公平性感知的差分隐私协同过滤

Zhenhuan Yang
zyang@etsy.com
Etsy Inc 美国纽约州布鲁克林

Yingqiang Ge
yingqiang.ge@rutgers.edu
罗格斯大学 美国新泽西州新不伦瑞克

Congzhe Su
csu@etsy.com
Etsy Inc 美国纽约州布鲁克林

Dingxian Wang
dingxianwang@etsy.com Etsy Inc 美国纽约州布鲁克林

Xiaoting Zhao
xzhao@etsy.com
Etsy Inc 美国纽约州布鲁克林

Yiming Ying
yying@albany.edu 奥尔巴尼大学，纽约州立大学，美国纽约州奥尔巴尼

arXiv:2303.09527v1 [cs.IR] 16 Mar 2023

摘要

近年来，差分隐私指导算法在隐私保护机器学习任务中的应用日益广泛。然而，众所周知，使用这类算法需要在算法公平性方面做出权衡。具体而言，我们通过实证观察发现，采用差分隐私随机梯度下降（DP-SGD）训练的经典协同过滤方法，会对不同用户参与度的用户群体产生差异影响。进而，这会使原本不公平的模型对非活跃用户产生更严重的偏见。为解决上述问题，我们提出**DP-Fair**，一种面向协同过滤算法的两阶段框架。具体而言，该框架将差分隐私机制与公平性约束相结合，在保护用户隐私的同时确保公平推荐。基于亚马逊数据集以及从 Etsy（最大的电子商务平台之一）收集的用户历史记录日志进行的实验结果表明，与原始差分隐私随机梯度下降相比，我们提出的方法在浅层和深层推荐模型上，无论总体准确率还是用户群体公平性方面，都表现更优。

CCS 概念

• **计算方法** → **机器学习**；• **信息系统** → **推荐系统**；**信息检索**；**协同过滤**。

关键词

协同过滤、公平性、差分隐私

ACM 参考格式：

Zhenhuan Yang、Yingqiang Ge、Congzhe Su、Dingxian Wang、Xiaoting Zhao 和 Yiming Ying。2023。《公平性感知的差分隐私协同过滤》。载于 *ACM Web Conference 2023* 伴随论文集（*WWW '23 Companion*），*2023*年*4*月*30*日至*5*月*4*日，美国德克萨斯州奥斯汀。ACM，美国纽约州纽约市，共5页。https://doi.org/10.1145/3543873.3587577

允许免费为个人或课堂用途制作本作品全部或部分内容的数字或纸质副本，前提是不得为营利或商业利益而制作或分发副本，且副本须包含本声明及首页上的完整引文。对于本作品中由作者以外的其他人拥有版权的组成部分，必须尊重其版权。注明出处后允许摘录。若要以其他方式复制、重新发布、发布到服务器或分发至列表，则须事先获得明确许可和/或支付费用。请向 permissions@acm.org 申请许可。
WWW '23 Companion，*2023*年*4*月*30*日至*5*月*4*日，美国德克萨斯州奥斯汀© 2023 版权所有归所有者/作者所有。出版权已授权给 ACM。ACM ISBN 978-1-4503-9419-2/23/04...$15.00 https://doi.org/10.1145/3543873.3587577

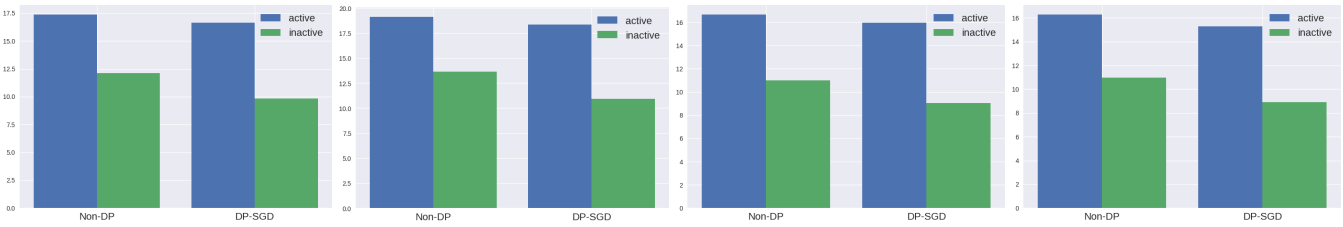
1 引言

随着电子商务的迅猛发展，消费者越来越依赖在线平台满足购物需求。传统的基于协同过滤（CF）的推荐模型会利用用户过去的互动（如评分和点击）来学习嵌入表示。然而，使用此类用户历史记录可能会泄露用户的敏感信息。先前研究表明，对手可以根据推荐列表推断目标用户的实际评分 [13, 19]，或判断该用户是否在数据库中 [3, 24]。为防止隐私泄露，差分隐私（DP） [5, 6] 是上述研究中广受欢迎的一种机制，因为它能够提供理论上可量化的隐私保证。具体而言，由于具有良好的可扩展性，适用于大型神经网络，因此常采用差分隐私随机梯度下降（DP-SGD） [1, 25, 27]。

尽管差分隐私随机梯度下降成功实现了隐私保护，但也可能带来某些权衡。近期发现，其中一种权衡是：在图像分类任务中，差分隐私模型造成的效用下降对代表性不足的子群体影响尤为显著 [2, 4]。在推荐系统领域，联邦学习等其他隐私保护机制也会引发公平性问题 [17, 20, 30]。因此，一个自然的问题是：保护用户隐私是否会导致不同用户群体之间的推荐性能存在差异?

具体而言，在本研究中，我们关注活跃度不同的用户群体之间受到的不公平对待，这是公平性感知推荐领域普遍关注的问题 [9, 15, 16,26, 29]。通常，在训练模型时，与平台互动更频繁的用户比活跃度较低的用户贡献的数据更充足。由于协同过滤的基本原理 [8, 12, 18, 23],，训练得到的推荐系统可能会对非活跃用户产生偏差 [7, 15, 28]。因此，活跃度较低的用户更可能获得较低的归一化折损累计增益（NDCG），也就是说，他们得到的推荐结果不尽如人意；图1中的非差分隐私柱状图展示了这一点。更糟的是，我们的实证发现表明，引入差分隐私后，这种准确率偏差会进一步加剧。具体而言，与活跃用户相比，非活跃用户的NDCG受到的影响更大；在四个数据集上，差分隐私随机梯度下降设置下，活跃用户与非活跃用户之间以NDCG差距衡量的不公平对待程度，比非差分隐私设置下高出15%至20%。这也被称为“穷者愈穷”现象（马太效应），如 [2, 10]所述。

为此，我们提出了一种面向协同过滤方法的新型两阶段框架，以解决效用退化和



(a) 家居与生活 **(b) 手工用品与工具** **(c) 杂货与美食** **(d) 美容**

图1：不同数据集上的 NDCG@10（%）（参见第 4 节的描述）中活跃用户（蓝色）与非活跃用户（绿色）的表现。在每个子图中，左侧标有"非差分隐私"的两根柱表示采用标准随机梯度下降训练的神经矩阵分解模型。右侧两根柱表示使用差分隐私随机梯度下降和 $\epsilon = 1$训练的模型。

DP-SGD 导致的不公平性加剧问题，称为 **DP-Fair**。具体而言，DP-Fair 的第一阶段分别对用户和物品嵌入进行噪声扰动，从而在提供隐私保证的同时，相比原始差分隐私随机梯度下降提升效用。第二阶段中，我们通过求解一个整数规划问题，应用后处理步骤，在最终推荐列表上实现用户群体公平性。我们在亚马逊基准数据集以及从 Etsy 收集的用户历史日志上进行实验，结果表明，在整体推荐性能和用户侧群体公平性方面，我们提出的算法优于基于原始差分隐私随机梯度下降的协同过滤方法。

2 预备知识

2.1 协同过滤模型

设 $\mathrm{U} = \{u_1, \cdots u_{n_1}\}$ 和 $\mathrm{V} = \{v_1, \cdots v_{n_2}\}$ 分别为用户集和物品集。设 $\mathrm{H}_u = \{v \in \mathrm{V}\}$ 表示用户 u 产生过正向交互的物品集合。值得注意的是，我们将所有交互都视为二元隐式反馈（例如，有点击时取 1）。评分等显式反馈 r（例如 1–5）会转换为 1，否则 $r > 3$ 转换为 0。H 表示所有 H_u 的集合。设 n 为正向交互总数，即 $|\mathrm{H}| = n$。令$\mathbf{x}_u \in \mathbb{R}^{n_1}$、$\mathbf{x}_v \in \mathbb{R}^{n_2}$ 分别表示用户 u 和物品 v 的独热编码。令 $\mathbf{z}_u = U\mathbf{x}_u \in \mathbb{R}^{d_1}$、$\mathbf{z}_v = V\mathbf{x}_v \in \mathbb{R}^{d_2}$ 表示相应的潜在嵌入。我们还用W表示任何其他潜在的特征提取参数，并用 $\Theta = (U、V、W)$ 表示所有可学习参数。

协同过滤潜因子模型 $f_\Theta : \mathrm{U} \times \mathrm{V} \to \mathbb{R}$ 在给定学习排序损失 l 后，学习推断隐式反馈模式。在本文中，我们关注经典的贝叶斯个性化排序（BPR）[23] 损失，如下所示，

$$l(f_\Theta) = \sum_{u,v,v'} -\log \sigma(f_\Theta(\mathbf{x}_u, \mathbf{x}_v) - f_\Theta(\mathbf{x}_u, \mathbf{x}_{v'})) + \frac{\lambda}{2}\|\Theta\|^2,$$

其中 σ 是 sigmoid 函数，λ 是正则化参数，$v \in \mathrm{H}_u$，而 $v' \in \mathrm{H}_u^- = \mathrm{V} \backslash \mathrm{H}_u$表示用户 u未提供隐式反馈的物品。由于正向交互通常更加稀疏，我们稍微宽泛地使用该符号，让 H^- 也表示其均匀子采样集合，使得 $|\mathrm{H}^-| = |\mathrm{H}| = n$。学习完成后，根据前 k 个排序分数 $\{f_\Theta(u,\mathrm{V})\}$，为每个用户 u 生成推荐列表 $\mathrm{R}_u^k = \{v \in \mathrm{V}\}$。

2.2 差分隐私与差分隐私随机梯度下降

我们首先介绍差分隐私的定义，具体如下。

定义 1. 对于任意 $\epsilon, \delta > 0$，如果（随机化）算法 A 满足：对于所有相邻的数据集D, D'（它们至多相差一个样本），以及由 A 生成的所有可能输出集 Θ，均有

$$\mathbb{P}[\mathcal{A}(D) \in \Theta] \leq \exp(\epsilon)\mathbb{P}[\mathcal{A}(D') \in \Theta] + \delta.$$

其中 ϵ 表示隐私预算（数值越小，隐私保证越强），δ表示隐私保证失效概率的容忍度。在实践中，通常要求 $\delta \ll \frac{1}{n}$。由于可以从交互数据中推断用户的敏感信息，因此此处的私有数据集为 $D = \mathrm{H} \cup \mathrm{H}^-$。

在每次迭代 t中，差分隐私随机梯度下降会对接收到的梯度 G_t进行带有某个界限 C 的梯度范数裁剪，并添加方差为σ^2的高斯噪声，然后基于新梯度 $\tilde{G}_t$对模型参数 Θ_t 执行常规随机梯度下降。若随机采样大小为 m的批次 $\mathrm{B}_t \subseteq \mathrm{H}$，则对于每个样本$S \in \mathrm{B}_t$，差分隐私随机梯度下降的执行方式如下：

$$\bar{G}_t(S) = G_t(S)/\max\{1, \|G_t(S)\|_2/C\} \quad (1)$$

$$\tilde{G}_t(S) = \bar{G}_t(S) + \mathcal{N}(0, \sigma^2\mathbf{I}) \quad (2)$$

2.3 用户侧公平性

令 R_u^K表示最初向用户 u 提供的推荐列表。我们采用 Kamishima 等人 [14], 最初提出的非均等不公平性度量，其定义如下：

定义 2. 给定一个推荐评估指标 M，针对群体 A 和 B 的用户群体公平性定义如下：

$$\mathbb{E}_u[\mathcal{M}(\mathcal{R}_u^K)|u \in A] = \mathbb{E}_u[\mathcal{M}(\mathcal{R}_u^K)|u \in B].$$

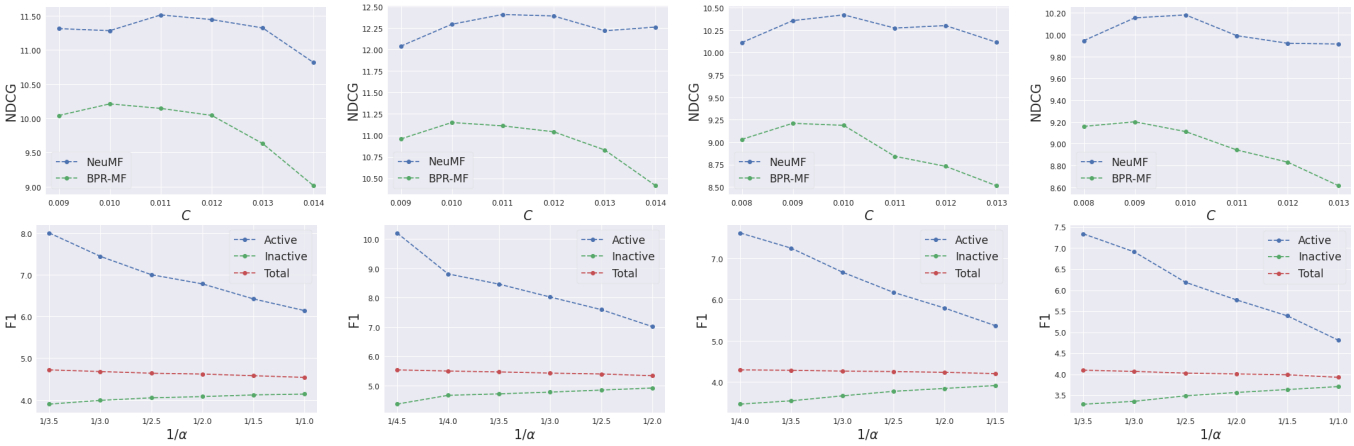
在实证中，用户群体公平性通过以下方式衡量：

$$\mathcal{F}_U(\mathcal{R}^K; A, B) = \left|\frac{1}{|A|}\sum_{u \in A}\mathcal{M}(\mathcal{R}_u^K) - \frac{1}{|B|}\sum_{u \in B}\mathcal{M}(\mathcal{R}_u^K)\right|.$$

由于用户活跃度不同，推荐系统通常对历史互动较少的用户表现较差。如 [2]所示，将差分隐私纳入模型后，这种偏差可能会进一步放大。遵循经典的二八定律，我们根据用户的参与活动选取排名前 20% 的用户作为高频／活跃群体 A，其余用户则归为低频／非活跃群体 B。

3 公平性感知的差分隐私协同过滤

我们提出的框架 DP-Fair 分为两个阶段。给定一个私有数据集 D，其中用户提供了隐私预算（ϵ，δ），第一阶段



(a) 家居与生活 **(b) 手工用品与工具** **(c) 杂货与美食** **(d) 美容**

图 2：上：不同裁剪界限 C 下的 NCDG 性能结果。下：不同公平性约束 α 级别下的 F1 分数性能结果。

算法 1 DP-Fair

1: **输入：** 私有数据集 D $=$H $\cup$H^{-}；隐私参数 ϵ， δ；迭代次数 T；学习率 $\{\eta_t\colon t \in [T]\}$，小批量大小 m；初始点 Θ_0。2: 阶段一：私有训练3: **对于** $t = 0$ 到 $T-1$ **执行**4: 从大小为 m的 D 中随机抽取一个批次 B$_t \subseteq$ 5: **对于** 每个样本 $S \in$ B$_t$ **执行** 6: 按照公式 (1) 和 (2) 分别对˜ ˜ ˜用户梯度和物品梯度应用差分隐私随机梯度下降步骤 $G_t(S) = (G_u(S)$ $G_v(S))$ 7: **结束循环**8: 更新 $\Theta_{t+1} = \Theta_t - \frac{\eta_t}{m} \sum_{S \in B_t} \tilde{G}_t(S)$ 9: **结束循环**10: 阶段二：公平性重排序

11 f_Θ — K :

12 3

基于……进行排序，并返回长推荐列表 R：通过整数规划求解公式 ( ) 13: **输出：** 新的短推荐列表 R^k

该阶段采用差分隐私随机梯度下降训练贝叶斯个性化排序，并提供隐私保证。值得注意的是，在第 6 行，我们将统一差分隐私随机梯度下降替换为分别作用于用户和物品的差分隐私随机梯度下降。由于训练过程中用户梯度和物品梯度的尺度可能不同，这一过程可以避免不必要的范数裁剪和噪声添加 [13, 22] 。总的来说，与原始差分隐私随机梯度下降相比，这种定制的差分隐私步骤对梯度的扰动更小，因此能带来更好的效用。

在第二阶段，为了缓解不公平待遇，我们采用了一种后处理方法 [15]。在第 12 行，一旦获得 top-K 排名列表 R^K，我们便在用户群体公平性约束下，通过最大化预测分数之和对其重新排序

$$\max_{\mathcal{R}^k} \sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{R}_u^k} f_\Theta(\mathbf{x}_u, \mathbf{x}_v) \tag{3}$$

$$\text{s.t.} \quad \mathcal{F}_U(\mathcal{R}^k; A, B) \le \alpha \text{ and } \mathcal{R}_u^k \subseteq \mathcal{R}_u^K \quad \forall u \in \mathcal{U}.$$

可以将 R^K 视为一个二元矩阵 $R \in \mathbb{R}^{n_1 \times K}$，其中 $R_{u,\ v} = 1$ 表示向用户 u推荐了物品 v。因此，优化问题 (3) 可以转化为 0 − 1 整数规划问题。对于排序指标 M，我们选择常用的 F1分数。与归一化折损累计增益相比，F1分数的计算效率更高，因为它避免了位置折扣效应。值得注意的是，由于此方法仅基于推荐列表进行后处理，因此不会破坏学习得到的参数 f_Θ [5]的差分隐私保证。

4 实验

4.1 实验设置

4.1.1 **数据集。** 我们使用了两个不同的数据来源。首先，我们采用一个基准数据集，即亚马逊评论数据集（5-core），其中包含杂货与美食 和美容 类别的商品评论 [11]。由于这两个类别的数据均通过评分编码为显式反馈，我们将其转换为二元反馈。其次，我们从最大的电子商务平台之一 Etsy 上的家居与生活 和手工用品与工具两个类别中收集并抽样了为期一个月的用户历史日志。在这两个数据集中，我们都将用户的点击视为正向反馈。

Name	# 用户	# 商品	# 交互次数	稀疏度
家居与生活	6,538	3,924	100,855	99.61%
工艺用品与工具	4,488	5,569	159,445	99.36%
杂货与美食	14,681	8,713	151,254	99.88%
美容	22,363	12,101	198,502	99.93%

表 1：数据集统计信息。

4.1.2 **实现细节。** 在本实验中，我们对数据集进行随机 8:1:1 划分，分别构建训练集、验证集和测试集。我们同时考虑浅层和深层推荐模型，即 BPR-MF [23] 和神经矩阵分解 [12]。为找到最佳裁剪界限，我们按照 McMahan 等人 [21] 提出的差分隐私调优策略，通过预训练进行调优。我们设置预排序器 $K = 20$ 和重排序器 $k = 10$。我们固定隐私参数 $\delta = \frac{1}{n^{1.5}}$，并使用 Opacus[1] 模块执行差分隐私随机梯度下降步骤。我们还使用 Gurobi[2] 求解器求解公式 (3) 中的重排序问题。

[1]https://opacus.ai/ [2] https://www.gurobi.com

模型	指标	算法	$\epsilon = \infty$ 总计 ↑	活跃。↑	不活跃.↑	F_U ↓	$\epsilon = 10$ 总计 ↑	活跃.↑	不活跃.↑	F_U ↓	$\epsilon = 1$ 总计 ↑	活跃.↑	不活跃.↑	F_U ↓
家居与生活														
贝叶斯个性化排序矩阵分解	NDCG	差分隐私随机梯度下降	11.57	15.17	10.73	4.44	10.77	14.93	9.67	5.25	10.21	14.44	9.15	5.29
		DP-Fair	**12.04**	14.18	**11.52**	**2.57**	**11.19**	14.09	**10.44**	**3.64**	**10.61**	13.55	**9.87**	**3.68**
	F1	差分隐私随机梯度下降	4.85	7.17	4.39	2.77	4.56	6.96	3.90	2.97	4.31	6.84	3.67	3.17
		DP-Fair	**4.89**	6.57	**4.51**	**2.07**	**4.61**	6.38	**4.11**	**2.27**	**4.33**	6.30	**3.83**	**2.47**
神经矩阵分解	NDCG	差分隐私随机梯度下降	12.24	17.36	11.12	6.21	11.61	16.73	10.17	7.19	11.19	16.64	9.82	6.82
		DP-Fair	**12.93**	16.12	**12.13**	**3.99**	**12.05**	15.78	**11.12**	**4.66**	**11.59**	15.01	**10.74**	**4.27**
	F1	差分隐私随机梯度下降	5.03	7.65	4.37	3.28	4.78	7.47	4.10	3.36	4.58	7.61	3.82	3.79
		DP-Fair	**5.08**	7.28	**4.53**	**2.75**	**4.81**	7.20	**4.21**	**2.99**	**4.64**	6.96	**4.05**	**2.91**
手工用品与工具														
贝叶斯个性化排序矩阵分解	NDCG	差分隐私随机梯度下降	12.63	16.51	11.72	4.79	11.76	16.28	10.57	5.72	11.15	15.75	10.00	5.75
		DP-Fair	**13.59**	16.02	**13.01**	**3.01**	**12.64**	15.91	**11.80**	**4.11**	**11.98**	15.32	**11.15**	**4.17**
	F1	差分隐私随机梯度下降	5.30	7.89	4.80	3.09	4.99	7.30	4.27	3.61	4.71	7.48	4.02	3.45
		DP-Fair	**5.36**	7.22	**4.95**	**2.16**	**5.06**	7.01	**4.52**	**2.49**	**4.75**	6.91	**4.21**	**2.70**
神经矩阵分解	NDCG	差分隐私随机梯度下降	13.57	19.12	12.37	6.74	12.88	18.37	11.32	7.05	12.41	18.35	10.93	7.41
		DP-Fair	**14.55**	17.76	**13.65**	**4.11**	**13.56**	17.62	**12.51**	**5.11**	**13.16**	17.49	**12.08**	**5.40**
	F1	差分隐私随机梯度下降	5.91	9.01	5.14	3.86	5.62	8.98	4.78	4.19	5.39	8.71	4.56	4.15
		DP-Fair	**5.98**	8.55	**5.33**	**3.21**	**5.64**	8.41	**4.95**	**3.46**	**5.46**	8.44	**4.71**	**3.73**
杂货与美食														
贝叶斯个性化排序矩阵分解	NDCG	差分隐私随机梯度下降	10.65	14.24	9.80	4.45	9.90	14.04	8.81	5.23	9.21	13.57	8.12	5.45
		DP-Fair	**11.34**	13.88	**10.85**	**3.03**	**10.54**	13.27	**9.70**	**3.58**	**9.79**	12.64	**9.08**	**3.56**
	F1	差分隐私随机梯度下降	4.22	6.04	3.79	2.26	4.03	5.92	3.52	2.39	3.78	5.74	3.28	2.46
		DP-Fair	**4.28**	5.78	**3.91**	**1.87**	**4.09**	5.71	**3.67**	**2.05**	**3.82**	5.44	**3.41**	**2.03**
神经矩阵分解	NDCG	差分隐私随机梯度下降	11.40	16.66	10.23	6.43	10.82	16.08	9.36	6.72	10.42	15.98	9.03	6.95
		DP-Fair	**11.87**	15.31	**11.01**	**4.29**	**11.06**	14.94	**10.09**	**4.85**	**10.73**	14.69	**9.74**	**4.95**
	F1	差分隐私随机梯度下降	4.64	7.01	4.05	2.96	4.41	6.76	3.76	3.25	4.23	6.79	3.59	3.20
		DP-Fair	**4.67**	6.66	**4.23**	**2.43**	**4.42**	6.39	**3.87**	**2.52**	**4.26**	6.62	**3.66**	**2.96**
美容														
贝叶斯个性化排序矩阵分解	NDCG	差分隐私随机梯度下降	10.43	13.66	9.63	4.02	9.68	13.61	8.68	4.93	9.20	13.16	8.22	4.94
		DP-Fair	**11.22**	13.15	**10.75**	**2.39**	**10.43**	13.13	**9.75**	**3.39**	**9.76**	12.20	**9.14**	**3.06**
	F1	差分隐私随机梯度下降	4.02	5.50	3.65	1.85	3.82	5.62	3.37	2.25	3.54	5.38	3.08	2.29
		DP-Fair	**4.10**	5.20	**3.73**	**1.47**	**3.85**	5.49	**3.44**	**2.04**	**3.56**	5.09	**3.18**	**1.91**
神经矩阵分解	NDCG	差分隐私随机梯度下降	11.22	16.27	10.21	5.05	10.65	15.25	9.24	7.03	10.18	15.28	8.90	6.39
		DP-Fair	**11.75**	15.51	**10.98**	**4.52**	**11.09**	14.80	**9.98**	**4.82**	**10.47**	14.42	**9.48**	**4.94**
	F1	差分隐私随机梯度下降	4.51	6.78	3.98	2.79	4.25	6.66	3.61	3.05	3.99	6.44	3.37	3.07
		DP-Fair	**4.54**	6.31	**4.10**	**2.21**	**4.27**	6.24	**3.78**	**2.46**	**4.03**	6.21	**3.49**	**2.72**

表 2：推荐性能结果。评估指标基于前 10 项预测计算。结果以百分比（%）报告，箭头表示有利方向。最佳结果以粗体突出显示。

4.2 实验分析

4.2.1 **主要结果。** 根据表 2，可以得出若干结论。首先，结果报告了三种不同隐私预算设置下的表现：非隐私设置（$\epsilon = \infty$）、宽松隐私预算设置（$\epsilon = 10$）和严格隐私预算设置（$\epsilon = 1$）。与预期一致，也与已有文献的结论相符，随着隐私约束收紧，以归一化折损累计增益和 F1 分数衡量的效用会下降。其次，无论隐私预算如何 ϵ，我们提出的 DP-Fair 算法在归一化折损累计增益和 F1 分数方面的整体效用均优于差分隐私随机梯度下降。这一提升归因于我们定制的噪声添加和梯度裁剪技术；该技术同样适用于非隐私设置，此时该算法等同于采用重排序步骤的随机梯度下降。强制实现公平性似乎会提升效用，这看起来有悖直觉，但我们的结果表明，这是因为非活跃用户的效用得到提升；他们占所有用户的 80%，从而提高了整体表现。最后，我们观察到，相较于差分隐私随机梯度下降，DP-Fair 显著缩小了 F_U 差距。这可归因于 DP-Fair 中的后处理步骤，该步骤从$K = 20$列表中识别出 $k = 10$ 公平性感知物品。

4.2.2 **超参数影响。** 在本实验中，我们将隐私预算 ϵ 固定为 1。我们的首要目标是考察裁剪界限的选择。为简化分析，我们施加$C_u =$ $C_v = C$ 2, 。根据图中呈现的结果，我们得出结论：C过小或过大都会对 NCDG 产生负面影响。当裁剪参数过小时，平均裁剪后的梯度可能会产生偏差。相反，增大范数界限 C 会导致向梯度中添加更多噪声。此外，我们还研究了公平性水平选择的影响。随着公平性要求变得更加严格，活跃群体的性能会下降，而非活跃群体的性能则会提升。

5 结论

在本文中，我们通过实证观察发现，在经典的基于协同过滤的推荐模型中引入差分隐私随机梯度下降，会扩大活跃用户与非活跃用户之间的不公平差距。我们提出了一种定制的差分隐私梯度映射，并结合整数规划方案，以提升其对活跃用户和非活跃用户的公平性。在真实世界电商数据集上的实验表明，无论是效用还是公平性指标，DP-Fair 的表现都优于差分隐私随机梯度下降。

参考文献

[1] Martin Abadi、Andy Chu、Ian Goodfellow、H Brendan McMahan、Ilya Mironov、Kunal Talwar 和 Li Zhang。2016。采用差分隐私的深度学习。载于*2016*年*ACM SIGSAC*计算机与通信安全会议论文集。308–318。[2] Eugene Bagdasaryan、Omid Poursaeed 和 Vitaly Shmatikov。2019。差分隐私对模型准确率具有差异影响。神经信息处理系统进展32 (2019)，15479–15488。[3] Arnaud Berlioz、Arik Friedman、Mohamed Ali Kaafar、Roksana Boreli 和 Shlomo Berkovsky。2015。将差分隐私应用于矩阵分解。载于第九届*ACM*推荐系统会议论文集。107–114。[4] Zhiqi Bu、Jinshuo Dong、Qi Long 和 Weijie J Su。2020。采用高斯差分隐私的深度学习。《哈佛数据科学评论》2020，23 (2020)，10–1162。[5] Cynthia Dwork、Frank McSherry、Kobbi Nissim 和 Adam Smith。2006。在隐私数据分析中校准噪声与敏感度。载于密码学理论会议。Springer，265–284。[6] Cynthia Dwork、Aaron Roth 等。2014。差分隐私的算法基础。*Found. Trends Theor. Comput. Sci.* 9，3-4 (2014)，211–407。[7] Zuohui Fu、Yikun Xian、Ruoyuan Gao、Jieyu Zhao、Qiaoying Huang、Yingqiang Ge、Shuyuan Xu、Shijie Geng、Chirag Shah、Yongfeng Zhang 等。2020。基于知识图谱的公平性感知可解释推荐。载于第*43*届国际*ACM SIGIR*信息检索研究与发展会议论文集。69–78。[8] Yingqiang Ge、Shuchang Liu、Zuohui Fu、Juntao Tan、Zelong Li、Shuyuan Xu、Yunqi Li、Yikun Xian 和 Yongfeng Zhang。2022。可信推荐系统综述。*arXiv*预印本 *arXiv:2207.12515*（2022）。[9] Yingqiang Ge、Shuchang Liu、Ruoyuan Gao、Yikun Xian、Yunqi Li、Xiangyu Zhao、Changhua Pei、Fei Sun、Junfeng Ge、Wenwu Ou 和 Yongfeng Zhang。2021。迈向推荐中的长期公平性。载于第*14*届*ACM*国际网络搜索与数据挖掘会议论文集。445–453。[10] Yingqiang Ge、Shuya Zhao、Honglu Zhou、Changhua Pei、Fei Sun、Wenwu Ou 和 Yongfeng Zhang。2020。理解电子商务推荐系统中的回音室。载于第*43*届国际*ACM SIGIR*信息检索研究与发展会议论文集（中国，线上会议）*(SIGIR '20)*。计算机协会，纽约，美国，2261–2270。https://doi.org/10 1145 3397271 3401431 。/。[11] Ruining He 和 Julian McAuley。2016。起伏：利用单类协同过滤对时尚趋势的视觉演变进行建模。载于第*25*届国际万维网会议论文集。507–517。[12] Xiangnan He、Lizi Liao、Hanwang Zhang、Liqiang Nie、Xia Hu 和 Tat-Seng Chua。2017。神经协同过滤。载于第*26*届国际万维网会议论文集。173–182。[13] Jingyu Hua、Chang Xia 和 Sheng Zhong。2015。差分隐私矩阵分解。载于*IJCAI*。[14] Toshihiro Kamishima、Shotaro Akaho 和 Jun Sakuma。2011。通过正则化方法实现公平性感知学习。载于*2011*年*IEEE*第*11*届国际数据挖掘会议研讨会。电气电子工程学会，643–650。[15] Yunqi Li、Hanxiong Chen、Zuohui Fu、Yingqiang Ge 和 Yongfeng Zhang。2021。推荐中的以用户为导向的公平性。载于万维网会议论文集 *2021*。624–632。[16] Yunqi Li、Yingqiang Ge 和 Yongfeng Zhang。2021。推荐系统中机器学习公平性的教程。收录于第*30*届*CIKM*论文集。[17] Shuchang Liu、Yingqiang Ge、Shuyuan Xu、Yongfeng Zhang 和 Amelie Marian。2022。公平性感知联邦矩阵分解。收录于第*16*届*ACM*推荐系统会议论文集。168–178。[18] Shuchang Liu、Fei Sun、Yingqiang Ge、Changhua Pei 和 Yongfeng Zhang。2021。生成式推荐列表的变异控制与评估。收录于*2021*年万维网会议论文集 （斯洛文尼亚卢布尔雅那）*(WWW '21)*。计算机协会，纽约，美国纽约州纽约市，436–448。https://doi.org/10 1145 3442381 3449864 。/。[19] Ziqi Liu、Yu-Xiang Wang 和 Alexander Smola。2015。快速差分隐私矩阵分解。收录于第*9*届*ACM*推荐系统会议论文集。171–178。[20] Kiwan Maeng、Haiyu Lu、Luca Melis、John Nguyen、Mike Rabbat 和 Carole-Jean Wu。2022。迈向公平的联邦推荐学习：刻画系统异质性与数据异质性的相互依赖关系。收录于第*16*届*ACM*推荐系统会议论文集。156–167。[21] H Brendan McMahan、Galen Andrew、Ulfar Erlingsson、Steve Chien、Ilya Mironov、Nicolas Papernot 和 Peter Kairouz。2018。为迭代训练过程添加差分隐私的一般方法。*arXiv*预印本 *arXiv:1812.06210*（2018）。[22] H Brendan McMahan、Daniel Ramage、Kunal Talwar 和 Li Zhang。2017。学习差分隐私循环语言模型。*arXiv*预印本 *arXiv:1710.06963*（2017）。[23] Steffen Rendle、Christoph Freudenthaler、Zeno Gantner 和 Lars Schmidt-Thieme。2012。BPR：基于隐式反馈的贝叶斯个性化排序。*arXiv*预印本 *arXiv:1205.2618*（2012）。[24] Hyejin Shin、Sungwook Kim、Junbum Shin 和 Xiaokui Xiao。2018。用于推荐的本地差分隐私增强型矩阵分解。*IEEE*知识与数据工程汇刊30，9（2018），1770–1782。[25] Zhenhuan Yang、Shu Hu、Yunwen Lei、Kush R Vashney、Siwei Lyu 和 Yiming Ying。2022。用于极小极大问题的差分隐私随机梯度下降-上升算法。收录于人工智能不确定性会议。机器学习研究论文集，2192–2202。[26] Zhenhuan Yang、Yan Lok Ko、Kush R Varshney 和 Yiming Ying。2022。极小极大AUC公平性：具有可证明收敛性的高效算法。*arXiv*预印本 *arXiv:2208.10451*（2022）。[27] Zhenhuan Yang、Yunwen Lei、Siwei Lyu 和 Yiming Ying。2021。用于具有非光滑损失的成对学习的随机梯度下降的稳定性与差分隐私。收录于人工智能与统计国际会议。机器学习研究论文集，2026–2034。[28] Sirui Yao 和 Bert Huang。2017。超越均等性：协同过滤的公平性目标。神经信息处理系统进展 30（2017），2921–2930。[29] Meike Zehlike、Ke Yang 和 Julia Stoyanovich。2021。排序公平性：综述。*arXiv*预印本 *arXiv:2103.14000*（2021）。[30] Zhitao Zhu、Shijing Si、Jianzong Wang 和 Jing Xiao。2022。Cali3F：校准型快速公平联邦推荐系统。收录于*2022*年国际神经网络联合会议（*IJCNN*）。电气电子工程师学会，1–8。