

多轮社交广告序列影响最大化

付冰洋¹, 张龙姣¹, 史麒麟^{1,2*}, 王泽宇¹, 王 灿¹, 宋明黎¹

(1. 浙江大学 计算机科学与技术学院, 浙江 杭州 310027; 2. 浙大城市学院 计算机与计算科学学院, 浙江 杭州 310015)

摘要: 现有的序列广告推荐研究主要关注用户对广告的偏好, 未充分考虑广告间的正向关系。从广告间的关联出发, 将广告网络 and 用户网络同时纳入考量, 构建了基于触发模型的多轮广告序列推荐影响力最大化模型。提出了基于广告边的多轮反向影响力采样贪心策略, 以提升广告平台收益, 并证明了这一方法具有严格的理论下界保证。实验表明, 与现有最优方法相比, 该方法的广告传播影响力收益平均提升了 35%, 显著增强了广告推荐效果, 为广告序列推荐提供了新的解决方案。

关键词: 社交网络; 触发模型; 影响力最大化; 广告推荐

中图分类号: TP301.6 **文献标志码:** A **文章编号:** 1001-2486(2025)03-010-11



Multi-round social advertising sequence influence maximization

FU Bingyang¹, ZHANG Longjiao¹, SHI Qihao^{1,2*}, WANG Zeyu¹, WANG Can¹, SONG Mingli¹

(1. College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China;

2. School of Computer and Computing Science, Hangzhou City University, Hangzhou 310015, China)

Abstract: Existing research on sequential ad recommendations mainly focuses on user preferences for advertisement, insufficiently considering positive relationships between ads. Starting from the associations between ads, incorporates both ad networks and user networks into consideration, a multi-round social advertising influence maximization model based on triggering model was constructed. An ad edge based greedy strategy based on multi-round reverse influence sampling was proposed to enhance platform revenue, with theoretical proofs of its strict lower bound guarantee. Experiments show that compared to existing optimal methods, the proposed method increases the average ad propagation influence revenue by 35%, significantly enhancing ad recommendation effectiveness, providing a new solution for ad sequence recommendations.

Keywords: social network; triggering model; influence maximization; advertising recommendation

随着移动设备的普及和信息网络的飞速发展, 人们上网越来越便利, 在社交网络中的沟通交流因打破了时间和空间的限制而备受欢迎。相比于传统广告, 社交广告可以通过用户的“点赞”和“转发”等行为快速获得大量曝光。因此在社交网络上投放广告成为社交平台的一个重要盈利方式, 广告序列推荐方向吸引了一系列研究。一些相关研究注重用户对广告的兴趣^[1-2], 假设用户之间彼此独立, 在浏览广告序列的过程中, 用户根据对广告的兴趣有概率停止浏览, 这些工作倾向于把更能吸引兴趣的广告排在前面, 让用户持续浏览以获取更多收益。Shi 等^[3]的另一部分研究则假定用户会浏览到平台推荐的所有广告, 且认

为广告之间有正向联系, 两个有关联的广告被用户接受会有额外收益。

然而, 这些工作忽略了社交网络中用户传播的影响。在社交广告的背景下, 用户并非孤立, 用户接受广告的行为也会进一步暴露给其他的用户, 从而影响其他用户的选择。因此也需要考虑社交网络中用户间的信息传播作用, 这一点在影响力最大化领域有着广泛的研究。社交网络影响力最大化(influence maximization, IM)问题^[4]针对市场营销过程中的“口碑效应”(word-of-mouth)和“病毒营销”(viral marketing)等传播原理进行研究, Kempe 等^[5]将其模型化为算法问题, 即在用户网络中选取一个最多 k 个节点的种子集合,

收稿日期: 2024-10-30

基金项目: 国家自然科学基金资助项目(62372399); 浙江大学上海高等研究院繁星科学基金资助项目(SN-ZJU-SIAS-001)

第一作者: 付冰洋(2001—), 女, 山东淄博人, 硕士研究生, E-mail: todobe@zju.edu.cn

*通信作者: 史麒麟(1991—), 男, 江苏无锡人, 副研究员, 博士, E-mail: shiqihao321@zju.edu.cn

引用格式: 付冰洋, 张龙姣, 史麒麟, 等. 多轮社交广告序列影响最大化[J]. 国防科技大学学报, 2025, 47(3): 10-20.

Citation: FU B Y, ZHANG L J, SHI Q H, et al. Multi-round social advertising sequence influence maximization[J]. Journal of National University of Defense Technology, 2025, 47(3): 10-20.

向它们投放广告,通过网络传播,最大化受影响的用户数量。Kempe 等引入了广泛使用的独立级联(independent cascade, IC)模型^[6]、线性阈值(linear threshold, LT)模型^[7]和触发(triggering model, TR)模型,证明了影响力最大化问题在这三种模型上都是 NP-hard 的,并给出了近似比为 $1 - 1/(e - \epsilon)$ 的贪心算法。后续的工作针对贪心算法中的采样开销进行优化。一部分研究^[8-9]利用模型的性质,通过减少模拟计算次数来提升模拟效率;另一部分研究采用启发式算法进行针对性优化^[10-11],但缺乏效率和精度保证。主流的方法基于 Borgs 等^[12]提出的反向影响力采样(reverse influence sampling, RIS)进行加速。在此基础上 Tang 等提出了 TIM/TIM+^[13]和 IMM^[14]算法,将时间复杂度优化到 $O((k(m+n)\log n)/\epsilon^2)$ 。为了进一步减少采样次数,优化处理效率,最新工作采用动态调试策略确定反向传播次数^[15]。

以上工作都只考虑了单轮次的影响传播过程, Sun 等^[16]首次考虑实际生活中的多轮传播过程,建模了多轮影响力传播模型, Lozano-Osorio 等^[17]在此基础上提出智能邻域搜索策略进一步提高了多轮 IM 问题的可扩展性。

影响力最大化问题虽然利用社交网络上的用户传播提升商品推广的影响力,但大多都默认用户一定会接受推荐给他们的广告,也不会考虑广告推荐顺序的差异。为解决这一挑战, Tang^[18]首次将社交广告序列推荐和影响力最大化问题结合起来,假定用户逐个按顺序到来,为每一个到来的用户都推荐一组序列广告,用户会根据自己的兴趣可能接受部分广告,也有可能停止浏览。他们将广告先按用户在网络上传播的期望收益与用户放弃继续浏览概率的比值进行排序,然后使用动态规划算法选取最终的广告序列,综合考虑了用户兴趣和网络传播。这一工作也是考虑影响力最大化的广告序列推荐问题目前比较好的解决方案。

目前的研究工作大多忽略了广告之间的关联,没有考虑到用户对广告的接受程度很可能也受其他广告的影响。产品之间的关联可能会激发用户对广告内容的兴趣。例如,一个用户看到了羽毛球拍的广告并且购买了它,这个用户接受羽毛球广告的概率会大大提升。可以利用广告的这种关联增益效应,提升用户兴趣,并利用社交网络传播连锁反应,促进广告在社交网络上产生更高的效益,避免在同等成本下推送无关广告。

在广告序列推荐问题上首次将广告网络 and 用户网络同时纳入模型,在已有的影响力最大化问

题和广告推荐问题的基础上,综合用户对广告的偏好、广告之间的促进作用以及用户网络的信息传递,并考虑到广告推荐不是一次性过程,加入多轮推荐的模式,建立能够同时为一组用户服务的多轮广告序列影响力最大化问题模型。提出基于边的贪心算法,取得了更好的推荐效果,使得平台的收益最大化。真实数据集上的对比实验验证了算法的有效性,提出的算法较现有最优算法平均提升了 35% 的广告收益。

1 问题描述

多轮广告序列影响最大化问题的最终目标为,对于一组具有相似特性的特定用户,基于广告之间的关联,向他们推荐相同的多轮序列广告,经过社交网络传播,达到广告收益的最大化。

1.1 符号表示

用 U_s 表示需要推送广告这组特定用户的集合。定义广告网络 $N = (V, W)$, 其中 V 代表可推荐的广告集合; $W \subseteq V \times V$ 为广告边,即广告和广告之间的有向边,自环代表广告本身的接受概率,其他边代表两个广告之间具有促进关系。例如,在同一个广告序列中先后向用户 u 推荐了两个不同的广告 v_1 和 v_2 , 并且 v_1 和 v_2 之间有一条有向边,那么如果用户接受了 v_1 ,则会有更大的概率接受广告 v_2 。但不同用户对于广告的接受程度,以及接受了一个广告对接受另一个广告的促进程度可能不尽相同,使用函数 $w: U_s \times W \rightarrow \mathbb{R}$ 来衡量这一程度。对于 $U_s \times W$ 中的每个元素 $(u, (v_1, v_2))$ 都有函数值 $w(u, (v_1, v_2))$:

1) 当 $v_1 = v_2$ 时, $w(u, (v_1, v_2))$ 表示无其他广告影响时用户 u 接受广告 v_2 的概率;

2) 当 $v_1 \neq v_2$ 时, $w(u, (v_1, v_2))$ 表示用户 u 接受了广告 v_1 再看到广告 v_2 时接受概率提高的倍数。

例如,用户 u 接受广告 v_1 和广告 v_2 的概率分别为 $w(u, (v_1, v_1))$ 和 $w(u, (v_2, v_2))$, 如果用户已经接受了 v_1 , 并且 v_1 和 v_2 之间存在一条有向边 (v_1, v_2) , 那么用户 u 接受广告 v_2 的概率就会增加 $w(u, (v_1, v_2)) \times w(u, (v_2, v_2))$ 。当 v_2 有多个入边邻居被接受时,真实的增益难以计算,但多个广告增益的结果最差也不会差过单个广告的增益,因此采用一种较为保守的假设,即 v_2 只会被增益幅度最大的边增益一次。这要求函数 w 至少满足以下两个性质: ① $\forall u, v, 0 \leq w(u, (v, v)) \leq 1$; ② $\forall u, v_1, v_2, w(u, (v_1, v_2)) \leq$

$$1/w(u, (v_2, v_2)) - 1.$$

1.2 传播模型

在影响力传播的研究中, IC 模型和 LT 模型是较为常用的两种模型, 都是触发模型的一种特例^[5]。为保证本方法的全面性, 将使用更为通用的触发模型作为传播模型的基础。定义社交网络 $G = (U, E)$, 其中 U 代表用户集合, $E \subseteq U \times U$ 为用户与用户之间的有向边集, 每条边关联一个概率 b_{u_1, u_2} 。信息的传播会在离散时刻上进行。初始时刻, 种子集合 S 中的节点会被激活, 每个节点 u 会根据其入边概率选择一个触发集合 $T(u)$ 。如果在某时刻, 未被激活的节点的触发集中有节点已被激活, 则该节点将会被激活。传播会持续到没有新的节点被激活。种子集合 S 的影响力为最终被激活点的期望数量。

触发模型可以用活跃边图 (live-edge graph) 上的方法等价表示。给定一组触发集合 $\{T(u) \mid u \in U\}$, 可以构造一个活跃边图 $L = (U, E(L))$, 其中 $E(L) = \{(u_1, u_2) \mid u_2 \in U, u_1 \in T(u_2)\}$, 每条边 (u_1, u_2) 都被称为一条活跃边 (live-edge)。令 $\Gamma(G, S)$ 代表图 G 中可以被 S 到达的点的集合。由于触发模型和活跃边图的等价性, 触发模型中种子集合 S 的影响力等价于 $\mathbb{E}[\|\Gamma(L, S)\|]$, 该期望值取决于生成的活跃边图 L 的分布。

1.3 问题定义

定义 1 多轮广告序列影响最大化问题是指, 给定用户网络 $G = (U, E)$ 、生成触发集合的概率分布、广告网络 $N = (V, W)$ 、用户组 U_s 、函数 $w: U_s \times W \rightarrow \mathbb{R}$ 、轮数 T 、序列长度限制 k 、广告被接受的收益函数 $r: V \rightarrow \mathbb{R}$, 推荐 T 轮广告序列 $\sigma_1^*, \sigma_2^*, \dots, \sigma_T^*$, 每轮至多 k 个广告, 使得平台推荐广告的总收益最大化, 即:

$$\sigma^* = \bigcup_{t=1}^T \{\sigma_t^*\} = \arg \max_{\sigma: |\sigma_t| \leq k, \forall t \in [T]} \rho(\sigma) \quad (1)$$

其中, $\rho(\sigma)$ 代表平台推荐广告带来的总收益。

定义如下符号来方便描述: $\sigma_{t,i}$ 表示一组解 σ 中第 t 轮广告序列中的第 i 个广告。使用符号 $E(\cdot)$ 来表示广告序列中生成的边集, 即

$$E(\sigma_t) = \{(\sigma_{t,i}, \sigma_{t,j}, t) \mid i \leq j, (\sigma_{t,i}, \sigma_{t,j}) \in W\} \quad (2)$$

$$E(\sigma) = \bigcup_{t=1}^T E(\sigma_t) \quad (3)$$

其中, $E(\cdot)$ 表示序列的生成边集。则 $\rho(\sigma)$ 可以等价地转化为边的函数:

$$\rho(\sigma) = h(E(\sigma)) = \sum_{v \in V} f(E(\sigma), v) \cdot r(v) \quad (4)$$

其中: $r(v)$ 为每有一个用户接受了广告 v , 平台可以获得的广告收益; $f(E(\sigma), v)$ 代表经过 T 轮推荐和传播后, 用户网络中接受广告 v 的人数期望值, 即

$$f(E(\sigma), v) = \mathbb{E}[\|\bigcup_{t=1}^T \Gamma(L_t, S_{v,t})\|] \quad (5)$$

其中: L_t 为在用户网络上随机生成的活跃边图; $S_{v,t}$ 为 U_s 的子集, 代表第 t 轮时用户组 U_s 中接受了广告 v 的用户集合。 $f(E(\sigma), v)$ 的值取决于 $L_1, L_2, \dots, L_T$ 和 $S_{v,1}, S_{v,2}, \dots, S_{v,T}$ 的分布。

为了得到 $S_{v,t}$ 的分布, 需要计算 U_s 中用户对于被推送的广告序列 σ_t 中每个广告的接受概率, 用 $p_{u,v,t}$ 代表用户 u 在第 t 轮接受广告 v 的概率, 则:

$$p_{u,v,t} = \mathbb{I}\{(v, v, t) \in E(\sigma_t)\} \cdot w(u, (v, v)) + \sum_{(v', v, t) \in E(\sigma_t)} w(u, (v, v)) \cdot w(u, (v', v)) \cdot P_{\max}(u, v', v, t) \quad (6)$$

其中: $\mathbb{I}$ 的含义为如果满足括号内条件值为 1, 否则值为 0; $P_{\max}(u, v', v, t)$ 代表 (v', v) 对 v 的增益是所有 v 的入边中最大那条的概率; $p_{u,v,t}$ 为用户 u 自身愿意接受 v 的概率与其他广告对 v 的增益之和。如果多个广告同时对 v 有增益, 则选择增益最大的广告。 $P_{\max}$ 可以表示为

$$P_{\max}(u, v', v, t) = p_{u,v',t} \cdot \prod_{\substack{(v_0, v, t) \in E(\sigma_t) \\ v_0 \neq v, v' \\ w(u, (v_0, v)) > w(u, (v', v))}} (1 - p_{u,v_0,t}) \quad (7)$$

即入边 (v', v) 对 v 产生最大增益的概率。 (v', v) 能对 v 造成最大增益的前提是 v' 被用户 u 接受, 并且与 v 有连接关系的其他增益大于 (v', v) 的广告没有被 u 接受。通过上述过程可以获取 $S_{v,t}$ 的分布, 进而求得 $h(E(\sigma))$, 即在选定广告序列 σ 后产生的期望收益。

多轮广告序列影响最大化问题的目标为找到最佳的 σ 使得推荐广告总收益 $\rho(\sigma)$ 最大化。由于 $\rho(\sigma)$ 可以等价地转化为边集合函数 $h(E(\sigma))$, 且为了边贪心策略的需要, 接下来将使用 h 作为需要最大化的目标函数。

2 算法设计

针对多轮广告序列影响最大化问题, 本节提出了基于广告边的贪心策略。算法框架如图 1 所示。算法会进行 T 轮推荐, 每轮使用边贪心策

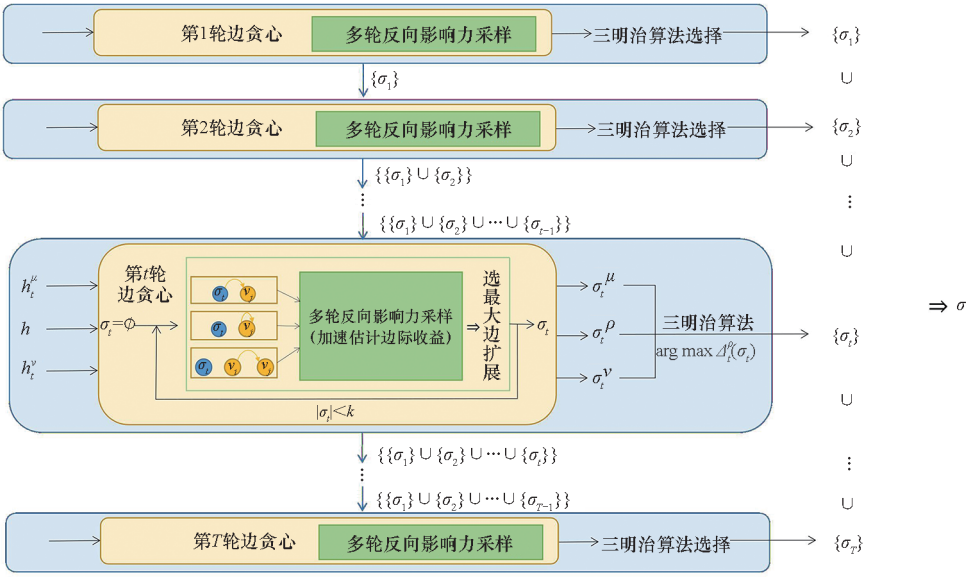


图1 算法框架

Fig. 1 Algorithm framework diagram

略结合前面的推荐结果将广告序列扩充至长度 k , 使用三明治算法选出每轮最优的序列, 结合 T 轮推荐序列形成最终结果。相比已有算法, 算法的创新之处在于: ①使用边贪心策略利用广告间正向关系; ②使用多轮反向影响力采样方法以快速估计边际收益; ③提出三明治算法解决目标函数不具备子模性的挑战。

2.1 基于边的贪心策略

对于多轮广告序列影响最大化问题, 算法 1 给出了基于边的贪心策略。在进行每一轮的广告选择时, 相较于其他直接对广告节点进行选择的算法^[1-2, 18], 算法 1 所示的边贪心策略更加注重广告边 (即广告间的促进关系) 的作用, 在广告之间具有强增益作用的广告网络中, 边贪心的算法可以有效挖掘出广告边的价值。同时, 算法将用户对广告本身的接受程度设置为自环边的边权, 与其他边一起竞争, 也不会忽略掉节点本身的价值, 即用户对广告的偏好。

算法 1 将从空集开始, 逐轮选择每一轮的广告序列。用 σ_t 表示第 t 轮推荐的广告序列, 初始使 σ_t 为空。用 ε 来表示尾部不在 σ_t 中的边的集合, 用 $h((v_i, v_j, t) | E) = h(E \cup \{(v_i, v_j, t)\}) - h(E)$ 表示广告边 (v_i, v_j, t) 在经过用户网络传播后的边际收益。每次贪心地选择边际收益最高的一条边 $(v_i, v_j) \in \varepsilon$, 如果边的头部 v_i 不在 σ_t 里, 或者 $v_i = v_j$, 即该边是自环, 则只把 v_j 加入 σ_t 的末尾, 否则就把 v_i 和 v_j 按顺序依次加入 σ_t 的末尾, 第 11 行和第 13 行中的 $\oplus$ 代表将该符号右边的节

点连接到左侧序列的末尾。将每轮选择的 σ_t 合并到 σ 中得到最终的结果。该结果具有合理的下界保证。

算法 1 边贪心算法

Alg. 1 Edge greedy algorithm

输入: 用户网络 $G = (U, E)$, 广告网络 $N = (V, W)$, 时间轮 T , 序列长度限制 k

输出: σ

1. $\sigma \leftarrow \emptyset$
2. **for** $t = 1$ **to** T
3. $\sigma_t \leftarrow ()$
4. **while** $|\sigma_t| \leq k - 2$
5. $\varepsilon = \{(v_i, v_j) \in W | v_j \notin \sigma_t\}$
6. **if** $\varepsilon = \emptyset$, 跳出循环
7. **end if**
8. $\forall (v_i, v_j) \in \varepsilon$ 计算 $h((v_i, v_j, t) | E(\sigma \cup \{\sigma_t\}))$
9. $(v_i, v_j) = \operatorname{argmax}_h((v_i, v_j, t) | E(\sigma \cup \{\sigma_t\}))$
10. **if** $v_i = v_j$ 或 $v_i \in \sigma_t$
11. $\sigma_t = \sigma_t \oplus v_j$
12. **else**
13. $\sigma_t = \sigma_t \oplus v_i \oplus v_j$
14. **end if**
15. **end while**
16. $\sigma = \sigma \cup \{\sigma_t\}$
17. **end for**
18. 返回 σ

定理 1 用 σ^* 表示多轮社交广告序列影响

力最大化的最优解。如果函数 h 具有单调性和子模性,边贪心算法满足:

$$\frac{\rho(\sigma)}{\rho(\sigma^*)} \geq 1 - e^{-\frac{1-e^{-\left(1-\frac{1}{k}\right)}}{2d_{in}+1}} \quad (8)$$

其中, d_{in} 为广告网络中节点的最大入度。

证明:算法可以从轮次的层面,看成对每个时间轮 t ,都贪心地选择一个 σ_t 来最大化 $\rho(\{\sigma_1, \sigma_2, \dots, \sigma_{t-1}, \sigma_t\}) - \rho(\{\sigma_1, \sigma_2, \dots, \sigma_{t-1}\})$ 。每一轮内部再选择边际收益最大的边加入进来最终形成 σ_t 。如果 h 是单调子模的,那么每一轮内的目标函数 $h_t(E') = h(E' | E(\{\sigma_1, \sigma_2, \dots, \sigma_{t-1}\}))$ 也是单调子模的。对于目标函数为单调子模函数的边贪心算法,可证明其近似比^[19]

$$\text{至少为 } \frac{1 - e^{-\left(1-\frac{1}{k}\right)}}{2d_{in}+1}。$$

对于贪心的每一步,如果无法找到最优的答案,但是能够找到一个近似比至少为 α 的答案,那么对于一个子模的问题来说,这样的贪心算法也可以达到 $1 - e^{-\alpha}$ 的近似比^[18, 20],应用该结论即可得到定理 1。□

2.2 多轮反向影响力采样

为了快速估算广告边在社交网络 G 中造成的影响力,即算法 1 的步骤 8,即计算 (v_i, v_j) 的边际收益,设计了多轮反向影响力采样方法。

对于社交网络 G 中的每条边 (u_1, u_2) ,将以 b_{u_1, u_2} 的概率被保留,以 $1 - b_{u_1, u_2}$ 的概率被删去,这样生成得到的子图 g 被称为 G 的一个采样。

选择 G 中的一个节点 u 作为起始节点, g 中所有能够到达 u 的点集被称为一个反向可达集,在多轮的情况下,每一轮都独立地随机生成一个 u 的反向可达集。定义多轮反向可达集 $\mathcal{R}_u = \bigcup_{t=1}^T \mathcal{R}_{u,t}$,其中 $\mathcal{R}_{u,t}$ 代表第 t 轮中 u 的反向可达集中的节点与时间 t 的二元组的集合。

定义一个随机多轮反向可达集 $\mathcal{R}$ 为随机选择一个 u 作为起始节点的多轮反向可达集。对于一个广告网络 N 以及一组广告序列 σ ,用 $S_v = \{(u, t) | u \in S_{v,t}\}$ 来表示每个时间轮 t 中会接受并转发广告 v 的那些特定用户,其中 $S_{v,t}$ 为上文提到的第 t 轮时用户组 U_S 中接受了广告 v 的用户集合。 S_v 的影响力 $f(E(\sigma), v)$ 计算方法如下:

$$f(E(\sigma), v) = n_u \cdot \mathbb{E}[\mathbb{I}[S_v \cap \mathcal{R} \neq \emptyset]] \quad (9)$$

由式(9)可见, $f(E(\sigma), v)$ 的值取决于 S_v 和 $\mathcal{R}$ 的分布。需要注意的是,对于 f 来说, $f(E(\sigma), v)$ 中的 $E(\sigma)$ 并不需要一定是 σ 的生成边集,可以替换为任意满足拓扑性质的边集,例如

在算法进行过程中,需要计算 $E(\sigma)$ 的某些子集的期望收益,后文中使用 E' 来表示这一参数,即:

$$h(E') = \sum_{v \in V} f(E', v) \cdot r(v) \quad (10)$$

用多轮反向影响力采样方法估计 h 值的过程如算法 2 所示。 S_v 的分布可以通过 $p_{u,v,t}$ 得出。如步骤 12 所示,对于生成的每一个多轮反向可达集 $\mathcal{R}$,可以直接计算出 $S_v \cap \mathcal{R} = \emptyset$ 的概率,进而得出 $S_v \cap \mathcal{R} \neq \emptyset$ 的期望。

算法 2 多轮反向影响力采样方法

Alg. 2 Multi-round reverse influence sampling method

输入:用户网络 $G = (U, E)$,已选的边集 E' ,特殊用户组 U_S ,时间轮 T ,参数 ε, ℓ ,收益函数 r

输出: $\bar{h}$

1. 对所有 u, v, t , 计算 $p_{u,v,t}$
2. $R = 0, LB = 0$
3. **for** 所有出现在 E' 的广告 v
4. $LB = LB + \sum_{u \in U_S} \left[1 - \prod_{t=1}^T (1 - p_{u,v,t}) \right] \cdot r(v)$
5. $R = \max\{R, r(v) \cdot Tk\}$
6. **end for**
7. $\theta = \frac{\left(2 + \frac{2}{3}\varepsilon\right)n_u(\ell \ln n + \ln(2T))R}{\varepsilon^2 h(E')}$
8. 生成 θ 个随机多轮反向可达集
9. **for** 所有出现在 E' 的广告 v
10. **for** $i = 1$ **to** θ
11. $\mathcal{R} \leftarrow$ 第 i 个被生成的多轮反向可达集
12. $prob = \prod_{(u,t) \in \mathcal{R}, u \in U_S} (1 - p_{u,v,t})$
13. $sum = sum + (1 - prob)$
14. **end for**
15. $f_v = sum \cdot n/\theta$
16. $\bar{h} = \bar{h} + f_v \cdot r(v)$
17. **end for**
18. 返回 $\bar{h}$

利用鞅论^[14, 21]理论对反向可达集估计信息传播的置信区间和置信度进行分析。多轮反向影响力采样方法满足:

引理 1 对于任意的 $\varepsilon > 0, \ell > 0$, 如果 $\theta \geq \frac{\left(2 + \frac{2}{3}\varepsilon\right)n_u(\ell \ln n + \ln(2T))R}{\varepsilon^2 h(E')}$, 则

$$\Pr[|\bar{h}(E') - h(E')| \geq \varepsilon h(E')] \leq \frac{1}{n^\ell T} \quad (11)$$

其中, $R = Tk \cdot \max_{v \in V} \{r(v)\}$, 代表每个采样值的上界。 $h(E')$ 是所求, 不能精准获取其值, 但是可

以找到一个尽可能大的下界,使得在保证 $1 - 1/(n^\ell T)$ 的置信度的同时,采样的次数尽可能少,提升时间效率。对于 U_s 中的用户,至少能计算没有传播的情况下接受广告 v 的人数期望,即

$$h(E') = \sum_{v \in V} f(E', v) \cdot r(v) \\ \geq \sum_{v \in V} \sum_{u \in U_s} [1 - \prod_{t=1}^T (1 - p_{u,v,t})] \cdot r(v) \triangleq LB \quad (12)$$

在算法中,使用 LB 作为 $h(E')$ 的一个下界计算 θ ,并进行 θ 次反向采样,结合对 S_v 分布的计算,最终有 $1 - 1/(n^\ell T)$ 的概率,得出函数 h 估计值的偏差与真实值相比不超过 ε 。

由引理 1 可知,使用多轮反向影响力采样方法计算边际收益的边贪心算法对于每一轮 t ,都有至少 $1 - 1/(n^\ell T)$ 的概率找到一个近似比为 $\frac{1 - e^{-(1-\frac{1}{k})}}{2d_{in} + 1} - \varepsilon$ 的近似结果。那么将所有轮合并起来,有至少 $1 - 1/n^\ell$ 的概率,算法在每一轮得到的答案都能满足该近似比。

2.3 三明治算法

满足定理 1 的一个重要前提条件是函数 h 必须同时满足单调性和子模性^[20]。函数具有单调性是指如果一个集合函数 $f: V \rightarrow \mathbb{R}$, 对于任意的集合 $A \subseteq V$ 和元素 $v \in V \setminus A$ (其中,“ $\setminus$ ”表示 V 和 A 的差集), 都有 $f(A) \leq f(A \cup \{v\})$ 。函数具有子模性是指如果一个集合函数 $f: V \rightarrow \mathbb{R}$, 对于任意两个集合 $A \subseteq B \subseteq V$ 和元素 $v \in V \setminus B$, 都有 $f(A \cup \{v\}) - f(A) \geq f(B \cup \{v\}) - f(B)$ 。

引理 2 前文定义的关于边集的函数 h 具有单调性。

直观上,加入一条边 (v_1, v_2, t_0) 后,只会改变第 t_0 轮中广告 v_2 被 U_s 中用户接受的概率,而且边的作用只会使该广告被接受的概率增加,至少不会减少,并且不会对其他广告的接受程度产生影响。 U_s 中用户更愿意接受 v_2 并分享出去,更有可能被更多的人看到,至少不会使平台的收益降低。因此函数 h 显然是单调的。但是函数 $h(E')$ 并不满足子模性,这很容易通过举反例得到。

h 虽然满足单调性,却不满足子模性。图 2 展示了 h 不满足子模性的一个具体实例。节点 A, B, C 代表广告,自环边的数字代表用户愿意接受该广告的概率,非自环边代表用户接受边的头部节点对尾部节点的增益倍率。先把问题简化为只有一个轮次 t , 并且用户网络中只有一个用户节点 u , 所有广告单次点击收益均为 1。在此例中,为了方便,用图中边上的小写字母代表边两侧

的节点和时间的三元组,如 a 表示 (A, A, t) 。

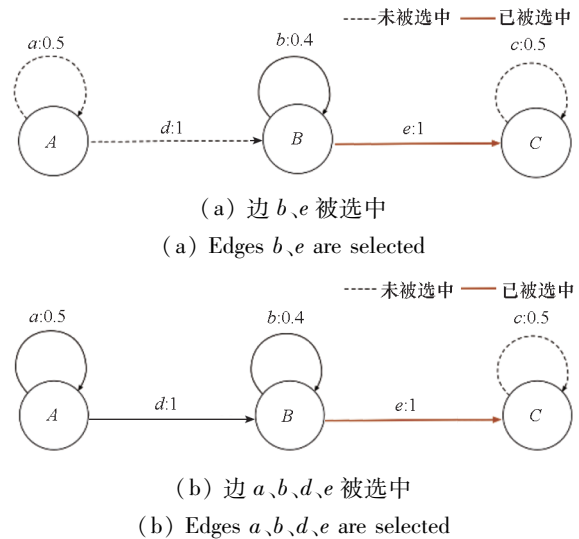


图 2 函数 h 不满足子模性举例

Fig. 2 Example of the non-submodularity function h

图 2(a) 计算边 e 的边际收益 $h(e | \{b\}) = h(\{b, e\}) - h(\{b\}) = 0.2$ 。图 2(b) 中的图结构与图 2(a) 中完全一致,但是多选择了两条边 a, d , 同样地,计算边 e 的边际收益为 $h(\{a, b, d, e\}) - h(\{a, b, d\}) = 1.4 - 1.1 = 0.3$ 。 $h(e | \{b\}) \leq h(e | \{a, b, d\})$, 显然不满足子模性。

对于满足单调性不满足子模性的目标函数,为了保证算法具有合理的近似比,需要引入文献[22]中提到的三明治算法来解决,令:

$$\Delta_t^l(\sigma_t) = \Delta h_t^l(E(\sigma_t)) = h(E(\bigcup_{i=1}^{t-1} \sigma_i \cup \sigma_t)) - h(E(\bigcup_{i=1}^{t-1} \sigma_i)) \quad (13)$$

$$\Delta_t^\mu(\sigma_t) = \Delta h_t^\mu(E(\sigma_t)) = h_t^\mu(E(\bigcup_{i=1}^{t-1} \sigma_i \cup \sigma_t)) - h_t^\mu(E(\bigcup_{i=1}^{t-1} \sigma_i)) \quad (14)$$

$$\Delta_t^v(\sigma_t) = \Delta h_t^v(E(\sigma_t)) = h_t^v(E(\bigcup_{i=1}^{t-1} \sigma_i \cup \sigma_t)) - h_t^v(E(\bigcup_{i=1}^{t-1} \sigma_i)) \quad (15)$$

在每一轮 t , 找到两个与 h 定义在同样定义域上的子模函数 h_t^μ 和 h_t^v , 满足对于定义域上的任意 $E'_t \subseteq \{(v_1, v_2, t_0) | (v_1, v_2, t_0) \in E', t_0 = t\}$, 都有 $\Delta h_t^\mu(E') \leq \Delta h_t(E') \leq \Delta h_t^v(E')$ 。即 Δh_t^μ 处处是 Δh_t 的下界, Δh_t^v 处处是 Δh_t 的上界, 并且如果将原问题中的 h 替换为 h_t^μ 和 h_t^v , 每一轮得到的结果近似比至少不差于原问题。

定义函数 h_t^μ 和 h_t^v , 除计算第 t 轮的 $p_{u,v,t}$ 以外, 其余部分与函数 h 一致, 只需重新定义用户在第 t 轮对所推荐广告的接受概率计算部分, 即:

$$p_{u,v,t}^\mu = \mathbb{I}\{(v, v, t) \in E(\sigma_t)\} \cdot w(u, (v, v)) \quad (16)$$

$$p_{u,v,t}^v = \mathbb{I}\{(v,v,t) \in E(\sigma_t)\} \cdot w(u,(v,v)) + \max_{(v',v,t) \in E(\sigma_t)} \{w(u,(v',v))\} \cdot w(u,(v,v)) \quad (17)$$

直观地理解函数 h_t^μ 和 h_t^v 限制了当前轮在边的增加导致的某些边对其他点增益增大且其余条件都相同的情况下, 函数 h_t^μ 在计算某个广告被接受的概率时, 不考虑边的增益, 只计算每个节点本身的价值, 相当于去除了边的增强作用, 最终得到的结果一定小于原函数值。函数 h_t^v 则是默认前面的广告被接受的概率均为 1, 增大了前序广告对该广告增益的概率, 最终得到的结果一定大于原函数值, 因此有 $\Delta h_t^\mu(E_t') \leq \Delta h_t(E_t') \leq \Delta h_t^v(E_t')$ 。

引理 3 函数 h_t^μ 和 h_t^v 在其定义域上都是单调并子模的。

如算法 3 所示, 每一轮分别使用 h_t^μ 、 h 和 h_t^v 替换原来的 h 来运行三次贪心算法, 得到三个结果 σ_t^μ 、 σ_t^ρ 和 σ_t^v , 然后在其中选择一个原目标函数值最大的作为最终结果 σ_t^{sand} 。在每一轮中使用三明治算法并将结果加入答案, 得到整体的 σ 。

算法 3 三明治算法

Alg. 3 Sandwich algorithm

输入: 用户网络 $G = (U, E)$, 广告网络 $N = (V, W)$, 时间轮 T , 序列长度 k , 函数 $h, h_t^\mu, h_t^v: 2^{|E|} \times T \rightarrow \mathbb{R}$
输出: σ

1. $\sigma \leftarrow \emptyset$;
2. **for** $t = 1$ **to** T
3. 使用 h_t^μ, h, h_t^v 分别代替 h 执行算法 2 中的步骤 3 ~ 18, 得到 $\sigma_t^\mu, \sigma_t^\rho, \sigma_t^v$
4. $\sigma_t^{\text{sand}} = \arg \max_{\sigma_t \in \{\sigma_t^\mu, \sigma_t^\rho, \sigma_t^v\}} \Delta_t^\rho(\sigma_t)$
5. $\sigma = \sigma \cup \{\sigma_t^{\text{sand}}\}$
6. **end for**
7. 返回 σ

定理 2 三明治算法保证对任意 $1 \leq t \leq T$, 使用多轮反向影响力采样的边贪心算法有:

$$\Delta_\rho(\sigma_t^{\text{sand}}) \geq \max \left\{ \frac{\Delta_\rho(\sigma_t^\mu)}{\Delta_v(\sigma_t^\mu)}, \frac{\Delta_\mu(\sigma_t^*)}{\Delta_\rho(\sigma_t^*)} \right\} \cdot \left(\frac{1 - e^{-(1-\frac{1}{k})}}{2d_{\text{in}} + 1} - \varepsilon \right) \cdot \Delta_\rho(\sigma_t^*) \quad (18)$$

其中, σ_t^* 代表 Δ_ρ 在第 t 轮的最优解, 证明见文献[22]。

2.4 准确性与复杂度分析

定理 3 对任意的 $\varepsilon > 0$ 和 $\ell > 0$, 使用多轮反向影响力采样和三明治算法的边贪心算法得到的

σ 至少有 $1 - \frac{1}{n^\ell}$ 的概率满足:

$$\frac{\rho(\sigma)}{\rho(\sigma^*)} \geq 1 - e^{-\min_{t \in [T]} \left\{ \max \left[\frac{\Delta_\rho(\sigma_t^\mu)}{\Delta_v(\sigma_t^\mu)}, \frac{\Delta_\mu(\sigma_t^*)}{\Delta_\rho(\sigma_t^*)} \right] \right\} \cdot \left(\frac{1 - e^{-(1-\frac{1}{k})}}{2d_{\text{in}} + 1} - \varepsilon \right)} \quad (19)$$

证明: 结合定理 1、引理 2、引理 1 和引理 3 可得, 有至少 $1 - 1/n^\ell$ 概率每一轮都能得到近似比为 $\max \left\{ \frac{\Delta_\rho(\sigma_t^\mu)}{\Delta_v(\sigma_t^\mu)}, \frac{\Delta_\mu(\sigma_t^*)}{\Delta_\rho(\sigma_t^*)} \right\} \cdot \left(\frac{1 - e^{-(1-\frac{1}{k})}}{2d_{\text{in}} + 1} - \varepsilon \right)$ 的近似结果。易知, 对将每个 σ_t 作为元素的集合函数 ρ , 该问题仍然是单调子模的, 套用定理 1 证明中提到的 $1 - e^{-\alpha}$ 的结论, 即可得定理 3。□

纵观算法过程, 使用了多轮反向影响力采样方法和三明治算法的边贪心时间复杂度为 $O\left(\frac{(T^2 k^2 m_v n_s + m_u) T n_u (\ell \ln n_u + \ln(2T))}{\varepsilon^2} \cdot \frac{R}{LB_{\min}}\right)$, 其中 $m_v = |W|$ 代表广告网络的边数, $m_u = |E|$ 代表用户网络的边数, $n_u = |G|$ 为用户网络节点数, $n_s = |U_s|$ 为特殊用户集合大小, R 与 LB 的含义与引理 2 中一致, $LB_{\min}$ 代表可能的 LB 的最小值。算法需进行 T 轮, 每轮 k 次选边, 选边需遍历大小为 m_v 的广告边集, 在估算一条边的边际收益时, 需要遍历 θ 个多轮反向可达集, 统计采样结果需要计算已经被选择的广告, 不会超过 Tk 个, 只要统计有哪些特殊用户在该广告出现的那一轮接受了这个广告即可, 因此统计其结果的时间为 Tkn_s , 总复杂度为 $O(T^2 k^2 m_v n_s \theta)$ 。对于生成采样的部分, 沿用上一次生成过的采样结果不会影响其估算的准确性, 故只需按照最大的 θ 生成一次即可。生成一个采样的时间是 Tm_u , 这部分时间复杂度为 $O(\theta Tm_u)$ 。综上, 根据引理 2, 代入 θ , 总时间复杂度即为: $O(Tkm_v \cdot Tk(n_u + m_u)\theta) = O\left(\frac{T^2 k^2 m_v (n_u + m_u) n_u (\ell \ln n_u + \ln(2T))}{\varepsilon^2} \cdot \frac{R}{LB_{\min}}\right)$ 。

3 实验结果

3.1 实验数据与设置

实验在 Amazon 评论数据集^[23]上进行。该数据集主要记录了用户对已经购买的商品的评论信息。对于网络结构不可见的图, 文献[24]提供了一种生成近似网络的方法, 商品网络 and 用户网络的生成遵循该方法。如果买过商品 i 后又买了商品 j 的人数超过特定阈值, 就在商品网络中加入一条由 i 到 j 的有向边, 并对每个商品加入自环; 如果两个用户购买相同商品的次数超过特定值,

就在用户网络中为他们增加有向边。在 TR 模型中,为每条边在 $[0,1]$ 中生成一个随机权重 b_{u_1,u_2} 。对于每一个商品 v ,在 $\{0.01,0.02,0.03,0.04,0.05\}$ 集合中取一个随机值作为 $r(v)$ 。用户 u 对商品 v 的接受程度用在 $[0,1]$ 范围内随机生成的 $w(u,(v,v))$ 表示。对非自环边, v_1 对 v_2 的促进程度用在 $[0,1/w(u,(v_2,v_2)) - 1]$ 范围内随机生成的 $w(u,(v_1,v_2))$ 表示。对于特殊用户集合 U_s ,现实中用户的影响力可能存在巨大差距,为了体现用户区分度以及算法对用户网络传播的考虑,分别从用户网络中度数较大和度数较小的节点中随机选择 5 个,取其并集作为 U_s 。

Amazon 评论数据集对于不同的商品类别分开统计。按照用户网络和广告网络的稠密程度以及节点与边的规模挑选不同类别,生成了 4 个不同大小的数据集,数据集的具体参数见表 1。

表 1 数据集参数
Tab. 1 Parameters of the data set

数据集名称	用户数量	用户边数量	商品数量	商品边数量
Gift Cards	1 477	50 000	306	13 122
Software	1 573	20 000	2 032	107 429
Magazine Subscriptions	854	36 566	1 125	17 381
Video Games	3 230	300 000	1 613	202 573

实验在操作系统 Ubuntu 18.04.6 LTS 下运行,CPU 型号为 Intel (R) Xeon (R) Gold 6254 CPU @ 3.10 GHz,运行内存为 314 G。

实验任务是在给定的用户网络和商品网络中为用户推荐多轮广告序列。基于此任务设计的实验,算法的目标是为用户组推荐 T 轮长度至多为 k 的广告序列,最大化目标函数 $\rho(\sigma)$,即广告平台的期望收益。最后会对结果序列进行 10 000 次推荐过程和影响力传播过程的蒙特卡罗模拟,取平均值作为评价结果的标准。通过与已有算法对比验证提出方法的有效性。本次实验中设置 $T=5,k=10$ 。为兼顾实验效果和效率,除探究 θ 对结果的影响实验外,设置 $\theta=1\,000$,即在估计边际收益时进行 1 000 次多轮反向影响力采样。

3.2 对比实验结果

选择两种已有方法和一种平凡算法作对比。

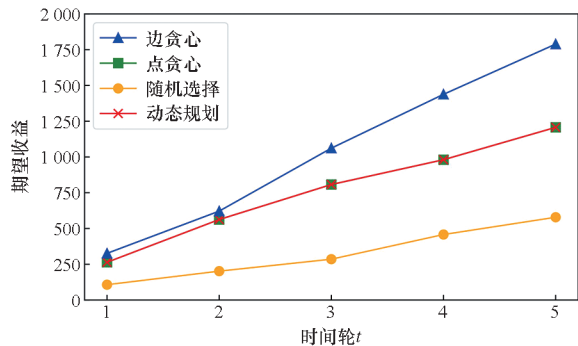
边贪心:使用多轮反向影响力采样估计边际期望收益,并使用三明治算法取三个结果中最好的序列作为最终结果。

点贪心:不考虑广告之间的促进作用。使用多轮反向影响力采样估计点的边际期望收益。将影响力最大化领域中经典的贪心算法^[5]应用于广告网络中,作为广告节点的选择方法。

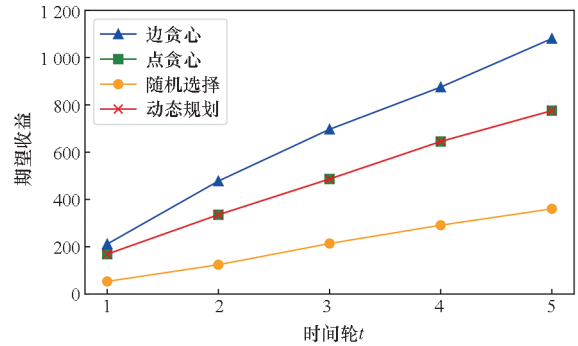
动态规划:目前该领域最优秀的算法^[18]之一,该算法将影响力最大化和广告序列结合起来,将所有广告节点按照期望收益排序,采用动态规划选择广告序列。由于该算法没有考虑多轮情况,实验时将使用多轮反向影响力最大化采样方法为该算法估计广告收益。

随机选择:平凡算法,随机选择广告推荐。

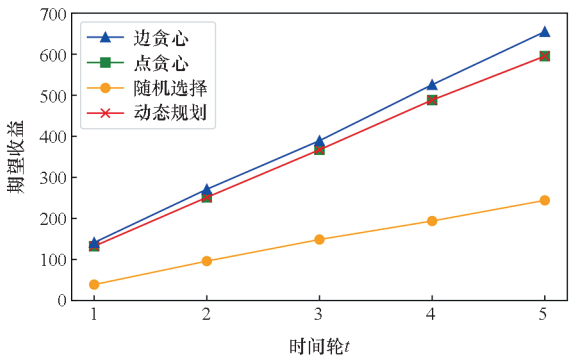
实验证明了边贪心算法能够更有效利用广告之间的促进作用,比已有的最优动态规划方法在 Amazon 评论数据集上平均提升 35%。如图 3 所示,点贪心算法和动态规划算法基本重合,因为在不考虑广告边收益的情况下,广告节点的收益基本不受其他节点影响,因此无论是动态规划算法还是点贪心算法对于该问题都会按照节点收益从大到小选择节点序列,从而导致结果基本一致。两者都没有考虑广告边对结果的影响,因此在广告网络较为稠密的三个数据集上边贪心算法明显更优,期望收益高出至少 40%。而在广告网络较为稀疏的 Magazine Subscriptions 数据集上,两者与边贪心算法的效果比较接近,边贪心算法只领先 10%,但也远远优于平凡算法。广告网络稀疏



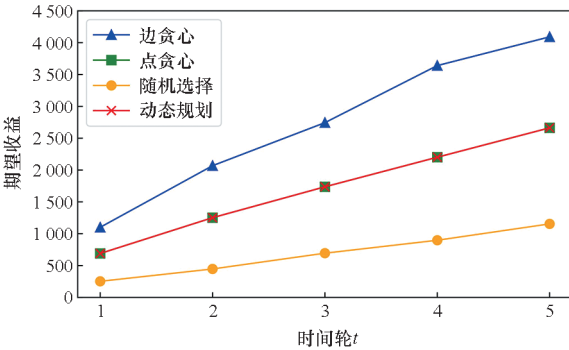
(a) Gift Cards 数据上 t -期望收益
(a) t -expected revenue on Gift Cards



(b) Software 数据上 t -期望收益
(b) t -expected revenue on Software



(c) Magazine Subscriptions 数据上 t -期望收益
(c) t -expected revenue on Magazine Subscriptions



(d) Video Game 数据上 t -期望收益
(d) t -expected revenue on Video Game

图 3 对比实验结果

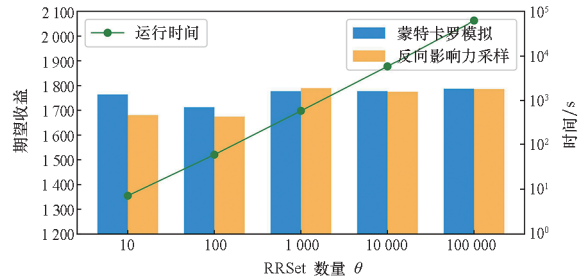
Fig. 3 Comparative experimental results

时,广告之间的联系较弱,广告节点的权重相对提升。这也说明边贪心算法在稠密广告网络上更加适用。但边贪心算法会综合考虑广告之间的联系和广告本身的权重,并且三明治算法的下界函数也相当于进行了点贪心,所以即使在广告网络稀疏的情况下至少不会差于点贪心算法。

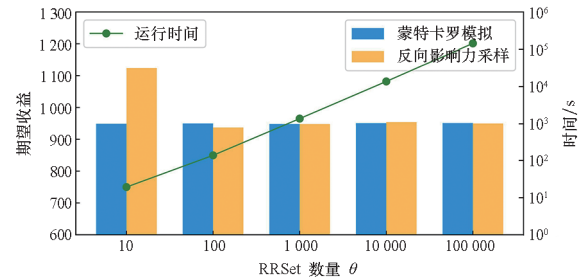
3.3 参数实验

算法中多轮反向影响力采样部分有两个参数 ε 和 ℓ ,可以通过调整这两个参数来控制算法的准确度和运行效率。 ε 越小, ℓ 越大,算法在估计每条边的边际收益时就越准确,同时运行时间也会越长。观察算法的流程,可以发现调整 ε 和 ℓ 的最终效果都只是改变了 θ 的大小,因此分别设置 $\theta = 10$ 、 $\theta = 100$ 、 $\theta = 1\,000$ 、 $\theta = 10\,000$ 、 $\theta = 100\,000$ 运行边贪心算法,记录分析其多轮反向采样得到的目标函数估计值,对结果进行 10 000 次蒙特卡罗模拟得到期望收益以及算法总运行时间。如图 4 所示,随着 θ 的增大,最终得到的蒙特卡罗模拟期望收益变化并不明显,说明即使是较小的 θ ,也能得出比较好的效果。但 θ 过小时,多轮反向采样的结果与蒙特卡罗模

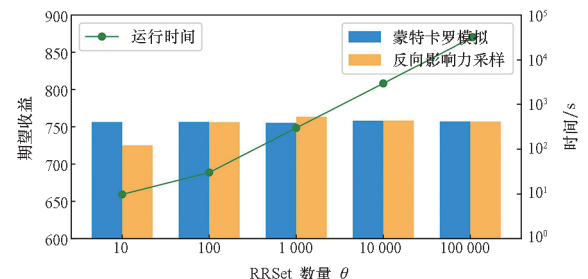
拟结果可能相差较大, θ 达到 1 000 以上后,两者基本持平,并且推荐的广告序列基本没有发生变化,结果数据波动较小。而算法的运行时间会随着 θ 的增大线性增长。因此在实验中,为了兼顾算法的效果、稳定性和运行效率,使用了 $\theta = 1\,000$ 作为反向采样次数。在实际应用中,也可以根据需求牺牲部分置信度进一步缩小 θ 的大小来满足效率要求。



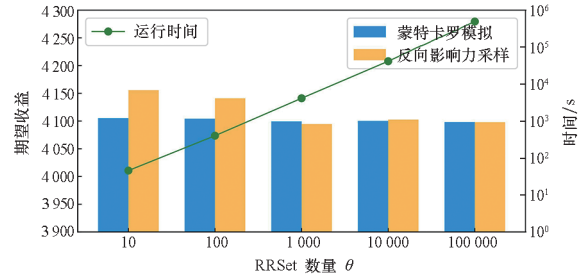
(a) Gift Cards 数据上 θ -期望收益
(a) θ -expected revenue on Gift Cards



(b) Software 数据上 θ -期望收益
(b) θ -expected revenue on Software



(c) Magazine Subscriptions 数据上 θ -期望收益
(c) θ -expected revenue on Magazine Subscriptions



(d) Video Games 数据上 θ -期望收益
(d) θ -expected revenue on Video Games

图 4 参数实验结果

Fig. 4 Parameter experiment results

3.4 消融实验

考虑用户网络传播的影响、广告边的关联、多轮传播等因素,除边贪心和点贪心之外,额外增加三组对比方法进行消融实验:

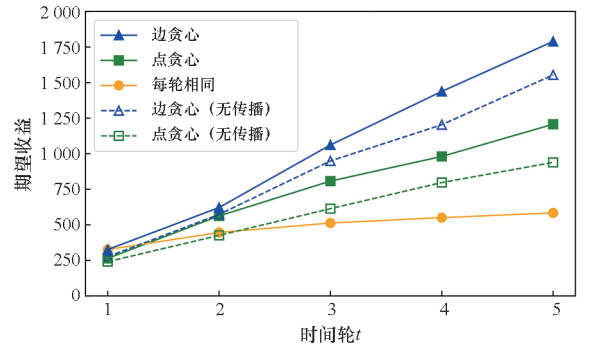
边贪心(无传播):去除边贪心算法中多轮反向影响力采样部分,不考虑用户网络的信息传播。

点贪心(无传播):去除点贪心算法中多轮反向影响力采样部分,不考虑用户网络的信息传播。

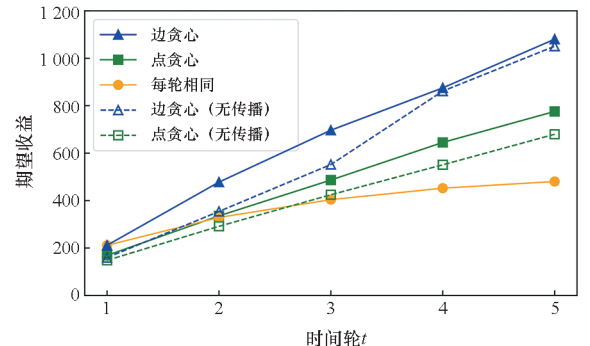
每轮相同:第一轮使用边贪心算法选取广告序列,后续每一轮推荐与第一轮相同的广告序列。

如图5所示,每轮推荐相同广告在开始的两个时间轮还比较有竞争力,这侧面说明了边贪心算法选出的序列比较优秀。但由于不能在后面的轮次推荐更高价值的新序列,期望收益只有微小增长。点贪心忽略了广告边的促进作用,期望收益也不如边贪心算法。

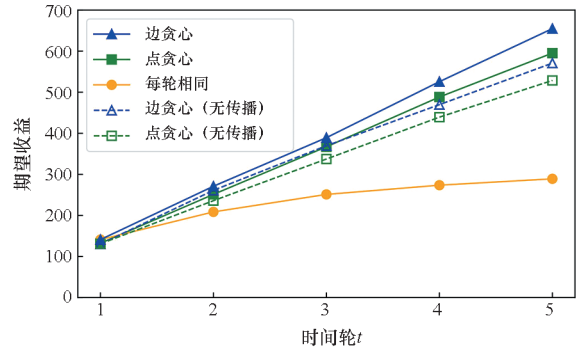
图5中两条虚线展示了不考虑用户网络影响力传播的情况。类似地,在用户网络较为稠密的三个数据集上,无传播的算法都表现较差。而在用户网络较为稀疏的 Software 数据集上,无传播的边贪心算法与有传播的边贪心算法效果较为接近。无传播的算法不考虑传播结果,就无法区分影响力大的用户和影响力小的用户,可能部分用户虽然对广告序列的接受程度很高,但无法造成



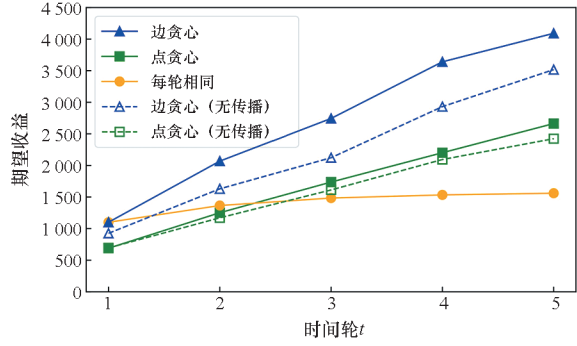
(a) Gift Cards 数据上 t -期望收益
(a) t -expected revenue on Gift Cards



(b) Software 数据上 t -期望收益
(b) t -expected revenue on Software



(c) Magazine Subscriptions 数据上 t -期望收益
(c) t -expected revenue on Magazine Subscriptions



(d) Video Game 数据上 t -期望收益
(d) t -expected revenue on Video Game

图5 消融实验结果

Fig. 5 Ablation experiment results

大量影响导致最终结果较差。在用户网络稀疏的图上,用户影响力差距较小,会降低影响力传播对结果的贡献,但无法完全消除。有传播的算法还是优于无传播的算法。用户网络越复杂,用户之间的影响力区别越大,有传播的算法越有优势。

4 结论

将广告序列推荐问题和影响力最大化问题结合,提出了多轮广告序列影响力最大化问题。综合考虑了用户对广告的偏好、广告之间的关联对用户接受广告程度的促进作用和用户网络上的信息传递,建立了新的多轮广告序列推荐影响力最大化问题模型以及解决该问题的贪心算法,使用多轮反向影响力采样技术提升算法效率,用三明治算法解决了目标函数不具有子模性的问题,并证明了算法的近似性和时间复杂度。基于现实数据集的实验表明,基于多轮反向影响力采样的贪心算法能够帮助平台推荐收益更大的多轮广告序列,较现有最优方法平均提升 35%,为多轮广告序列推荐影响力最大化问题提供合理的近似解。

参考文献(References)

[1] KEMPE D, MAHDIAN M. A cascade model for externalities

- in sponsored search [C]//Proceedings of the International Workshop on Internet and Network Economics, 2008: 585 – 596.
- [2] TANG S. Robust advertisement allocation [C]//Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, 2017: 4419 – 4425.
- [3] SHI Q H, FU B Y, WANG C, et al. Robust sequence networked submodular maximization [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2023: 15100 – 15108.
- [4] RICHARDSON M, DOMINGOS P. Mining knowledge-sharing sites for viral marketing [C]//Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2002: 61 – 70.
- [5] KEMPE D, KLEINBERG J, TARDOS É. Maximizing the spread of influence through a social network [C]//Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD'03, 2003: 137 – 146.
- [6] GOLDENBERG J, LIBAI B, MULLER E. Talk of the network: a complex systems look at the underlying process of word-of-mouth [J]. *Marketing Letters*, 2001, 12: 211 – 223.
- [7] GRANOVETTER M. Threshold models of collective behavior [J]. *American Journal of Sociology*, 1978, 83(6): 1420 – 1443.
- [8] KUMAR S, GUPTA A, KHATRI I. CSR: a community based spreaders ranking algorithm for influence maximization in social networks [J]. *World Wide Web*, 2022, 25(6): 2303 – 2322.
- [9] 王璿, 张瑜, 周军锋, 等. 基于社交网络的影响力最大化算法 [J]. *通信学报*, 2022, 43(8): 151 – 163.
WANG X, ZHANG Y, ZHOU J F, et al. Influence maximization algorithm based on social network [J]. *Journal on Communications*, 2022, 43(8): 151 – 163. (in Chinese)
- [10] KUMAR S, LOHIA D, PRATAP D, et al. MDER: modified degree with exclusion ratio algorithm for influence maximisation in social networks [J]. *Computing*, 2022, 104(2): 359 – 382.
- [11] LOZANO-OSORIO I, SÁNCHEZ-ORO J, DUARTE A. An efficient and effective GRASP algorithm for the budget influence maximization problem [J]. *Journal of Ambient Intelligence and Humanized Computing*, 2024, 15: 2023 – 2034.
- [12] BORGS C, BRAUTBAR M, CHAYES J, et al. Maximizing social influence in nearly optimal time [C]//Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, 2014: 946 – 957.
- [13] TANG Y Z, XIAO X K, SHI Y C. Influence maximization: near-optimal time complexity meets practical efficiency [C]//Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, 2014: 75 – 86.
- [14] TANG Y Z, SHI Y C, XIAO X K. Influence maximization in near-linear time: a martingale approach [C]//Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, 2015: 1539 – 1554.
- [15] 邓心惠, 宾晟, 孙更新. 基于反向可达集的影响力最大化算法 [J]. *计算机工程*, 2022, 48(1): 60 – 68, 74.
DENG X H, BIN S, SUN G X. Influence maximization algorithm based on reverse reachable set [J]. *Computer Engineering*, 2022, 48(1): 60 – 68, 74. (in Chinese)
- [16] SUN L C, HUANG W R, YU P S, et al. Multi-round influence maximization [C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018: 2249 – 2258.
- [17] LOZANO-OSORIO I, SÁNCHEZ-ORO J, DUARTE A. Multi-round influence maximization: a variable neighborhood search approach [C]//Proceedings of the International Conference on Variable Neighborhood Search, 2022: 112 – 124.
- [18] TANG S J. When social advertising meets viral marketing: sequencing social advertisements for influence maximization [C]//Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence, 2018: 176 – 183.
- [19] MITROVIC M, FELDMAN M, KRAUSE A, et al. Submodularity on hypergraphs: from sets to sequences [C]//Proceedings of the International Conference on Artificial Intelligence and Statistics, 2018: 1177 – 1184.
- [20] NEMHAUSER G L, WOLSEY L A, FISHER M L. An analysis of approximations for maximizing submodular set functions: I [J]. *Mathematical Programming*, 1978, 14: 265 – 294.
- [21] CHUNG F, LU L Y. Concentration inequalities and martingale inequalities: a survey [J]. *Internet Mathematics*, 2006, 3(1): 79 – 127.
- [22] LU W, CHEN W, LAKSHMANAN L V S. From competition to complementarity: comparative influence diffusion and maximization [J]. *Proceedings of the VLDB Endowment*, 2015, 9(2): 60 – 71.
- [23] HOU Y P, LI J C, HE Z K, et al. Bridging language and items for retrieval and recommendation [EB/OL]. (2024 – 03 – 06) [2024 – 10 – 15]. <https://arxiv.org/abs/2403.03952>.
- [24] GOMEZ-RODRIGUEZ M, LESKOVEC J, KRAUSE A. Inferring networks of diffusion and influence [J]. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2012, 5(4): 1 – 37.