第三十四次AAAI人工智能会议（AAAI-20）。

**在随机多臂匪徒问题中实现公平性。**

# Vishakha Patil, Ganesh Ghalme, Vineet Nair, Y. Narahari

印度科学研究所，印度班加罗尔。

{patilv, ganeshg, vineet, narahari}@iisc.ac.in

### 摘要

我们研究了随机多臂强盗问题的一个有趣的变体，我们称之为FAIR-MAB问题，在这个问题中，除了最大化预期奖励总和的目标外，算法还需要确保在任何时候，每只手臂至少被拉动了预先指定的部分次数。我们研究了*学习*和*公平性之间的*相互作用，用一个预先指定的向量表示保证拉动的分数。我们设计了一个*公平性感知的遗憾*，我们称之为*r*-Regret，它考虑到了上述公平性约束，并以一种自然的方式扩展了传统的遗憾概念。我们的主要贡献是通过两个参数获得一类FAIR-MAB算法的全面特征：不公平性容忍度和用作黑盒的学习算法。对于这类算法，我们提供了一个公平性保证，无论学习算法的选择如何，它都会随着时间的推移而统一保持。此外，当学习算法是UCB1时，我们证明我们的al- githm在足够大的时间范围内实现了恒定的*r*-Regret。最后，我们用传统的后悔概念分析了*公平性的代价*。We conclude by experimen- tally validating our theoretical results.

# 1 介紹

多臂匪徒（MAB）问题是不确定环境下连续决策的经典框架。从Robbins(1952)的开创性工作开始，多年来，已经有大量的工作被开发出来，以解决这个问题的理论方面和实际应用问题；关于MAB问题的教科书阐述见(Bubeck和Cesa-Bianchi，2012；Lattimore和Szepesva´ri，2018；Slivkins，2019)。事实上，对MAB问题及其众多变体的研究仍然是在线学习和强化学习等多个领域的核心追求。在MAB设置中，每一轮决策者（一个在线算法）都面临着*k个*选择，这些选择对应着未知的（对算法而言）奖励分布。每一个选择被称为一个臂，当决策者拉动一个特定的臂时，她

从相应的（先验未知的）分布1中获得奖励。决策者的目标是使预期累积的报酬最大化。

通过一系列的手臂拉动，也就是说，如果这个过程重复了*T*轮，那么在每一轮中，决策者都会选择一个手臂，目标是最大化总的预期

奖励2。

文献中对MAB问题的几种变体进行了广泛的研究。多篇论文研究了带有附加约束的MAB问题，其中包括带有背包约束的匪徒（Badanidiyuru，Kleinberg和Slivkins，2013），带有预算约束的匪徒（Xia等，2015），睡觉的匪徒（Kleinberg，Niculescu-Mizil和Sharma，2010；Chatterjee等，2017）等。在本文中，我们考虑了FAIR-MAB，这是MAB问题的一个变体，除了最大化累积预期奖励，算法还需要确保均匀地（即在每一轮结束时）每只手臂被拉动至少一个预先确定的分数。这给算法带来了额外的约束，我们将其称为*公平性约束*，用向量*r*∈*Rk来表示*。

从形式上看，给定向量*r*的每个分量*ri*规格为臂*i*的*公平性配额*，在线算法必须确保对于所有时间步*长t*（即均匀），每个臂*i*在*t*轮中至少被拉动*ri t*次。在线算法的目标是最小化预期遗憾，同时满足每个臂的公平性要求。在这种情况下，预期的遗憾，我们称之为*r*-Regret，是根据最优*公平*策略计算的（见定义4）。我们注意到，这个问题的难点在于在每一轮结束时满足这些公平性约束，特别是当算法事先不知道时间范围时，也能保证公平性。值得注意的是，目前的工作为MAB问题的约束变体的长线工作做出了贡献（Badanidiyuru，Kleinberg和Slivkins，2013；Kleinberg，Niculescu-Mizil和Sharma，2010；Xia等，2015）。

|· ∫

1不拉的手臂不给任何奖励。

版权所有c

2020年，促进艺术发展协会。

2我们研究标准设置，其中T是不预先知道的。

智能（www.aaai.org）。reserved. 在线算法的所有权利。

上面描述的公平性约束自然而然地捕捉到了许多现实世界的环境，其中手臂的拉扯对应于在具有特定配额（quotas）的代理之间分配资源。确保每个人都能得到一定的小规模分配保证的目标，有时与任何学习算法的分类目标--效率最大化的目标相冲突。然而，在许多应用中，为了保证公平性，分配规则必须考虑这样的约束。对可用资源的最小权利保证了个人的特权。为了具体化，我们接下来介绍一个激励性的例子。

美国住房和城市发展部最近起诉Facebook从事住房歧视，根据性别、种族、宗教等属性定向投放广告，这些都是美国受保护的类别。

法3。脸书的算法决定应该向某一特定用户显示哪些广告，但由于其试图实现的目标，无意中最终导致犯罪。例如，如果算法了解到它可以通过向更多的男性相比于女性显示广告来获得更多的收入，那么它最终会对女性产生歧视。所提出的FAIR-MAB模型保证了男性和女性至少在广告展示总数的一个预先指定的分数上显示广告，从而维护了机会平等 的基本权利。在某种程度上，最小分数保证也为所选择的分配规则提供了一种道德上的正当性，在无知的面纱（Rawls，1971）下评价其公平性，其中，分配规则被认为是 个人之间的假设协议，不知道他们可能具有的自然能力和环境上的进步和偏见，即在原始位置上的社会约定分配（详细讨论请参考（Freeman，2019；Heidari等人，2018））。

这项工作中的公平性模型自然而然地捕捉到了许多资源分配的情况，例如搜索引擎上的赞助广告，每个广告商应该保证一定比例的拉动，以避免广告空间的垄断；人群外包，每个人群工作者被保证一部分任务，以诱导参与；以及无线通信环境，接收器必须确保对每个发送者的最低服务质量。(Li, Liu, and Ji 2019)的工作包含了对这些应用的详细讨论。我们在第7节讨论了其他关于公平性的相关工作。

**我们的贡献。**我们首先在第2节中定义了FAIR -MAB的概率。任何FAIR-MAB算法都是基于两个标准来评估的：它所提供的公平性保证和它的*r*-Regret。我们所考虑的公平性概念要求在每一轮之后满足公平性约束条件，而*r*-Regret概念则是公平性约束条件的自然延伸。

遗憾的概念，它被定义为对一个人的遗憾。

*T*对算法来说是未知的。我们注意到，即使地平线*T*是已知的，直观的方法是多次拉动每个手臂以满足其公平性要求，但并不能保证每轮结束时的公平性。读者可以参考附录中的扩展版本。

sion4.作为我们的主要贡献，在第3节中，我们定义了一类FAIR-MAB算法，称为FAIR-LEARN，由两个参数Char- acterized：不公平性容忍度和用作黑盒的学习算法。我们证明FAIR-LEARN中的任何算法在任何时间*t都能*满足公平性约束，因此FAIR-LEARN的公平性保证在一段时间内都是统一的，与学习算法的选择无关。我们在这里注意到，我们的元算法FAIR-LEARN，允许任何MAB算法作为黑盒插入。这个简单而优雅的框架可以实现在任何现有的MAB算法之上，以确保公平性与可量化的遗憾损失。我们算法的实际适用性是本工作的一个显著特点。

当学习算法为UCB1时，我们证明了FAIR-UCB算法的次对数*r*-Regret边界。此外，对于足够大的*T，*我们看到FAIR- UCB会产生恒定的*r*-Regret。然后，我们在第4节中评估了FAIR-MAB的公平性成本与*传统的*no- tion的遗憾。在第6节中，我们通过提供去尾实验结果来验证我们的理论瓜拉尼-tees来结束。特别是，我们比较了FAIR-UCB与LFG算法提出的性能(Li，Liu，and Ji 2019)，这是最接近本论文的工作。我们在这里指出，我们获得了一个更强的公平性保证，它在任何时候都是有效的，不像LFG的渐进公平性保证。我们还证明了一个更好的遗憾约束，它对问题实例参数有一定的依赖性。第7节提供了一个详细的比较。

# 2 该模型

在本节中，我们正式定义了FAIR-MAB问题，公平性的概念，以及本工作中使用的*r*-regret概念。

## FAIR-MAB问题

FAIR-MAB问题 的 一个实例 是一个 元组。

(*T，*[*k*]*，*(*μi*)*i*∈[*k*]*，*(*ri*)*i*∈[*k*])，其中*T* 为时间范围。

[*k*] = 1*，*2*，...，k*是臂的集合，*μi* [0*，*1] 代表与臂*i*相关的奖励分布*i*的平均值，(*ri*)*i* [*k*] 代表公平性约束向量。在FAIR-MAB环境中，公平性约束以向量*r*=(r1*，*r2*，...，r*Σ*k*)的形式对算法进行外显，其中*ri*∈[0*，1/k*)，对于所有*i*∈[*k*]。

∈

D

{ }∈

∈

因此

*i*∈[*k*]*ri<1*，且ri*表示*mini-。

∈

满足公平性约束的最优政策。统一的

我们所寻求的形式时间公平保证，确保公平性

即使在*地平线不可知的情况下*，那也是在时间地平线上的

3[https://www.technologyreview.com/s/613274/facebook-](http://www.technologyreview.com/s/613274/facebook-)

algorithm-discriminates-ai-bias/。

臂*i*[*k*]在*T*中被拉动的最大倍数

轮，对于任何*T* 。我们认为*ri*[0*，1/k*)是一致的。

与*相称性*的概念，其中，保证任何。

4 扩充版可在https://arxiv.org/abs/1907。

10516

臂大于其比例分数，即

*1/k*，本身就*不公平*。

在每一轮*t*中，FAIR-MAB算法都会拉出一条手臂。

如果一个算法的后悔率为零，我们称之为最优算法。很容易看出，上述遗憾的概念并不能完全量化FAIR-MAB算法的性能。

*it*[*k*]并领取奖励*Xi*。

∈

*t*

奖励分配是

∼ *Dit* 。我们假设

rithm，因为这里的最优政策没有考虑到公平性约束。另外，请注意，传统的后悔

每个臂*i*的*Bernoulli*(*μi*)

∈

[*k*].这个假设在不损失通用性的情况下成立，因为人们可以使用(Agrawal和Goyal 2012)中提供的扩展，将支持在[0,1]上的广义分布的MAB问题简化为Bernoulli re-wards的MAB问题。请注意，*μ* = (*μ*1*，μ*2*，...，μk*)的真实值对算法来说是*未知的*。在本文中，我们在不失一般性的前提下假设*μ*1 *> μ2 >* ... *...> μk，臂1*称为最优*臂。*接下来，我们对本文中使用的公平性和后悔的概念进行形式化。

## 公平的概念

在 FAIR-MAB 环境中，公平性感知的运筹帷幄可以是 *O*(*T* )（详见第 4 节）。因此，我们首先说明我们认为作为基线的公平性感知运营政策。

**观察1.**FAIR-MAB算法A是最优的，如果

A满足以下条件：如果|*riT* ∫- *α > 0*，则Ni,*T = 。*

|对所有*i*ƒ=1，*riT*∫- *α*，否则*Ni,T*=0。

从观察1我们可以得到，一个知道*μ*值的最优FAIR-MAB算法必须准确地下次最优臂*ri T α*次，才能满足公平性约束，并在其余回合即*T* - *i*ƒ=1|*ri* -*T* ∫+(*k*-1)*α*回合下最优臂(臂1)。遗憾

Σ

| · ∫−

让*N*

*i,t*

表示手臂*i*在*t*中被拉动的次数。

的算法与这样一个最优策略进行比较，该算法的

满足FAIR-MAB环境下的公平性约束。

轮。我们首先提出（Li，Liu，and Ji 2019）提出的公平性定义，然后提出本文考虑的更强的公平性概念。

A

≥

**Deﬁnition 1.**(Li，Liu，and Ji 2019)A FAIR-MAB algorithm

A *t*

**定义4.**给定一个公平性约束向量*r*=(r1*，*r2*，...，rk*)和*不公平性容忍度α*0，FAIR-MAB算法的公平性感知*r*-Regret被定义为：。

称之为（渐近）公平，如果lim inf

→∞

对于所有*i*∈[*k*]，*Ni,t* Σ≤0。

*r*

AΣ*ri* -

RA(*T* ) = Σ

Δ*i* -

.E[*Ni,T*] - max.0*,* |*ri* - *T* ∫- *α*ΣΣ

*t*

我们把上述公平的概念称为*渐进式公平*。

*i*∈[*k*]

(2)

*性*。我们现在定义一个更强的公平性概念，它适用于所有回合，并以系统中允许的*不公平性容忍度*为参数，用con- stant *α*≥0表示。

**定义2.**给定一个*不公平容忍度α*≥0，如果对所有*t*≤*T*且对所有臂*i*∈[*k*]而言，|*rit*∫- *Ni,t*≤*α，则*称FAIR- MAB算法A是*α*-公平的。

特别是，如果上述保证对*α*=0成立，那么我们称FAIR-MAB算法为*公平*。请注意，我们的公平性保证在时间跨度上均匀成立，并且对于算法的任何手臂拉（*it*）*t*序列都是如此。因此，它比(Li, Liu, and Ji 2019)中的保证要强得多，后者只保证渐进公平性(Deﬁnition 1)。注意，对于任何给定常数*α* 0，*α*公平性（Deﬁnition 2）意味着渐近公平性。

≤

≥

## 遗憾的概念

在MAB环境下，最优策略是在每一轮中拉动最优手臂的策略。MAB算法的遗憾被定义为最优策略的累积奖赏与算法的累积奖赏之差。

**定义3.**一个MAB算法在*T*轮之后的预期遗憾被定义为。

A

Σ

Δ*i* - E[*Ni,T*](1)

RA(*T* ) =

上述定义中的max(0*, ri T α*)是指最优算法为满足公平性约束而拉动*第i*臂的次数。另外，一个不公平*的*算法的*r*-Regret也可能是负数，但这是一个不可行的解。一个学习算法在超过*riT α*轮的情况下拉动一个次优臂*i*，每多拉动一次就会产生Δ*i* = *μ*1 *μi的*re-regret。在设计FAIR- MAB问题的最优算法时，技术上的困难在于对次优臂*i*=1的数量*Ni,T riT*的约束：在任何时间*T，为了使*算法公平，我们希望*Ni,T riT*至少是*α*，而为了使遗憾最小化，我们希望*Ni,TriT*接近*α*。

# 3 公平的MAB算法3 框架

−

|∫−

- − |∫

- − |∫

| · ∫−

| - | − ∫ƒ

在本节中，我们提供了我们提出的一类FAIR-MAB算法的框架。我们的元算法FAIR- LEARN在算法1中给出。 这项工作的关键结果是下面的定理，它保证FAIR- LEARN是*α*-公平的（见Deﬁnition 2），与学习算法Learn( )的选择无关。 需要注意的是，对于FAIR-LEARN的任何手臂拉动序列，公平性保证在时间跨度上均匀成立。

-·

**定理1．***对于给定的α*≥0*，对于任何给定的公平性约束向量r* = (r1*，*r2*，...，rk*)*，其中ri*∈[0*，*1 )*为*

*所有i*∈[*k*]*，*FAIR-LEARN*为α。*

*k*

*-公平，无论选择如何*

*i*∈[*k*]

其中Δ*i*＝*μ*1 -*μi*，*Ni,T*表示臂*i*∈[*k*]在*T*轮中被A拉动的次数。

*的学习算法*Learn（-*）。*

定理1的证明在第5节中给出。当*α*=0时，上述定理中的瓜拉尼也是成立的，因而

算**法1：**FAIR-LEARN(公平学习)

**Theorem 4.**FAIR-UCB*的预期遗憾由以下公式给出。*

**输入。**[*k*]*，*(*ri*)*i*∈[*k*]*，α。* 0, Learn( )

≥·

### 1初始化。

**2** *Ni,*0=0，对所有*i*∈[*k*]而言。

Σ

Σ

R(*T* ) ≤

*i*∈*S*(*T* )

(*ri* - *T* - *α*) - Δ*i* + 。

*i S*(*T* )

∈

*i*=1

8 ln *T/*Δ*i*

**3** 对所有*i*[*k*]来说，*Si,*0=0，其中*Si,t*=*第i*臂在*t*轮中的总奖励。

∈

**4对于***t*=1*，*2*，...***做**

+ (1+*π*2/3) - Δ*i*

*i*∈[*k*]

Σ

.,

(其中*S*(*T* )= {*i* | (*ri* - *T* - *α*) *< 8* ln T/*Δ2*}。)

*i*

5Deﬁne：*A*(*t*)=。

**6**拉臂

⎧⎨ *i*∈[*k*]

*i* . *ri* - (*t* - 1) - *Ni,t*-1 *> α.*

定理4捕捉到了后悔的明确权衡。

*它*=

氩气

...*ri -* (*t* - 1) - *Ni,t*-1Σ 如果*A*(*t*) ∅

和公平性，以*不公平性容忍度为*参数

*α*.如果*S*(*T* )=我们有，遗憾是*O*(ln *T* )。这个

∅

2

⎩↪Sm\_23A9↩Learn(*Nt, St*) 否则

意味着*对*所有的i1*来说* 如果*α>riT -* 8 ln T/*Δi*，则

**7**更新参数*Nt*和*St*

的遗憾是 *O*(ln *T* )。但是，如果 *S*(*T* ) ƒ= ∅，那么对于每个

*i*∈*S*(*T* )，就会产生一个相当于*riT* -*α的*额外遗憾，在

**8结束**

*α*=0的FAIR-LEARN是*公平的*。 特别是，当学习算法Learn( ) = UCB1时，我们称这个算法为FAIR- UCB。我们提供FAIR- UCB的*r*-Regret边界。

-·

**Theorem 2.**FAIR-UCB*的r-Regret由以下公式给出。*

在这种情况下，遗憾是 *O*(*T* )。我们用第6节的模拟来补充这些结果。

# 5 理论结果5 的证明

### 定理1的证明

在每一轮*t*之后（以及在*t*+1轮之前），我们考虑的是

*k*+1组，M1*,t,*M2,*t,...,Mk,t*和*St*，如下所述。

*π*2

*r*

.

RFAIR-UCB(*T*)≤ 1+3

* Σ Δ*i*
  + arm *i*∈*Mj,t* ↪So\_21D0↩⇒*α*+(*k*-*j*)

(*k*-*j*+1) ，∀*j*∈[*k*]。

*k*

≤ *rit* - *Ni,t < α +。*

+

Σ

*i*∈*S*(*T* )

*i*ƒ=1

Δ*i* -

*i*∈[*k*]

8 ln *T*

*α* +1

*α*+*k*-1

*k*

α + 2

*k*

α + 1

*k*

*α*

Σ

.

|  |  |
| --- | --- |
|  | |
| M1*,t* |  |
|  | |
| *Mk*-1*,t* |  |
| *Mk,t* |
| *St* |  |

2

Δ

*i*

* .*ri* - *T* - *α*ΣΣ

*k*

* + arm *i*∈*St*↪So\_21D0↩⇒*rit* - *Ni,t < α。*

*其中S*(*T* )= ,*i*∈[*k*].*r* - *T* - *α <。*  ,

*i*

2

*.特别是*

*对于足够大的T，Rr*

M1*,t*

Δ2

. 8 ln *Ti*

.

Σ

(*T* ) ≤

1+ *π* - ΣΔ *.*

FAIR-UCB

*3i*∈[*k*]*i*

定理2在第5节中证明。观察到，如果*S*(*T* )=

*Mk,t*

*Mk*-1*,t*

ƒ

FAIR-UCB 的 *r*-Regret 是亚对数，如果 *S*(*T* )

∅

=则*r*-后悔是常数。我们在定理3 5中证明FAIR-UCB的无分布后悔。

∅

**Th**√**eorem 3.** FAIR-UCB*的无分布r-Regret为*

图1：武器的分割

*O*( *T* ln *T* )*。*

在本节的最后，我们指出，由于公平竞争法是一个很好的法律框架，因此，我们必须要有一个公平的法律框架。

设*Vj,t*=！*jMA,t*，对所有*j*∈[*k*]。那么，下面是

ness guarantee of FAIR-LEARN holds without any loss in LEARN( ), this framework can easily make operational in practice.

*A*=1

-·

# 公平的4 成本

到目前为止，我们的遗憾保证一直是以*r*-Regret为标准，但现在我们以传统的遗憾概念来评估*公平性的代价*。特别是，我们用*不公平性容忍度*来展示传统遗憾和公平性之间的权衡。

5定理3和4的证明可以在本文的扩展版中找到。

定理保证了算法的公平性，是证明的核心。该定理的证明由该定理的证明可直接得出。

**Lemma 1.***对于t*≥1*，我们有*

1. *Vk,t* !*St* = [*k*]
2. *|对所有j*∈[*k*]*来说，Vj,t*|≤*j。*

Lemma 1中的条件1保证了在任何时间*t* 1，*k* + 1个集合M1*,t,*M2*,t,...,Mk,t,St*形成了臂集[*k*]的分区。因此，算法在第(*t*＋1)-轮拉出的臂来自这些*k*＋1集中的一个。作为Lemma 1的证明的一部分，在观察2中，我们表明，如果*it*+1是在(*t*+1)-轮被拉动的手臂，那么在*t*+1之后

≥

(*k*-*j*+1)

轮回*it*+1∈*Mk,t*+1 !*St*+1.同样在观察3中*werit* - *Ni,t < α +。*那么在第*t*+1轮之后，我们有

*k*

证明如果第(*t*+1)-轮中没有拉动臂*i*∈*Mj,t，*那么在*t*+1轮后，臂*i*∈*Mj*-1*,t*+1 !*Mj,t*+1

∈

*ri*(*t*+1) - *N*

*i,t*+1

= *rit* - *N*

*i,t*

+ *ri*

对于所有*j*[2*，k*]。我们注意到Lemma中的两个条件是

1在第一轮之后为真，那么观察结果2也为真。

*< α + (*k *-* j + *1*)+r

*淇*

和3共同确保这些条件在以下情况下仍然有效

(*k*-*j*+1) 1

因此，*对于*所有的t 1，所有的武器i [*k]都*满足rit *Ni,t < α +1*，*这 Ni,t α .*特别是，*对于所有的*t 1，*对于* 所有的i [k]，我们有rit Ni*,t* α*，*由定义2证明FAIR-LEARN是α-*fair* 。Q

|∫−≤ ≥∈

≥ |∫−≤

∈−

*< α +*

= *α* +

+

*k k*

(*k* - (*j* - 1) + 1)

*k*

*Lemma 1的证明*：我们从两个互补的

和*rit* - *Ni,t* + *ri* ≥ *α* + *k*-*j* + *ri* ≥ *α* + *k*-*j。*

观察，然后用归纳法来证明该定律。

**观察2.**让*i*成为FAIR-LEARN拉动的手臂，在

*k k*

=⇒*i*∈*Mj*-1*,t*+1 !*Mj,t*+1

绕*t*+1。

1. 如果*i*∈*St*，则*i*∈*St*+1
2. 如果对一些*j*∈[*k*]，*i*∈*Mj,t，*则*i*∈*Mk,t*+1 !*St*+1

通过上述观察，我们用归纳法完成了对该定理的证明。

感应基本情况（*t*=1）。让i1成为在*t*=时被拉动的手臂。

1.那么

*证明。情况1：i*∈*S*

在*t*+1回合，我们有

*t* =⇒ *rit* - *Ni,t*

*< α.*然后在

ri1 *t* - Ni1 *,*1 = ri1 - 1 *< 0* ≤ α。

=⇒i1 ∈S1

对于所有*i* ƒ= *i* ，我们有*r t* - *N*= *r < 1* ≤ α *+ 1* =⇒ 。

| | ∈−

∈! !||≤−

1

*i*

*i,*1

*i*

*k*

*k*

*ri*(*t*+1) - *Ni,t*+1 = *rit* + *ri* - *Ni,t* - 1；

*< α -* (1 - ri)

*< α(*因为1 - ri *> 0)*

=⇒*i*∈*St*+1

*i* S1 *Mk,*1.因此，*Vk,* 1 S1=[*k*]，*Vk,*1 *k* 1，对所有*j*[*k 1*]，*Vj,*1=0。因此，外稃的条件（1）和（2）成立。

归纳步骤。假设在*t*轮之后条件成立，我们证明在*t*+1轮之后条件成立。

*案例1：it*+1∈*St*.由观察2可知*it*+1∈。

*情况2：*对于一些*j*∈[*k*]，*i*∈*Mj,t*=⇒*rit* - *N。*

*i,t <*

*St*+1.由观察3可知，对于任何臂*i*ƒ=?

*it*+1，*i*∈*St*+1 !*Mk,t*+1。因此，*Vk,t+*1 !*St*+1 = [*k*]．

| | ∈ − | | ≤−

*α*+（*k*-*j*+1） 。那么在*t*+1轮之后，我们有

*k*

*ri*(*t*+1) - *Ni,t*+1 = *rit* + *ri* - *Ni,t* - 1；

(*k*-*j*+1)

对于所有*j* [*k* 1]，*Vj,t+* 1=0，且*Vk,t*+*1k* 1。因此，在第*t*+1轮之后，外稃中的条件（1）和（2）成立。

*情况2：it*+1∈*Ma,t*，对于一些*a*∈[*k*]。

*< α +*- (1 - ri) *k (*)

j1

*it+*1∈*Ma,t* =⇒ *it*+1∈*Va,t*

*< α -* k *+ k +* ri=⇒*Vj,t*|=0 对于所有*j*∈[1*,a* - 1]，如果*a > 1*的话

*< α +* ri

1

*< α +*

*k*

(由于*ri*

*< 1 )*

根据观察2，我们知道*it*+1∈*St*+1 !*Mk,t+*1，根据观察3，我们推断*Vj*-1*,t*+1 = *Vj,t* ｛*it*+1｝。

=⇒*i*∈*Mk.t*+1 !*St*+1

*k*

**意见3.** 让*i*[*k*]是任何在时间上没有被拉动的手臂。

∈

t + 1.

1. 如果*i*∈*St*，则*i*∈*St*+1 !*Mk,t*+1
2. 如果对*j*∈[2*，k*]，*i*∈*Mj,t，*则*i*∈*Mj*-1*,t*+1 !*Mj,t*+1

对于所有*j*∈[2*，k*]。另外：

|*Vj,t* {*it*+1}| ≤ *j* - 1 对于所有*j*∈[*a，k*]

=⇒*Vj,t*+1|≤*j* 对所有*j*∈[*k*]来说

另外，*Vk,t*+*1St*+1=[*k*]。因此，在第*t*+1轮之后，条件(1)和(2)的外稃成立。

!

### 定理2的证明

*证明。案例1：iSt*

∈

在*t*+1回合，我们有

=⇒*rit* - *N*

*i,t*

*< α.*然后在

FAIR-UCB的后悔分析建立在UCB1的后悔分析基础上，我们在ex- tended版本的附录中给出了UCB1的后悔分析。在附录A中，我们还介绍了nota-。

*ri*(*t*+1) - *Ni,t+*1 = *rit* - *Ni,t* + *ri*(*Ni,t*+1 = *Ni,t*)

1

在本证明中使用的 tions。臂*i*的平均值的UCB1估计表示为*μ*¯*i*(*t*)=*μ*ˆ*i,Ni,t*-1(*t* - 1)+*ct,Ni,t*-1 。

*< α +* ri

*k*

*< α +*

*k*

(由于*ri*

*< 1 )*

其中*μ*ˆ*i,N*

臂的

.

*i,t*-1

(*t* - 1)是平均数的经验估计。

=⇒*i*∈*St*+1 !*Mk,t*+1

*情况2：*对于一些*j*∈[2*，k*]=⇒*α*+*k*-*j*≤的*情况，i*∈*Mj,t。*

*k*

当它在*t*-1轮中被拉动*Ni,t*-1次时，*i。*

*c*=2 ln *t*是手臂*i*的确定区间。

*t,Ni,t*-1

*Ni,t*-1

在第*t*轮时，与UCB1算法的分析类似。

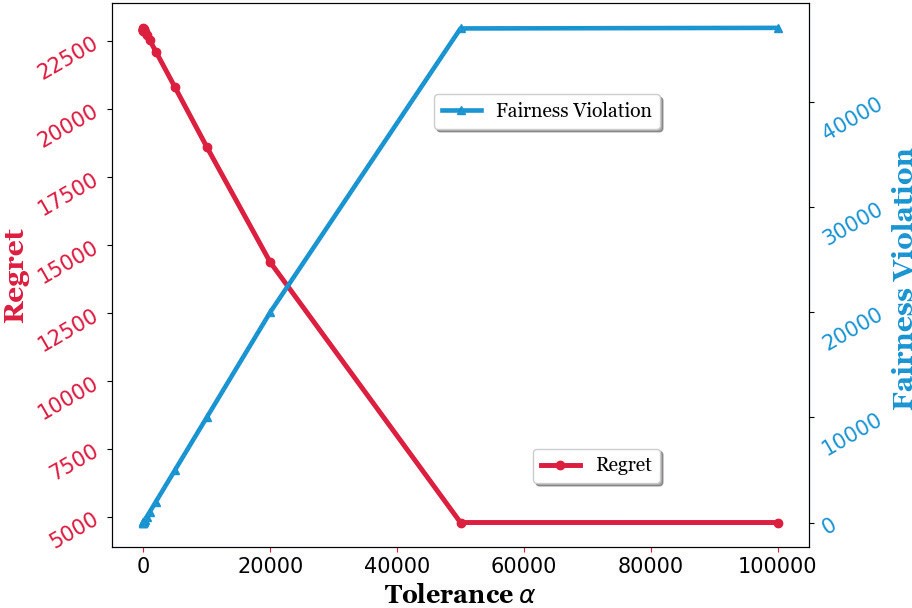
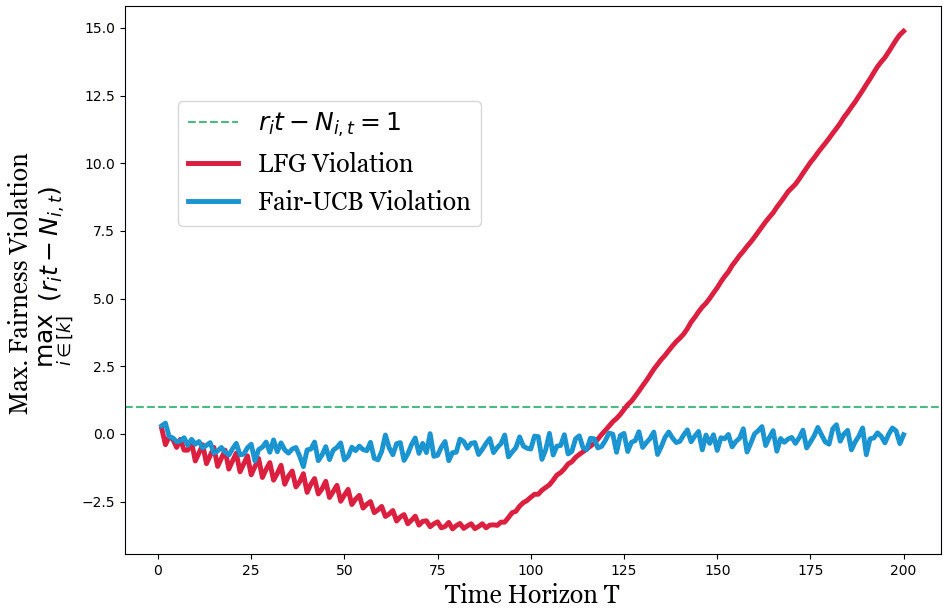
 

图2：公平的成本

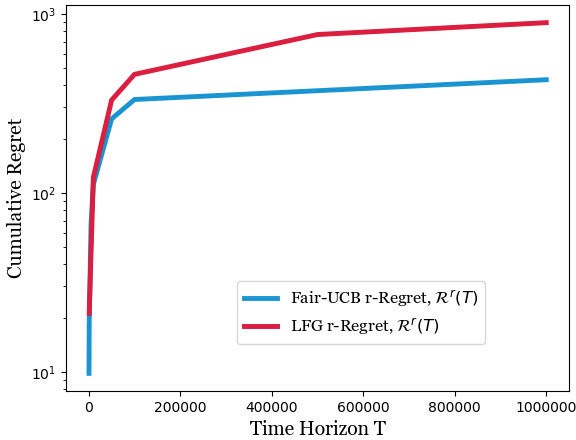
图4：公平性保证

假设*S*(*T* )=*i*∈[*k*] *ri* - *T* - *α < 8* ln T 。那么从上面讨论的两种情况来看，我们可以得出结论，即

*i*

Δ2

,.,

*π*2

*r*

.

RFAIR-UCB(*T*)*≤*1+3。

Σ

Δ*i*

Σ

* Σ

*i*∈[*k*]

Δ2

*i*

+Δ*i*

*i*∈*S*(*T* )*,i*ƒ=1

* ∫ 8 ln *T -* .

*i*

* *T - α*Σ,

*r*

因此，R

FAIR-UCB

(*T* )=*O*(Σ

*i*=1

ln *T* ）。)Q

图3：*r*-Regret保证

Δ*i*

我们对次优臂被拉动的预期次数进行上界。我们通过考虑两种情况来实现这一点，这两种情况取决于满足公平性约束的次优臂被拉动的次数。

情况1：设*i*ƒ= 1，*ri* - *T* - *α* ≥ 8 ln *T* 。那么

*i*

Δ2

*T*

. Σ

# 6 实验结果

在本节中，我们展示了模拟的结果，估价我们的理论结论。首先，我们通过显示后悔和公平性与*不公平性容忍度α*之间的权衡来表示公平性的成本。 其次，我们通过与(Li, Liu, and Ji 2019)的算法(称为Learning with Fairness Guarantee(LFG))作为基线进行比较，来评估我们算法在*r*-Regret和公平性保证方面的性能。请注意，在图3中，累积遗憾是以对数比例绘制的。的ra-。

E[*Ni,T*]*≤ri - T - α* +。

*t*=1

*≤*. *ri - T - α*Σ

∞

*t*

*t*

*{it* = *i, Ni,t*-1 *≥ ri - T - α}。*

本文的扩展版中讨论了实例参数选择背后的tionale。

### 取舍：公平与遗憾。

Σ Σ Σ。



,我们考虑以下FAIR-MAB实例：*k*=10。

+ *μ*ˆ1,s1 (*t*)+ *ct,*s1 *≤ μ*ˆ1*,si* (*t*)+ *ct,si)*

*μ* = 0*.*8，且*μ* = *μ* - Δ ，其中Δ = 0*.01i*，且

*t*=1 s1 =1 *si*=*ri-T* -*α。*

*i 1i1i*

*k*

(源于附录A，定理6)

∈

由于*ri* - *T* - *α* ≥ 8 ln *T*，从Theo-的证明中可以看出。

Δ2

*r* = *(0.*05, *0.*05*,...,* 0*.*05) [0, 1] .我们显示结果

为*T*=106。图2显示了传统遗憾与最大公平性vio-之间的权衡。

*i*

在附录A中提醒6，E[*Ni,T*]≤*ri* - *T* - *α* +。

.1+ *π*2 Σ.

等于最大值

*i*∈[*k*]*rit* - *N*

*i,t*

，相对于*α*，这个

因此，E[*N*

*i,T*

]- *ri*

* *T* - *α*Σ ≤ . 1+ *π*2 Σ.

3特别是抓住了*公平的成本*。可以看出，

后悔减少，最大公平违规行为增加

分别随着*α*的增加而增加，直到达到*α*的阈值。

情况2：设*i*=1，*ri*-*T<8* ln T。

3

*i*

Δ2

那么，附录A中定理6的证明就可以被批准。

.Σ

适应性强，表明E[*N*]≤8 ln *T* +1+*π*2 。

Δ

对于小于这个阈值的*α*值，公平性CONS-1.B.A.的数值为0。

紧张导致一些次优武器被拉动的次数超过了确定其平均报酬所需的次数。

因此

E[*Ni,T*] -

.*ri* - *T* - *α*Σ ≤

8 ln *T*

2

Δ

*i*

*i,*T2

*i*

+ 1+ *π*

Σ

2

.

3

3

* *ri* - *T* - *α*Σ

有足够的可信度。另一方面，对于*α*的值

超过这个临界点，遗憾就会急剧减少，而

我们从经典的UCB1算法中恢复对数遗憾。请注意，在这种情况下，*α*的阈值是依赖于问题的。

**比较。FAIR-UCB与LFG：**与我们最接近的工作是(Li，Liu，and Ji 2019)的工作，他们的算法叫*Learning with Fairness Guarantee*（LFG）。

Talebi and Proutiere 2018）研究了资源配置约束的匪优化问题。

**与（李、刘、季2019）比较**√**n**。除了

在下面的模拟结果中，以其为基线。

证明了一个*O*( *T* ln *T* ) 无分布的*r*-Regret边界，因为

我们在比较*r*-Regret时考虑的仿真参数与前一实例相同。图3是FAIR-UCB和LFG的时间与*r*-Regret的关系图。请注意，FAIR-UCB和LFG在算法遭受的*r*-Regret方面表现相当。同时，模拟结果也验证了我们理论上的对数*r*-Regret约束结果。

接下来我们比较FAIR-UCB与LFG的公平性保证。我们考虑一个实例，*k*=3，*μ*=(0*.*7*，*0*.*5*，*0*.*4)，*r*=(0*.*2*，*0*.*3*，*0*.*25)，并且，*α*=0.图为

4显示了时间与最大公平性违规的关系图。观察到FAIR-UCB的公平性保证是一致的。

在(Li, Liu, and Ji 2019)中，我们展示了一个*O*(ln *T* ) *r*-Regret边界，它对实例参数具有ﬁner依赖性。我们的公平性保证在一段时间内均匀地保持，因此比LFG的渐进公平性保证强得多。此外，由于我们的公平性保证与FAIR-LEARN中使用的学习算法无关，因此对于(Li, Liu, and Ji 2019)中考虑的set- ting来说，它是成立的。

**与(Celis et al. 2018)比较**。(Celis等人，2018年)最近的一项工作考虑了一个个性化的新闻源设置，在任何时间*t*，对于给定的上下文(用户)，手臂(即要显示的广告)从一个分布中取样。

*t*

在时间跨度*T*的形式上。请注意，虽然公平

*p*在军火（广告）的集合[*k*]上，公平性达到了

仿真结果表明，LFG的公平性违约率似乎越来越高，但它确实在某一时刻重新降到了零，从而保证了非对称性的公平性。综上所述，仿真结果重新证明了我们理论上对FAIR-LEARN的公平性和*r*-Regret的保证，特别是FAIR-UCB。

# 7 相关工作

旨在通过机器学习算法确保决策公平性的研究工作激增，如分类算法（Agarwal等人，2018；Narasimhan，2018；Zafar等人。2017a；2017b）、回归算法（Berk等2017；Rezaei等2019）、排名和推荐系统（Singh和Joachims 2019；Beu- tel等2019；Singh和Joachims 2018；Celis、Straszak和Vishnoi 2017；Zehlike等2017）等。即使在在线学习的背景下也是如此，特别是在MAB环境下。下面我们陈述这些相关的著作。

(Joseph等人，2016)提出了一个UCB算法的变体，以确保他们所谓的择优公平性，即无论算法对每个臂的平均报酬的确信程度如何，一个臂永远不会优于一个更好的臂。这保证了每个臂的个体公平性（见（Dwork等人，2012）），同时实现了次线性后悔的效率。(Liu et al. 2017)的工作旨在确保"待遇平等"，即相似的个体得到相似的待遇。(Gillen等人，2018)考虑了对未知公平度量的个体公平性保证。

以上讨论的论文将传统的累积报酬最大化目标与模拟满足一些附加约束条件的目标相结合。已经考虑了带有其他附加约束条件的MAB概率-勒姆。例如，（Badanidiyuru，Kleinberg，和Slivkins 2013；Immorlica等人2018）研究了带有knapsack con- straints的MAB，其中特定手臂可以被拉动的次数受到一些预算的限制。(Xia et al. 2015; Amin et al. 2012; Tran-Thanh et al. 2014)的作品考虑了MAB问题，其中拉动每个手臂有一些成本，学习者有一个固定的预算。Lattimore、Crammer和Szepesva´ri的工作（Lattimore，Crammer和Szepesva´ri 2014；2015。

通过确保每个手臂上的预设概率质量，将允许的分布集限制在简单集的子集上。(Celis et al. 2018)中的算法当应用于我们所考虑的经典随机多臂匪徒环境时，仅在算法对随机拉动手臂的*expec- tation*上确保任何时间的公平性。相比之下，我们的算法（定理1）提供了更强的确定性的任何时间公平性保证。此外，我们还提供了一个明确的公平性和后悔之间的权衡（以不公平性tol- erance *α*为例）。另外，我们算法的计算开销仅为*O*(1)，而(Celis等人2018)中的算法需要在每一轮中解决LP。我们还注意到，我们的模型可以直接调整以捕获(Celis et al. 2018)中的设置。

# 8 讨论和今后的工作

本文所考虑的约束条件是通过保证每个臂在任何时候都有最小的拉动比例来实现 公平性。在很多情况下，这种公平性约束是不可缺少的，在这种情况下，*r*-Regret概念比较了任何在线al- githm的预期损失和同样满足这种公平性约束的最优算法的预期损失。我们提出的元算法FAIR-LEARN的一个重要特征是它所提供的均匀时间公平性保证与所使用的学习算法无关。我们还通过评估传统遗憾和公平性之间的权衡，以不公平性容忍度参数的方式，阐明了满足这种公平性约束的成本。公平性的几个概念，如差异影响、统计学上的均等性、均等化的几率等。(Barocas, Hardt, and Narayanan 2018)在文献中得到了广泛的研究。如（Blum等2018；Blum和Lykouris 2019；Bechavod等2019）在在线学习框架中融入这样的公平性概念，是一个令人兴奋的未来方向。

# 鸣谢

我们感谢Siddharth Barman教授的有益见解，并帮助我们改进了论文的表述。我们还

感谢Krishna Gummadi教授为我们指出了Face-book诉讼案。Ganesh Ghalme感激地承认塔塔信托的旅行支持。最后，我们感谢Anony- mous审稿人的有益意见。

# 参考文献

Agarwal, A.; Beygelzimer, A.; Dud´ık, M.; Langford, J.; and Wal- lach, H. 2018.A reductions approach to fair classiﬁcation.*arXiv preprint arXiv:1803.02453*.

Agrawal, S., and Goyal, N. 2012.Analysis of thompson sampling for the multi-armed bandit problem.In *Conference on Learning Theory*, 39-1.

Amin, K.; Kearns, M.; Key, P.; and Schwaighofer, A. 2012.Bud- get optimization for sponsored search:mdps中的删减学习。 *arXiv preprint arXiv:1210.4847*。

Badanidiyuru，A.；Kleinberg，R.；和Slivkins，A.2013。带背包的匪徒。In *Annual Symposium on Foundations of Com- puter Science*, 207-216.

Barocas, S.; Hardt, M.; and Narayanan, A. 2018.*Fairness and Machine Learning*. fairmlbook.org.

Bechavod, Y.; Ligett, K.; Roth, A.; Waggoner, B.; and Wu, Z. S. 2019.Equal opportunity in online classiﬁcation with partial feed- back. *arXiv preprint arXiv:1902.02242*.

Berk, R.; Heidari, H.; Jabbari, S.; Joseph, M.; Kearns, M.; Morgen- stern, J.; Neel, S.; and Roth, A. 2017.A convex framework for fair regression.*arXiv preprint arXiv:1706.02409*.

Beutel, A.; Chen, J.; Doshi, T.; Qian, H.; Wei, L.; Wu, Y.; Heldt, L.; Zhao, Z.; Hong, L.; Chi, E. H.; et al. 2019.Fairness in recom- mendation ranking through pairwise comparisons.*arXiv preprint arXiv:1903.00780*.

Blum, A., and Lykouris, T. 2019.Advancing subgroup fairness via sleeping experts.*arXiv preprint arXiv:1909.08375*.

Blum，A.；Gunasekar，S.；Lykouris，T.；和Srebro，N. 2018。On preserving non-discrimination when combining expert advice.In *Advances in Neural Information Processing Systems*, 8376-8387.

Bubeck, S., and Cesa-Bianchi, N. 2012.Regret analysis of stochas- tic and nonstochastic multi-armed bandit problems.*Foundations and Trends§*R *in Machine Learning* 5（1）：1-122.

Celis, L. E.; Kapoor, S.; Salehi, F.; and Vishnoi, N. K. 2018.An algorithmic framework to control bias in bandit-based personaliza- tion.*arXiv preprint arXiv:1802.08674*.

Celis, L. E.; Straszak, D.; and Vishnoi, N. K. 2017.Ranking with fairness constraints.*arXiv preprint arXiv:1704.06840*.

Chatterjee, A.; Ghalme, G.; Jain, S.; Vaish, R.; and Narahari, Y. 2017.Analysis of thompson sampling for stochastic sleeping ban- dits.In *Uncertainty in Artiﬁcial Intelligence*.

Dwork，C.；Hardt，M.；Pitassi，T.；Reingold，O.；和Zemel，R.2012。Fairness through awareness.In *Theoretical Computer Sci- ence Conference*, 214-226.

弗里曼，S. 2019年。原始立场。In Zalta, E. N. ed., *The Stanford Encyclopedia of Philosophy*.Metaphysics Research Lab, Stanford University.

Gillen, S.; Jung, C.; Kearns, M.; and Roth, A. 2018.Online learning- ing with an unknown fairness metric.In *Advances in Neural Infor- mation Processing Systems*.

Heidari，H.；Ferrari，C.；Gummadi，K.；和Krause，A. 2018。Fair- ness behind a veil of ignorance:A welfare analysis for automated decision making.In *Advances in Neural Information Processing Systems*.1265–1276.

Immorlica, N.; Sankararaman, K. A.; Schapire, R.; and Slivkins,

A.2018.Adversarial bandits with knapsacks.*arXiv preprint arXiv:1811.11881*.

Joseph, M.; Kearns, M.; Morgenstern, J. H.; and Roth, A. 2016.学习中的公平性。Classic and contextual bandits.In *Advances in Neural Information Processing Systems*.

Kleinberg, R.; Niculescu-Mizil, A.; and Sharma, Y. 2010.睡觉专家和强盗的后悔界限。*Machine learning* 80(2- 3):245-272.

Lattimore，T.，和Szepesva´ri，C. 2018. Bandit algorithms.

*预印本*：

Lattimore, T.; Crammer, K.; and Szepesva´ri, C. 2014. Opti- mal resource allocation with semi-bandit feedback. *arXiv preprint arXiv:1406.3840*.

Lattimore, T.; Crammer, K.; and Szepesva´ri, C. 2015. Linear multi-resource allocation with semi-bandit feedback.In *Advances in Neural Information Processing Systems*.

Li, F.; Liu, J.; and Ji, B. 2019.Combinatorial sleeping bandits with fairness constraints.In *IEEE Conference on Computer Communi- cations*.

Liu, Y.; Radanovic, G.; Dimitrakakis, C.; Mandal, D.; and Parkes,

D.C. 2017.Calibrated fairness in bandits.*arXiv preprint arXiv:1707.01875*.

Narasimhan, H. 2018.Learning with complex loss functions and constraints.In *International Conference on Artiﬁcial Intelligence and Statistics*, 1646-1654.

Rawls, J. 1971。*A theory of justice*.哈佛大学出版社。

Rezaei, A.; Fathony, R.; Memarrast, O.; and Ziebart, B. D. 2019.Fair logistic regression:An adversarial perspective.*CoRR* abs/1903.03910.

Singh, A., and Joachims, T. 2018.Fairness of exposure in rank- ings.In *International Conference on Knowledge Discovery & Data Mining,* 2219-2228.

Singh, A., and Joachims, T. 2019.Policy learning for fairness in ranking.*arXiv preprint arXiv:1902.04056*.

Slivkins, A. 2019.Introduction to multi-armed bandits. *arXiv preprint arXiv:1904.07272*.

Talebi, M. S., and Proutiere, A. 2018.Learning proportionally fair allocations with low regret.*Proceedings of the ACM on Measure- ment and Analysis of Computing Systems* 2（2）：36.

Tran-Thanh, L.; Stavrogiannis, L.; Naroditskiy, V.; Robu, V.; Jen- nings, N. R.; and Key, P. 2014.Efﬁcient regret bounds for online bid optimisation in budget-limited sponsored search auctions.

Xia，Y.；Li，H.；Qin，T.；Yu，N.；and Liu，T.-Y.2015.Thompson sampling for budgeted multi-armed bandits.In *International Joint Conference on Artiﬁcial Intelligence*.

Zafar，M. B.；Valera，I.；Gomez Rodriguez，M.；和Gummadi，K. P. 2017a.Fairness beyond disparate treatment & disparate impact:Learning classiﬁcation without disparate mistreatment.In *Interna- tional Conference on World Wide Web*，1171-1180.

Zafar, M. B.; Valera, I.; Rogriguez, M. G.; and Gummadi, K. P. 2017b.公平性约束。Mechanisms for fair classiﬁcation.In *Artiﬁcial Intelligence and Statistics*，962-970.

Zehlike，M.；Bonchi，F.；Castillo，C.；Hajian，S.；Megahed，M.；和Baeza-Yates，R.2017。Fa\*ir:A fair top-k ranking algorithm.In *ACM Conference on Information and Knowledge Management*, 1569-1578.