户外广告中的标签特定遗憾最小化问题

Dildar Ali*、Abishek Salaria†、Ansh Jasrotia†、Suman Banerjee*
* 印度理工学院贾姆穆分校，印度， †国立理工学院斯利那加分校，印度{2021
rcs2009,suman.banerjee}@iitjammu.ac.in， {2023nitsgr185,2023nitsgr161}@nitsri.ac.in

arXiv:2603.06405v1 [cs.DB] 6 Mar 2026

摘要——**近年来，户外广告凭借较高的投资回报率，已成为最有效的营销手段之一，日益受到欢迎。电商企业会联系影响力提供方，通过标签（广告内容）、影响力需求和预算来实现有效广告投放。影响力提供方的目标是妥善分配标签，并在预算约束内满足所需的影响力需求。如果影响力提供方提供了所需影响力，广告主就会全额付款；否则，他们会根据其所提供影响力的比例部分付款。然而，如果影响力提供方提供的影响力超过广告主的需求，就无法获得额外报酬。因此，无论哪种情况，影响力提供方都会蒙受损失，这种损失被形式化为遗憾。影响力提供方的目标是使总遗憾最小化。本文将这一问题形式化为一个组合优化问题，并称之为户外广告中的标签特定遗憾最小化（TRMOA）。我们证明，TRMOA 是 NP难的，且无法在常数因子范围内近似。我们考虑的遗憾模型是非单调且非子模的，因此简单的贪心方法效果不佳。我们提出一种公平感知贪心轮询方法，以降低遗憾，并在广告主之间实现均衡分配。为提高可扩展性和解的质量，我们进一步提出随机贪心和随机局部搜索算法。我们使用真实世界的轨迹和广告牌数据集，对所有提出的方法及基准方法进行了实验，以展示这些解决方案的有效性和效率。**

索引词—**遗憾最小化、广告牌广告位、标签、广告主、影响力提供方、户外广告。**

I. 引言

INFLUENCE 广告领域已广泛研究影响力最大化问题。人们采用多种广告方式来最大化影响力，例如社交媒体、广告牌、电视和报纸。其中，广告牌广告是最有效的户外广告（OOH）技术之一，投资回报率更高。在现实生活中，广告主若需要影响力，就会向影响力提供方购买。影响力提供方向客户提供影响力并获得利润。谷歌和 Facebook 等大多数互联网巨头都利用自身平台作为影响力提供方投放产品广告，并获得巨额利润。大多数关于广告的文献都从客户的角度，研究如何通过社交网络和广告牌广告最大化影响力。针对影响力提供方视角的研究则为数不多。不同于现有文献，我们从影响力提供方的角度研究这个问题，旨在最大化其利润。在下文中

本文中，我们交替使用以下术语：影响力提供方和主机、广告和标签。基于主机的影响力最大化比基于客户的优化更难，正如我们稍后将要说明的那样。本文选择标签特定影响力模型，并在户外广告中描述我们的问题；这一问题已在现有文献中得到广泛研究 [1]–[3]。本文的普遍适用性将在后文中说明。

a) 我们的观察： 在现有关于户外广告场景的文献中，研究目标主要分为两类：(a) 帮助单个或多个广告主最大化其总影响力[1]–[7]；(b) 在多广告主场景中，最小化总遗憾并最大化影响力提供方的利润 [8]–[13]。现有文献均采用基于用户的影响力模型。然而，对于现实世界中的影响力提供方而言，更具挑战性的是日常管理多个广告主。广告主拥有预算、需求以及一组标签。影响力提供方的任务是在给定预算内选择标签子集，并向广告主分配标签特定影响力；否则，广告主将不会支付全额付款。

b) 我们的问题： 受这一观察启发，我们从影响力提供方的角度研究广告位分配问题。影响力提供方拥有许多广告牌广告位，并将其分配给广告主。每个广告主都希望获得一个广告位子集，使其总影响力满足自身需求。只有当所有广告主的需求都得到完全满足时，影响力提供方才能获得最高付款。在此场景下，我们采用遗憾模型 [3] 来指导影响力提供方的分配决策。我们确定了两种会直接影响利润的遗憾来源：未满足遗憾和过度遗憾。当影响力提供方未能满足某个广告主的需求时，就会产生未满足遗憾；而当向广告主分配的影响力超过其需求时，就会产生过度遗憾。这两种遗憾相互独立，也可能同时发生，如例1所示。

示例1. 假设某影响力提供方拥有五个广告牌广告位 $\mathcal{BS} = \{s_1, s_2, s_3, s_4, s_5\}$，每个广告位的影响力如表*I*所示。三位广告主 $\mathcal{A} = \{a_1, a_2, a_3\}$联系该影响力提供方，各自提出所需的影响力 $\mathcal{I}_i$ 和付款金额 $\mathcal{L}_i$，如表*II*所示。为服务这些广告主，影响力提供方为每位广告主分配一组互不重叠的广告牌 $\mathcal{S}_i$。

首次部署的策略-*I*如表 *III*所示。影响力供给与需求之间的差值为

表I：广告牌广告位信息。

$\mathcal{BS}$	s_1	s_2	s_3	s_4	s_5
$\mathcal{I}(s_i)$	4	5	3	6	2

表II：广告商信息。

$\mathcal{A}$	a_1	a_2	a_3
$\mathcal{I}_i$	6	7	8
$\mathcal{L}_i$	\$9	\$12	\$18

计算为 $\mathcal{X} = \mathcal{I}(\mathcal{S}_i) - \mathcal{I}_i$。广告主 a_1 供给过剩，导致过度遗憾；而广告主 a_2 和 a_3则供给不足，造成未满足遗憾。另一种部署方式——策略-*II*如表*IV*所示。尽管存在一些细微偏差，但该策略的总遗憾较低，因此更可取。

表III：策略 I

$\mathcal{A}$	a_1	a_2	a_3
$\mathcal{S}_i$	s_2, s_5	s_4	s_1, s_3
满足	Y	N	N
$\mathcal{X}$	1	-1	-1

表 IV：策略 II

$\mathcal{A}$	a_1	a_2	a_3
$\mathcal{S}_i$	s_1, s_5	s_2, s_3	s_4
满足	Y	Y	N
$\mathcal{X}$	0	1	-2

请注意，影响力提供方可能只遭受未满足遗憾（a_3 见表IV），也可能只遭受过度遗憾（a_2 见表 IV），或者既没有过度遗憾也没有未满足遗憾（a_1 见表IV）。为简化此处的说明，我们通过简单汇总 $\mathcal{I}(\mathcal{S})$ 来计算影响力 $\mathcal{I}(s_i)$；然而，在实际情况下，影响力 $\mathcal{I}(\mathcal{S})$是根据定义1中描述的影响力模型计算的。我们将该问题表述为户外广告中的标签特定遗憾最小化（TRMOA）问题。

c) 普遍适用性： 尽管本研究的动机源于户外广告，但我们的框架适用于广泛的影响力驱动的分配问题，包括在线广告、推荐系统和资源供给平台；在这些场景中，分配不足和分配过度都会造成损失。

d) 我们的贡献： 总而言之，本研究作出了以下几项关键贡献：

- 我们从影响力提供方的视角出发，提出户外广告中的标签特定遗憾最小化（TRMOA）问题，并明确对未满足的影响力和过量影响力进行建模。
- 我们证明了该问题的计算困难性，凸显了寻找最优解所面临的挑战。
- 我们提出一种自适应方法来选择影响力标签，以降低计算开销。
- 我们提出三种高效算法：两种基于贪心的方法，以及一种兼顾遗憾最小化、公平性和可扩展性的随机局部搜索方法。
- 我们通过在真实世界的广告牌和轨迹数据集上进行大量实验，验证了所提出方法的有效性及其实用价值。

e）论文结构：本文其余部分组织如下。第II节回顾相关文献。第III 节介绍必要的背景，并正式定义问题。第IV节介绍所提出的方法。第V节报告实验设置和结果。最后，第VI 节总结全文，并概述未来研究方向。

II. 相关工作

A. 影响力最大化

随着手持电子设备日益普及，大规模轨迹数据变得更易获取。因此，户外广告中的影响力最大化问题引起了研究人员的关注 [7], [9], [14]–[16]。大多数影响力最大化研究都是从社交网络广告的视角展开的 [9], [16]–[18]。在广告牌广告研究中 [3], [7],, 作者采用略有不同的影响力度量，以满足广告主视角下多样化的业务需求。在这些模型中，只有当广告牌距离受众的出行轨迹足够近时，才会对受众产生影响。具体而言，研究[1], [2], [7] 在考虑重叠的同时，以交通流量衡量影响力；而 [3] 则利用展示次数触发影响力指标，避免对在多个广告牌上看到同一广告的用户重复计数。这些研究与我们的 TRMOA 之间的关键区别在于目标不同。我们关注同时服务多位广告主的主机，旨在满足每位广告主需求的同时，将主机的整体损失降至最低。相比之下，以往研究仅关注最大化单个广告主的影响力。此外，还有一些略有关联的研究。研究 [19], [20]开发可视化工具，协助广告主选择广告牌；而 [6] 则利用社交媒体数据提升定向户外广告的效果。

B. 遗憾最小化

在社交病毒式营销（SVM）场景中，Twitter 和 Facebook 等影响力提供方平台会在社交网络上推广广告，并从广告主处收取费用，通常采用每次互动成本（CPE）模式。在这种模式下，广告主会根据其广告带来的每次点击或互动向平台付费 [21]–[23]。近期关于 SVM 的研究中，一条重要研究方向聚焦于最小化影响力提供方的遗憾 [5], [17], [24]。在此情形下，每位广告主都会指定所需的影响力水平。如果实际达到的影响力未达到该需求，广告主只需为实际获得的影响力付费。反之，如果实际达到的影响力超过了需求，广告主则无需为超出的部分付费。因此，影响力交付不足和超额交付都会给影响力提供方造成遗憾 [11]–[13], [25]。其他基于 CPE 模式的研究则聚焦于收益最大化，其中总收益定义为向所有广告主收取的付款总额。在这种情形下，每位广告主的付款等于与激活用户相关的 CPE 值之和 [17], [24]。在广告领域，遗憾最小化主要在两种情境中得到研究：SVM 和户外广告市场遗憾最小化（TR-MOA）问题。这两种情境在商业模式和影响机制上都有根本区别。首先，付款结构不同。在 SVM 中，付款依据互动情况；在某些情况下，如果广告主的需求未得到满足，影响力提供方可能无法获得

任何付款 [8],[13]。其次，影响力模型各不相同。在 TRMOA 中，影响力由地理邻近性驱动；只有当用户亲自经过广告牌时，才会受到影响，而且影响力不会在用户之间传播 [2], [8], [12], [13]。相比之下，SVM 依赖独立级联和线性阈值等概率扩散模型，影响力会通过网络中的社交关系传播 [26]–[28]。因此，SVM 研究主要关注对虚拟社交网络中的影响力传播进行估计，而 TRMOA 则侧重于用户与广告牌之间的空间交互。这些差异使得这两种情境下的基础优化问题存在根本区别。

III. 预备知识与问题建模

A. 轨迹、广告牌与标签数据库

轨迹数据库记录了移动个体随时间变化的位置。在我们的设定中，轨迹数据库 $\mathcal{D}$由形如 $(\mathcal{U}^{'}$，loc $[t_1, t_2])$ 的元组组成，表示个体群组 $\mathcal{U}^{'}$ 在时间区间 $[t_1, t_2]$内位于位置 loc。数据库 $\mathcal{D}$ 包含 m 个这样的元组。对于任意元组 $d \in \mathcal{D}$，令 t_u 表示与该元组关联的个体集合。令 $\mathcal{U} = u_1, u_2, \ldots, u_r$ 表示轨迹数据库所覆盖的所有个体的集合。形式上，$\mathcal{U}$ 定义为 $\mathcal{D}$中所有元组里出现的个体的并集，即 $\mathcal{U} = \bigcup_{t \in \mathcal{D}} t_u$，这意味着，如果某个个体至少出现在数据库的一个元组中，该个体就属于 $\mathcal{U}$ 。类似地，令 $\mathcal{L}$ 表示轨迹数据库覆盖的位置集合。如果某个位置至少出现在 $\mathcal{D}$的一个元组中，该位置就属于 $\mathcal{L}$。最后，令 $[T_1, T_2]$表示轨迹数据库的整体时间跨度。对于每个元组 $d \in \mathcal{D}$，其关联的时间区间 $[t_1, t_2]$ 都位于这一全局时间范围内，即 $[t_1, t_2] \subseteq [T_1, T_2]$。

广告牌数据库 $\mathcal{B}$ 包含城市中部署的广告牌信息。$\mathcal{B}$ 中的每条记录都表示为一个元组（b_{id}，loc，slot_duration，cost），其中b_{id} 表示广告牌的唯一标识符，loc 表示其位置，而 slot_duration 指定每个广告时段的时长和费用。假设有 ℓ 个广告牌 $\mathcal{B} = \{b_1, b_2, \ldots, b_\ell\}$，且每个广告牌的运行时间段为 $[T_1, T_2]$，并假设 $T = T_2 - T_1$。此外，假设所有广告牌均按时段分配用于展示广告，且每个时段的时长固定，记为 Δ。令 $\mathcal{BS}$ 表示所有广告牌时段的集合；即 $\mathcal{BS} = \{(b_i, [t_j, t_j + \Delta]) : i \in [k]$ 且 $j \in \{T_1, 1+\Delta, 1+2\Delta, \ldots, T_2-\Delta+1\}, cost\}$。现在，若 $|\mathcal{BS}| = h$，则 $h = k \cdot \frac{T}{\Delta}$。接下来，标签数据库包含商业客户提供的广告内容信息。具体而言，标签数据库 $\mathcal{T}$由标签构成，每个标签都有一个唯一标识符。在我们的设定中，$\mathcal{B}$ 中的每个广告牌都关联一个轨迹数据库，并且我们只考虑其位置与轨迹数据中至少一个位置相重合的广告牌。下面我们将介绍广告牌广告。

B. 广告牌广告

给定一组广告牌时段集合 $\mathcal{S}$， $\mathcal{S}$的影响力可表示为 $\mathcal{I}$（$\mathcal{S}$）。正如先前研究所述，在广告牌广告的影响力最大化 [1], [2], [4], [7]和遗憾最小化 [8],[10], [11], 问题中，有多种计算影响力的方法。我们采用与先前研究相同的度量方法。接下来，定义 1 中给出了广告牌时段影响力的定义1。

定义 1 (广告牌位影响力)。 给定广告牌广告位的一个子集 $\mathcal{S} \in \mathcal{BS}$， $\mathcal{S}$ 的影响力记作 $\mathcal{I}(\mathcal{S})$，其定义为方程 *1* 中轨迹数据库内所有用户的影响概率之和。

$$\mathcal{I}(\mathcal{S}) = \sum_{u \in \mathcal{D}} [1 - \prod_{b \in \mathcal{S}} (1 - Pr(u, b))] \tag{1}$$

这里，影响力函数 $\mathcal{I}(\cdot)$ 将每个广告牌广告位子集映射到相应的影响力值，即 $\mathcal{I} : 2^{\mathcal{BS}} \to \mathbb{R}_0^+$，其中 $\mathcal{I}(\emptyset) = 0$。

引理 1。 影响力函数 $\mathcal{I}(\cdot)$非负、单调且次模。

如前所述，当一个人在广告牌时段看到广告时，他会以一定概率受到影响。大多数现有研究 [3], [4], [7], [8]假设这一概率与广告内容无关，因此忽略了具体广告（例如其标签）的影响。然而，实际上，影响力在很大程度上取决于广告内容。例如，对酒类或软饮料不感兴趣的人不太可能受到此类广告的影响。另一方面，如果同一个人对电影感兴趣，那么与电影相关的广告就更有意义，因为受到广告宣传的影响后，这些广告可能会促使他去看电影。

因此，在本文中，我们明确将标签视为影响用户行为的关键因素。令 $\mathcal{T}_{u_i}$ 表示与用户 u_i相关联的标签集合。考虑 $\mathcal{U}$中的所有用户，相关标签的总体集合为 $\mathcal{T} = \bigcup_{u_i \in \mathcal{U}} \mathcal{T}_{u_i}$，且每个标签 $x \in \mathcal{T}$，我们用 $Pr(u \mid x)$表示影响力概率，其形式化定义见定义 2。

定义 2 (标签特定影响力)。对于用户 $u \in \mathcal{U}$ 和标签子集 $\mathcal{T}^{'} \subseteq \mathcal{T}$，标签特定影响力概率为

$$Pr(u \mid \mathcal{T}^{'}) = 1 - \prod_{x \in \mathcal{T}^{'}} \big(1 - Pr(u \mid x)\big). \tag{2}$$

接下来，我们在定义3中定义一组广告牌广告位的标签特定影响力。

定义 3 （广告牌广告位的标签特定影响力）。 对于一组广告牌广告位 $\mathcal{S} \subseteq \mathcal{BS}$ 和标签 $\mathcal{T}^{'} \subseteq \mathcal{T}$，标签特定影响力定义为

$$\mathcal{I}(\mathcal{S} \mid \mathcal{T}^{'}) = \sum_{u \in \mathcal{U}} (1 - \prod_{b \in \mathcal{S}} \big(1 - Pr(u, b \mid \mathcal{T}^{'})\big)). \tag{3}$$

函数 $\mathcal{I}(\cdot)$ 综合反映选定标签和广告牌广告位的影响，将每一对 $(\mathcal{T}^{'}, \mathcal{S})$ 映射为一个非负值，即 $\mathcal{I}: 2^{\mathcal{T}} \times 2^{\mathcal{BS}} \rightarrow \mathbb{R}_0^+$。

C. 问题定义

假设有 n 个广告主 $\mathcal{A} = a_1, a_2, ..., a_n$ 以及一个影响力提供方 $\mathcal{P}$。每个广告主 a_i 都会提交一个营销活动请求，指定标签$x_1, x_2, ..., x_k$上的影响力需求 σ_i 以及付款金额 u_i。影响力提供方维护着一个包含所有 $i \in [n]$的广告主数据库 $\mathbb{A} = (a_i, \sigma_i, \langle x_1, x_2, ..., x_k \rangle, u_i)$。令 $\mathcal{W}_{a_i}$ 表示分配给广告主 a_i 的广告位集合。根据付款规则，如果 $\mathcal{I}(\mathcal{W}_{a_i}) \geq \sigma_i$，那么 $\mathcal{P}$ 将收到全额付款 u_i；否则，只会收到部分（按比例）付款。我们重点关注两种重要情况：

- **未满足遗憾：** 如果提供的影响力低于所需需求 $\mathcal{I}(\mathcal{W}_{a_i}) < \sigma_i$，提供方就只能收到部分付款，从而造成损失。我们将此未满足遗憾称为未满足遗憾。
- 过度遗憾：</style}如果提供的影响力超过所需量 $\mathcal{I}(\mathcal{W}_{a_i}) > \sigma_i$，超出的影响力不会带来额外收入。此外，这些多余的影响力本可重新分配给需求尚未得到满足的其他广告主。我们将此称为过度遗憾</style}。

现在，结合这两种损失，我们在定义4中正式定义遗憾模型。

定义 4 （遗憾模型）。对于已分配广告牌广告位的广告主 $a_i \in \mathcal{A}$ $\mathcal{W}_{a_i}$，其遗憾记作 $\mathcal{R}_{a_i}$，定义如下：

$$\mathcal{R}_{a_i} = \begin{cases} u_i \cdot (1 - \delta \cdot \frac{\mathcal{I}(\mathcal{W}_{a_i})}{\sigma_i}), & if\ \sigma_i > \mathcal{I}(\mathcal{W}_{a_i}) \\ u_i \cdot \frac{\mathcal{I}(\mathcal{W}_{a_i}) - \sigma_i}{\sigma_i}, & otherwise \end{cases}$$

这里，$\frac{\mathcal{I}(\mathcal{W}_{a_i})}{\sigma_i}$ 表示需求满足比例，而δ 是未满足需求的惩罚比率。

a) 广告位分配：本研究的目标是分配广告位，以满足广告主的特定标签的影响力需求，并使影响力提供方的总遗憾最小化。我们考虑以下两个约束：

特定标签的影响力需求：此约束表示，对于任意广告主 $a_i \in \mathcal{A}$，都应满足其对特定标签影响力的需求。

互不相交约束：对于任意两个广告主 a_x 和 a_y，设 $\mathcal{W}_{a_x}$ 和 $\mathcal{W}_{a_y}$ 为其分配。互不相交约束表示 $\mathcal{W}_{a_x} \cap \mathcal{W}_{a_y} = \emptyset$。

现在，令 $\mathcal{G}$ 为所有可能的分配集合，其中n 表示广告主。将一个可能的分配记为$\{\mathcal{W}_{a_1}, \mathcal{W}_{a_2}, \ldots, \mathcal{W}_{a_n}\}$。现在，我们在定义 5中对其进行定义。

定义 5 (广告位的可行分配)。如果向广告主分配广告位满足标签特定的影响力需求和互不相交约束，则该分配是可行的。

如前所述，我们的主要目标是最小化影响力提供方的总遗憾。因此，我们在定义6中定义总遗憾的概念。

定义 6 （总遗憾）。与分配 $\mathcal{Z} = \{\mathcal{W}_{a_1}, \mathcal{W}_{a_2}, \ldots, \mathcal{W}_{a_n}\}$ 相关的总遗憾可表示为$\mathcal{R}$（$\mathcal{Z}$），并定义为方程4中所有广告主相关遗憾的总和。

$$\mathcal{R}(\mathcal{Z}) = \sum_{a_i \in \mathcal{A}} \mathcal{R}_{a_i} \tag{4}$$

现在，我们在定义7中定义 TRMOA问题。

定义 7 (标签特定遗憾最小化问题)。给定广告牌广告位 $\mathcal{BS}$、轨迹信息 $\mathcal{D}$和广告主 $\mathcal{A}$，该问题的目标是创建一种分配，使影响力提供者的总遗憾最小化。数学上可表示为：

$$\mathcal{Z}^{OPT} = \underset{\mathcal{Z}_i \in \mathcal{G}(\mathcal{Z})}{argmin}\ \mathcal{R}(\mathcal{Z}_i) \tag{5}$$

现在，从计算角度来看，该问题可写为如下形式：

TRMOA问题

输入： 广告牌广告位 $\mathcal{BS}$、影响力函数 $\mathcal{I}()$、轨迹数据库 $\mathcal{D}$、广告主数据库 $\mathbb{A}$。**问题：** 找出槽位的最优分配 $\mathcal{Z}^{OPT} = \{\mathcal{W}_{a_1}, \mathcal{W}_{a_2}, \ldots, \mathcal{W}_{a_n}\}$，使总体遗憾最小。

D. 问题难度。

定理1。户外广告中的标签特定遗憾最小化（*TRMOA*）问题是*NP*难的，且无法在任何常数因子范围内进行近似。

证明。我们通过从数值三维匹配（N3DM）问题归约，证明TRMOA问题的难度；已知该问题是NP完全的。设 X、Y、Z 为三个整数多重集，每个集合包含 n 个元素。设 b为一个界，使得 $b = \frac{\sum X + \sum Y + \sum Z}{n}$。N3DM判定问题询问是否存在一个完美匹配$M \subseteq X \times Y \times Z$，使得每个元素恰好出现一次，且对于每个三元组（$x_i$、$y_i$、$z_i$），$\in M$、$x_i + y_i + z_i = b$。给定一个N3DM实例，我们按如下方式构造一个TRMOA实例。（1）创建 $3n$ 个广告牌广告位，并将其划分为三个互不相交的集合：D_1、D_2、D_3，每个集合的大小为 n。（2）映射 $X \rightarrow D_1$、$Y \rightarrow D_2$、$Z \rightarrow D_3$中的每个元素。（3）对于每个广告位 s_i，构造一个互不相交的用户集，使得根据定义3，$\mathcal{I}(\{s_i\} \mid \mathcal{T}) = w_i$，其中 w_i 为映射后的整数。由于不同广告位对应的用户互不相交，完整的标签特定影响函数仍为：$\mathcal{I}(\mathcal{S} \mid \mathcal{T}) = \sum_{s_i \in \mathcal{S}} w_i$，同时保持非负性、单调性和次模性。（4）引入一个足够大的常数 c。修改广告位影响力如下：$\forall\ s_i \in D_1$：$\mathcal{I}(s_i) \leftarrow c + w_i$，$\forall\ s_j \in D_2$：$\mathcal{I}(s_j) \leftarrow 3c + w_j$，$\forall\ s_k \in D_3$：$\mathcal{I}(s_k) \leftarrow 9c + w_k$。（5）创建 n 个具有相同 $\sigma = b + 13c$的广告主。

我们观察到 $(c + w_i) + (3c + w_j) + (9c + w_k) = 13c + (w_i + w_j + w_k)$。当 c 足够大时，未能从 D_1, D_2, D_3 中各恰好选择一个槽位，都会产生

与 σ 严格不同的影响力。因此，当且仅当 每位广告主都从每个集合中恰好获得一个槽位时，遗憾为零。

正确性。 如果 TRMOA 的总遗憾为零，那么对于每个广告主 a_i： $\mathcal{I}(\mathcal{S}_i) = \sigma$。根据 c 的系数支配性，每个 $\mathcal{S}_i$ 必须恰好包含来自D_1、 D_2、 D_3的一个元素。移除 $13c$ 偏移量后，得到 $w_i + w_j + w_k = b$。因此，存在一个有效的 N3DM 匹配。如果存在有效的 N3DM 匹配，则将每个三元组分配给一个不同的广告主。每个广告主获得的影响力恰好为 σ，因此遗憾为零。

近似困难性。令 OPT^N 表示 N3DM 的最优值（即违反的三元组数量），令 OPT^T 表示 TRMOA 中的最小遗憾。我们有 $\text{OPT}^N = 0$，$\text{OPT}^T = 0$。假设存在一个多项式时间算法，能够以任意常数因子 τ近似 TRMOA。那么，$\mathcal{R}(\mathcal{S}) \leq \tau \cdot \text{OPT}^T$。因此，若 $\text{OPT}^N = 0$，则 $\mathcal{R}(\mathcal{S}) = 0$。若 $\text{OPT}^N > 0$，则 $\mathcal{R}(\mathcal{S}) > 0$。由此，我们可以通过检查遗憾是否为零，在多项式时间内求解 N3DM，这意味着 P =NP。因此，TRMOA 是 NP难的，且无法在任意常数因子内近似。 □

IV. 所提出的解决方法

A. 自适应影响力标签选择（*AITS*）

在我们的问题背景下，选择标签是后续所有解决方案的首要要求。在 AITS 方法中，以轨迹数据库 $\mathcal{D}$、标签集 $\mathcal{T}$ 和自适应停止阈值（$\omega = 0.01$）作为输入。该方法采用贪心策略，逐步构建影响力标签集。在每次迭代中，它会针对当前选择，选取边际影响力增益最大的标签 $\Delta(t \mid \mathcal{T}') = \mathcal{I}(\mathcal{T}' \cup \{t\}) - \mathcal{I}(\mathcal{T}')$。当最佳边际增益低于当前影响力的 ω 倍时，流程便会结束，以避免加入影响微不足道的标签。该算法输出一组标签 $\mathcal{T}'$，使其对轨迹的整体影响力最大化。

算法1： 自适应影响力标签选择

输入： 标签集 $\mathcal{T}$，轨迹数据 $\mathcal{D}$, ω
输出： 选定的标签集 $\mathcal{T}'$

$\mathcal{T}' \leftarrow \emptyset$;
当为真时执行
  $t^* = \text{取最大值}_{t \in \mathcal{T} \setminus \mathcal{T}'} \Delta(t \mid \mathcal{T}')$;
  如果 $\Delta(t^* \mid \mathcal{T}') < \omega \cdot \mathcal{I}(\mathcal{T}')$ **则**
    中断
  $\mathcal{T}' \leftarrow \mathcal{T}' \cup \{t^*\}$;
返回 $\mathcal{T}'$;

a) 复杂度分析。： 我们分析算法1的时间复杂度和空间复杂度。初始化步骤（第1行）需要 $\mathcal{O}(1)$ 时间。在每次迭代中，第3 行扫描所有剩余标签，以识别具有最大边际影响力增益的标签，需要 $\mathcal{O}(|\mathcal{T}|^2 \cdot c)$ 时间，其中 c 是轨迹数据库中元组的数量。第4行的停止条件以及第6 行的更新各需要 $\mathcal{O}(1)$ 时间。若算法在终止前选择 $|\mathcal{T}'|$ 个标签，则总时间复杂度为 $\mathcal{O}(|\mathcal{T}'| \cdot |\mathcal{T}|^2 \cdot c)$。额外的空间复杂度为 $\mathcal{O}(|\mathcal{T}|)$。

B. 面向标签特定遗憾最小化的公平性感知轮询贪心算法（*BG*）

算法2 提出了一种公平性感知的贪心框架，用于最小化影响力提供方的总遗憾。首先，按照预算需求比对广告主进行排序 $\frac{u_i}{\sigma_i}$。对于每个广告主，算法首先使用算法1细化标签集，仅保留影响力最大的标签，以剔除影响较小的标签。随后，按照轮转方式在细化后的标签集上进行分配，确保标签间公平性，并避免任何单一标签占据主导地位。在每一步中，算法都会贪心地选择单位成本下遗憾减少最大的广告牌时段，并明确优化归一化遗憾目标。该过程持续进行，直到满足广告主特定的影响力需求或资源耗尽。最终的分配实现了遗憾最小化和预算效率，并在广告主和标签之间公平分配广告位，从而得到全局平衡的解决方案。

算法2： 公平性感知的标签特定遗憾最小化轮转贪心标签特定遗憾最小化算法

数据： 轨迹数据库 $\mathcal{D}$、广告牌广告位信息 $\mathcal{BS}$、广告主信息 $\mathcal{A}$，影响力函数 $\mathcal{I}(\cdot)$，标签信息 $\mathcal{T}$
结果： 将广告牌广告位 $\mathcal{Y}$ 分配给广告主—标签对，使总遗憾

初始化：
$\mathcal{Y} \leftarrow \{\mathcal{S}_{i,t} = \emptyset \mid a_i \in \mathcal{A},\ t \in \mathcal{T}_i\}$;
$\mathcal{Z} \leftarrow \{Z_1, Z_2, \ldots, Z_{|\mathcal{A}|}\}$
按 $\frac{u_i}{\sigma_i}$降序排列广告主 $a_i \in \mathcal{A}$ —;
对于每个广告主 $a_i \in \mathcal{A}$ **执行**
  计算精炼标签 $\mathcal{T}'_i \leftarrow AITS(\mathcal{T}_i, \mathcal{D})$;
  令 $\mathcal{T}'_i = \{t_{i,1}, t_{i,2}, \ldots, t_{i,|\mathcal{T}'_i|}\}$ 为 a_i的标签集;
  初始化标签指针 $k \leftarrow 0$;
  当 $\exists\, t \in \mathcal{T}'_i$ 满足 $\mathcal{I}(\mathcal{Z}_i) < \sigma_i$ 且 $\mathcal{BS} \neq \emptyset$ **执行**
    $t \leftarrow t_{i,k}$;
    $s^* \leftarrow$取最大值$_{s \in \mathcal{BS}} \frac{\mathcal{R}(\mathcal{Z}_i) - \mathcal{R}(\mathcal{Z}_i \cup \{s\})}{\sigma(\{s\})}$;
    $\mathcal{S}_{i,t} \leftarrow \mathcal{S}_{i,t} \cup \{s^*\}$;
    $\mathcal{Z}_i \leftarrow \mathcal{Z}_i \cup \{s^*\}$
    $\mathcal{BS} \leftarrow \mathcal{BS} \setminus \{s^*\}$;
    $k \leftarrow (k+1)$ 模 $|\mathcal{T}_i|$;
  end
结束
返回 $\mathcal{Y}$;

a) 复杂度分析。： 现在，我们分析算法2的时间和空间需求。按 $\frac{u_i}{\sigma_i}$ 对广告主排序需要 $\mathcal{O}(|\mathcal{A}|\log |\mathcal{A}|)$ 时间。对于每位广告主a_i，自适应影响力标签选择过程需要 $\mathcal{O}(|\mathcal{T}'_i| \cdot |\mathcal{T}_i|^2 \cdot c)$ 时间，其中 c 表示轨迹元组的数量。在分配阶段，每个广告牌时段最多被选择一次，每次选择都会扫描剩余时段，因此最坏情况下的总复杂度为 $\mathcal{O}(|\mathcal{BS}|^2)$。因此，总体时间复杂度为 $\mathcal{O}(|\mathcal{A}| \log |\mathcal{A}| + \sum_{a_i \in \mathcal{A}} |\mathcal{T}'_i| \cdot |\mathcal{T}_i|^2 \cdot c + |\mathcal{BS}|^2)$。额外空间复杂度为 $\mathcal{O}(|\mathcal{BS}| + |\mathcal{A}| \cdot |\mathcal{T}|)$。

C. 面向公平性的标签特定遗憾最小化随机轮转贪心算法（*RG*）

算法3 在影响力和预算约束下，将广告牌广告位分配给广告主-标签对，以最小化总遗憾。广告主按其归一化预算-需求比排序；对于每个广告主，使用算法

1获得经过优化的影响力标签集。分配按标签轮流进行，每一步都会随机抽取一部分可用广告位，并选择单位成本遗憾减少最大的广告位。采样使用$\binom{|\mathcal{BS}|}{k}$对数$\frac{1}{\epsilon}$，其中$k$是每次迭代中可用广告位的$|\mathcal{BS}|$ 10%。从该子集中，选择单位成本遗憾减少最大的广告位。该采样策略遵循子模优化的随机贪心框架[29]，，在以高概率保持解的质量的同时，显著降低计算复杂度。总的来说，该算法能够实现预算高效、公平且可扩展的分配，适用于大规模广告牌数据集。

算法 3: 公平性感知的随机轮转贪心算法，用于特定标签的遗憾最小化

数据: 轨迹数据库 $\mathcal{D}$、广告牌广告位信息 $\mathcal{BS}$、广告主信息 $\mathcal{A}$、影响力函数 $\mathcal{I}(\cdot)$、标签信息 $\mathcal{T}$
结果: 一种将广告牌广告位分配给广告主—标签对的分配 $\mathcal{Y}$，以最小化总遗憾

初始化:
$\mathcal{Y} \leftarrow \{\mathcal{S}_{i,t} = \emptyset \mid a_i \in \mathcal{A},\ t \in \mathcal{T}_i\}$;
$\mathcal{Z} \leftarrow \{Z_1, Z_2, \ldots, Z_{|\mathcal{A}|}\}$
将广告主 $a_i \in \mathcal{A}$ 按以下指标降序排序 $\frac{u_i}{\sigma_i}$;
对于每个广告主 $a_i \in \mathcal{A}$ **执行**
  计算精炼标签 $\mathcal{T}_i^{'} \leftarrow AITS(\mathcal{T}_i, \mathcal{D})$;
  令 $\mathcal{T}_i^{'} = \{t_{i,1}, t_{i,2}, \ldots, t_{i,|\mathcal{T}_i^{'}|}\}$ 为 a_i的标签集;
  初始化标签指针 $k \leftarrow 0$;
  当 $\exists t \in \mathcal{T}_i^{'}$ 使得 $\mathcal{I}(\mathcal{Z}_i) < \sigma_i$ 且 $\mathcal{BS} \neq \emptyset$ **执行**
    $t \leftarrow t_{i,k}$;
    $\mathcal{BS}^{'} \leftarrow$ 通过抽取 s个随机元素得到的随机子集来自 $\mathcal{BS} \setminus \mathcal{Z}_i$;
    $s^* \leftarrow$ 参数 $\max_{s \in \mathcal{BS}^{'}} \frac{\mathcal{R}(\mathcal{Z}_i) - \mathcal{R}(\mathcal{Z}_i \cup \{s\})}{\sigma(\{s\})}$;
    $\mathcal{S}_{i,t} \leftarrow \mathcal{S}_{i,t} \cup \{s^*\}$;
    $\mathcal{Z}_i \leftarrow \mathcal{Z}_i \cup \{s^*\}$
    $\mathcal{BS} \leftarrow \mathcal{BS} \setminus \{s^*\}$;
    $k \leftarrow (k+1) \bmod |\mathcal{T}_i|$;
  end
结束
返回 $\mathcal{Y}$, $\mathcal{BS}$;

a) 复杂度分析：我们分析算法3的时间复杂度和空间复杂度。按 $\frac{u_i}{\sigma_i}$ 对广告主排序需要 $\mathcal{O}(|\mathcal{A}| \log |\mathcal{A}|)$ 时间。对于每个广告主 a_i，自适应影响力标签选择过程需要 $\mathcal{O}(|\mathcal{T}_i^{'}| \cdot |\mathcal{T}_i|^2 \cdot c)$ 时间，其中 c 表示轨迹元组的数量。在分配过程中，每次迭代仅评估剩余广告牌广告位中大小为 s 的采样子集，因此每次选择需要 $\mathcal{O}(s)$ 时间。由于每个广告牌广告位最多分配一次，总分配成本为 $\mathcal{O}(|\mathcal{BS}| \cdot s)$。因此，总体时间复杂度为 $\mathcal{O}(|\mathcal{A}| \log |\mathcal{A}| + \sum_{a_i \in \mathcal{A}} |\mathcal{T}_i^{'}| \cdot |\mathcal{T}_i|^2 \cdot c + |\mathcal{BS}| \cdot s)$。额外的空间复杂度为 $\mathcal{O}(|\mathcal{BS}| + |\mathcal{A}| \cdot |\mathcal{T}|)$。

D. 面向标签特定遗憾最小化的随机局部搜索算法（*RLS*）

算法4 提出了一种随机局部搜索框架，以进一步改进针对标签的广告牌分配，从而降低总遗憾。算法首先为所有广告主-标签对初始化空分配，并根据 $\frac{u_i}{\sigma_i}$对广告主进行排序。随后，使用算法3构造初始解，以提供高质量的起点。接着，算法执行 N 次随机局部搜索迭代。

在每次迭代中，使用通过算法1获得的优化影响力标签，逐个广告主重新分配广告牌广告位。分配过程以轮询方式遍历标签，每一步都从剩余广告位中均匀随机地选取一个广告牌广告位，从而鼓励探索多样的分配模式。每次迭代后，都会评估所构造解的总遗憾，并与迄今为止找到的最优解进行比较。如果发现更优解，当前分配就会取代当前最优解。最后，使用算法3 过程，将所有剩余的未分配广告牌广告位贪心地分配给最优解，以进一步降低遗憾。算法返回所有迭代中的最佳分配，从而有效平衡探索与利用，以最小化遗憾。

算法 4: 随机局部搜索算法，用于标签特定遗憾最小化

数据: 轨迹数据库 $\mathcal{D}$，广告牌广告位信息 $\mathbb{BS}$，广告主信息 $\mathcal{A}$，影响力函数 $\mathcal{I}(\cdot)$，样本量 s，标签信息 $\mathcal{T}$
结果: 将广告牌广告位分配给广告主—标签对，使其最小化总遗憾

初始化:
$\mathcal{M} \leftarrow \{\mathcal{S}_{i,t} = \emptyset \mid a_i \in \mathcal{A},\ t \in \mathcal{T}_i\}$;
$\mathcal{Z} \leftarrow \{Z_1, Z_2, \ldots, Z_{|\mathcal{A}|}\}$
按降序排列广告主 $a_i \in \mathcal{A}$ $\frac{u_i}{\sigma_i}$;
$\mathcal{S}^{best}, \mathcal{BS}^{'} \leftarrow \mathrm{RG}(\mathcal{BS}, \mathcal{T}, \mathcal{A}, \mathcal{Z})$;
对于 $iter \leftarrow 1$ **至** N **执行**
  $\mathcal{BS}^{'} \leftarrow \mathcal{BS}$;
  对于每个广告主 $a_i \in \mathcal{A}$ **执行**
    计算精炼标签 $\mathcal{T}_i^{'} \leftarrow AITS(\mathcal{T}_i, \mathcal{D})$;
    令 $\mathcal{T}_i^{'} = \{t_{i,1}, t_{i,2}, \ldots, t_{i,|\mathcal{T}_i^{'}|}\}$ 是 a_i的标签集;
    初始化标签指针 $k \leftarrow 0$;
    当 $\exists t \in \mathcal{T}_i^{'}$ 使得 $\mathcal{I}(\mathcal{Z}_i) < \sigma_i$ 且 $\mathcal{BS} \neq \emptyset$ **执行**
      $t \leftarrow t_{i,k}$;
      $s^* \leftarrow$ 随机选择一块广告牌 $o \in \mathcal{BS}^{'}$;
      $\mathcal{S}_{i,t} \leftarrow \mathcal{S}_{i,t} \cup \{s^*\}$;
      $\mathcal{Z}_i \leftarrow \mathcal{Z}_i \cup \{s^*\}$
      $\mathcal{BS}^{'} \leftarrow \mathcal{BS}^{'} \setminus \{s^*\}$;
      $k \leftarrow (k+1)$ 模 $|\mathcal{T}_i|$;
    如果 $\mathcal{R}(\mathcal{Z}) < \mathcal{R}(\mathcal{S}^{best})$ **则**
      $\mathcal{S}^{best} \leftarrow \mathcal{Z}$;
      $\mathcal{BS}^{''} \leftarrow \mathcal{BS}^{'}$;
如果 $\mathcal{BS}^{''} \neq \emptyset$ **则**
  $\mathcal{S}^{best} \leftarrow RG(\mathcal{BS}^{''}, \mathcal{T}, \mathcal{A}, \mathcal{S}^{best})$ 跳过中的第 6 行算法3
返回 $\mathcal{S}^{best}$;

a) 复杂度分析：我们分析算法 4 的时间复杂度和空间复杂度。按 $\frac{u_i}{\sigma_i}$ 进行广告主排序需要 $\mathcal{O}(|\mathcal{A}| \log |\mathcal{A}|)$ 时间。初始随机贪心构造耗费 $\mathcal{O}(|\mathcal{BS}| \cdot s)$ 时间。随后，算法执行 N次局部搜索迭代；在每次迭代中，对每个广告主a_i，自适应影响力标签选择过程的运行时间为$\mathcal{O}(|\mathcal{T}_i^{'}| \cdot |\mathcal{T}_i|^2 \cdot c)$，其中 c 表示轨迹元组的数量，随后进行随机槽位分配，二者合计需要 $\mathcal{O}(|\mathcal{BS}|)$ 时间。因此，总时间复杂度为 $\mathcal{O}(|\mathcal{A}| \log |\mathcal{A}| + |\mathcal{BS}| \cdot s + N \cdot \sum_{a_i \in \mathcal{A}} (|\mathcal{T}_i^{'}| \cdot |\mathcal{T}_i|^2 \cdot c + |\mathcal{BS}|))$。额外空间复杂度为 $\mathcal{O}(|\mathcal{BS}| + |\mathcal{A}| \cdot |\mathcal{T}|)$。

V. 实验评估

我们尝试了不同的参数设置，以确定这些算法在多大程度上能够最小化遗憾以及所需的计算时间。由于这是关于广告牌广告中标签特定遗憾最小化的首项工作，目前没有可直接比较的现有方法。因此，我们对下文的两个研究问题（RQ）进行了广泛分析，以便为不同真实场景中的遗憾最小化提供实用见解。

- RQ1：当所有广告主的总需求远低于、接近或超过影响力提供方的最大供给时，会发生什么？
- RQ2：哪种类型的广告主有助于影响力提供方最小化遗憾：少数需求量大的大型广告主，还是大量需求量小的小型广告主？

A. *TRMOA*问题中的挑战

TRMOA问题面临以下几项挑战。

- 该问题是NP难且不可近似的，因此无法为大型现实实例高效计算最优解。该问题会同时惩罚分配不足（未满足遗憾）和分配过多（过度遗憾）。其目标是减少总遗憾，但减少其中一种遗憾往往会增加另一种，因此难以实现平衡分配。

- 影响力取决于选定标签和分配的广告位，因此会形成庞大的组合搜索空间和复杂的相互依赖关系。此外，多个广告主会争夺互不重叠的广告位，满足某个广告主的需求可能会对其他广告主产生负面影响，需求量高时尤其如此。此外，大规模的轨迹和广告牌数据集会显著增加计算成本，因此需要高效、可扩展的启发式解。

- 由于这是首项针对标签的遗憾最小化研究，目前没有可供比较的现有方法。我们将自己的方法与作为基线的“随机”方法进行比较。

B. 实验设置

1）数据集：我们使用两个真实世界的轨迹数据集进行了实验：纽约市（NYC）和洛杉矶（LA）的签到数据。纽约市共记录了 227,428 次签到[1]，时间范围为2012年4月12日至2013年2月16日。洛杉矶[2] 数据集包含从15条街道收集的74,170条用户记录，其中包括街道名称、GPS位置、时间戳及相关信息。我们从 Lamar Advertising Company 抓取了广告牌数据[3], 。该公司是全球最大的户外广告服务商之一。纽约市广告牌数据集包含716个广告牌（1,031,040个广告位），而洛杉矶数据集包含1,483个广告牌（2,135,520个广告位）。

[1]https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page [2] https://github.com/Ibtihal-Alablani [3] http://www.lamar.com/InventoryBrowser

2）关键参数：我们的实验使用了若干关键参数，包括供需比 α、平均个体需求比 β、距离阈值 γ以及未满足惩罚比 δ。表中总结了所有关键参数V。

表 V：关键参数

参数	取值
α	40%，60%，80%，**100%**，120%
β	1%，2%，**5%**，10%，20%
ϵ	**0.01**, 0.05, 0.1, 0.15, 0.2
ω	**0.01**, 0.05, 0.1, 0.15, 0.2
δ	0, 0.25, **0.5**, 0.75, 1
γ	$25m$，$50m$，**100米** $125m$，$150m$

a) 供需比 (α)：这是总影响力需求与影响力供给之比，即 $\alpha = \sigma^d/\sigma^s$，其中 $\sigma^d = \sum_{i=1}^{|\mathcal{A}|}\sigma_i$ 指总需求，σ^s 指总影响力供给，即 $\sigma^s = \sum_{b\in\mathcal{BS}}\sigma(b)$。

b) 平均个体需求比 (β)：这是广告主的平均个体需求占影响力供给的百分比，即 $\beta = \frac{\sigma^d/|\mathcal{A}|}{\sigma^s}$。通过控制$\beta$的取值，我们可以控制广告主的影响力需求。

c) 广告商影响力需求 (σ)：每位广告主的需求由参数 α 和 β生成，并计算为 $\sigma = \lfloor\psi\cdot\sigma^s\cdot\beta\rfloor$，其中 $\psi\in[0.8,1.2]$ 是随机选取的，以引入广告主需求的变动性。

d) 广告商标签 ($\mathcal{T}$)：在所有实验中，我们假设每位广告主至少有 100 个、至多有 500 个标签。

e) 广告商付款 (λ)：遵循广泛采用的设定 [5], [24], ，可以将广告主的付款表示为 $\lambda = \lfloor\eta\cdot\sigma\rfloor$，其中 $\eta\in[0.9,1.1]$。

f) 广告牌广告位成本：所有户外广告公司（如 LAMAR）都不会公开广告位的确切租赁价格；不过，文献 [7], [8], [11] 将广告位成本视为与其影响力成正比。因此，我们采用与 $\lfloor\tau\times\mathcal{I}(s)/10\rfloor$相同的设定，其中 $\mathcal{I}(s)$是槽位 s 的影响力，而 τ是在 0.9 到 1.1 之间随机选取的，用以模拟波动。

g) 惩罚比率(δ)：当广告主不满意时，惩罚比率 $\delta\in[0,1]$ 控制付款比例。

h) 环境配置：所有算法均使用 Visual Studio Code 以 C++ 实现。实验在一台基于 Ubuntu 的台式机上进行，该计算机配备 64 GB 内存和一颗 Xeon(R) 3.50 GHz 处理器。

C. 实验结果与讨论。

1) 纽约市数据集上的实验：实验讨论分为以下四种情景。

a) 案例***1***：低 α，低 β 图 ***1*** ***(a,b,c,f,g,h,k,ℓ,m)***：对应案例1，我们有一种 $\alpha\leq 80\%$和 $\beta\leq 5\%$的情形。这指的是全局需求和单个广告主需求都较低的情况。影响力提供方拥有少量影响力需求较低的广告主。我们有三个主要观察结果。**首先**，当 α 增加时，所有所提出的方法的过度影响力都会降低，而未满足遗憾则会增加。这是因为，随着 α 增加，广告主的需求会上升，而由于大多数广告主很早就得到满足，因此会导致过度遗憾。**其次**，在所提出的方法中，

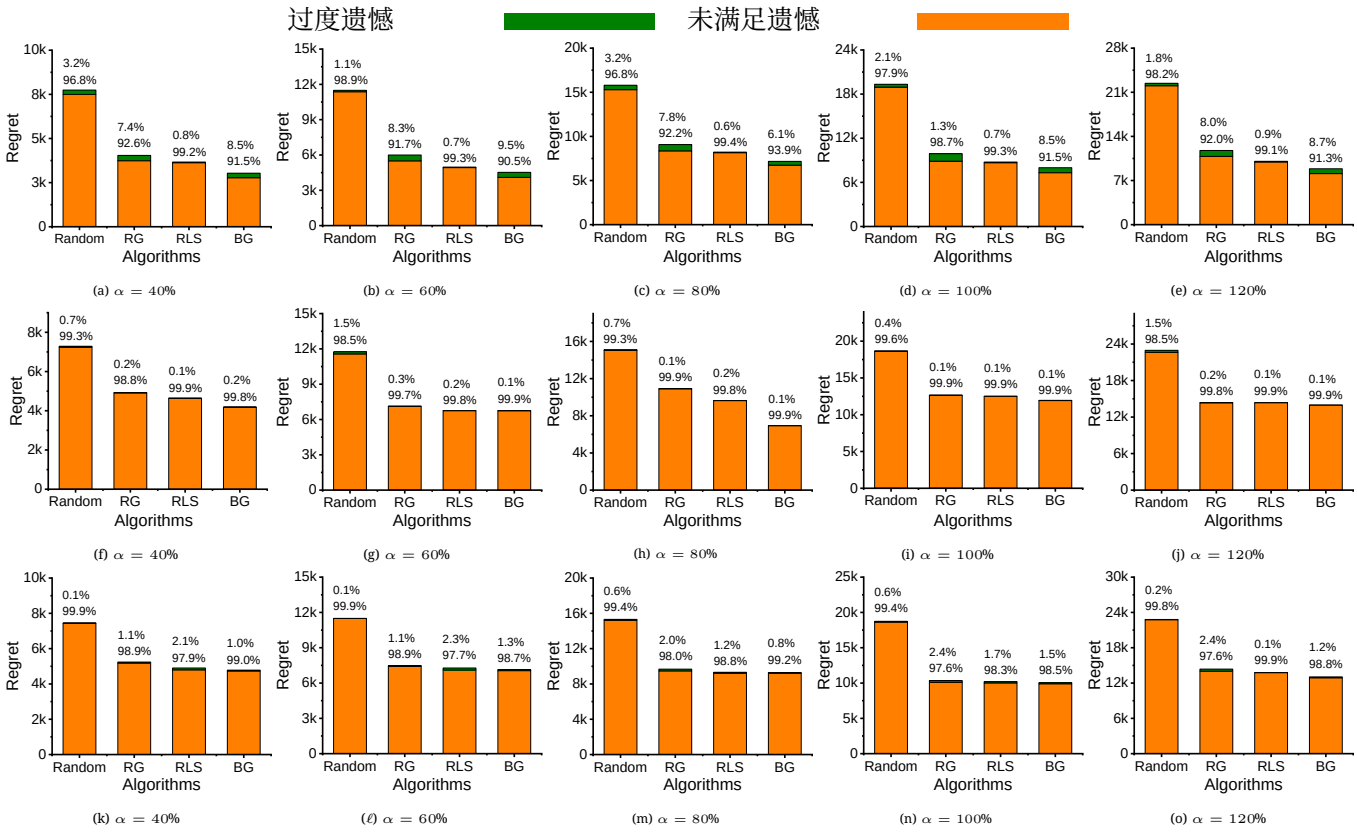


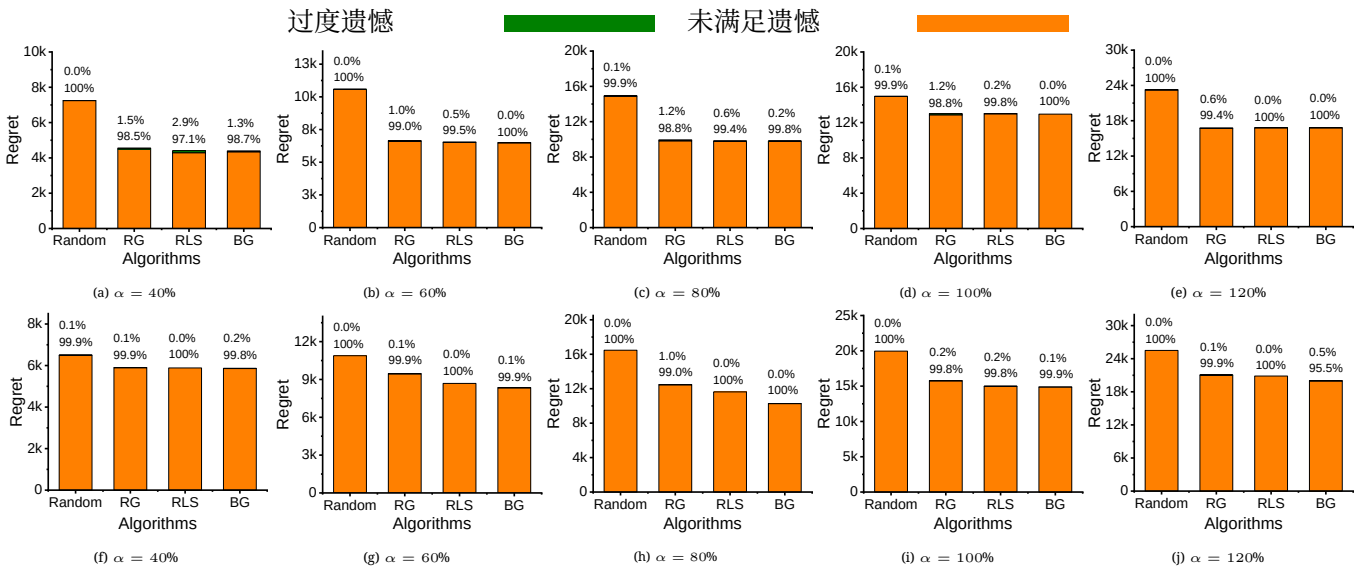
图 1：变化的供需比 α当 $\beta = 1\%$ $|\mathcal{A}| = 100$ $(a,\ b,\ c,\ d,\ e)$，当 $\beta = 2\%$ $|\mathcal{A}| = 50$ $(f,\ g,\ h,\ i,\ j)$，当$\beta = 5\%$ $|\mathcal{A}| = 20$ $(k,\ \ell,\ m,\ n,\ o)$（纽约市数据集）

就过度遗憾和未满足遗憾而言，‘BG’的表现优于‘RG’和‘RLS’。这是因为‘RG’和‘RLS’在向广告主分配广告位时采用了随机化。**第**三，与所提出的‘BG’、‘RG’和‘RLS’相比，‘随机’方法产生的总遗憾最高。然而，供需比（α）非常低，而且由于广告主有特定标签的影响力需求，影响力提供方无法满足所有广告主的需求。

b) 情况***2***：低 α，高 β，图***2 (a,b,c,f,g,h)***：对应情况2，我们有 $\alpha \leq 80\%$ 和 $\beta \geq 10\%$。这表示一种全局影响力需求保持较低、而个体需求较高的情况。影响力提供方的广告主数量较少，但每位广告主的需求较高。我们有两个主要观察结果。**首先**，随着全局需求保持较低且个体需求增加，所有算法的总体遗憾都会降低。随着 β 增加，广告主数量减少，但广告主的个体需求增加。也就是说，影响力供给接近广告主的需求，因此过度遗憾会下降。**其次**，个体需求越高，广告主就能投放更多广告位，而“RLS”可以通过向广告主分配广告位来探索更大的邻域搜索空间。

c) 情况***3***：高 α，低 β 对应图***1 (d,e,i,j,n,o)***：对应情况3，我们有 $\alpha \geq 100\%$ 和 $\beta \leq 5\%$。这表示一种全局需求高、但个体影响力需求低的情形。影响力提供方拥有大量需求较小的广告主。我们有两个主要观察结果。**首先**，在全局需求极高（$\alpha \geq 100\%$）时，没有任何算法能够满足所有广告主。考虑 $\alpha \geq 100\%$且 $\beta = 1\%$的情况。总体遗憾包含过度遗憾，尽管并非所有广告主都完全得到满足。如今，随着 β增加，过度遗憾会下降，只剩下未满足遗憾。**其次**，当 $\alpha = 100\%$时，过度遗憾较少，而未满足遗憾较多。由于这些广告主的标签特定影响力需求，影响力提供方无法满足所有广告主。尽管影响力提供方还有许多剩余广告位，但无法为这些广告主安排广告位。需要注意的一点是，“RLS”方法满足的广告主数量几乎与“BG”相同，而“RG”和“随机”满足的广告主数量则少于“RLS”。

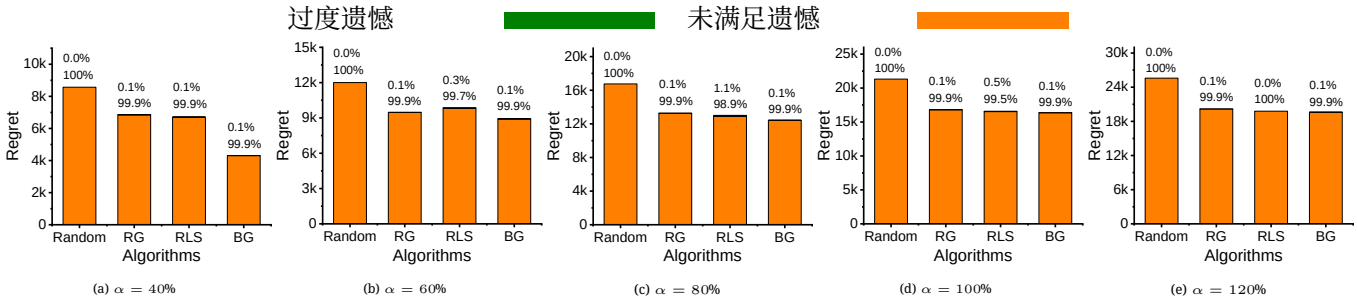
*d)*情形***4***：高 α、高 β 图***2 (d,e,i,j)***：对应情形4，我们有 $\alpha \geq 100\%$ 和 $\beta \geq 10\%$。这指的是全球需求和个体需求都非常高的情形。影响力提供方拥有少量广告主，每位广告主的个体需求都很高。我们有两个主要观察结果。**首先**，当 α 和 β 的值较大时，所有算法的未满足遗憾都更高。



(a) $\alpha = 40\%$ (b) $\alpha = 60\%$ (c) $\alpha = 80\%$ (d) $\alpha = 100\%$ (e) $\alpha = 120\%$

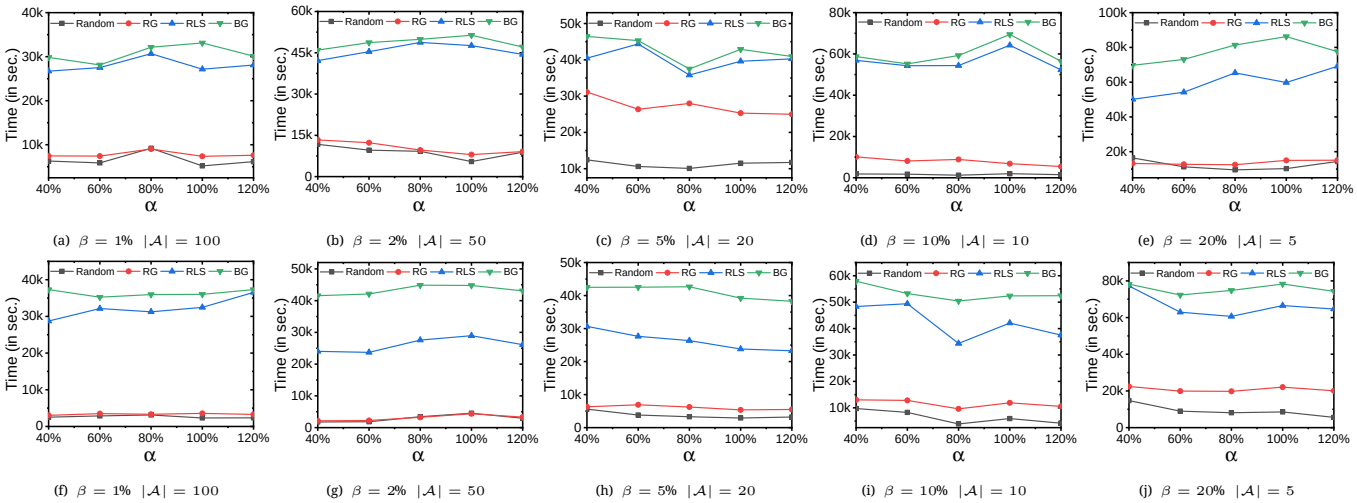
(f) $\alpha = 40\%$ (g) $\alpha = 60\%$ (h) $\alpha = 80\%$ (i) $\alpha = 100\%$ (j) $\alpha = 120\%$

图 2：变化的 供需比， α 当 $\beta = 10\%$ $|\mathcal{A}| = 10$ 时，当 $\beta = 20\%$ $|\mathcal{A}| = 5$ 时（纽约市数据集 ）



(a) $\alpha = 40\%$ (b) $\alpha = 60\%$ (c) $\alpha = 80\%$ (d) $\alpha = 100\%$ (e) $\alpha = 120\%$

图 3：变化的供需比， α 当 $\beta = 5\%$ $|\mathcal{A}| = 20$ （洛杉矶数据集）



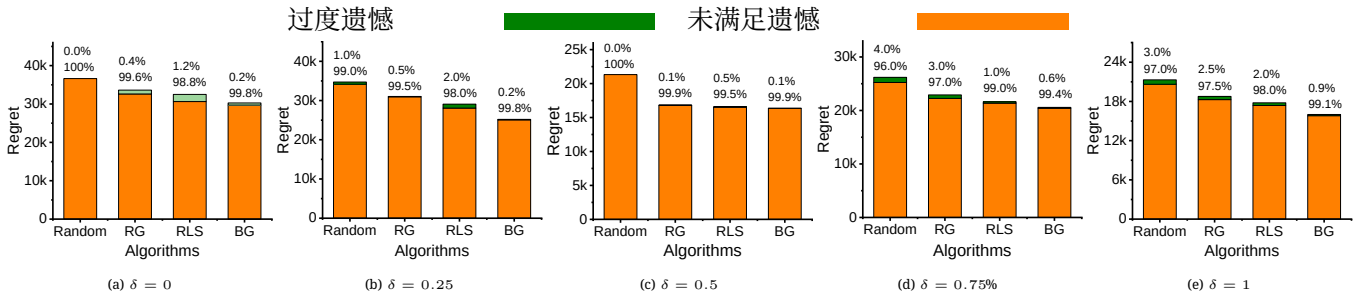
(a) $\beta = 1\%$ $|\mathcal{A}| = 100$ (b) $\beta = 2\%$ $|\mathcal{A}| = 50$ (c) $\beta = 5\%$ $|\mathcal{A}| = 20$ (d) $\beta = 10\%$ $|\mathcal{A}| = 10$ (e) $\beta = 20\%$ $|\mathcal{A}| = 5$

(f) $\beta = 1\%$ $|\mathcal{A}| = 100$ (g) $\beta = 2\%$ $|\mathcal{A}| = 50$ (h) $\beta = 5\%$ $|\mathcal{A}| = 20$ (i) $\beta = 10\%$ $|\mathcal{A}| = 10$ (j) $\beta = 20\%$ $|\mathcal{A}| = 5$

图4：效率测试 a， b， c， d， e(纽约市数据集)， f， g， h， i， j(洛杉矶数据集)

‘RG’、‘RLS’ 和 ‘BG’ 的总体遗憾几乎相同，而‘随机’的总体遗憾更高，其中大部分来自未满足遗憾。**其次**，与当 $\alpha = 120\%$ 时的供给，即影响力提供方没有足够的广告位来满足广告主的需求。

2)洛杉矶数据集上的实验： 图 3 展示了洛杉矶数据集上的实验结果。我们观察到



(a) $\delta = 0$ (b) $\delta = 0.25$ (c) $\delta = 0.5$ (d) $\delta = 0.75\%$ (e) $\delta = 1$

Fig. 5：当 $\alpha = 100\%$ $\beta = 5\%$ $|\mathcal{A}| = 20$ 变化时的遗憾（δ）（洛杉矶数据集）

洛杉矶数据集中的广告位比纽约市数据集更多。洛杉矶的广告位影响力较低，平均影响力也较低。与纽约市相比，洛杉矶的广告位影响力分布更加均匀。此外，不同广告位之间的影响力重叠较小。我们有两个主要观察结果。**首先**，洛杉矶数据集的结果与纽约市的结果相当相似，但所有算法中过度遗憾所占的比例都更低。这主要是因为洛杉矶的广告位影响力较低，且广告位之间的重叠较少，因此影响力提供方更容易以更高的效率和准确度进行分配。**其次**，由于洛杉矶的广告位更多，且每个广告牌的平均影响力较低，‘RLS’可以利用这些广告位探索更多潜在的分配。因此，与纽约市观察到的趋势相似，‘RLS’能够有效减少洛杉矶的过度影响力。

重新审视 RQ1 和 RQ2。根据我们对这些案例的观察，我们回答此前提出的两个研究问题。(1) 当全局需求较低时，总遗憾来自过度影响力和未满足影响力。过度影响力也是总体遗憾的重要组成部分。在这种情况下，广告主选择至关重要。当总需求接近供给时，过度影响力会减少，而未满足影响力会增加。当总需求超过影响力提供方的供给时，总遗憾较高，其中很大一部分源于未满足影响力。(2) 当全局需求非常高或超过供给时，遗憾主要由未满足惩罚造成。在这种情况下，拥有大量中等需求的广告主效果最佳，因为这能在分配广告牌时提供更大的灵活性，同时漏掉单个大型广告主所产生的惩罚仍相对较小。因此，拥有大量影响力需求较低的广告主将有利于影响力提供方。

3) 效率研究： 在户外广告中，对于在一个城市拥有一千多个广告牌广告位、且每天有数位广告主前来寻求影响力的影响力提供方而言，效率十分重要。因此，我们在不同的 α 和 β取值下开展了效率研究。结果展示于图4 中，所使用的是纽约市和洛杉矶数据集。我们有三个主要观察结果。**首先**，在所提出的方法中，‘BG’的运行时间长于‘RG’和‘RLS’。这是因为在每次分配广告位的迭代中，‘BG’都会进行过度遗憾计算。另一方面，‘RG’和‘RLS’在遗憾计算过程中采用随机化，因此耗时更少。

‘随机’方法由于采用均匀随机分配，运行时间最短。**其次**，随着 α增加，所有算法的计算时间都会增加，这是因为广告主的需求增加，影响力提供方需要分配更多广告位。**第三**，随着 β 增加，广告主的个体需求增加。因此，所有算法的运行时间也会增加，并且‘RLS’在分配过程中需要探索更多搜索空间。

4) 其他参数研究： 我们讨论实验中使用的其他参数。**首先**，距离参数 γ 决定广告牌时段能够影响轨迹的范围。随着 γ增大，槽位的影响力也会增强，因为槽位能够影响更多轨迹。在实验中，我们将 γ 设为100米。**其次**，采样参数 ϵ决定随机样本集的大小。当 ϵ 增大时，样本集规模会缩小，‘RG’和‘RLS’的运行时间也会缩短。同时，解的质量也会下降。如果ϵ 值减小，样本集规模会增大，因此解的质量会更好。我们在所有实验中均将 $\epsilon = 0.01$设为默认值。 **第**三，惩罚比率 δ用于控制广告主不满意时，付款中惩罚金额所占的比例。当 δ 非常小时，影响力提供方会承受更大的遗憾。随着 δ 增大，如图5所示，所有算法的遗憾都会减少。我们将$\delta = 0.5$ 设为所有实验的默认设置。

VI. 结论

在本研究中，我们从影响力提供方的角度研究了户外广告中的标签特定遗憾最小化（TRMOA）问题。不同于传统的影响力最大化方法，我们的模型同时考虑了特定标签的影响力需求下未满足的遗憾和过度遗憾。我们证明了该问题是NP难问题，并提出了三种可扩展的启发式算法，每种算法都采用了自适应标签选择。在真实世界的纽约市和洛杉矶数据集上的实验表明，我们的方法能够显著降低总遗憾，同时保持较高的计算效率。研究结果还为了解供需动态和广告主需求分布如何影响遗憾行为提供了实用信息。本研究为现实户外广告系统中的遗憾感知分配策略奠定了基础。

参考文献

[1] P. Zhang、Z. Bao、Y. Li、G. Li、Y. Zhang 和 Z. Peng，“基于轨迹的影响力广告牌投放”，载于第24届ACM SIGKDD知识发现与数据挖掘国际会议论文集，2018年，第2748–2757页。[2] D. Ali、S. Banerjee 和 Y. Prasad，“使用剪枝次模图进行影响力广告牌版位选择”，载于高级数据挖掘与应用国际会议。施普普林格，2022年，第216–230页。[3] Y. Zhang、Y. Li、Z. Bao、S. Mo 和 P. Zhang，“优化户外广告的曝光次数”，载于第25届ACM SIGKDD知识发现与数据挖掘国际会议论文集，2019年，第1205–1215页。[4] D. Ali、S. Banerjee 和 Y. Prasad，“使用空间聚类和剪枝次模图进行影响力广告牌版位选择”，arXiv预印本 arXiv:2305.08949，2023年。[5] C. Aslay、F. B. L. V. Lakshmanan 和 W. Lu，“激励式社交广告中的收益最大化”，VLDB基金会论文集，第10卷，第11期，2017年。[6] J. Lai、T. Cheng 和 G. Lansley，“基于带地理标签的社交媒体数据改进定向户外广告”，GIS年刊，第23卷，第4期，第237–250页，2017年。[7] P. Zhang、Z. Bao、Y. Li、G. Li、Y. Zhang 和 Z. Peng，“迈向最优户外广告选址：当预算约束遇到移动轨迹”，ACM数据知识发现汇刊（TKDD），第14卷，第5期，第1–32页，2020年。[8] Y. Zhang、Y. Li、Z. Bao、B. Zheng 和 H. Jagadish，“最小化影响力提供方的遗憾”，载于2021年国际数据管理会议论文集，2021年，第2115–2127页。[9] P. Sharma、D. Ali 和 S. Banerjee，“社交媒体广告中的遗憾最小化”，载于2024年IEEE大数据国际会议（BigData）。IEEE，2024年，第473–478页。[10] D. Ali、A. K. Bhagat、S. Banerjee 和 Y. Prasad，“广告牌广告中的遗憾最小化高效算法（学生摘要）”，载于AAAI人工智能会议论文集，第37卷，第13期，2023年，第16148–16149页。[11] D. Ali、S. Banerjee 和 Y. Prasad，“区域影响力约束下广告牌广告中的遗憾最小化”，载于第39届ACM/SIGAPP应用计算研讨会论文集，2024年，第329–336页。[12] ——，“迈向无遗憾的广告牌版位分配”，国际数据科学与分析杂志，第1–25页，2024年。[13] ——，“区域影响力约束下广告牌广告中的遗憾最小化”，arXiv预印本 arXiv:2402.01294，2024年。[14] S. Wang、Z. Bao、J. S. Culpepper 和 G. Cong，“轨迹数据管理、分析与学习综述”，ACM计算调查（CSUR），第54卷，第2期，第1–36页，2021年。[15] Y. Zheng，“轨迹数据挖掘：概述”，ACM智能系统与技术汇刊（TIST），第6卷，第3期，第1–41页，2015年。[16] E. Tardos、D. Kempe 和 J. Kleinberg，“最大化社交网络中的影响力传播”，载于ACM SIGKDD知识发现与数据挖掘国际会议，2003年，第137–146页。[17] P. Banerjee、W. Chen 和 L. V. Lakshmanan，“基于效用驱动影响力扩散模型最大化社交网络中的福利”，载于2019年国际数据管理会议论文集，2019年，第1078–1095页。[18] S. Banerjee、M. Jenamani 和 D. K. Pratihar，“社交网络中的影响力最大化综述”，Knowledge and Information Systems，第62卷，第3417–3455页，2020年。[19] T. Guo、K. Feng、G. Cong 和 Z. Bao，“为交互式可视化探索高效选择地图上的地理空间数据”，载于2018年国际数据管理会议论文集，2018年，第567–582页。[20] D. Liu、D. Weng、Y. Li、J. Bao、Y. Zheng、H. Qu 和 Y. Wu，“SmartAdP：用于选择广告牌位置的大规模出租车轨迹可视分析”，IEEE可视化与计算机图形学汇刊，第23卷，第1期，第1–10页，2016年。[21] Facebook Business。（2025）每次帖子互动成本。Meta Platforms, Inc. 访问日期：2026年1月。 [在线]。可访问：https://en-gb.facebook.com/business/help/1514027528775302

[22] B. Haven，“营销的新关键指标：互动”，Marketing，第1–15页，2007年。

[23] H. Wang、Q. Meng、J. Fan、Y. Li、L. Cui、X. Zhao、C. Peng、G. Chen 和 X. Du，“社会影响力确实重要：信息流广告中的用户行为预测”，载于 *AAAI*人工智能会议论文集，第 34 卷，第 01 期，2020 年，第 246–253 页。[24] C. Aslay、W. L. F. Bonchi、A. Goyal 和 L. V. Lakshmanan，“病毒式营销遇上社交广告：以最小遗憾进行广告分配”， *VLDB*基金会论文集，第 8 卷，第 7 期，2015 年。[25] C. Aslay、W. Lu、F. Bonchi、A. Goyal 和 L. V. Lakshmanan，“病毒式营销遇上社交广告：以最小遗憾进行广告分配”， *arXiv* 预印本 *arXiv:1412.1462*，2014 年。[26] N. Barbieri、F. Bonchi 和 G. Manco，“主题感知的社会影响力传播模型”， *Knowledge and Information Systems*，第 37 卷，第 3 期，第 555–584 页，2013 年。[27] S. Bian 、 Q. Guo 、 S. Wang 和 J. X. Yu，“ 大规模社交网络上预算约束的影响力最大化高效算法”， *VLDB*基金会论文集，第 13 卷，第 9 期，第 1498–1510 页，2020 年。[28] Q. Guo、S. Wang、Z. Wei 和 M. Chen，“ 重新审视影响力最大化：通过收紧界限实现高效的反向可达集生成”，载于 *2020* 年 *ACM SIGMOD*数据管理国际会议论文集，2020 年，第 2167–2181 页。[29] B. Mirzasoleiman、A. Badanidiyuru、A. Karbasi、J. Vondrak 和 A. Krause，“比惰性贪心更懒惰”，2014 年。 [在线]。可访问：https://arxiv.org/abs/1409.7938