising Game Theory Decision-Making Mechanism for Overlapping Seeds in Geo-Social Network

Yu Yaxin and Wang Lei

(College of Computer Science and Engineering, Northeastern University, Shenyang 110169)

Abstract Social advertising (or social promotion), one of the most important application for influence maximization (IM), has become a hot industry. The purpose of social advertising is to identify a seed set of k nodes for maximizing the expected influence spread to make businesses promote products by utilizing the cascade effect. However, most existing works concerning influence maximization problem were confined to behaviors that were observed mostly in the virtual world and neglected the location information. In fact, the distance between users can also affect propagation probability during information propagation. Thus, a problem of location-aware influence maximization (LAIM) in geo-social network is formally defined in this paper. Further, a location-aware influence maximization algorithm based on greedy frame is proposed, which introduces the promotion location information into the existing IM definition and solves the problem that influence spread is not in line with actual demand in traditional IM. Due to the overlapping seeds problem caused by inevitable competition, some of the selected nodes may not work well in practice. Thus we first conduct a decision-making game from overlapping seed's perspective to obtain Nash equilibrium by analyzing the payoff matrix of overlapping seeds and their neighbors so that the overlapping ratio of seed set and influence loss will be reduced. Finally, comprehensive experiments demonstrate the effectiveness of our algorithm and strategy.

Key words geo-social network; location-aware influence maximization (LAIM); overlapping seeds; game theory; Nash equilibrium

摘 要 作为社交影响最大化问题的重要应用之一,社交广告(或社交营销)已成为一个热门行业.其目标就是寻找 k 个最具影响力的种子节点,使产品公司利用成员间推介的“级联”效应推销产品.然而,由于所用数据集的限制,大多数现有影响力最大化问题的研究成果只能用于分析用户在虚拟世界中的行为,忽略了位置信息所起的作用.在信息传播过程中,用户间的距离也会对传播概率造成影响.因此,对地理社交网络中的位置敏感的影响力最大化(location-aware influence maximization, LAIM)问题进行了定义,并提出一种贪婪框架下考虑位置的影响力最大化算法,该算法将营销位置信息引入现有影响力

收稿日期:2018-01-31;修回日期:2018-11-29
基金项目:国家重点研发计划项目(2016YFC0101500);国家自然科学基金项目(61871106)

This work was supported by the National Key Research and Development Plan of China (2016YFC0101500) and the National Natural Science Foundation of China (61871106).

最大化(influence maximization, IM)问题定义中,解决了传统 IM 中由于缺少位置信息所导致的传播范围与实际需求不符问题,此外,鉴于同一领域不可避免的竞争会引发种子重叠现象,从而导致种子个体不能实现预期传播范围,立足重叠种子角度,旨在对公司选择进行决策博弈并找到纳什均衡点,从而降低了种子集合的重叠率与影响力损失.实验结果验证了贪婪框架下考虑位置的影响力最大化算法和重叠种子下决策博弈策略的有效性.

关键词 地理社交网络;位置敏感的影响力最大化问题;重叠种子;博弈论;纳什均衡

中图法分类号 TP399

随着 Twitter、Facebook、微博等社交应用的兴起,社交网络已成为信息传播的重要媒介,许多公司也开始将其作为广告推广和产品营销的主要工具.与传统大众营销方式不同,作为数字广告领域中发展最快的行业,社交广告不仅关注消费者的固有价值,还注重其在信息传播中的作用.在现实世界中,人们对于是否购买一种产品或接受一种想法多受其朋友、同事或同龄人的影响.因此一些公司会采用免费试用或利益回报的方式在整个网络中寻找意见领袖,即寻找最具影响力的种子,以便通过这些种子个体的社交影响力来扩大产品或广告的影响范围,从而获得更大收益.这种基于“口碑”^[1]的营销方式,利用成员间推介的“级联”效应使产品的潜在客户不断增多.与影响力最大化(influence maximization, IM)问题^[2-3]类似,社交广告要解决的问题是,在特定传播模型下,找到 k 个最具影响力的节点集合,从而最大化产品或广告的影响范围.近年来,病毒营销^[4]中的影响力最大化问题被广泛研究,但由于所用数据集的限制,大多数现有研究成果只能用于分析用户在虚拟世界中的行为,而忽略了位置信息的作用.幸运的是,随着无线通信设备和定位技术的发展,人们可以通过网络随时分享位置和想法,由此产生了海量的时空数据,从而使得在传统社交网络中加入空间这一维度成为可能.

举个例子,许多现实世界中的营销活动都增加了对营销位置^[5-6]这一因素的考量,比如,图 1 解释了与营销位置有关的用户进行信息传播的过程,其中 u_1 和 u_2 分别代表 2 个传播种子,他们可以继续向自己的下层传播,如黑线所示.当营销活动主办方在位置 V 进行营销活动时,距离 V 较近的用户(如图 1 中的 u_1),相比距离 V 较远的用户(如图 1 中的 u_2),往往更有可能成为营销活动的成功推广者或参与者,因为较近的距离是影响决策的重要因素之一.此时,距离营销位置不同的用户应给予不同权重的考虑.类似地,在信息传播过程中,与 u_1 距离较近的

朋友(圆圈所示),因同处于营销位置附近而更可能受其影响成为潜在顾客.进一步而言,用户间的影响概率也应根据二者间的物理距离进行动态推演,但遗憾的是,现有 IM 模型由于未考虑空间位置信息,因此不能直接用于解决该问题.



Fig. 1 Location-aware promotion

图 1 考虑位置信息的营销活动示意图

基于上述问题,本文首先提出一种位置敏感独立级联模型(distance-aware independent cascade, DAIC),并基于该模型对地理社交网络中的影响力最大化(location-aware influence maximization, LAIM)问题进行了重新定义,进而提出一种在贪婪框架下的位置敏感影响力最大化算法 LA_Greedy(location-aware greedy).

值得注意的是,大多数 IM 问题的研究工作是在“忽略竞争”情况下展开^[7-9],即假定 1 家公司可以在网络中独立于其他公司进行种子挑选,但实际上几家公司常常在同一市场相互竞争(例如宝马、奔驰、奥迪在汽车市场中的竞争),并在同一网络挑选种子,由此,针对竞争网络的影响力最大化问题得到了广泛关注与研究^[10-11].虽说如此,研究者在解决竞争环境下的 IM 问题时,做出了许多不切实际的假设^[12-13],比如 1 家公司可以在知道竞争对手的种子挑选结果后再做出最有利本公司的选择.然而在现实世界中,公司的营销策略和种子选择结果都是重要的商业秘密,不可能轻易地被他人获知.因此,同一领域的 2 家公司常常要在无任何辅助信息情况下完成种子挑选工作,这往往会导致种子重叠现象^[14],

使得一些种子个体不能实现预期的传播范围,从而造成影响力的损失^[15].

考虑图 2(a)中的无竞争网络,为了扩大新款车型的影响,假定公司甲选择节点 A 和节点 B 作为初始推广者,期待利用 A 和 B 的个人影响扩大该款汽车的市场影响力,进而得到更多消费者青睐,并最终可以影响到 7 人.在图 2(b)所示的竞争网络中,假设甲公司竞争对手即乙公司,也要推出一款新车型,在无法获知竞争对手甲公司的选择结果下,同样可将节点 A 和节点 B 作为推销甲公司的车的初始推广者,继而在信息传播过程中,二级推广者 C 将受到 A 和 B 的共同影响.但由于受到现有广告传播机制的限制,C 往往只会选择一款汽车进行推荐,因为公司只与初始推广者(即种子个体)存在利益交互(如免费试用或根据最终影响范围给予一定利益回报),二级推广者 C 仅仅靠与种子个体之间的往来关系决定是否继续进行推荐.因此相对自由的二级推广者在做推荐选择时会受到更多因素的影响(如对重叠种子的信任程度,甚至可能直接被其他竞争公司选为其产品或广告的推广者).

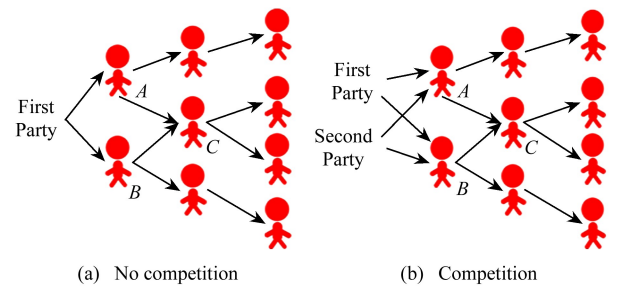


Fig. 2 Information propagation in social network
图 2 社交网络信息传播示意图

综上所述,无论是重叠种子个体(overlapping seeds)A 与 B 的出现,还是二级推广者 C 不确定的选择结果,都会使同一网络中相互竞争的 2 家公司不能达到预期的影响力传播范围.同样地,作为被公司寄予厚望的种子个体也会因其推广效果不佳而受到相应损失.为解决该问题,本文通过分析重叠种子与其邻居节点的支付矩阵,首次为重叠种子个体提供了选择公司的博弈决策机制,使重叠种子在博弈中找到纳什均衡点,进而决策出将为其推销产品的最佳公司,相应地,该博弈决策机制也降低了种子集合的重叠率和影响力的损失.

综上所述,本文主要贡献有 3 个方面:
1) 提出了一种距离敏感独立级联模型 DAIC,该模型将用户间物理距离引入独立级联模型(independent cascade, IC)中,量化了距离因素对传

播概率的影响,解决了地理社交网络中用户间影响力概率的动态推演问题.

2) 提出了一种贪婪框架下位置敏感影响力最大化算法 LA_Greedy,该算法基于 DAIC 模型将营销物理位置引入影响力最大化(IM)的定义中,解决了传统 IM 由于未考虑位置信息所导致的影响力传播范围与实际可用范围间的距离偏差问题.

3) 提出了一种面向重叠种子的竞争博弈决策机制(game decision-making mechanism, GDMM),该机制首次立足广告重叠种子角度,通过分析重叠种子与其邻居节点的支付矩阵,对竞争选择进行决策博弈,最终找到纳什均衡点,从而降低了种子重叠率和影响力损失.

1 基于 DAIC 模型的 IM 问题

本节首先介绍距离敏感独立级联模型 DAIC,然后基于该模型重新定义了地理社交网络中的影响力最大化问题.为清楚起见,表 1 给出了本文用到的一些符号及其含义:

Table 1 Notation of Symbols	
表 1 符号及含义	
Symbol	Signification
G	Social/geographic social network
V	Network user set
E	User relationship set
u, v	Network user u, v
$\langle u, v \rangle$	Directed edge $\langle u, v \rangle$
l	Marketing location
S	Seed set
p_{uv}	Consider the probability of the physical distance between the user u and v
$I(S)$	Influence spread of the set v
$w(v, l)$	The weight of the node at position $\beta=0$
$\sigma(S)$	The influence spread in the context of considering the set S

1.1 距离敏感独立级联模型 DAIC

网络用户间影响概率的大小是决定传播范围的关键因素之一,同一节点其传播能力在不同的影响概率下具有很大差异.在 IC 模型中,用户间的影响概率通常被随机指定,但在携带位置信息的地理社交网的营销活动中,用户间影响概率还会受到彼此间物理距离的作用,因此,不能直接套用 IC 模型对影响概率值的随机指派方法.基于此,本文提出一种距离敏感独立级联模型 DAIC,旨在有效量化位置

距离对传播概率的影响.直观地,当用户 u 要在位置 l_u 进行产品推荐时,在其朋友中,与之距离较近的用户会因同处于营销位置附近而更可能被影响,从而成为潜在顾客,因此有理由认为, u 对邻居 v 的影响力会随着彼此间物理距离的增加而减弱,即用户间位置距离与影响概率成反比.下面,对用户间影响概率的理论计算推演过程做详细阐述.

不失一般性,假定地理社交网络中 u 对 v 的影响概率为 p_{uv} , u 的地理位置为 l_u , $p_{u \rightarrow v}(l, l_u)$ 表示当用户 v 处在位置 l 时 u 对 v 的影响概率,简写为 $p(l)$.令 $dl > 0$ 表示一个极小的距离增量,则当用户 v 处在位置 $l + dl$ 时, u 对 v 的影响概率为 $p(l + dl)$,影响概率的改变量则为 $dp = p(l + dl) - p(l)$.如前所述,节点 u 对 v 的影响力会随着距离的增加而减弱,于是有 $dp = p(l + dl) - p(l) < 0$,表示对应于距离增量 dl 的影响概率损失.与放射性衰变^[16-17]类似(微粒越大,衰变越快),理论上, dp 正比于 $p(l)$:

$$\frac{p(l + dl) - p(l)}{dl} \propto p(l). \quad (1)$$

等式形式:

$$\frac{dp}{dl} = \frac{p(l + dl) - p(l)}{dl} = -\frac{p(l)}{\tau}, \quad (2)$$

其中,负号表示 $p(l)$ 随着距离的增加而减小,比例常数为 τ ,整理式(2),得:

$$\frac{dp}{dl} = -\frac{p}{\tau}. \quad (3)$$

整理式(3),得:

$$\frac{dp}{p} = -\frac{dl}{\tau}. \quad (4)$$

将式(4)两边同时积分,得:

$$\int \frac{dp}{p} = -\int \frac{dl}{\tau}. \quad (5)$$

整理式(5),得:

$$\ln(p) = -\frac{l}{\tau} + c. \quad (6)$$

整理式(6),得:

$$e^{\ln(p)} = e^{-\frac{l}{\tau} + c} = e^c e^{-\frac{l}{\tau}}, \quad (7)$$

其中, c 为积分常数,整理式(7),得:

$$p(l) = e^c e^{-\frac{l}{\tau}}. \quad (8)$$

令 $l(u) = 0$,则在式(8)中, $\Delta = l - l_u = l$ (在二维空间中 Δ 表示 l 与 l_u 之间的欧氏距离),即式(8)从理论上证明了 u 对 v 的影响概率随着距离的增加呈指数衰减.通过简化,定义用户间位置敏感的影响概率计算公式为

$$p_{uv} = e^{-\alpha l}. \quad (9)$$

与 IC 模型类似,DAIC 模型的信息传播过程如下:给定初始活跃集合 A ,当不活跃节点 u 在时间步 t 变为活跃状态后,该节点将尝试激活每个处在不活跃状态的邻居.但无论激活结果如何,在以后的时间步中, u 都不会再对其邻居产生任何影响.若成功,则邻居节点 v 在时间步 $t + 1$ 变为活跃状态.当节点 v 有多个邻居节点变为活跃状态时,这些节点可以按任意顺序尝试激活 v .当不再存在激活可能时,传播过程结束.

1.2 基于 DAIC 模型的影响力最大化问题

在社交网络研究中,通常将有向图 $G(V, E)$ 作为抽象研究对象,其中 V 为节点的集合, E 为边的集合.

定义 1. 影响力传播范围.给定一个网络 $G(V, E)$ 和一个集合 $S \subset V$,集合 S 的影响力传播范围 $I(S)$ 代表 S 最终影响的人数,也可通过 $I(S) = \sum_{v \in V} p(S \rightarrow v)$ ^[18] 进行计算.

定义 2. 位置敏感影响力传播范围.给定一个地理社交网络 $G(V, E)$ 和一个营销位置 l ,以及集合 $S \subset V$.考虑位置的影响力传播范围

$$\sigma(S) = \sum_{v \in V} p(S \rightarrow v) w(v, l),$$

其中, $w(v, l)$ 表示节点 v 对应营销位置 l 的权值.

如上所述,营销位置附近的用户应该具有更高的权值,即 $w(v, l)$ 与 v 和 l 间的距离成反比.为了简化,本文令 $w(v, l) = c e^{-\beta d(v, l)}$,其中 $c > 0$ 表示节点能够达到的最大权值, $d(v, l)$ 为节点 v 与营销位置 l 间的欧氏距离, $\beta > 0$ 为衰减速率.

问题 1. 影响力最大化(IM)问题.给定一个社交网络 $G(V, E)$ 和种子个数 k ,求一个节点集合 $S \subset V, |S| = k$,使得以 S 作为初始种子集合时,在特定信息传播模型下具有最大的影响力传播范围,即对于任意大小为 k 的节点集合 $S', I(S) \geq I(S')$,集合 S 为

$$S = \arg \max_{S' \subset V, |S'| = k} I(S'). \quad (10)$$

问题 2. 位置敏感影响力最大化(LAIM)问题.给定一个地理社交网络 $G(V, E)$ 、营销位置 l 及种子个数 k ,要求找到一个节点集合 $S \subset V, |S| = k$,使得以 S 作为初始种子集合在 DAIC 模型上进行影响力最大化传播时,其位置敏感的影响力传播范围 $\sigma(S)$ 最大,即对于任意大小为 k 的节点集合 $S', \sigma(S) \geq \sigma(S')$,集合 S 为

$$S = \arg \max_{S' \subset V, |S'| = k} \sigma(S'). \quad (11)$$

2 位置敏感影响力最大化算法 LA_Greedy

算法 1. LA_Greedy 算法.

输入: 地理社交网络 G 、种子集合大小 k 、营销位置 l ;

输出: 种子集合 S .

- ① Off-line compute p_{uv} for $\langle u, v \rangle \in E$; /* 计算每对用户间影响概率 p_{uv} */
- ② compute $w(v, l)$ for $v \in V$; /* 为用户分配不同权值 */
- ③ $S \leftarrow \emptyset$; /* 给定种子集合 S , 初始为空 */
- ④ for $i=1$ to k do /* 根据种子个数 k 进行循环计算 */
- ⑤ $u \leftarrow \arg \max_{w \in V \setminus S} (\sigma(S \cup \{w\}) - \sigma(S))$; /* 获得最大边际收益节点 u */
- ⑥ $S \leftarrow S \cup \{u\}$; /* 将 u 加入种子集合 S */
- ⑦ end for
- ⑧ return S .

在位置敏感影响力最大化问题 LAIM 中, 当权值函数 $w(v, l) = ce^{-\beta d(v, l)}$ 的参数 $c=1, \beta=0$ 时, 网络中所有节点的权值都变为与位置无关的常量 1, 这时 LAIM 就等价于 IM, 因此, LAIM 问题是 NP-难的^[19]. 类似地, 可以证明计算位置敏感的影响力传播范围也是一个 NP-难问题^[20]. 基于影响力传播函数的性质, 本文提出一种贪婪框架下位置敏感影响力最大化算法 LA_Greedy, 算法 1 给出了其详细步骤.

步骤 1. 以网络个体间的位置距离为输入, 利用式(9)计算每对用户 u 和 v 间的影响概率 p_{uv} ;

步骤 2. 根据网络中各用户 v 与特定营销推广位置 l 之间的距离, 为用户分配不同的权值, 即 $w(v, l) = ce^{-\beta d(v, l)}$.

步骤 3. 给定种子集合 S , 初始为空, 根据种子个数 k 进行循环运算.

步骤 4. 基于目标函数 $\sigma(S)$, 在每次迭代过程中, 将具有最大边际收益的节点 u 加入到集合 S 中.

步骤 5. 循环结束, 将集合 S 作为初始种子集合返回.

3 重叠种子个体的博弈决策机制 GDMM

现实世界同一领域中不可避免的竞争往往引发种子重叠现象, 导致一些种子个体不能实现预期的

传播范围, 造成影响力损失. 通过分析重叠种子与其邻居节点的支付矩阵, 本文首次为重叠种子个体制定选择公司的博弈决策机制 GDMM, 使其在博弈中找到纳什均衡点并作出最佳公司选择决策. GDMM 博弈过程如下.

根据重叠种子个体和二级推广者做出不同决策时的利益得失情况, 分别给出双方的支付矩阵 S_{Seed} 与 $N_{Neighbor}$:

$$S_{Seed} = \begin{pmatrix} S_{Aincome}^A \gamma_1 & -S_{Aloss}^B \gamma_2 \\ -S_{Bloss}^A \gamma_3 & S_{Bincome}^B \gamma_4 \end{pmatrix}, \quad (12)$$

$$N_{Neighbor} = \begin{pmatrix} 0 & N_{Abenefit}^A \gamma_5 - N_{Floss}^A \gamma_6 \\ N_{Bbenefit}^B \gamma_7 - N_{Floss}^B \gamma_8 & 0 \end{pmatrix}, \quad (13)$$

其中, $\gamma_k \in [0, 1] (k=1, 2, \dots, 8)$ 为博弈分析的参数因子, 主要取决于个体间的信任程度, 可以根据具体要求进行调整.

$S_{Aincome}^A$ 表示种子个体选择 A 公司, 邻居节点也受其影响为 A 推荐产品或传播广告时, 种子从 A 公司获得的收益 (如免费试用新品或得到约定的回报).

S_{Aloss}^B 表示种子个体选择 A 公司, 邻居节点受其他因素影响转而为 B 推荐产品或传播广告时, 种子个体遭受的损失 (如未达到预期影响范围阻碍了与 A 公司日后的合作, 从而造成一定的经济损失).

S_{Bloss}^A 表示种子个体选择 B 公司, 邻居节点受其他因素影响转而为 A 推荐产品或传播广告时, 种子个体遭受的损失 (如未达到预期影响范围阻碍了与 B 公司日后的合作, 从而造成一定的经济损失).

$S_{Bincome}^B$ 表示种子个体选择 B 公司, 邻居节点也受其影响为 B 推荐产品或传播广告时, 种子从 B 公司获得的收益 (如免费试用新品或得到约定的回报).

$N_{Bbenefit}^B$ 表示种子个体选择 A 公司, 邻居节点 (二级推广者) 受其他因素影响或直接被竞争公司选中为 B 推荐产品或传播广告时, 二级推广者获得的收益.

$N_{Abenefit}^A$ 表示种子个体选择 B 公司, 邻居节点 (二级推广者) 受其他因素影响或直接被竞争公司选中为 A 推荐产品或传播广告时, 二级推广者获得的收益.

N_{Floss}^B 表示种子个体选择 A 公司, 邻居节点 (二级推广者) 受其他因素影响转而为 B 推荐产品或传播广告时, 二级传播者受到的损失. (由于公司

只与种子存在利益关系,故该损失是对 2 个体间信任缺失的量化).

$N_{Floss\ B}^A$ 表示种子个体选择 B 公司,邻居节点(二级推广者)受其他因素影响转而为 A 推荐产品或传播广告时,二级传播者受到的损失.(由于公司只与种子存在利益关系,故该损失是对 2 个体间信任缺失的量化).

为计算方便,对 Neighbor 在原支付矩阵的基础

$$\begin{aligned} E_{Neighbor}(P_{Seed}, P_{Neighbor}) &= P_{Neighbor} N_{Neighbor} P_{Seed}^T = \\ (y, 1-y) &\begin{pmatrix} \mathbf{0} & N_{Abenefit\ B}^A \gamma_5 - N_{Floss\ B}^A \gamma_6 \\ N_{Bbenefit\ A}^B \gamma_7 - N_{Floss\ A}^B \gamma_8 & \mathbf{0} \end{pmatrix} \begin{pmatrix} x \\ 1-x \end{pmatrix} = \\ &x N_{Bbenefit\ A}^B \gamma_7 - x N_{Floss\ A}^B \gamma_8 - x y N_{Bbenefit\ A}^B \gamma_7 + x y N_{Floss\ A}^B \gamma_8 + \\ &y N_{Abenefit\ B}^A \gamma_5 - y N_{Floss\ B}^A \gamma_6 - x y N_{Abenefit\ B}^A \gamma_5 + x y N_{Floss\ B}^A \gamma_6. \end{aligned} \tag{14}$$

对式(14)关于 y 求偏导,可得重叠种子个体最优一阶条件:

$$\begin{aligned} \frac{\partial E_{Neighbor}(P_{Seed}, P_{Neighbor})}{\partial y} &= -x N_{Bbenefit\ A}^B \gamma_7 + \\ &x N_{Floss\ A}^B \gamma_8 + N_{Abenefit\ B}^A \gamma_5 - N_{Floss\ B}^A \gamma_6 - \\ &x N_{Abenefit\ B}^A \gamma_5 + x N_{Floss\ B}^A \gamma_6 = 0. \end{aligned}$$

解得 x :

$$x = \frac{N_{Abenefit\ B}^A \gamma_5 - N_{Floss\ B}^A \gamma_6}{N_{Bbenefit\ A}^B \gamma_7 - N_{Floss\ A}^B \gamma_8 + N_{Abenefit\ B}^A \gamma_5 - N_{Floss\ B}^A \gamma_6}.$$

即在博弈过程中,重叠种子个体选择公司的最佳混合策略为 $P_{Seed} = (x, 1-x)$.可以看出,重叠种子个体选择 A 公司的概率 x 与其邻居节点被 A 公司选中时获得的收益有关,邻居节点从 A 公司获得的收益越多,重叠种子选择 A 公司的概率越大;反之,则更倾向于选择 B 公司.

4 实验数据集与性能测试

4.1 实验数据与测试环境

为验证所提算法和策略的有效性,利用 Twitter 官方提供的 API 接口采集了 2016 年 10 月内 1 周的数据.经过清洗得到有效节点 6 046 个、节点间关系数据 14 487 条.由于所采数据的用户地理位置比较紧密(各节点距中心位置的平均距离约为 72.99 km).为了检测所提算法在不同类型测试集上的效果,又模拟了具有 12 000 个节点、28 320 条边的位置较为疏远(各节点距中心位置的平均距离约为 341.03 km)的数据集,具体情况如表 2 所示.

在实验中,采用 Java 语言编程,CPU 为 Intel Core i5-4590 3.3 GHz,内存为 8 GB,操作系统为 Windows 7.

上进行了转置.通过简单的画线法可以得出该博弈模型不存在纯策略均衡,但可求出混合策略的纳什均衡.假定重叠种子个体以概率 x 选择 A 公司,以概率 $1-x$ 选择 B 公司,即重叠种子个体的混合策略为 $P_{Seed} = (x, 1-x)$;二级推广者以概率 y 选择 A 公司,以概率 $1-y$ 选择 B 公司,即二级推广者的混合策略为 $P_{Neighbor} = (y, 1-y)$.则重叠种子个体的预期支付函数为

Table 2 Experimental Data Set

表 2 实验数据集

Data Type	Node Number	Number of Relationships Between Nodes	Average Distance of the Node from the Center Position/km
Twitter	6 046	14 478	72.99
Analog Data	12 000	28 320	341.03

4.2 参数设置

实验中,营销位置 l 与各节点的公司选择通过随机指定.设用户间影响概率计算公式 $n = |V|$ 的参数 α 为经验值 0.02,权值函数 $w(v, l) = c e^{-\beta d(v, l)}$ 中 $c = 10, \beta = 0.02$.通常每个重叠种子个体对应多个邻居,故在博弈分析中将邻居得失的均值作为支付函数参数, $\gamma_k \in [0, 1] (k = 1, 2, \dots, 8)$ 暂设为 1.

4.3 实验结果与性能分析

为体现社交网用户之间关系紧密程度,本文用网络聚类系数来评测, C_v 表示节点聚类系数, C_G 表示网络聚类系数,具体计算:

$$C_v = \frac{2|E_v|}{k_v(k_v - 1)}, \tag{15}$$

其中, $|E_v|$ 表示网络用户 v 的邻居节点间的实际连接数, k_v 为用户 v 的邻居个数.

$$C_G = \frac{1}{n} \sum_{v \in V} C_v, \tag{16}$$

其中, $V = \{v_1, v_2, \dots, v_n\}$ 表示网络 G 中的所有个体集合, $n = |V|$ 为个体数目.

通过计算发现,与具有相同节点数和边数的一个完全随机化网络相比,采集的 Twitter 数据所构建的社交网有着明显较高聚类系数,说明 Twitter 网用户间关系比较紧密,有利于信息传播.

为验证 GDMM 策略对重叠种子在进行公司选择时的指导效果,分别针对真实 Twitter 数据以及模拟数据,在 DAIC 与 IC 这 2 种传播模型下对种子集合的重叠率进行了测试(此处将 Greedy 与 LA_Greedy 算法分别作为 A 和 B 这 2 家公司挑选种子时所用的策略予以指定),实验结果分别如图 3 和图

4 所示,其中,横坐标代表种子个数 k ,纵坐标代表种子集合的重叠率.从图 3 和图 4 中可以看出,尽管随着 k 的增大,2 家公司的种子集合的重叠率都呈变大趋势,但相比于没有使用纳什平衡的方法,GDMM 策略却可以明显降低种子集合的重叠率,从而更好地保证传播效果.

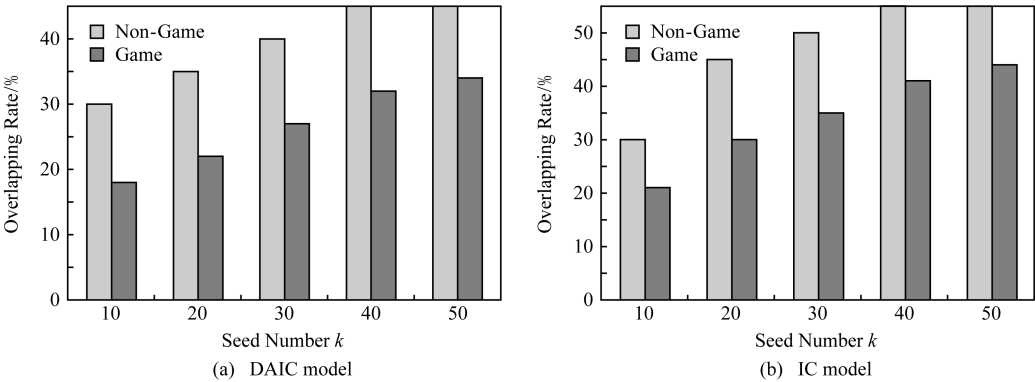


Fig. 3 Overlapping rate of Twitter data based on different models
图 3 Twitter 数据在不同传播模型下的种子重叠率

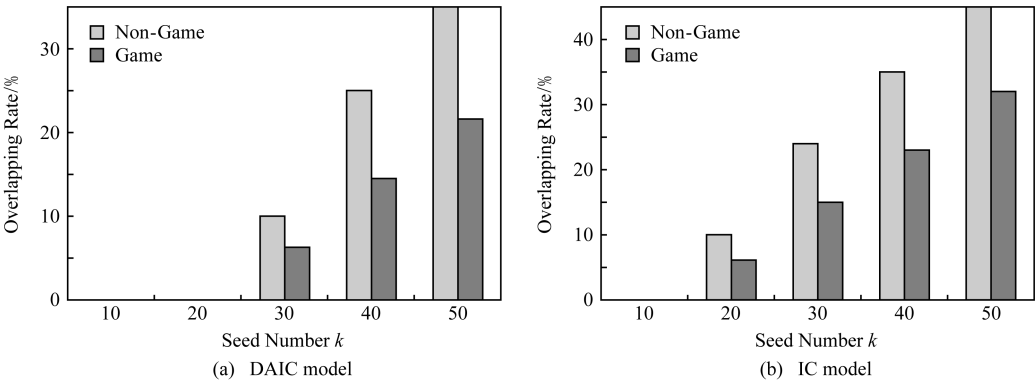


Fig. 4 Overlapping rate of simulated data based on different models
图 4 模拟数据在不同传播模型下的种子重叠率

为验证 LA_Greedy 算法的影响力传播范围,实验对种子集合的影响力传播过程进行了 20 000 次

模拟,然后将此过程所得平均值作为种子节点最终的影响力范围,并体现在影响人数和影响人数权值这

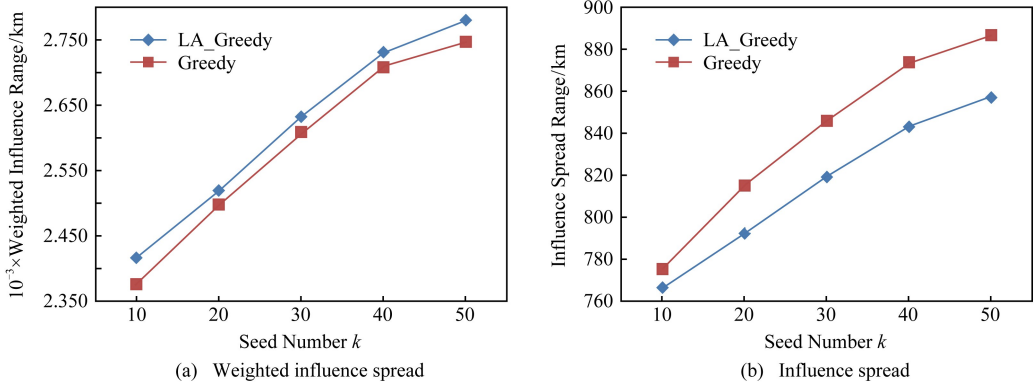


Fig. 5 Influence spread of Twitter data based on DAIC model
图 5 Twitter 数据在 DAIC 模型下的影响力传播

2 个指标上.图 5 和图 6 为 Greedy 和 LA_Greedy 这 2 种算法在 Twitter 数据集上分别在 DAIC 模型和 IC 模型下的测试结果,其中横坐标代表种子个数 k ,纵坐标代表传播过程结束后种子集合的影响范围和加权影响范围.

首先,从图 5 和图 6 中可以看出,随着 k 的增加,种子集合的影响范围与加权影响范围也随之增加,但当 k 增至一定规模后,其增长速度开始变得缓慢.这是因为随着具有较大影响力的节点不断地被选入到集合后,新加入种子的边际影响力也越来越小.

其次,从图 5 和图 6 中还可以看出,虽然通过 LA_Greedy 算法所求种子集合的最终影响人数略少于 Greedy 算法,但是 LA_Greedy 算法所影响节点的加权影响范围却高于 Greedy 算法,这说明 LA_Greedy 算法确实更加适用于解决地理社交网络中位置敏感的影响力最大化问题,即通过 LA_Greedy 算法所找到的种子集合可以影响更多网络中权值较大的用户,也就是营销位置附近的潜在顾客,解决了传统 IM 中由于缺少位置信息所导致的传播范围与实际需求不符问题.

图 7 和图 8 则给出了 Greedy 和 LA_Greedy 这 2 种算法在模拟数据集上分别在 DAIC 模型和 IC 模型下的测试结果.与 Twitter 数据类似,横坐标代表种子个数 k ,纵坐标代表传播过程结束后种子集合的影响范围和加权影响范围.分别对比图 5(a)与图 7(a)、图 5(b)与图 7(b),可以看出,用户在地理位置比较疏远的模拟数据集上的种子集合的影响范围与加权影响范围,要远远小于用户地理位置比较紧密的 Twitter 数据集上的传播结果,这是因为模拟数据集中各节点距中心位置的平均距离约为 341.03 km,类似于日常生活中相距较远的 2 个城市之间的距离,因此当给定营销位置后,网络中各用户的权值与用户间的影响概率的计算结果都非常小,导致种子集合的影响范围不理想.不过,此结果也符合日常生活中人们的行为习惯,即当用户受到线上影响而进入线下环节后,是否参与线下体验将取决于用户和推荐产品的位置特征,这是因为用户的消费行为受到地理位置影响,存在对消费位置的偏好,此偏好由用户的生活习惯和交通能力的因素决定,体现了用户的日常生活范围,推荐产品的位置越超出用户的

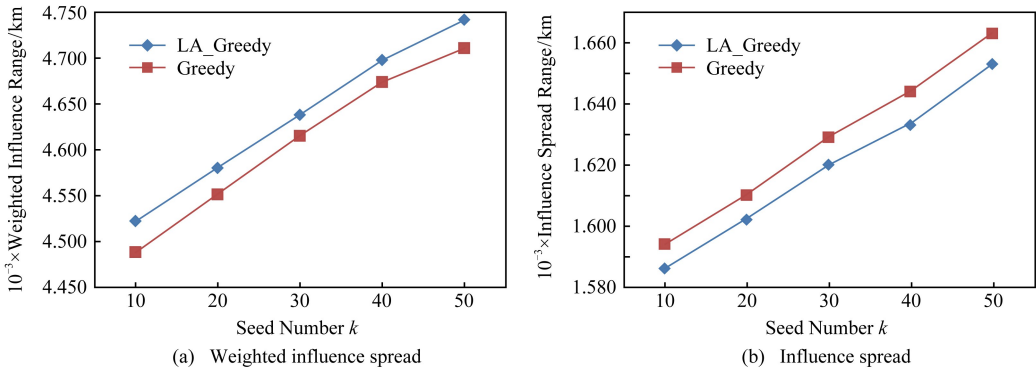


Fig. 6 Influence spread of Twitter data based on IC model

图 6 Twitter 数据在 IC 模型下的影响力传播

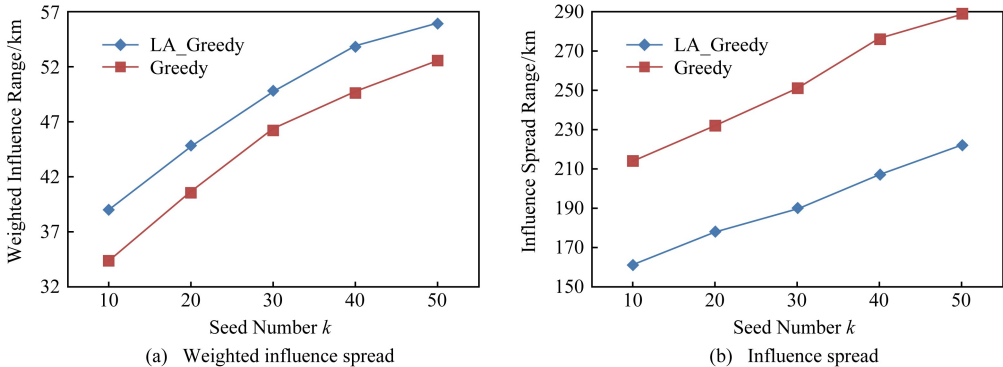


Fig. 7 Influence spread of simulated data based on DAIC model

图 7 模拟数据在 DAIC 模型下的影响力传播

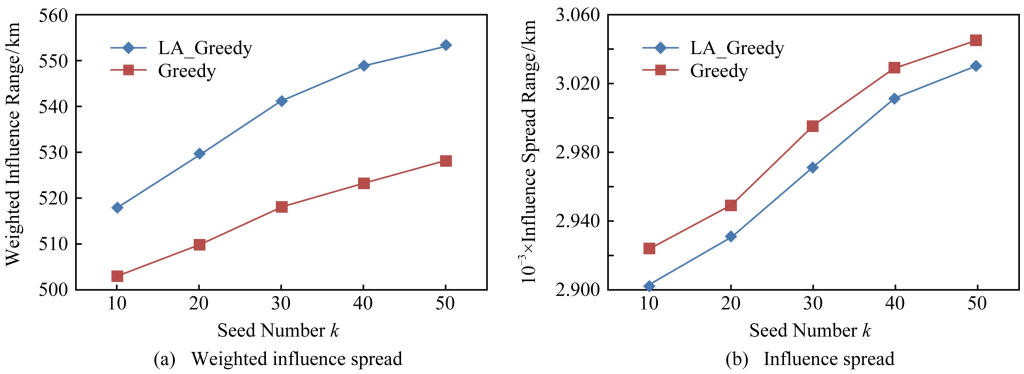


Fig. 8 Influence spread of simulated data based on IC model

图 8 模拟数据在 IC 模型的实验结果

日常生活范围,用户去尝试的概率就会越低,这也说明所提算法确实可以用于日常生活中针对某一特定范围内的营销推广活动。

5 总 结

基于放射性衰变定律,首先对地理社交网中用户间影响概率的计算公式进行了理论推导,并提出了一种距离敏感的 DAIC 传播模型,进一步,基于该模型又提出了贪婪框架下位置敏感的影响力最大化算法 LA_Greedy.实验证明,在求解位置敏感的影响力最大化问题时,LA_Greedy 算法与传统 Greedy 算法相比,前者可以吸引更多营销位置附近的用户,确保了用户的精准推荐.此外,当重叠种子根据所提的 GDMM 机制进行公司选择时,种子集合的重叠率确实得以明显下降,从而保证了有效的传播效果.未来工作将考虑把 Twitter 以外更多其他社交网数据纳入研究,并加强对部分垂直领域的数据分析处理,以体现本文工作可扩展性和实际应用价值。

参 考 文 献

[1] Steffes E M, Burgee L E. Social ties and online word of mouth[J]. Internet Research, 2013, 19(1): 42-59

[2] Domingos P, Richardson M. Mining the network value of customers[C] //Proc of the 7th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2001: 57-66

[3] Kempe D, Kleinberg J. Maximizing the spread of influence through a social network [C] //Proc of the 9th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2003: 137-146

[4] Aslay C, Lu Wei, Bonchi F, et al. Viral marketing meets social advertising: Ad allocation with minimum regret [J]. Proceedings of the Vldb Endowment, 2014, 8(7): 814-825

[5] Li Guoliang, Chen Shuo, Feng Jianhua, et al. Efficient location-aware influence maximization [C] //Proc of ACM SIGMOD Int Conf on Management of Data. New York: ACM, 2014: 87-98

[6] Zhu Wenyuan, Peng Wenchih, Zheng Kai, et al. Modeling user mobility for location promotion in location-based social networks [C] //Proc of the 21st ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2015: 1573-1582

[7] Goyal A, Lu Wei, Lakshmanan L V S. An efficient algorithm for influence maximization under the linear threshold model [C] //Proc of the 11th IEEE Int Conf on Data Mining. Piscataway, NJ: IEEE, 2012: 211-220

[8] Chen Wei, Wang Yajun, Yang Siyu. Efficient influence maximization in social networks [C] //Proc of ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2009: 199-208

[9] Wang Yaxuan, Wang Hongzhi, Li Jianzhong, et al. Efficient influence maximization in weighted independent cascade model [M] //Database Systems for Advanced Applications. Berlin: Springer, 2016: 49-64

[10] Bharathi S, Kempe D, Salek M. Competitive influence maximization in social networks [M] //Internet and Network Economics. Berlin: Springer, 2007: 306-311

[11] Carnes T, Nagarajan C, Wild S M, et al. Maximizing influence in a competitive social network: A follower's perspective [C] //Proc of the 9th Int Conf on Electronic Commerce. New York: ACM, 2007: 351-360

[12] Wu Hong, Liu Weiwei, Yue Kun, et al. Maximizing the spread of competitive influence in a social network oriented to viral marketing [M] //Web-Age Information Management. Berlin: Springer, 2015: 516-519

[13]

Li Hui, Bhowmick S S, Cui Jiangtao, et al. GetReal: Towards realistic selection of influence maximization strategies in competitive networks [C] //Proc of ACM SIGMOD Int Conf on Management of Data. New York: ACM, 2015: 1525-1537

[14]

Yu Zhiyong, Zhang Daqing, Yu Zhiwen, et al. Participant selection for offline event marketing leveraging location-based social networks [J]. IEEE Transactions on Systems Man & Cybernetics Systems, 2015, 45(6): 853-864

[15]

Zeng Yifeng, Chen Xuefeng, Cong Gao, et al. Maximizing influence under influence loss constraint in social networks [J]. Expert Systems with Applications, 2016, 55: 255-267

[16]

Krane K S, Ruby L, Williams R E. Post use review: Introductory nuclear physics [J]. American Journal of Physics, 1990, 58(2): 191-192

[17]

Ginzburg V L, Levin L M, Rabinovich M S, et al. Chapter I—Molecular physics and thermodynamics [M] //Molecular Physics, Thermodynamics, Atomic and Nuclear Physics. Amsterdam: Elsevier, 1965: 115-176

[18]

Hu Jie, Meng Kun, Chen Xiaomin, et al. Analysis of influence maximization in large-scale social networks [J]. ACM SIGMETRICS Performance Evaluation Review, 2014, 41(4): 78-81

[19]

Wang Xiaoyang, Zhang Ying, Lin Xuemin, et al. Distance-aware influence maximization in geo-social network [C] //Proc of the 32nd Int Conf on Data Engineering. Los Alamitos, CA: IEEE Computer Society, 2016: 1-12

[20]

Chen Wei, Wang Chi, Wang Yajun. Scalable influence maximization for prevalent viral marketing in large-scale social networks [C] //Proc of the 16th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2010: 1029-1038



Yu Yaxin, born in 1971. PhD, associate professor, MS supervisor. Member of IEEE, ACM and CCF. Her main research interests include data mining and social network.



Wang Lei, born in 1992. MSc. His main research interests include data mining, social network, machine learning and Web mining.

文章撤销声明

经反映并调查证实,本刊 2017 年 54 卷 1 期刊登的文章“网络媒体大数据流异构多模态目标识别策略”(作者:文孟飞、刘伟荣、胡超),与《中南大学学报(自然科学版)》2016 年 47 卷 5 期刊登的文章“一种基于深度学习的异构多模态目标识别方法”(作者:文孟飞、胡超、刘伟荣)存在较多雷同。

根据《计算机研究与发展》相关规定,即日起撤销文章“网络媒体大数据流异构多模态目标识别策略”,并追究主要作者的责任,特此声明!

《计算机研究与发展》编委会
《计算机研究与发展》编辑部
2019 年 5 月