

变,所有算法变化趋势基本一致.最初阶段数据比较稳定,所有算法都保持了较高和平稳的准确率,本文算法并无明显优势.随着数据量的增大,概念漂移次数的增多,所有算法的分类准确率都有所下降,其中 HT 算法下降的比较严重,且波动较大.当在 250 K, 500 K 和 750 K 处发生概念突变时,所有算法的准确率都急剧下降,而本文算法维持了较高和平稳的准确率.本文算法准确率的平均值比 HT 算法高出 20% 左右.这是由于本文算法能快速捕捉到概念变化,并建立新的分类器,从而能及时应对这种变化.由于数据中添加了 10% 的噪音,表明了本文算法具有抗噪声能力.

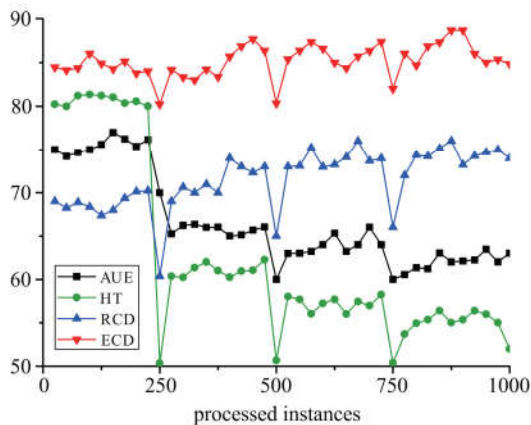


图 2 SEA 数据集上分类准确率比较

Fig.2 Accuracy on the SEA dataset

真实的数据流环境中概念变化具有不可预知性和不确定性,因此更能验证算法的泛化能力.在 Elist 上,准确率的变化情况如图 3 所示,所有算法的准确率曲线均出现不同程度的波动,这表明该数据集中存在概念漂移现象.由图 3 可知,本文算法的

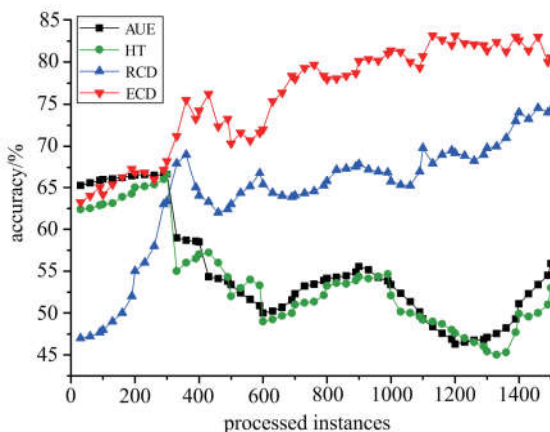


图 3 Elist 数据集上分类准确率比较

Fig.3 Accuracy on the Elist dataset

准确率曲线相对平稳,能充分利用历史分类器所拥有的知识来解决数据流概念漂移呈现周期性变化的问题,表明本文算法受到数据中概念漂移影响最小,对真实的数据环境有较好的适应性.

通过实验对比分析,可得出以下结论:①本文算法在包含重复概念的数据集上具有明显优势;②在保持较高分准确率情况下,消耗相对较少的时间;③对噪声具有一定的健壮性.

4 结论

本文主要研究如何利用历史概念对应的模型来分类预测概念重现的数据流,以减少模型更新的代价,从而加快分类预测的速度和分类准确率.本文提出了一种基于信息熵的概念检测模型,不仅能检测到概念漂移而且能识别重复概念.只有当检测到有新概念时才构建新的分类器并放入分类器池中,防止重复概念出现导致的重复训练的问题,减少模型更新频率,提高模型实时分类能力和分类效果.通过在人工合成数据集和真实数据集上与经典的数据流算法进行性能分析对比,实验结果表明,本文模型不仅能够应对多种类型概念漂移,提升分类模型抗噪声能力,并在保证较高分准确率前提下,消耗更少的时间代价.

目前,数据流学习算法主要针对有标记的数据进行处理,未考虑数据带有缺失属性值或不完整类标识的情况,因此如何在动态数据流环境下利用少量的标签实例信息指导无标签实例的标记,进而开展概念漂移检测与分类是值得进一步研究的问题.

参考文献 (References)

- [1] COHEN L, AVRAHAMI-BAKISH G, LAST M, et al. Real-time data mining of non-stationary data streams from sensor networks[J]. *Information Fusion*, 2008, 9(3): 344-353.
- [2] WANG H X, FAN W, YU P S, et al. Mining concept-drifting data streams using ensembles classifiers[C]// *Proceedings of 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM Press, 2003: 226-235.
- [3] ELWELL R, POLIKAR R. Incremental learning of concept drift in nonstationary environments[J]. *IEEE Transactions on Neural Networks*, 2011, 22(10): 1517-1531.
- [4] GAMA J. *Knowledge Discovery from Data Streams* [M]. New York: CRC Press, 2010.

- [5] WIDMER G, KUBAT M. Learning in the presence of concept drift and hidden contexts [J]. *Machine Learning*, 1996, 23(1): 69-101.
- [6] TSYMBAL A. The problem of concept drift: Definitions and related work[R]. Technical Report, Department of Computer Science, Trinity College, Dublin, Ireland, 2004.
- [7] GAMA J, ŽLIOBAITĖ I, BIFET A, et al. A survey on concept drift adaptation [J]. *ACM Computing Surveys*, 2014, 46(4): 231-238.
- [8] 亓开元, 赵卓峰, 房俊, 等. 针对高速数据流的大规模数据实时处理方法[J]. *计算机学报*. 2012, 35(3): 477-490.
- QI Kaiyuan, ZHAO Zhuofeng, FANG Jun, et al. Real-time processing for high speed data stream over large scale data[J]. *Chinese Journal of Computers*, 2012, 35(3): 477-490.
- [9] GAMA J, MEDAS P, CASTILLO G, et al. Learning with drift detection [C]// *Proceedings of the 17th Brazilian Symposium on Artificial Intelligence*. Berlin: Springer-Verlag, 2004: 286-295.
- [10] BAENA-GARCÍA M, CAMPO-ÁVILA J D, FIDALGO R, et al. Early drift detection method[C]// *Proceedings of the 4th International Workshop on Knowledge Discovery from Data Streams*. New York: ACM Press, 2006: 77-86.
- [11] NISHIDA K, YAMAUCHI K. Detecting concept drift using statistical testing[C]// *Proceedings of the 10th International Conference on Discovery Science*. Sendai, Japan: Springer-Verlag, 2007: 264-269.
- [12] BIFET A, GAVALDÀ R. Learning from time-changing data with adaptive windowing [C]// *Proceedings of the 7th SIAM International Conference on Data Mining*. Philadelphia, PA: SIAM Press, 2007: 443-448.
- [13] ROSS G J, ADAMS N M, TASOULIS D K, et al. Exponentially weighted moving average charts for detecting concept drift [J]. *Pattern Recognition Letters*, 2012, 33(2): 191-198.
- [14] RAMAMURTHY S, BHATNAGAR R. Tracking recurrent concept drift in streaming data using ensemble classifiers [C]// *Proceedings of the 6th International Conference on Machine Learning and Applications*. Cincinnati, USA: IEEE Press, 2007: 404-409.
- [15] KATAKIS I, TSOUMAKAS G, VLAHAVAS I. Tracking recurring contexts using ensemble classifiers: An application to email filtering[J]. *Knowledge and Information Systems*, 2010, 22(3): 371-391.
- [16] YANG Y, WU X D, ZHU X Q. Mining in anticipation for concept change: Proactive-reactive prediction in data streams [J]. *Data Mining and Knowledge Discovery*, 2006, 13(3): 261-289.
- [17] GAMA J, KOSINA P. Recurrent concepts in data streams classification[J]. *Knowledge and Information Systems*, 2014, 40(3): 489-507.
- [18] GONÇALVES P M, DE BARROS R S M. RCD: A recurring concept drift framework [J]. *Pattern Recognition Letters*, 2013, 34(9): 1018-1025.
- [19] DATAR M, GIONIS A, INDYK P, et al. Maintaining stream statistics over sliding windows [J]. *SIAM Journal on Computing*, 2002, 31(6): 1794-1813.
- [20] DASU, T, KRISHNAN S, VENKATASUBRAMANIAN S, et al. An information-theoretic approach to detecting changes in multi-dimensional data streams [C]// *Proceedings of the 38th Symposium on the Interface of Statistics, Computing Science, and Applications*. Pasadena, USA: IEEE Press, 2006: 1-24.
- [21] BIFET A, HOLMES G, KIRKBY R, et al. MOA: Massive online analysis [J]. *Journal of Machine Learning Research*, 2010, 11(2): 1601-1604.
- [22] DOMINGOS P, HULTEN G. Mining high-speed data streams[C]// *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Boston, USA: ACM Press, 2000: 71-80.

移动广告用户传播能力评价与覆盖优化算法

徐婉茹, 杨盘隆

(解放军理工大学通信工程学院, 江苏南京 210007)

摘要:移动广告的分发效果对于广告商和用户来说都是相当重要的事情.目前对于高效率的广告分发特别是对用户的轨迹和预算的研究较为匮乏.为了获得有效可行的移动广告分发策略,提出了以位置为中心的移动众包网络的概念,代替传统的以用户为中心的网络和平台,其中,位置信息对于广告分发起到至关重要的作用.为此重点研究考虑有兴趣区域的覆盖(interested area coverage, IAC)策略下的移动广告用户选择问题.对于以位置为中心的研究需要考虑每个用户的时空特性,并需要有效地计算有兴趣的覆盖区域,资金预算的约束使这一问题更加难以解决.为应对上述挑战,首先,考虑到对位置敏感的移动广告应用程序时,提出了被证明是 NP-hard 的有预算约束的用户选择问题;其次,对问题的子模特性进行了探究,提出了一个简单而有效的具有近似比 $(1 - \frac{1}{e})$ 的启发式算法;最后,大量的仿真结果表明,该方案使移动广告的传播效应有效提高了 120%.

关键词:移动广告;用户位置;区域覆盖;时空特性;用户选择;子模问题

中图分类号: TP311 **文献标识码:** A **doi:** 10.3969/j.issn.0253-2778.2017.07.010

引用格式: 徐婉茹, 杨盘隆. 移动广告用户传播能力评价与覆盖优化算法[J]. 中国科学技术大学学报, 2017, 47(7): 323-330.

XU Wanru, YANG Panlong. Mobile user propagation capability evaluation and coverage optimization algorithm[J]. Journal of University of Science and Technology of China, 2017, 47(7): 323-330.

Mobile user propagation capability evaluation and coverage optimization algorithm

XU Wanru, YANG Panlong

(School of Communications Engineering, PLA University of Science and Technology, Nanjing 210007, China)

Abstract: The distribution efficiency of mobile advertising is extremely important for both advertisers and users. Few studies have been conducted on efficient ad delivery, especially user tracing and the budget. In order to obtain a feasible and effective mobile advertising distribution policy, the concept of location-centric mobile crowdsourcing network was presented to replace the traditional user-centric networks and platforms, in which the location information for advertizing distribution plays a crucial role. Therefore, the user selection under the interested area coverage(interested area coverage, IAC) region was mainly focused upon. However, research centering on location information we need requires the consideration of the temporal characteristics of each user, and effective calculation of the ICA. The problem will be more difficult to solve when considering the budget constraint. To address these challenges, considering the

收稿日期: 2016-08-28; **修回日期:** 2016-12-08

基金项目: 国家自然科学基金重点项目(61232018), 江苏省自然科学基金杰出青年基金(BK20150030)资助.

作者简介: 徐婉茹, 女, 1992年生, 硕士生. 研究方向: 数据挖掘. E-mail: xwr88023@gmail.com

通讯作者: 杨盘隆, 博士/教授. E-mail: panlongyang@gmail.com

location sensitive mobile advertising applications, and a user selection solution was proposed, which was proved to be an NP-hard budget-constrained problem. Then, the submodularity problem was explored and a simple and effective heuristic was presented whose approximate ratio is $(1 - \frac{1}{e})$. A large number of simulation results indicate that our solution improved the propagation effect of mobile advertising effectively by 120%.

Key words: mobile advertisement; user location; area coverage; spatial-temporal feature; user selection; submodularity

0 引言

随着智能手机和各种移动应用的普及,对于逐渐增多的移动用户来说移动广告已经变得越来越流行和重要.与传统的网络广告不同^[1-2],移动广告更侧重于用户的移动性,即用户的时空特性和位置信息.这样,广告能在适当的时间和地点准确地定向到目标用户,可以节省大量的资金和时间.然而,随着移动用户数目的增加,大量的具有激励机制的用户却难以被考虑进来.

最近,移动用户正逐渐参与到众包模式中.众包模式指的是用户可以被整合到无需干预的协同工作中.在移动众包网络中,有两种典型的工作模式,分别是用户为中心和以平台为中心的工作模式.对于以用户为中心的模式,用户是机制设计和任务执行中的关键因素,因为用户需要作出是否接受任务和执行任务的决策.以平台为中心的模式中,任务是由平台通过高度可信的全局信息来进行分配.上述的工作模式并不适用于移动广告应用程序,因为在移动广告中用户位置是更为敏感的信息,尤其是在增加了优惠券,折扣等激励机制的情况.

现实中,人们更关心的是以位置为中心的移动广告.例如,一些热门地点可能是一个咖啡厅、餐厅,或附近的购物区.为此,我们提出移动广告主要关注的以位置为中心的模式,研究的问题就是,广告任务应该在所有热门区域高效地被执行.也就是说,要实现有兴趣区域的覆盖(interested area coverage, IAC).换句话说,被选择的一组用户要有效地覆盖传播移动广告的热门区域.

移动广告对于商店的宣传非常重要.对商家来说,在有限的预算条件下提高广告传播的准确性是提高收入的重要方法.移动广告的定向研究包括位置^[3-4]、终端、内容^[5]和人群^[6]等.由于移动设备如智能手机已经变成了人们随时随地必须携带的物品之

一,基于位置和终端的移动广告渐渐地吸引了更多人的注意.基于内容和人群的移动广告能提高用户对广告的兴趣,因此增加了广告传播的范围.

子模问题是一个广泛使用的理论,在信息收集^[7]和病毒式营销^[8]等领域已经被广泛应用.文献^[9]讨论了子模优化算法特别是最大化子模问题.文献^[10]介绍了有预算限制的最大化覆盖问题并提出了近似比为 $(1 - \frac{1}{e})$ 的算法来进行解决.文献^[11]

研究了有限的无线热点数量约束条件下的最优分配问题,与传统的最大化子模问题不同的是,所有的路由器被要求连结成一个连通网络,并且路由器的位置不同会造成额外的耗费.

为解决上述问题,我们将面临以下两个主要挑战:①被选择的移动用户的数量和覆盖效果之间的权衡.我们需要设计一个保证区域覆盖效果的智能的用户选择方法.②对覆盖区域而言,我们需要充分考虑用户的实际路线.我们要对用户的时空属性加以考虑和处理.

本文提出了计算有效的保证覆盖热门区域的接近最优的移动用户选择方案.首先,我们提出了热门区域覆盖的问题,并证明了它是 NP 难问题,即证明了热门区域覆盖问题的子模特性.这样,我们提出了一个可以实现与最优结果有较好近似比的启发式算法.本文的贡献可以总结为以下 3 点:

(I) 我们提出了在移动广告传播过程中发挥重要作用的热门区域覆盖问题,其中对带有时空特性的用户实际路线也进行了深入探究.

(II) 我们正式定义了热门区域覆盖问题,并且证明了它是 NP 难问题.为了解决这个问题,我们证明了它的子模特性,并提出了一个可以实现与最优结果有较好近似比的启发式算法.

(III) 我们做了大量的实验以验证所提出的方案,并且通过了用户实际路线的数据集进行验证.

1 基本原理

首先,我们引入子模问题的正式定义.

定义 1.1 给定一个有限集 E , 若函数 $f(\cdot)$ 在 E 的子集上满足 $f(X) + f(Y) \geq f(X \cup Y) + f(X \cap Y), \forall X, Y \subseteq E$, 则称该问题是子模问题.

为了更好地理解子模问题,我们接下来介绍子模问题的单调性.

定义 1.2 给定一个有限集 E , 若函数 $f(\cdot)$ 在 E 的子集上满足

$$f(X) \leq f(Y), \forall X \subseteq Y \subseteq E,$$

则称 $f(\cdot)$ 是单调的.

离散特性也能很好地解释子模问题.

定义 1.3 给定一个有限集 E 及在 E 的子集上的函数 $f(\cdot)$. 对于 $X \subseteq E, u \in E$, 令 $\Delta(u | X) = f(X \cup \{u\}) - f(X)$ 表示函数 $f(\cdot)$ 在集合 X 上关于 u 的增长量. 这样子模的正式定义即为, 如果函数 $f(\cdot)$ 满足收益递减规律, 它被称为子模问题. 对于 $X \subseteq Y \subseteq E, u \in E, u \notin Y$, 添加元素 u 到一个子集 Y 中的增长量不会大于加入同一元素到 Y 的子集 X 产生的增长量, 即为

$$\Delta(u | X) \geq \Delta(u | Y).$$

2 问题定义

2.1 网络模型

在我们的系统中, 假设有 m 个发布广告或者优惠券的商店 $L = \{l_1, l_2, \dots, l_m\}$, 其中 $l_j (j = 1, 2, \dots, m)$ 是一个商店的标示符. 这些商店分散在不同的位置 $P = \{p_{l_1}, p_{l_2}, \dots, p_{l_m}\}$. 另外, 有 t 种广告类型定义为 $A = \{a_1, a_2, \dots, a_t\}$. 商店 l_j 发布的广告类型为 A_{l_j} . 例如, 商店 l_1 发布三种广告 $A_{l_1} = \{a_1, a_2, a_3\}$.

此外, n 个移动用户的集合为 $V = \{v_1, v_2, \dots, v_n\}$, 通过选择移动用户来进行广告的分发. 每一个用户都有一个已知的移动轨迹, 其中包含了路线上涵盖的商店. 由于移动应用通常被授权可以记录用户所在的位置, 因此这些轨迹信息是可以使用的.

对于以位置为中心的移动应用广告来说, 优惠券、打折券这类的广告会分发给商店附近的用户. 我们假设每一个商店 l_j 都有一个半径为 r_l^j 的影响区域, 这个范围大小与部署在商店内的 AP 的通信距离有关, 因此在我们的系统中, 当移动用户进入到商店的影响区域后, 那么广告信息就能够发送给该用

户. 这对于商店和特定的广告来说, 即移动用户被覆盖了.

用户的位置信息被记录在用户的轨迹中, 我们用 $\vec{q}_{v_i} = (x_{v_i}, y_{v_i})$ 表示移动用户 v_i 的位置, 用 $\vec{p}_{l_j} = (x_{l_j}, y_{l_j})$ 表示商店 l_j 的位置. 如果移动用户处于商店的影响区域内, 即 $\|\vec{p}_{l_j} - \vec{q}_{v_i}\| \leq r_l^j$, 我们就说商店 l_j 和它的广告集 A_{l_j} 被用户 v_i 覆盖了. 我们用 C_{v_i} 表示被用户 v_i 覆盖的商店集合. 例如, 用户 v_1 覆盖了四个商店, 即 $C_{v_1} = \{l_1, l_2, l_5, l_6\}$, 并且该用户接收了这些商店推送的广告, 因此被用户集合 V 中的所有用户所覆盖的商店即为 $\bigcup_{v_i \in V} C_{v_i}$.

2.2 函数定义

定义 2.1 给定一个商店集合 $L = \{l_1, l_2, \dots, l_m\}$, 广告类型集合 $A = \{a_1, a_2, \dots, a_t\}$, 移动用户集合 $V = \{v_1, v_2, \dots, v_n\}$ 和它的子集 $S (S \subseteq V)$, 那么子集 S 所覆盖的移动广告数量 $f(S)$ 定义为

$$f(S) = \left| \bigcup_{l_j \in \bigcup_{v_i \in S} C_{v_i}} A_{l_j} \right|.$$

2.3 问题定义

本文的网络中, 帮助传播广告的移动用户会被奖励相应的优惠券或打折券. 换句话说, 每一个广告为促进自己的传播需要支出相应的耗费. 用 $c(a_k)$ 表示一个广告 a_k 的耗费, 相应的对用户 v_i 来说, 他帮助传播自己所覆盖的广告能够得到的奖励为

$$b(v_i) = \sum_{a_k \in \bigcup_{l_j \in C_{v_i}} A_{l_j}} c(a_k).$$

我们的目标是在有限预算 $B (B > 0)$ 的条件下, 使能够覆盖的不同种类的广告数最多. 问题定义如下:

定义 2.2 给定一个商店集合 $L = \{l_1, l_2, \dots, l_m\}$, 广告类型集合 $A = \{a_1, a_2, \dots, a_t\}$, 移动用户集合 $V = \{v_1, v_2, \dots, v_n\}$ 和一个预先定义的预算 B , 那么有预算的移动广告覆盖最大化问题即为选择这样一个用户集合 $S (S \subseteq V)$, 使得所覆盖的广告总数 $f(S)$ 最大化, 同时所有的耗费小于 B . 即为

$$\begin{aligned} & \max_{S \subseteq V} f(S) \\ & \text{s.t. } \sum_{v_i \in S} b(v_i) \leq B. \end{aligned}$$

相关研究^[9,11]已经得出预算限制的覆盖问题是 NP-hard 问题. 我们的问题与之不同, 需要同时考虑移动用户轨迹的时空特性进行优化, 这使问题变的更复杂. 如果一个 NP 问题是子模问题, 那么可以用具有 $(1 - \frac{1}{e})$ 近似比的算法来找到近优解.

2.4 子模性分析

我们先证明函数 $f(S)$ 是单调递增性子模问题,然后利用子模性质来获得问题的近优解。

引理 2.1 给定一个商店集合 L , 广告类型集合 A , 移动用户集合 $S(S \subseteq V)$, 那么移动广告覆盖数量 $f(S)$ 函数是单调递增子模问题。

首先,使 $f(\emptyset) = 0$. 我们从集合 V 中取任意一个子集 $S'(S \subseteq S' \subseteq V)$, 那么集合 S 和 S' 中的用户的路线轨迹覆盖的商店集合满足:

$$\bigcup_{v_i \in S} C_{v_i} \subseteq \bigcup_{v_i \in S'} C_{v_i}.$$

相应地,所覆盖的广告集合满足:

$$\bigcup_{l_j \in \bigcup_{v_i \in S} C_{v_i}} A_{l_j} \subseteq \bigcup_{l_j \in \bigcup_{v_i \in S'} C_{v_i}} A_{l_j},$$

那么我们得到

$$f(S) \leq f(S'), \forall S \subseteq S' \subseteq V,$$

因此 $f(S)$ 是单调递增的。

对于一个任意的用户 $u \in V \setminus S'$, 令

$$\Delta f(u | S) = f(S \cup \{u\}) - f(S),$$

$$\Delta f(u | S') = f(S' \cup \{u\}) - f(S').$$

如果将用户 u 分别加入到集合 S 和 S' 中,那么所覆盖的商店集合为

$$C = \bigcup_{v_i \in S \cup \{u\}} C_{v_i} = \bigcup_{v_i \in S} C_{v_i} \cup C_u,$$

$$C' = \bigcup_{v_i \in S' \cup \{u\}} C_{v_i} = \bigcup_{v_i \in S'} C_{v_i} \cup C_u;$$

因此得到

$$\bigcup_{v_i \in S} C_{v_i} \subseteq \bigcup_{v_i \in S'} C_{v_i}, C' \subseteq C.$$

那么将用户 u 分别加入到集合 S 和 S' , 所增长的覆盖广告数量满足:

$$\bigcup_{l_j \in C'} A_{l_j} \subseteq \bigcup_{l_j \in C} A_{l_j},$$

则可得到

$$\Delta f(u | S) \geq \Delta f(u | S').$$

即满足了增益递减规律^[9,12], 因此 $f(S)$ 是单调递增性子模问题。

3 算法描述

我们研究的问题是预算限制下的子模问题,即背包限制的最大化子模问题.本文提出了近优算法来解决以上问题(算法 3.1).

算法 3.1 预算限制下的最大化广告覆盖算法

输入:商店位置集合 $L = \{l_1, l_2, \dots, l_m\}$

移动用户集合 $V = \{v_1, v_2, \dots, v_n\}$

广告类型集合 $A = \{a_1, a_2, \dots, a_k\}$

最大预算 B

输出:被选择用户集合 S

for 用户 v_i 和商店 l_j do

if 商店 l_j 被用户 v_i 覆盖 then

for 商店 l_j 发布的广告 a_k do

if 用户 v_i 对广告 a_k 感兴趣 then

广告覆盖数量+1;

end

end

end

end

for 用户 v_i do

计算单位耗费覆盖的广告数 $h(i)$;

end

for 排序之后的 $h(i)$ do

if $b + h(i) \leq B$ then

将用户 v_i 加入到集合 S ;

$b = b + h(i)$;

end

end

事实上,我们已经证明了这个问题是单调递增性子模问题,这样就可以利用启发式贪心算法来实现更好的近似比,并获得最满意的用户选择方案。

当考虑到选择哪些用户进行广告传播时,我们首先需要知道哪些商店处于用户的轨迹路线上以及用户是否对这些广告感兴趣.更重要的是,如果用户被选择进行广告传播,那么该用户对他的酬劳也有一定的要求。

算法 3.1 中,我们首先输入商店的位置集合 $L = \{l_1, l_2, \dots, l_m\}$, 用户集合 $V = \{v_1, v_2, \dots, v_n\}$, 广告类型集合 $A = \{a_1, a_2, \dots, a_k\}$; 然后记录每一个用户的移动轨迹.如果一个移动用户 v_i 处于商店 l_j 的影响区域内并且对广告 A_{l_j} 感兴趣,我们就把用户 v_i 视为一个候选对象并且记录用户 v_i 的轨迹上所感兴趣的所有广告的数量.同时,根据用户覆盖的广告,用户会有一个相应的报酬 $b(i)$. 计算用户 v_i 的单位报酬所能覆盖的广告数量,根据商店所能提供的最大预算 B , 利用贪婪算法来进行用户的选择.当增加一个用户 u 到被选择用户集合 S 时,这些用户总的酬劳没有超过预算 B , 那么就增加用户 u 到被选择用户集合 S ; 如果总的酬劳超过了最大预算,算法停止,就得到了最终的被选择用户集合 S .

4 仿真实验

我们建立仿真模型并对本文提出的方法进行验证,包括从网络规模、商店数量和预算限制等不同方面进行实验验证.另外,对于覆盖效率的验证,我们采用所覆盖到的所有广告数量和支付给被选择用户的酬劳比进行分析。

4.1 实验设置

我们利用 MATLAB 建立了一个仿真环境来模仿商店区域和用户的移动轨迹.发布广告的商店随机分布在 $3000\text{ m} \times 3000\text{ m}$ 的区域中,用户可以自由到访该区域内的任意商店.我们给每个商店的影响区域半径设置为相同值 $r_s = 100\text{ m}$. 如果一个用户 i 经过一个商店的影响区域并且他对该商店发布的广告兴趣度超过了门限值 α_i , 那么该用户覆盖了这个商店和他感兴趣的广告. α_i 是用来描述不同用户对不同广告的不同兴趣程度,我们给每一个商店分配一个权重,这样用户就可以根据他所覆盖的商店来决定他应得的报酬.商店的权重值用来描述不同商店拥有的优惠券数量.

我们将本文所提算法与随机选择方法进行了比较,即随机选择用户构成目标集合 S .

4.2 预算限制的影响

根据直观经验,预算越大覆盖的广告数量会越多.我们设置用户数为 $n = 30$, 商店数为 $m = 50$, 广告类型数为 $A = 10$ 进行实验.实验结果如图 1-3 所示.为了便于理解,我们对预算进行了归一化处理.图 1 和图 3 的实验结果表明,随着预算越来越大,本文的算法表现出越好的性能,与随机选择方法结果的差距也越来越大.另外,由于随着预算增长,越来越多的用户被选择,本文的算法在覆盖效率方面仍表现出较好的性能.如图 2,随着预算的提高,广告覆盖效率有所降低,表明预算越多,选择的用户数也越多,在不超出预算的情况下我们可以雇佣更多的用户来获得高效的广告覆盖.

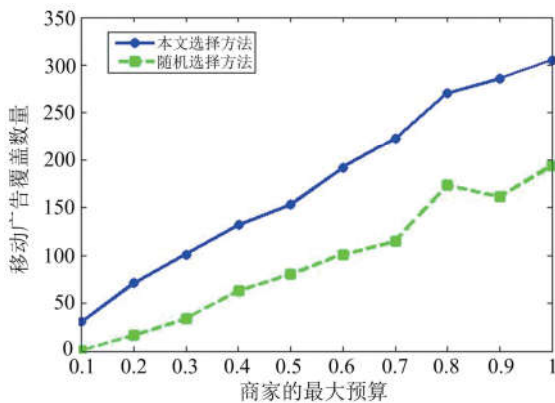


图 1 移动广告覆盖数量
Fig.1 The number of covered mobile advertisement

4.3 用户数量的影响

需要考虑的另外一个因素即移动用户的数量.本文在商店数为 $m = 50$, 广告类型数为 $A = 10$ 的情况下设置移动用户的数量分别为 $n = 30, 50, 100$ 进

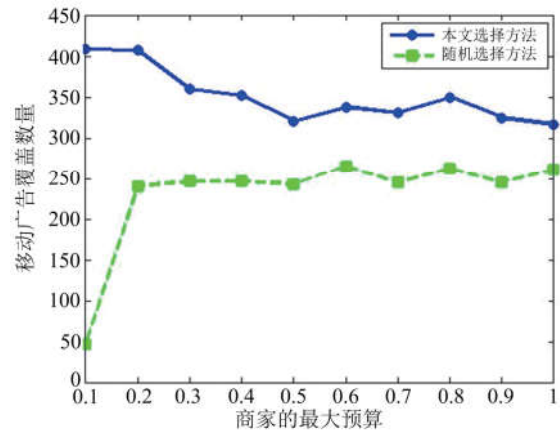


图 2 移动广告覆盖效率

Fig.2 The mobile advertisement coverage efficiency

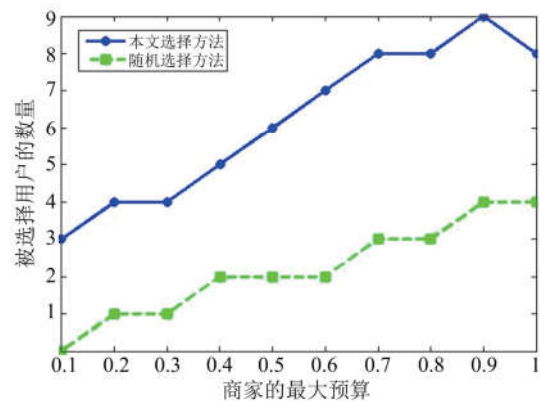


图 3 被选择用户的数量

Fig.3 The number of selected users

行对比实验.如图 4-6 所示,本文算法与随机选择方法相比仍有更好的性能.值得注意的是,移动广告覆盖效果几乎是随机选择方法的 2 倍.从图 4-6 还可以看出,与本文算法显著增长的表现性能相比,随机选择方法并没有太大的变化.这表明,当用户数目增多时,本文算法表现性能依旧很好.

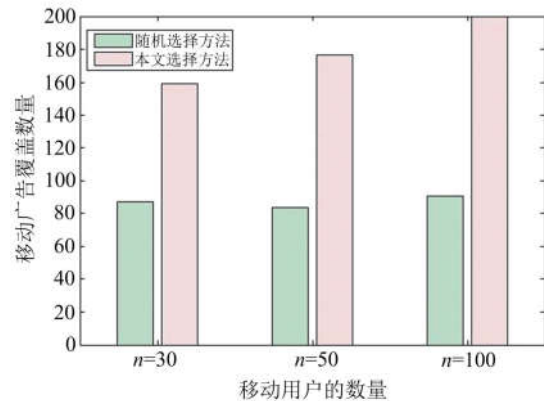


图 4 移动广告覆盖数量

Fig.4 The number of covered mobile advertisement

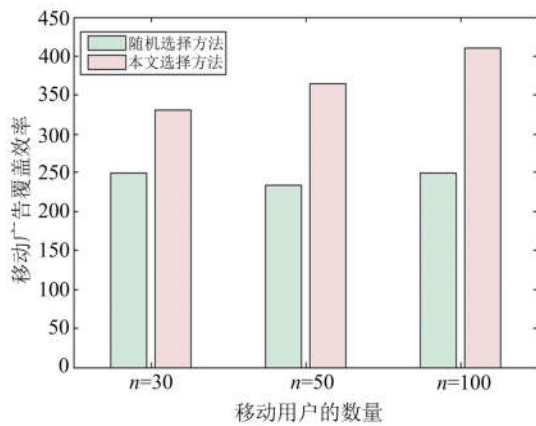


图 5 移动广告覆盖效率

Fig.5 The mobile advertisement coverage efficiency

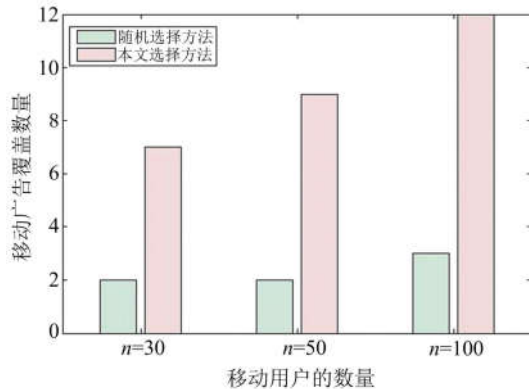


图 6 被选择用户的数量

Fig.6 The number of selected users

5 结论

本文提出了一种能够获得近优的有兴趣区域覆盖的移动用户选择的高效方案,并且还研究了如何在有限预算限制条件下最大化广告的覆盖问题.我们很好地利用了用户轨迹路线的时空特性并且验证了本文方法较好的性能.在以后的工作中,我们将利用用户的真实轨迹数据集进行全面的验证.

参考文献 (References)

- [1] BORGS C, CHAYES J, IMMORLICA N, et al. Dynamics of bid optimization in online advertisement auctions[C]// Proceedings of the 16th International Conference on World Wide Web. Banff, Canada: ACM Press, 2007: 531-540.
- [2] MAHDIAN M, NAZERZADEH H, SABERI A. Allocating online advertisement space with unreliable estimates [C]// Proceedings of the 8th ACM Conference on Electronic Commerce. San Diego: ACM Press, 2007: 288-294.
- [3] KÖLMEL B, ALEXAKIS S. Location based target advertising[C]// Proceedings of the First International Conference on Mobile Business, Athens, Greece: Springer, 2002: 1-7.
- [4] BANERJEE S, DHOLAKIA R R. Mobile advertising: Does location based advertising work? [J]. International Journal of Mobile Marketing, 2008, 3 (2): 68-74.
- [5] NATH S, LIN F X Z, RAVINDRANATH L, et al. SmartAds: Bringing contextual ads to mobile apps [C]// Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services. Taipei, China: ACM Press, 2013: 111-124.
- [6] NATH S. Madscope: Characterizing mobile in-app targeted ads [C]// Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services. Florence, Italy: ACM Press, 2015: 59-73.
- [7] KRAUSE A, GUESTIN C. Near-optimal observation selection using submodular functions[C]// Proceedings of the 22nd National Conference on Artificial Intelligence. Vancouver, Canada: AAAI Press, 2007, 2: 1650-1654.
- [8] KEMPE D, KLEINBERG J, TARDOS É. Maximizing the spread of influence through a social network[C]// Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Washington: ACM Press, 2003: 137-146.
- [9] KRAUSE A, GOLOVIN D. Submodular function maximization[M]// Tractability: Practical Approaches to Hard Problems. Cambridge: Cambridge University Press, 2014: 71-104.
- [10] KHULLER S, MOSS A, NAOR J S. The budgeted maximum coverage problem [J]. Information Processing Letters, 1999, 70(1): 39-45.
- [11] Kuo T W, Lin K C J, Tsai M J. Maximizing submodular set function with connectivity constraint: Theory and application to networks[J]. IEEE/ACM Transactions on Networking (TON), 2015, 23(2): 533-546.
- [12] Nemhauser G L, Wolsey L A, Fisher M L. An analysis of approximations for maximizing submodular set functions—I[J]. Mathematical Programming, 1978, 14(1): 265-294.