arXiv:2305.01470v1 [cs.LG] 2 May 2023

基于图上下文的随机上下文老虎机

Jittat Fakcharoenphol* Chayutpong Prompak†

2023年5月3日

摘要

我们将在线图预测问题自然地推广到随机上下文老虎机问题的一种形式，其中上下文是图中的顶点，而图的结构提供了有关上下文相似性的信息。更具体地说，给定一个图 $G = (V, E)$，其顶点集 V 表示带有 $unknown$ 顶点标签 y的上下文。在我们的随机上下文老虎机设定中，具有相同标签的顶点共享相同的奖励分布。图标签预测中实例难度的标准概念是割大小 f，其定义为端点具有不同标签的边的数量。对于线图和树，我们提出了一种算法，其遗憾界为 $\tilde{O}(T^{2/3}K^{1/3}f^{1/3})$，其中 K 是臂的数量。我们的算法依赖于 Zimmert 和 Seldin [在 AISTAT'19、JMLR'21]中提出的最优随机老虎机算法。当最佳臂优于其他臂时，遗憾可改进为 $\tilde{O}(\sqrt{KT \cdot f})$。后一种情况下的遗憾界与更一般情形下的其他最优上下文老虎机结果相当，但我们的算法易于分析、运行效率很高，并且不要求输入上下文序列满足独立同分布假设。该算法还可通过标准的随机生成树归约用于一般图。

1 引言

许多学习问题的侧信息可以表示为一个图。例如，由于社交网络用户的"类型"通常与其好友高度相关，社交网络服务提供商在完全了解社交网络图的情况下，可以利用图结构来帮助决定向用户展示哪则广告。它还可以利用图结构推断用户的各种属性，例如他们的生活方式、社会地位和营销价值。

用户类型预测可以建模如下。给定一个具有 n 个顶点的图 $G = (V, E)$、一个标签集$\mathcal{L}$，以及未知的顶点标签 $y : V \to \mathcal{L}$（表示某种顶点分类），由 Herbster、Pontil 和 Wainer [1]；Herbster 和 Pontil [2]；Herbster、Lever 和 Pontil [3]；以及 Cesa-Bianchi、Gentile 和 Vitale [4], 研究的在线标签预测问题按轮次进行：在每一轮中查询一个顶点，并由学习算法预测该顶点的标签。可用的图结构通过表明相邻顶点通常具有相同标签，为学习者提供线索。问题实例的复杂度度量是割大小 $f = f_G(y)$，其定义为边的数量 $(u, v) \in E$ 满足 $y(u) \neq y(v)$。我们注意到，虽然割大小的概念在图为树时非常自然，但在一般图中，割大小可能很大；例如，对于稠密图，割大小可以达到 $\Omega(n^2)$。以往的图预测结果保证了错误界，其对 f 呈线性依赖，对图的大小 n呈对数依赖。

在大多数情况下，我们无法获取顶点真实标签；更重要的是，我们的目标是做出决策，而不是弄清这些标签。此外，问题本身也可能存在不确定性。例如，同一位用户在多次被询问时，可能会给出不同的回答。

基于这一动机，我们将标签预测推广为一种随机上下文老虎机问题：上下文是图中的顶点，而图的结构提供了有关上下文相似性的信息。更具体地说，未知标签 y 定义了上下文组，同一上下文组中的顶点具有相同的奖励行为（由老虎机问题建模）。

*电子邮箱：jittat@gmail.com。泰国农业大学计算机工程系，泰国曼谷，邮编 10900。获泰国研究基金 RSA-6180074号资助。

†电子邮箱：chay.promp@gmail.com。泰国农业大学计算机工程系，泰国曼谷，邮编 10900。

我们关注奖励行为可表示为具有 K 个臂的随机多臂老虎机问题的情形。给定图 $G = (V, E)$（顶点标签 y未知），算法进行 T 轮次，并在每一轮 t 接收一个上下文 $u \in V$。算法必须选择一个臂 I_t，以最大化总奖励。奖励随机依赖于未知顶点标签 $y(u)$ 和臂 I。与标签预测问题不同，同一个顶点 u 可以被查询多次。我们只考虑 $[0,1]$ 奖励的情形。

本质上，我们处理的是线图。给定线图算法，Herbster、Lever 和 Pontil [3] 提出的一个简单归约，使我们能够处理具有 f 条割边的树，同时使用具有$2 \cdot f$ 条割边的线图。对于一般图，我们采用其随机生成树表示，例如 [4, 3, 5]中所述。我们在第4节陈述相关结果。

我们关于线图（和树）的主要结果如下。

定理 1. 考虑一个具有 n 个顶点和 f 条割边的线图或树。存在一种高效的老虎机算法，其遗憾界为 $\tilde{O}(T^{2/3} \cdot K^{1/3} \cdot f^{1/3})$，其中 T表示轮次数， K表示臂数。

我们的算法是一种非常简单的分治算法，它将上下文分层分解为多个层级。最初，我们的算法严重依赖 Audibert、Bubeck 和 Munos [6]提出的最佳臂识别算法 Successive Rejects，其遗憾界为 $\tilde{O}(T^{3/4} \cdot K^{1/4} \cdot f^{1/4})$。此处介绍的由 Zimmert 和 Seldin [7] 提出的、采用最优老虎机子程序的更简单且更优版本，是匿名审稿人2在 ALT'21 上建议的。

这仍然需要预先知道轮次总数 T。我们注意到， T 与 K 之间存在依赖关系，即 $K \leq T$，因此该界并未给出优于 $\Omega(\sqrt{KT})$ 的上界。上述界与分布无关，也就是说，它适用于任意奖励分布。

然而，在一种简单情形下，最佳臂的表现优于其他臂，Tsallis-INF 算法[~]具有更好的遗憾界，因此我们也得到了改进的 $O(\sqrt{KT \cdot f})$ 界。

该算法的一个显著特点是，它并不试图识别割边，因此不需要要求不同上下文组的行为有所不同。

此外，参数 f 实际上是"可观测"割大小：也就是说，如果某个顶点 u 没有作为上下文出现，我们可以自由地将其标签 $y(u)$ 设为任意值，而可观测割大小就是经过某种标签重命名后所能得到的最小割大小。

由于我们的工作非常简单（这是根据一位匿名审稿人的建议作出的决定），我们将相关工作的讨论和比较留到第 5 节。第 2 节给出了形式化设定。第 3 节介绍了线图的算法及其分析。我们简要讨论如何将线图的结果扩展到树和一般图，见第4节。

2 预备知识

2.1 设定

给定一个上下文集 $S = \{c_1, c_2, c_3, ..., c_N\}$ 和臂 1，2，.....， K。在每个时间步 $t = 1$，2，.....，玩家都会收到一个上下文 $c \in S$，并须选择一个臂 $I_t \in \{1, \ldots, K\}$，随后获得奖励 $X_{I_t,\ t} \in [0,1]$。目标是最大化奖励。对于上下文随机老虎机策略，奖励 $X_{I_t,\ t}$ 是在给定上下文 c 和臂 I_t 的条件下，从一个未知分布 $\nu^c_{I_t}$ 中独立于过去抽取的。本文考虑损失，其定义为 $\ell_{I_t,\ t} = 1 - X_{I_t,\ t}$，而非奖励。

该学习问题进行 T 轮次。对于每一轮次，算法将上下文映射到臂并最小化损失。经过 T 轮次后的总损失为 $\sum_{t=1}^{T} \ell_{I_t,\ t}$。

对于上下文 c 和臂 i，令 μ^c_i 为分布 ν^c_i的平均奖励。令 i^*_c 为上下文 c的最佳臂，即 $i^*_c = \arg\max_{i=1}^{K} \mu^c_i$。因此，对于上下文 c，最小期望损失记为 $loss^*_c$，其值为

$$1 - \mu^c_{i^*_c}.$$

因此，给定一系列上下文 s_1、s_2、.....、s_T，其最小期望损失记为 $loss^*$，则

$$loss^* = \sum_{t=1}^{T} loss^*_{s_t}.$$

此外，给定一张图 $G = (S,\ E)$，其顶点标签 y 未知：$S \to L$ 对于某个标签集L。我们有时将 S 中的上下文称为顶点。每个具有相同标签的上下文都共享相同的奖励分布；也就是说，对于每一对上下文 $c,\ c' \in S$，若 $y(c) = y(c')$，则对每个臂 i，分布 $v_i^c = v_i^{c'}$。请注意，我们无法访问这些标签 y。这些标签 y 将上下文划分为上下文组。当上下文 c 和 c' 共享相同标签时，我们称 c 和 c' 属于同一组。同一组中的上下文共享相同的最佳臂。

给定 图 $G = (S,\ E)$ 和顶点标签 y，若边 $(u,\ v) \in E$ 是割边，则 $y(u) \neq y(v)$。L et f 表示 n 条割边的数量。

在这项工作中，我们主要研究线图。利用标准技术，结果可推广到树和一般图（见第4节）。

2.2 多臂老虎机问题

由于特定上下文 c 下的底层问题是多臂老虎机的随机版本，本节将先回顾该问题及其求解算法。在此问题中，有K个臂；每个臂 i 都对应一个未知的概率分布 ν_i。在每一轮 $t = 1, \ldots, T$中，算法会选择一个臂 $I_t \in \{1, \ldots, K\}$，并获得从 ν_i中抽取的奖励 $X_{I_t,t}$。令 μ_i 为臂 i的均值；最佳臂 i^* 为 $\arg\max_i \mu_i$。令 $\mu^* = \mu_{i^*}$。该算法的伪遗憾为

$$T\mu^* - \mathbb{E}\left[\sum_{t=1}^{T} \mu_{I_t}\right].$$

当其他臂的期望奖励接近 μ_i时，学习会变得更加困难，因此我们定义$\Delta_i = \mu^* - \mu_i$ 来表示臂 i 与最佳臂的接近程度。与 [7], 中一样，我们假设存在唯一的最佳臂 i^*，其中 $\Delta_{i^*} = 0$，且对所有 $i \neq i^*$，$\Delta_i > 0$。令 $\Delta_{\min} = \min_{\Delta_i > 0} \Delta_i$。

我们的主要子程序是 [7], 提出的 Tsallis-INF 算法，该算法基于在线镜像下降，并采用幂为 $\alpha = 1/2$ 的 Tsallis熵正则化。该算法使用损失的无偏 ˆ ˆ 估计量 $\ell_{i,\ t}$ 用于臂 i 在时间 t 的损失，使得 $\mathbb{E}_{I_t}[\ell_{i,t}] = \ell_{i,\ t}$ 在内部成立。尽管 [7] 提出了这类估计量的两种替代方案，我们在此重点讨论基于重要性加权采样的方案，称为（IW）估计量，并仅陈述下文中针对这类估计量的结果。

定 理 2 ([7])。使用 $\alpha = 1/2$ （*IW*）估计量时，*Tsallis-INF* 的伪遗憾至多为 t

$$4\sqrt{KT} + 1,$$

其中，K 是臂的数量，T 是轮次的数量。此外，如果 $\Delta_{\min}$ 是常数，则伪遗憾可界定为

$$O(K \log T).$$

我们将 $\Delta_{\min}$ =Ω(1) 的情形称为简单情形。第 3 节将讨论这种情形。 2.

3 线图的算法及其分析

我们首先介绍一种用于线图的分治算法[1] 。对于 $1 \leq i < n$，顶点（上下文） c_i 与顶点 c_{i+1}相邻。为简单起见，假设 n 是 2 的幂，即令 $n = 2^L$。

该算法将上下文集分层分解为 $O(\log n)$ 层。对于每一层，我们将上下文集划分为若干个大小相同的集合，并将同一集合中的上下文视为来自同一个未知上下文组。在最高层，我们只有两个上下文集，每个集合的大小为$n/2$。随着层级降低，每个上下文集的大小呈指数级递减。我们首先在最高层运行多臂老虎机子程序。当针对上下文集 C' 的一个活跃子程序运行若干轮次后，我们终止该子程序，假设将上下文 C' 向下划分一层，并开始运行两个新的多臂老虎机子程序。

[1]我们再次指出，最初的算法比这里介绍的基于 Tsallis-INF 的算法更复杂，这一建议来自 ALT'21 的一位匿名审稿人。我们非常感谢这项建议。

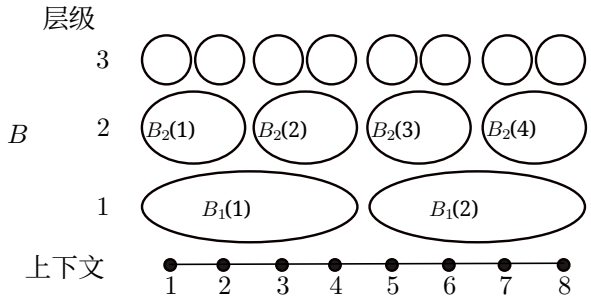


图 1：子程序和上下文。 $B_1(1)$ 负责上下文集 $\{1,2,3,4\}$。 $B_2(1)$ 和 $B_2(2)$ 是 $B_1(1)$ 的子节点。

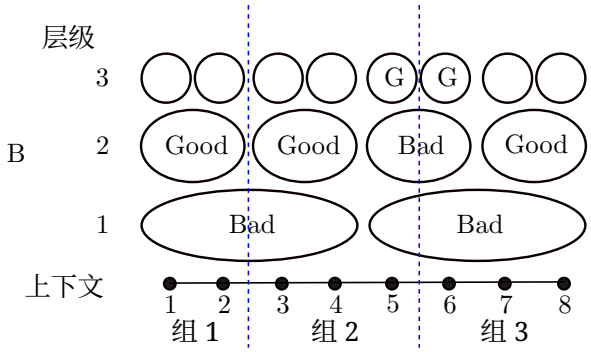


图 2：处于良好和糟糕情形下的子程序。

现在我们正式描述该算法。回想一下， $n = 2^L$。我们在 L 个层级上维护随机赌博机子程序，每个子程序都执行 Tsallis-INF 算法。对于层级 p，其中 $1 \le p \le L$，共有$n/2^{L-p}$ 个多臂老虎机子程序，分别记为 $B_p(1) \dots, B_p(n/2^{L-p})$。多臂老虎机子程序 $B_p(j)$ 本质上处理的是上下文

$$\{c_{2^{(L-p)(j-1)}}, \dots, c_{2^{(L-p)j}}\}$$

仿佛它们 来自同一组。我们还说，老虎机子程序 $B_p(j)$ 对上下文负责 ， ts

$$\{c_{2^{(L-p)(j-1)}}, \dots, c_{2^{(L-p)j}}\}.$$

对于层级 $1 \le p < L$，对于子程序 $B_p(j)$，我们称 $B_{p+1}(j_l)$ 和 $B_{p+1}(j_r)$ 为 $j_l = 2j - 1$ 和$j_r = 2j$的子节点 ，而 $j_l = 2j - 1$ 和$j_r = 2j$ 是 $B_p(j)$ 的子节点。请注意，它们分别负责上下文 $\{c_{2^{(L-p)(j-1)}}, \dots, c_{2^{(L-p)j}/2}\}$ 和 $\{c_{2^{(L-p)j}/2+1}, \dots, c_{2^{(L-p)j}}\}$。 （见图 1）。

有一个参数 D ，它定义了一个老虎机子程序在其负责的上下文上处理请求的轮次数。每个不在第 L 层的子程序 $B_p(j)$ 都会运行 Tsallis-INF 算法 D 轮，之后它会停用自身，并激活其子节点 $B_{p+1}(j_l)$ 和 $B_{p+1}(j_r)$。

3.1 线图的伪遗憾分析

我们主要分析线图的情形。一般情形（树和一般图）的分析可由标准归约得到（见第4节）。

对于每个老虎机子程序 $B_p(j)$，如果其负责的上下文来自同一个上下文组，我们就称该子程序处于 良好 情形。否则，我们称其处于不良 情形。（见图 2）。

令 $\mathcal{C}$ 表示某个老虎机子程序负责的所有上下文集所构成的集合。

我们首先界定处于不利情形时的多臂老虎机子程序数量。当负责上下文 $B_p(j)$ 的子图中包含割边时，多臂老虎机子程序 $B_p(j)$ 便处于不利情形。回忆一下，我们的分解包含 $O(\log n)$ 个层级；因此，一条割边可能出现在 $\mathcal{C}$ 中的 $O(\log n)$ 个上下文集中，并导致 $O(\log\ n)$ 个子程序处于不利情形。由于共有 f 条割边，处于不利情形的子程序数量为 $O(f \cdot \log n)$，从而得到以下引理。

引理 1。 处于不利情形的子程序数量至多为 $O(f \cdot \log n)$。

我们现在可以证明主定理（即定理 1 的另一种表述）。

定理 3. 若线图的割大小为 f，则伪遗憾至多为 $\tilde{O}(T^{2/3}K^{1/3}f^{1/3})$。

证明. 由于父子程序停用后仅有两个新的子程序会被激活，并且每个子程序在处理 D 个请求后便会停用，因此被激活的子程序数量至多为 $2T/D$。

根据引理 1，处于不良情形的子程序数量至多为 $O(f \cdot \log n)$。每个此类子程序最多会产生 D 的遗憾，总计为 $O(D \cdot f \cdot \log n)$。

根据定理 2，处于良好情形的每个子程序至多贡献

$$4\sqrt{KD} + 1,$$

因为它运行的轮次至多为 D。

结合这两个界，算法产生的遗憾至多为

$$O(D \cdot f \cdot \log n) + 2T/D \cdot O(\sqrt{KD}) = O(D \cdot f \log n + T\sqrt{K/D}).$$

取 $D = \tilde{\Theta}(T^{2/3}K^{1/3}/f^{2/3})$，则伪遗憾至多为

$$\tilde{O}(T^{2/3}K^{1/3}f^{1/3}),$$

如所述。 □

3.2 当最佳臂易于识别时

在本节中，我们假设 $\Delta_{\min}$ 是一个常数。在这种情况下，定理 2 的第二部分保证，在经过 D 轮次后，处于良好情形下的子程序的伪遗憾至多为 $O(K \log D)$。

采用与定理 3 相同的分析，我们得到伪遗憾可以界定为

$$O(D \cdot f \log n + (T/D) \cdot K \log D).$$

选择 $D = \Omega(\sqrt{TK/f})$ 后，伪遗憾变为

$$\tilde{O}(\sqrt{KT \cdot f}).$$

4 一般图

4.1 从树归约到线图

与图标签预测问题类似，Herbster、Lever 和 Pontil [3] 证明，当图是一棵树时，我们可以将该问题归约为图是线图（称为脊柱）的情形，代价是割大小增加至两倍。可以通过对该树进行深度优先搜索来构造脊柱图。我们陈述来自 [3] 第4节的以下引理。

引理 2 ([3])。**给定一棵树** (V, E) 及其顶点标签 y，存在一个路径图 $G = (V', E')$、一个映射 $g\colon V \to V'$，以及将顶点标签 y' 自然扩展到 V' 的方式，使得 G 在 y' 下的割大小至多为 T 在 y 下的割大小的两倍。

4.2 随机生成树方法

为处理一般图，我们直接采用随机生成树方法，如 Cesa-Bianchi、Gentile 和 Vitale [4]；Herbster、Lever 和 Pontil [3]；以及 Cesa-Bianchi、Gentile、Vitale 和 Zap-pella [5] 中所述。在这种情况下，学习问题的割大小参数变为随机生成树的期望割边数。

5 相关工作

本文的工作可与标准上下文老虎机中针对大量策略的研究结果进行比较。另有一些研究直接处理一组相关的老虎机问题，这些问题可以在线聚类，也可以划分为等价类。

5.1 与其他情境老虎机算法的比较

在考察老虎机的性能时，明确说明其性能所比较的集合非常有用。在情境老虎机设定中，对于给定的上下文集 $\mathcal{S}$，策略 $g: \mathcal{S} \to \{1, \ldots, K\}$ 是从上下文集到臂集的映射。设 Π 为可能的策略集合，且 $N = |\Pi|$。由于策略集合可能非常大，我们希望遗憾界关于 T 呈次线性，并关于 N 呈对数增长。注意，在图预测的背景下，如果割大小为 f，则可能的策略数量 N 可能大到 $\binom{n-1}{f} = \Omega(n^f)$。在这种设定下，对抗性设定中的情境老虎机算法，例如 Auer、Cesa-Bianchi、Freund 和 Schapire 提出的 Exp4 [8]，以及 Beygelzimer、Langford、Li、Reyzin 和 Schapire 提出的 Exp4.P [9],，其遗憾界为 $O(\sqrt{KT\log N}) = O(\sqrt{KT \cdot f \log n})$（[8] 的期望意义下，以及 [9] 的高概率意义下）。

由于我们更一般的结果（定理 1）的遗憾界依赖于 $T^{2/3}$，因此很可能是次优的。然而，当我们假设最佳臂远优于其他臂时（第 3.2 节），我们的遗憾界具有正确的阶，即 $\tilde{O}(\sqrt{KT \cdot f})$。

在一般情形下，对抗性老虎机算法 Exp4 [8]和 Exp4.P [9], 的表现优于本文提出的算法，但这些算法必须处理大小为 $N = \Omega(n^f)$ 的权重向量，这可能非常庞大。Beygelzimer 等人 [9] 还提出了算法 VE，该算法在独立同分布情形下可与具有有限 VC 维数 d 的策略竞争。就我们的情形而言，该算法同样效率低下，因为它必须针对大小为 $\Omega(\sqrt{T^f})$ 的策略集合运行，而 VC 维数至少为 f。

Langford 和 Zhang [10], 、Dudik 等人 [11], 以及 Agarwal 等人 [12] 等较新的研究，旨在降低运行时间和内存需求对 N 的依赖。他们不再保留庞大的权重向量，而是调用假设搜索预言机。Agarwal 等人 [12] 实现了最优遗憾界 $O(\sqrt{T\log(|\Pi|)})$，且只需调用 $O(\sqrt{T/\log(|\Pi|)})$ 次。其缺点是，这些方法都基于独立同分布假设，即假设每一轮的上下文和奖励均从固定分布中独立采样。虽然我们的工作采用随机模型，即给定上下文后奖励独立采样，但提供给算法的上下文可以由对手生成。

5.2 与老虎机问题中其他聚类结果的比较

已有研究探讨了线性老虎机的在线聚类，例如 [13, 14, 15, 16]。Cesa-Bianchi、Gentile 和 Zappella [13] 考虑了一个线性老虎机问题，其中社会关系由无向图 $G = (V, E)$建模，V 表示一组 n 用户，每位用户都有一个未知参数向量$u_i \in \mathbb{R}^d$。该图为参数 u_i提供了结构，即他们假设$\sum_{(i,j)\in E} \|u_i - u_j\|^2$相较于 $\sum_{i\in V} \|u_i\|^2$较小。学习过程按轮次进行。在每一轮 t，系统会呈现用户索引i_t和一组任意的上下文向量 $C_{i_t} = \{x_{t,1}, \ldots, x_{t,c_t}\}$，学习者必须选择一个动作 $\bar{x}_t \in C_{i_t}$，并获得奖励 $u_i^T \bar{x}_t$，其中还包含额外的次高斯噪声。Cesa-Bianchi 等人 [13] 维护一组 n 线性老虎机算法，并使用逆相关矩阵 M_t 在老虎机算法之间共享反馈。他们得到的遗憾界取决于 $\sqrt{nT}$ 和矩阵 M_t的对数行列式，该界可以是 $O(n)$。—— Gentile、Li 和 Zappella [14] 考虑了一个结构更明确的设置，其中用户可划分为 m 个未知簇，并且每一轮 t 的上下文向量 C_t均独立同分布地生成（其规模可以任意）。我们注意到，该设置与我们的工作密切相关，其中 $m = f + 1$。Gentile 等人 [14] 给出了一个遗憾界，该界取决于 $\sqrt{mT}$，并包含额外的 $O(n+m)$项；这些项关于 T为常数。Gentile 等人 [15] 最近的一项研究考虑了各种依赖数据的假设，以得到更紧的界，该界取决于$\sqrt{Tm}$和 $n\cdot$polylog(nT)。 ——

Maillard 和 Munos[17] 以及 Hong 等人 [18], 研究了潜在老虎机，其中上下文类型 $\mathcal{B}$ 被划分为 C 多个聚类 $\mathcal{C} = \{\mathcal{B}_c\}$，每个聚类都有已知的奖励分布。然而，学习者收到上下文类型 $b \in \mathcal{B}$时，并不知道包含它的聚类 $\mathcal{B}_c$ b。

这是一个难得多的问题。我们的设定提供了更多结构，即图，可用于指导顶点的聚类。不过，我们事先并不知道奖励分布。

5.3 其他相关工作

还有其他研究探讨了具有某些结构的上下文老虎机问题。例如，[19, 20],和 [21] 研究了上下文与臂之间的相似性。

图标签预测问题有许多扩展，例如 [1, 2, 3, 4, 22],和 [23]。

Alon 等人 [24] 考虑了反馈受反馈图约束时的多臂老虎机问题。

6 致谢

我们感谢COLT'19、ALT'22和IPL的审稿人提出的许多有益意见。尤其感谢ALT'22的审稿人，他们提出了在此呈现的一种更简单、更优的算法。两位作者均获泰国研究基金资助 [资助编号RSA-6180074]。

参考文献

[1] M. Herbster、M. Pontil、L. Wainer，《图上的在线学习》，载于：Machine Learning，《第二十二届国际会议论文集》（ICML 2005），德国波恩，2005年8月7日至11日，2005年，第305–312页。doi:10.1145/1102351.1102390。URL https://doi.org/10.1145/1102351.1102390

[2] M. Herbster, M. Pontil，使用感知机进行图上的预测，载于：神经信息处理系统进展第19卷，《第二十届神经信息处理系统年度会议论文集》，加拿大不列颠哥伦比亚省温哥华，2006年12月4日至7日，2006年，页577–584。网址 http://papers.nips.cc/paper/3100-prediction-on-a-graph-with-a-perceptron[3] M. Herbster, G. Lever, M. Pontil，大直径图上的在线预测，载于：神经信息处理系统进展第21卷，《第二十二届神经信息处理系统年度会议论文集》，加拿大不列颠哥伦比亚省温哥华，2008年12月8日至11日，2008年，页649–656。网址 http://papers.nips.cc/paper/3475-online-prediction-on-large-diameter-graphs[4] N. Cesa-Bianchi, C. Gentile, F. Vitale，标记树上的快速最优预测，载于：COLT 2009——第22届学习理论会议，加拿大魁北克省蒙特利尔，2009年6月18日至21日，2009年。网址 http://www.cs.mcgill.ca/%7Ecolt2009/papers/015.pdf#page=1[5] N. Cesa-Bianchi, C. Gentile, F. Vitale, G. Zappella，随机生成树与加权图预测，载于：第27届国际机器学习会议（ICML- 10）论文集，2010年6月21日至24日，以色列海法，2010年，页175–182。网址 http://www.icml2010.org/papers/298.pdf[6] J. Audibert, S. Bubeck, R. Munos，多臂老虎机中的最佳臂识别，载于：A. T. Kalai, M. Mohri（编），COLT 2010——第23届学习理论会议，以色列海法，2010年6月27日至29日，Omnipress，2010年，页41–53。网址 http://colt2010.haifa.il.ibm.com/papers/COLT2010proceedings.pdf#page=49[7] J. Zimmert, Y. Seldin，Tsallis-inf：一种适用于随机与对抗性老虎机的最优算法，《机器学习研究杂志》22 (28) (2021) 1–49。网址 http://jmlr.org/papers/v22/19-753.html[8] P. Auer, N. Cesa-Bianchi, Y. Freund, R. E. Schapire，非随机多臂老虎机问题，《SIAM计算杂志》32 (1) (2003) 48–77。doi:10.1137/S0097539701398375。网址 https://doi.org/10.1137/S0097539701398375

[9] A. Beygelzimer, J. Langford, L. Li, L. Reyzin, R. Schapire，具有监督学习保证的情境老虎机算法，载于：G. Gordon, D. Dunson, M. Dud´ık（编），第十四届人工智能与统计国际会议论文集，机器学习研究论文集第15卷，PMLR，美国佛罗里达州劳德代尔堡，2011年，页19–26。网址 http://proceedings.mlr.press/v15/beygelzimer11a.html[10] J. Langford, T. Zhang，情境多臂老虎机的周期贪心算法，载于：第二十届神经信息处理系统国际会议论文集，NIPS'07，Curran Associates Inc.，美国，2007年，页817–824。网址 http://dl.acm.org/citation.cfm?id=2981562.2981665[11] M. Dudik, D. Hsu, S. Kale, N. Karampatziakis, J. Langford, L. Reyzin, T. Zhang，情境老虎机的高效最优学习，载于：第二十七届人工智能不确定性会议论文集，UAI'11，AUAI出版社，阿灵顿，弗吉尼亚州，美国，2011年，页169–178。网址 http://dl.acm.org/citation.cfm?id=3020548.3020569[12] A. Agarwal, D. Hsu, S. Kale, J. Langford, L. Li, R. E. Schapire，驯服怪兽：一种快速简洁的情境老虎机算法，载于：第三十一届国际机器学习会议论文集——第32卷，ICML'14，JMLR.org，2014年，页II–1638–II–1646。网址 http://dl.acm.org/citation.cfm?id=3044805.3045075[13] N. Cesa-Bianchi, C. Gentile, G. Zappella，一群老虎机，载于：第二十六届神经信息处理系统国际会议论文集——第1卷，NIPS'13，Curran Associates Inc.，美国纽约州雷德胡克，2013年，页737–745。[14] C. Gentile, S. Li, G. Zappella，老虎机在线聚类，载于：E. P. Xing, T. Jebara（编），第三十一届国际机器学习会议论文集，机器学习研究论文集第32卷，PMLR，中国北京，2014年，页757–765。网址 https://proceedings.mlr.press/v32/gentile14.html[15] C. Gentile, S. Li, P. Kar, A. Karatzoglou, G. Zappella, E. Etrue，情境依赖的老虎机聚类，载于：D. Precup, Y. W. Teh（编），第三十四届国际机器学习会议论文集，机器学习研究论文集第70卷，PMLR，2017年，页1253–1262。网址 https://proceedings.mlr.press/v70/gentile17a.html[16] S. Li, W. Chen, S. Li, K.-S. Leung，老虎机在线聚类的改进算法，载于：第二十八届人工智能国际联合会议论文集，IJCAI-19，人工智能国际联合会议组织，2019年，页2923–2929。doi:10.24963/ijcai.2019/405。网址 https://doi.org/10.24963/ijcai.2019/405[17] O.-A. Maillard, S. Mannor，潜在老虎机，载于：E. P. Xing, T. Jebara（编），第三十一届国际机器学习会议论文集，机器学习研究论文集第32卷，PMLR，中国北京，2014年，页136–144。网址 https://proceedings.mlr.press/v32/maillard14.html[18]J. Hong, B. Kveton, M. Zaheer, Y. Chow, A. Ahmed, C. Boutilier，再探潜在老虎机，载于：H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, H. Lin（编），神经信息处理系统进展，第33卷，Curran Associates, Inc.，2020年，页13423–13433。网址 https://proceedings.neurips.cc/paper/2020/file/9b7c8d13e4b2f08895fb7bcead930b46-论文。pdf[19] R. Agrawal，《连续臂老虎机问题》，《SIAM控制与优化杂志》33（6）（1995）1926–1951。doi:10.1137/S0363012992237273。URL http://dx.doi.org/10.1137/S0363012992237273

[20] R. Kleinberg, A. Slivkins, E. Upfal, 《度量空间中的多臂老虎机》，载于：第四十届 ACM计算理论研讨会论文集，STOC '08，ACM，纽约州纽约，美国，2008年，第681–690页。doi:10.1145/1374376.1374475。URL http://doi.acm.org/10.1145/1374376.1374475[21] A. Slivkins, 《利用相似性信息的上下文老虎机》，机器学习研究期刊 15（1）（2014）2533–2568。[22] C. Gentile, M. Herbster, S. Pasteris, 《基于已知和未知图的网络化数据在线相似性预测》，载于：COLT 2013——第26届年度学习理论会议，2013年6月12日至14日，美国新泽西州普林斯顿大学，2013年，第662–695页。URL http://jmlr.org/proceedings/papers/v30/Gentile13.html[23] S. Pasteris, F. Vitale, C. Gentile, M. Herbster, 《关于相似性预测与成对聚类》，载于：算法学习理论，ALT 2018，2018年4月7日至9日，兰萨罗特岛，加那利群岛，西班牙，2018年，第654–681页。URL http://proceedings.mlr.press/v83/pasteris18a.html[24] N. Alon, N. Cesa-Bianchi, C. Gentile, S. Mannor, Y. Mansour, O. Shamir, 《具有图结构反馈的非随机多臂老虎机》，SIAM计算杂志 46（6）（2017年）1785–1826。doi: 10.1137/140989455。URL https://doi.org/10.1137/140989455