计算机科学
COMPUTER SCIENCE

# 基于拉格朗日对偶的小样本学习隐私保护和公平性约束方法

王静红，田长申，李昊康，王威

引用本文

王静红，田长申，李昊康，王威. 基于拉格朗日对偶的小样本学习隐私保护和公平性约束方法[J]. 计算机科学，2024，51(7)：405-412.

WANG Jinghong, TIAN Changshen, LI Haokang, WANG Wei. Lagrangian Dual-based Privacy Protection and Fairness Constrained Method for Few-shot Learning [J]. Computer Science, 2024, 51(7): 405-412.

---

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230500012

# 基于拉格朗日对偶的小样本学习隐私保护和公平性约束方法

王静红[1,2,3]　田长申[1,2,3]　李昊康[4]　王　威[1,3]
1 河北师范大学计算机与网络空间安全学院　石家庄 050024
2 河北师范大学河北省网络与信息安全重点实验室　石家庄 050024
3 河北师范大学供应链大数据分析与数据安全河北省工程研究中心　石家庄 050024
4 河北工程技术学院人工智能与大数据学院　石家庄 050020
(wangjinghong@126.com)

**摘　要**　小样本学习旨在利用少量数据训练并大幅提升模型效用,为解决敏感数据在神经网络模型中的隐私与公平问题提供了重要方法。在小样本学习中,由于小样本数据集中往往包含某些敏感数据,并且这些敏感数据可能有歧视性,导致数据在神经网络模型的训练中存在隐私泄露的风险和公平性问题。此外,在许多领域中,由于隐私或安全等,数据很难或无法获取。同时在差分隐私模型中,噪声的引入不仅会导致模型效用的降低,也会引起模型公平性的失衡。针对这些挑战,提出了一种基于Rényi差分隐私过滤器的样本级自适应隐私过滤算法,利用Rényi差分隐私以实现对隐私损失的更精确计算。进一步,提出了一种基于拉格朗日对偶的隐私性和公平性约束算法,该算法通过引入拉格朗日方法,将差分隐私约束和公平性约束加到目标函数中,并引入拉格朗日乘子来平衡这些约束。利用拉格朗日乘子法将目标函数转化为对偶问题,从而实现同时优化隐私性和公平性,通过拉格朗日函数实现隐私性和公平性的平衡。实验结果证明,该方法既提升了模型性能,又保证了模型的隐私性和公平性。

**关键词**:小样本学习;隐私与公平;Rényi差分隐私;公平性约束;拉格朗日对偶
**中图分类号**　TP309.3

# Lagrangian Dual-based Privacy Protection and Fairness Constrained Method for Few-shot Learning

WANG Jinghong[1,2,3], TIAN Changshen[1,2,3], LI Haokang[4] and WANG Wei[1,3]
1 College of Computer and Cyber Security, Hebei Normal University, Shijiazhuang 050024, China
2 Hebei Key Laboratory of Network and Information Security, Hebei Normal University, Shijiazhuang 050024, China
3 Hebei Provincial Engineering Research Center for Supply Chain Big Data Analytics & Security, Hebei Normal University, Shijiazhuang 050024, China
4 Artificial Intelligence and Big Data College, Hebei University of Engineering Science, Shijiazhuang 050020, China

**Abstract**　Few-shot learning aims to use a small amount of data for training and significantly improve model performance, and is an important approach to address privacy and fairness issues of sensitive data in neural network models. In few-shot learning, there is a risk of privacy and fairness issues in training neural network models due to the fact that small sample datasets often contain certain sensitive data, and that such sensitive data may be discriminatory. In addition, in many domains, data is difficult or impossible to access for reasons such as privacy or security. Also, in differential privacy models, the introduction of noise not only leads to a reduction in model utility, but also causes an imbalance in model fairness. To address these challenges, this paper proposes a sample-level adaptive privacy filtering algorithm based on the Rényi differential privacy filter to exploit Rényi differential privacy to achieve a more accurate calculation of privacy loss. Furthermore, it proposes a Lagrangian dual-based privacy and fairness constraint algorithm, which adds the differential privacy constraint and the fairness constraint to the objective function by introducing a Lagrangian method, and introduces a Lagrangian multiplier to balance these constraints. The Lagrangian multiplier method is used to transform the objective function into a pairwise problem, thus optimising both privacy and fairness, and achie-

到稿日期:2023-04-30　返修日期:2023-08-29
基金项目:河北省自然科学基金(F2021205014);河北省高等学校科学技术研究项目(ZD2022139);中央引导地方科技发展资金项目(226Z1808G);河北省归国人才资助项目(C20200340);河北师范大学博士基金项目(L2022B22)
This work was supported by the Natural Science Foundation of Hebei, China(F2021205014), Science and Technology Project of Hebei Education Department(ZD2022139), Central Guidance on Local Science and Technology Development Fund of Hebei Province(226Z1808G), Project Funded by the Introduction of Overseas Students in Hebei Province(C20200340) and Science Foundation of Hebei Normal University(L2022B22).
通信作者:王威(wangwei2021@hebtu.edu.cn)

ving a balance between privacy and fairness through the Lagrangian function. It is shown that the proposed method improves the performance of the model while ensuring privacy and fairness of the model.

**Keywords** Few-shot learning, Privacy and fairness, Rényi differential privacy, Fairness constraint, Lagrangian dual

## 1 引言

由于数据和模型中的固有风险，神经网络模型中隐私性和公平性问题日益突显[1]。当前，差分隐私技术被广泛应用于神经网络模型中的隐私保护[2]。差分隐私中噪声的引入在降低模型准确性的同时，也会加重模型公平性的失衡[3]。在小样本学习中[4]，受到数据集体量的限制，隐私性和公平性问题更为突出。因此，在保证小样本学习模型效用的同时，需要进一步探求模型中隐私性和公平性的联系。

本文通过拉格朗日对偶性来实现隐私和公平的小样本学习，利用拉格朗日对偶算法来平衡模型的隐私和公平性，利用Rényi差分隐私滤波器，提出了一种样本级自适应隐私过滤算法，以更准确地计算隐私损失。通过更严格的隐私损失计算，模型的公平性得到了一定程度的提高，该模型的形式与标准MAML算法相似。在内循环中使用了基于Rényi隐私过滤器的差分隐私梯度下降法，并考虑到了隐私和公平性。拉格朗日对偶法被用来确保隐私和公平的平衡。在外环中，通过同态加密和其他方法传输进行合作训练，从保护隐私和公平的角度出发，改善和平衡了小样本学习中的隐私、公平和模型效用。

本文研究内容如下：

1)提出了一种样本级的自适应隐私过滤算法，以更精确、计算隐私损失。该方法提供了更精确的严格的隐私损失计算。且由于对噪声添加的限制，Rényi差分隐私过滤器可以减小差分隐私方法对模型公平性的影响。

2)提出了基于拉格朗日对偶的隐私和公平性约束算法(PF-LD)，通过拉格朗日对偶性实现了隐私和公平的有效平衡。

3)本研究通过实验证明，在小样本学习中，利用本文方法可以在确定隐私配置下，使模型的隐私性和公平性达到最佳效果。

## 2 相关工作

本章梳理了隐私保护、公平性、小样本学习的相关知识以及研究现状。

### 2.1 Rényi差分隐私

Mironov[5]在标准差分隐私[6](Differential Privacy, DP)的基础上提出了Rényi差分隐私(Rényi Differential Privacy, RDP)。RDP是在引入Rényi散度时对标准差分隐私的自然松弛。RDP实现了对隐私预算更严格的限制，可以提供一种简单准确的方式用于跟踪运行过程中隐私损失的累加。同时，RDP将隐私预算的概念与高级组合定理的应用相结合，允许对组合定理进行严格的分析，适用于隐私保护算法和异构机制的组合。

在针对梯度扰动的研究中，Abadi等[7]提出了差分隐私随机梯度下降方法(Differential Privacy-Stochastic Gradient Descent, DP-SGD)，通过训练阶段的噪声注入达到隐私保护的目的。DP-SGD中采用了隐私会计(Moments Accountant, MA)，MA的核心思想是基于机制的敏感性，通过计算不同查询或操作对原始数据集的敏感性，结合差分隐私的复合性质，估计累积隐私泄露的上界。它提供了一种定量的方法来评估隐私保护，并支持在设计和优化差分隐私算法时进行隐私预测和调整。

为进一步限制隐私预算，降低噪声对模型的影响，Feldman等[8]提出了基于RDP的个体差分隐私过滤器(Rényi Filter)，实现了更加严格的个体隐私损失估计的核算方法。通过Rényi Filter可以获得针对每个个体的隐私泄露估计，从而更好地理解差分隐私机制在个体级别的隐私保护能力。它可以帮助完成隐私预测、隐私优化和隐私风险评估等任务。

**定义1**(差分隐私) 当随机算法$M$满足$(\varepsilon,\delta)$-DP时，则对于任意相邻数据集$D$和$D'$以及所有集合$S\in Range(M)$的输出，有：

$$\Pr[M(D)\in S]\leqslant \exp(\varepsilon)Pr[M(D')\in S]+\delta \tag{1}$$

定义中$\varepsilon$为隐私预算，差分隐私限制了随机算法的输出分布变化，其通过引入$\delta$变量来放宽差分隐私定义。

**定义2**(Rényi散度) 对于定义$R$上的概率分布$P$和$Q$，$\alpha$阶的Rényi散度为：

$$D_\alpha(P\parallel Q)\triangleq\frac{1}{\alpha-1}\log E\left(\frac{P(x)}{Q(x)}\right)^\alpha \tag{2}$$

**定义3**(Rényi差分隐私，RDP) 对于一个随机机制$f$：$D\mapsto R$，满足$\alpha$阶的Rényi差分隐私，对于任意的$D,D'\in D$，都满足于$(\alpha,\varepsilon)$-RDP。

$$D_\alpha(f(D)\parallel f(D'))\leqslant\varepsilon \tag{3}$$

Rényi差分隐私的一个重要的性质是可以转换为标准差分隐私。

**定理1** 如果算法$M$满足$(\alpha,\varepsilon)$-RDP，则$M$也满足$\left(\varepsilon+\dfrac{\log\left(\frac{1}{\delta}\right)}{\alpha-1},\delta\right)$-DP。

本文将Rényi差分隐私与神经网络模型更加有效地结合，从而使差分隐私模型在小样本学习中得到更好的表现。

### 2.2 公平性

Du等[9]从计算的角度概述了公平性在神经网络模型中存在偏差的问题，并关注模型中的可解释性。当前神经网络中公平性的概念大致可分为群组公平和个体公平，与个体公平相比，群组公平的研究更为广泛。神经网络中实现公平性主要采用数据预处理、公平性约束以及敏感属性处理等方法。

Zhang等[10]研究半监督学习是否有助于解决公平性问题，预处理阶段提出了一个公平性半监督学习框架，一种获得多个公平数据集的重采样方法。Zafar等[11]利用决策边界公平性度量设计公平分类器机制，该机制允许对公平程度进行细粒度控制，有助于减少模型对不同群体的偏见和不公平对待，并使得模型保持较高的准确性。在敏感性处理方法方面，

Liang 等[12]提出了一种基于双对抗学习的方法，通过公平分类器和领域分类器之间的对抗性训练，实现了公平分类和领域适应的结合。通过这种方法，能够在跨领域任务中实现公平分类，并降低数据来源的偏见对分类性能的影响。

在小样本学习与公平性结合的方向，主要使用了公平性约束。本文将常用的 DP(Demographic Parity)改为组公平(Group Fairness，GF)。

假设算法访问数据集 $D=(X,A,Y)$，其中 $X\in R^d$，$A$ 和 $F$ 分别表示特征、敏感属性和真实标签。

**定义 4**(组公平) 若预测到 $M$ 独立于 $A$，则模型满足组公平为：

$$\Pr[M(X)\in\tilde{y}|A=a]=\Pr[M(X)\in\tilde{y}],\forall a\in A,\tilde{y}\in Y \tag{4}$$

当 $\tilde{y}\in\{0,1\}$ 时，有：

$$E[M(X)|A=a]=E[M(X)],\forall a\in A \tag{5}$$

**定义 5**(几率相等) 几率相等[13](Equality of Opportunity，EO)要求预测 $M$ 有条件地独立于敏感属性 $A$ 和标签 $Y$，即：

$$\Pr[M(X)=\tilde{y}|A=a,Y=y]=\Pr[M(X)=\tilde{y}|Y=y],\quad\forall a\in A,\tilde{y}\in Y,y\in Y \tag{6}$$

当 $\tilde{y}\in\{0,1\}$ 时，有：

$$E[M(X)|A=a,Y=y]=E[M(X)|Y=y],\quad\forall a\in A,y\in Y \tag{7}$$

EO 要求分类器的预测结果在真实阳性率(True Positive Rate，TPR)和假阳性率(False Positive Rate，FPR)的条件下与敏感属性 $A$ 无关。

**定义 6**(反歧视水平) 设 $\gamma_a(D,f_\theta)$ 表示在一个基于公平度量的模型 $f_\theta$ 训练中，敏感属性群组 $a$ 阳性预测的概率(也可记为 $\gamma(D_0)$，表示总体；$\gamma(D_a)$，$a\neq0$ 表示敏感群组)。在使用数据集 $D$ 的模型 $f_\theta$ 上训练的反歧视水平 $\Gamma(D,f_\theta)$，可以通过群组间的差异来衡量。

$$\Gamma(D,f_\theta)=|\gamma_0(D,f_\theta)-\gamma_1(D,f_\theta)| \tag{8}$$

### 2.3 小样本学习

在小样本学习中，只需要少量样本便可使得模型适应新的任务[14-15]，目前已被广泛用于解决图像分类、回归、目标检测等问题。当前已有多个基准数据集衡量该领域的进展，包括 MiniImagenet[16]，Meta-Dataset[17]，Dax-Blicket 快速绑定数据集[18]。

小样本学习分为基于元学习和基于非元学习的方法，其中研究的大部分进展来自于基于元学习的方法。基于元学习的小样本学习可分为 3 种方法：基于度量[19]、基于优化[20]和基于模型[21]。Finn 等[22]提出了基于元学习的方法 MAML(Model-Agnostic Meta-Learning)，以快速泛化到新的任务。Abbas 等[23]针对 MAML 算法改进提出的 Sharp-MAML 方法，通过考虑模型的锐利度(Sharpness)来进行元学习，这是当前较为高效的元学习方法之一。

面对小样本学习隐私泄露和公平性失衡的风险。Li 等[24]研究了基于梯度的参数传递的差分隐私方法，提出了 DP GBML 算法，以实现模型的隐私保证。在小样本学习与公平性结合的方面，Slack 等[25]提出了公平预警和 Fair-MAML。Zhao 等[26]提出了原始-对偶公平元学习模型，通过次梯度对偶方法联合优化原始参数和对偶参数的初始化，实现公平的元学习。Jagielski 等[27]进一步研究了差分隐私中的公平机器学习。

### 2.4 拉格朗日对偶

拉格朗日对偶方法已被广泛用于优化约束性问题。拉格朗日对偶问题是通过构造一个拉格朗日函数来制定的，利用最大化对偶函数就可以得到原问题的最优解。

Fioretto[28]提出了一种深度学习的最优交流问题(OPF)方法。学习模型利用系统类似状态下可用的信息，以及双重拉格朗日方法，以满足 OPF 中存在的物理和工程约束。

一些研究已经将拉格朗日对偶法应用于隐私与公平平衡的问题。例如，Tran 等[29]提出了一个基于拉格朗日对偶方法的深度学习模型，它平衡了隐私保护和公平性，提高了模型的有效性和可用性。

根据先前研究的经验和推广，通过在拉格朗日函数中引入隐私和公平作为约束条件，隐私性和公平性的问题可以在小样本学习中得到平衡。

## 3 基于拉格朗日对偶的隐私和公平性约束算法

本文采用拉格朗日对偶方法，将公平性约束添加到神经网络的训练周期中，并提出了一种差分隐私和公平性约束的学习算法。本文的重点是在满足群体公平的前提下，使用差分隐私的概念防止敏感属性的泄露。

### 3.1 样本级自适应 Rényi 隐私过滤算法

当前研究使用单个组合算法中最差隐私损失，即最大损失预算。但是由于在梯度下降算法中，样本梯度范数远小于最大值，因此现有组合定理隐私预算的计算方式过于保守。为解决样本级隐私损失组合问题，需实现自适应选择隐私参数组合。本文利用 RDP 作为隐私过滤器实现完全自适应隐私组合，将其转换为 $(\varepsilon,\delta)$-DP 时，会在边界中产生一个额外的 $\sqrt{\log(1/\delta)}$ 因子。

**定理 2** 固定 $B\geqslant0$，$\alpha\geqslant1$。假设 $M_t$ 是 $(\alpha,B)$-RDP，其中 $\rho_t$ 是 $m_1,m_2,\cdots,m_{t-1}$ 中的任意函数。

若 $\sum_{t=1}^{k}\rho_t\leqslant B$，则自适应组合 $M_1,M_2,\cdots,M_{t-1}$ 满足 $(\alpha,B)$-RDP。

当给定个体隐私过滤器固定的隐私预算时，若个体隐私损失超过隐私预算，隐私过滤器将会自适应地删除数据。实验中无需跟踪整体的隐私损失，只需要对数据集中的单个样本进行估计。

若 $Data=(x_1,x_2,\cdots,x_n)$ 表示数据集，则用 $Data^{-i}=(x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_n)$ 表示除去 $x_i$ 的数据集。

**定义 7** 两个测度 $\mu$ 和 $\nu$ 之间的阶数 $\alpha\in(1,\infty)$，满足 $\mu\ll\nu$ 的 Rényi 散度表示为：

$$D_\alpha(\mu\|\nu)=\frac{1}{\alpha-1}\log\int\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right)^\alpha\mathrm{d}\nu \tag{9}$$

$D_\alpha(M(Data)\|M(Data^{-i}))$ 表示任意算法 $M$ 在 $Data$ 与 $Data^{-i}$ 两个数据集上的输出分布之间的距离。$m$ 为从 $Data$ 的输出分布中随机抽样，表示为 $m\sim M(Data)$。

$$D_{\alpha}^{\leftrightarrow}(\mu \| \nu)=\max(D_{\alpha}(\mu \| \nu) \| D_{\alpha}(\nu \| \mu)) \tag{10}$$

$D_{\alpha}^{\leftrightarrow}(\mu \| \nu)$表示 Rényi 散度两个方向的最大值。

**定理 3** 对于数据集 $Data$，在 $Data^{-i}$ 中，若有：

$$D_{\alpha}^{\leftrightarrow}(M(Data) \| M(Data^{-i}))\leqslant\rho \tag{11}$$

对于 $i\in n$，则认为随机算法 $M$ 满足$(\alpha,B)$-RDP。RDP 表示可以随时转换为 DP 表示，如定义 8 所示。

**定义 8** 若随机算法 $M$ 满足$(\alpha,\rho)$-RDP，则对于任意 $\delta\in(0,1)$，将 RDP 转换为 DP 表示为：

$$\left(\rho+\frac{\log(1/\delta)}{\alpha-1},\delta\right)\text{-DP}$$

**定义 9** 对于数据集 $Data$ 中的所有数据，若：

$$D_{\alpha}^{\leftrightarrow}(M(Data) \| M(Data^{-i}))\leqslant\rho \tag{12}$$

对于 $i\in n$，则称随机算法 $M$ 满足$(\alpha,\rho)$-IRDP。

**定义 10** 对于数据集 $Data$，在 $Data^{-i}$ 中，若有：

$D_{\alpha}^{\leftrightarrow}(M(Data) \| M(Data^{-i}))\leqslant\rho$，则称随机算法 $M$ 对于每个数据 $x_i$ 是$(\alpha,\rho)$-IRDP。

因此，为了满足 IRDP，随机算法 $M$ 需要数据集中所有数据点都满足 IRDP。

**定义 11** 固定 $B\geqslant0$，$\alpha\geqslant1$。对于任意随机算法序列 $M_1$，$M_2$，…，$M_k$，若 $\sum_{t=1}^{k}\rho_t\leqslant B$，则自适应组合随机算法 $M^{(k)}$ 满足：

$$D_{\alpha}^{\leftrightarrow}(M(Data) \| M(Data^{-i}))\leqslant\rho \tag{13}$$

其中，$\rho_1$，…，$\rho_k$ 定义为：

$$\rho_t=\max D_{\alpha}^{\leftrightarrow}(M_t(m_1,m_2,\cdots,m_{t-1},Data) \| M_t(m_1,m_2,\cdots,,m_{t-1},Data^{-i})) \tag{14}$$

**算法 1** 基于 Rényi 隐私过滤器的自适应组合算法

1. while k<N do
2. 计算 $\rho_{k+1}$
3. if $F_{\alpha,\beta}(\rho_1,\cdots,\rho_{k+1})$=Enough(足够)
4. 中止
5. end
6. 计算 $m_{k+1}=M_k(m_1,m_2,\cdots,m_k,Data)$
7. k←k+1
8. end

**定义 12**(RDP 过滤器) 固定参数 $\alpha\geqslant1$ 和隐私预算 $B$。若对任意随机算法序列 $(M_k)_{k=1}^{N}$ 和任意数据集对 $(Data_n,Data_n^{-i})$，自适应组合随机算法 $M^{(k)}$ 在给定算法 1 时满足：

$$D_{\alpha}^{\leftrightarrow}(M^{(k)}(Data_n) \| M^{(k)}(Data_n^{-i}))\leqslant\rho \tag{15}$$

即定义为 Rényi 隐私过滤器。

$$F_{\alpha,\beta}\rightarrow\{Continue,Enough\}$$

$$F_{\alpha,\beta}(\rho_1,\cdots,\rho_{k+1})=\begin{cases}Continue, & \text{if } \sum_{t=1}^{k}\rho_t\leqslant B\\ Enough, & \text{if } \sum_{t=1}^{k}\rho_t>B\end{cases} \tag{16}$$

**算法 2** 基于 RDP 过滤器的隐私梯度下降(IRDP-SGD)

1. 初始化参数 $\theta_1$
2. for t=1,2,…,$k_{max}$ do
3. ∀i，计算梯度 $g_t(X_i)\leftarrow\nabla_{\theta}l(\theta_t;X_i)$
4. ∀i，梯度裁剪 $\bar{g}_t(X_i)$
5. 添加噪声 $\tilde{g}_t(X_i)$
6. 梯度下降 $\theta_{t+1}\leftarrow\theta_t-\eta_t\tilde{g}_t$
7. end

差分隐私随机梯度下降算法在计算随机采样样本的梯度后，剪裁梯度范数并添加高斯噪声，并引入隐私会计跟踪隐私损失。随机梯度下降算法截止时，将统计并耗尽剩余的隐私预算。

**算法 3** 自适应协作方法

1. 初始化参数 $\theta_A$ 和 $\theta_B$，并交换公匙
2. for 任务 A 和任务 B do
3. 采样任意 batch 的任务 $T_{A,i}\sim p(T)$，$T_{B,i}\sim p(T)$
4. for 所有 $T_{A,i}$，$T_{B,i}$ do
5. 从 $T_{A,i}$和 $T_{B,i}$中采样 K 个数据，组成支撑集 $D_A$ 和 $D_B$
6. 使用 IRDP-SG 模块计算 $\theta_A$
7. 使用 SGD 更新参数 $\theta'_B$
8. 从 $T_{A,i}$和 $T_{B,i}$中采样 K 个数据，组成支撑集 $D_A'$和 $D_B'$
9. end for
10. 分别计算 $L_1$，$L_2$
11. 生成$[L_1]_B$，$[L_2]_B$ 并上传到中央服务器
12. end for
    计算掩码 $K_B$ 的全局损失$[L]_B$
13. for 任务 B do
14. 从中央服务器下载$[L]_B$，并解码为 L
15. 更新全局参数 $\theta_B=\theta_B-\beta\nabla_{\theta_B}L$
16. 掩码 $K_B$ 生成$[\theta_B]_A$ 并上传至中央服务器
17. end for
18. 从中央服务器下载$[\theta_B]_A$，并用 $\theta_A$ 解码$[\theta_B]_A$

### 3.2 基于自适应协作模型的总体流程

如图 1 所示，在基于自适应协作模型中展示了本地参数与全局参数(中央服务器)在每轮学习任务中的隐私迭代过程。设 $D_A$ 和 $D_B$ 分别为两个本地数据集，由于隐私问题，$D_A$ 和 $D_B$ 无法直接访问，模型中通过将本地参数上传到中央服务器聚合为全局参数。$D_A$ 通过 IRDP-SGD 模型获得本地隐私损失 $\theta_A$，并将其上传到中央服务器，$D_B$ 通过 SGD 模型获得本地隐私损失 $\theta_B$，并将其上传到中央服务器，然后中央服务器聚合为全局损失。$D_B$ 从中央服务器下载并利用全局损失进一步更新全局参数并再次上传，进而帮助 $D_A$ 获得更好的全局参数。在整个传输过程中采用同态加密对隐私参数进行加密，避免隐私泄露的风险。在总体模型中，不断循环迭代，直至模型收敛。

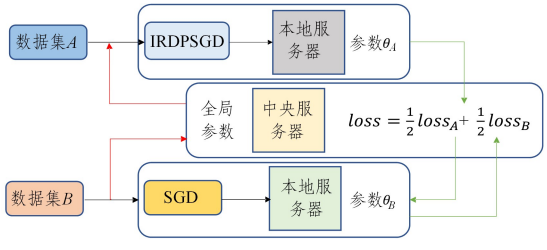



图 1 自适应模型的总体流程

Fig. 1 Overall flow of adaptive model

### 3.3 基于拉格朗日对偶的隐私和公平约束算法

本文将隐私保护和公平性作为优化问题的约束条件。引入拉格朗日乘子将约束条件纳入优化问题，每个约束条件都

被赋予一个拉格朗日乘子，用于权衡约束条件的重要性。根据约束条件和拉格朗日乘子，构建拉格朗日函数。

模型中拉格朗日函数综合了目标函数、隐私保护和公平性约束，并通过拉格朗日乘子对约束条件进行加权。对拉格朗日函数进行优化，使用拉格朗日对偶优化方法来解决问题。通过得到拉格朗日乘子的最优值，进而获得隐私和公平性的最佳权衡。

公平性度量（反歧视水平）定义为：

$$\Gamma(D,f_w)=|\gamma(D_0)-\gamma(D_1)| \tag{17}$$

在本模型中，将公平性度量应用在损失函数中作为约束条件，从而转变为经验风险最小化问题。本研究中利用拉格朗日对偶函数实现，用$|\Gamma|$表示约束，约束条件为：

$$\gamma(D_0)-\gamma(D_1)=0^{\mathrm{T}} \tag{18}$$

其中，$\theta$为差分隐私模型参数，$f$表示分类器，则优化问题表示为：

$$\begin{aligned}&\arg\min_{\theta} J(f_\theta,D)=\frac{1}{n}\sum_{i=1}^{n}L(f_\theta(x_i),y_i)\\&\text{s.t. } \gamma(D_0)-\gamma(D_1)=0^{\mathrm{T}}\end{aligned} \tag{19}$$

引入拉格朗日乘子，在拉格朗日松弛过程中，使用与约束$|\Gamma|$相关的拉格朗日乘子$\lambda_i\geqslant 0$将问题约束松弛为目标函数。当约束都被放宽后，拉格朗日函数变为：

$$L_\lambda(\omega)=J(f_\theta,D)+\lambda^T|\gamma(D_0)-\gamma(D_1)| \tag{20}$$

其中，$\lambda^{\mathrm{T}}=(\lambda_1,\cdots,\lambda_i)$。

使用拉格朗日函数，优化转换之后为：

$$\overline{\overline{\lambda}}=\arg\min L_\lambda(f_{\tilde{\theta}(\lambda)},D)_{\lambda\geqslant 0} \tag{21}$$

拉格朗日对偶将会找到最好的拉格朗日乘子，来得到最强拉格朗日松弛。首先，确定优化器的步长$\eta_t$和$s_t$，将拉格朗日乘子初始化。在一个epoch中，利用batch梯度下降在原始更新步骤优化参数$\theta$。利用拉格朗日乘子$\lambda_k$在优化步骤更新参数$\theta$。在对偶更新步骤，通过对偶上升更新拉格朗日乘子，最终得到最佳乘子，小于预定义上界$\lambda^{\max}$。

该框架通过提供约束梯度和约束违反的预期误差的界限来解决裁剪的偏差-方差权衡问题，然后通过最小化这些上界来校准裁剪边界。

**算法4** 基于拉格朗日对偶的隐私和公平性约束算法（PF-LD）

1. 初始化参数 $\theta_{A,t=1,i}$
2. for t=1,2,…,$k_{max}$ do
3. for mini-batch B do(采样概率为 q)
4. 计算公平梯度 $g_t(x_i)$
5. 梯度裁剪 $\tilde{g}_t(X_i)$
6. 添加噪声 $\widetilde{g}_t(X_i)$
7. 梯度下降 $\theta_{A,t+1}$
8. end for
9. 更新 $\lambda_{t+1}$
10. $\lambda_{t+1}\leftarrow\min(\lambda^{\max},\lambda_{t+1,i})$
11. end for

在拉格朗日对偶模块中输入$Data=(x_1,x_2,\cdots,x_n)$，学习率分别为$\eta_t$和$s_t$，batch大小为$B$，噪声方差$\sigma>0$，裁剪值$\Delta>0$，总步长$K_{\max}\in N$，$L_2$范数预算$B_{\mathrm{norm}}>0$，最大乘数值为$\lambda^{\max}$，采样概率为$q=\frac{B}{Data}$。在算法中，公平性约束的$A$表示敏感属性集，参数$\theta_A$表示算法3中的参数$\theta_A$。

### 3.4 算法中的隐私性保证

隐私要求是对原始和双步骤使用裁剪方法，并添加由约束项及其梯度的灵敏度校准的噪声来实现的。原始步骤仅对涉及敏感属性的约束梯度应用裁剪，因此对模型精度没有重大影响。

公平性度量的两个条件GF和EO可以认为，在每个群组中访问敏感数据$A$时，相比总体的约束，约束条件的等式表示为：

$$E_{D_0}[\gamma(D_0)]-E_{D_A}[\gamma(D_A)]=0 \tag{22}$$

其中，$D_0$为总体，$D_A$表示敏感属性群组。当用$B_0$和$B_A$分别表示每个batch的总体和其中的敏感属性群组时，有：

$$E_{D_0}[\gamma(B_0)]-E_{D_A}[\gamma(B_A)]=0 \tag{23}$$

损失函数的梯度的计算式由两项组成，在$t$时分别为原始梯度$\nabla_\theta l(\theta_t;x_i)$和公平项约束$\nabla_\theta(\lambda^{\mathrm{T}}|\gamma(D_0)-\gamma(D_1)|)$。在模型中利用引入的高斯噪声（如IRDP-SGD）来实现差分隐私。在IRDP-SGD的batch中以采样率$q$计算单个样本的梯度，并通过算法自适应地裁剪范数，引入噪声使隐私得到保证。

## 4 实验分析

### 4.1 公平性分析

数据集采用的是UCI Adult数据集和CeleA数据集。UCI Adult数据集为成年人收入预测，需要消除性别偏见，数据集中以性别为敏感属性。在CeleA数据集的众多属性中，本实验以波浪发型为敏感属性。在UCI Adult数据集中，使用模型为两层全连接层，隐藏层大小为200。实验中数据集以60%，20%和20%的比例划分为训练集、测试集和验证集。训练模型10次，其中batch size设置为1 000，利用在验证集上的效果选择模型，模型学习率为$1\times e^{-3}$，采用Adam优化器。

在CelebA数据集中，模型使用ResNet-18网络，后接两层全连接层进行预测。模型学习率为$1\times e^{-3}$，batch size设置为512，dropout为0.2，采用Adam优化器。

度量标准如下：

$$GF=\frac{\Pr(\tilde{y}=1|A=0)}{\Pr(\tilde{y}=1|A=1)} \tag{24}$$

$$ACC=\frac{True\ Positive+True\ Negative}{False\ Positive+False\ Negative+True\ Positive+True\ Negative} \tag{25}$$

$$\begin{aligned}EO=\{&\Pr(\tilde{y}=1|A=0,y=1)-\Pr(\tilde{y}=1|A=1,y=0)+\\&\Pr(\tilde{y}=1|A=0,y=0)-\Pr(\tilde{y}=1|A=1,y=0)\}\end{aligned} \tag{26}$$

表1列出了在Adult数据集上，用于对比实验的两种公平性方法（RNF[30]，EOR[31]），以及本研究提出的拉格朗日对偶方法L-D的GF效果。通过表中的数据可以看到，在实验中持续增大GF时，会导致模型准确率下降。除去初始化时RNF方法中最大GF值为0.862，本实验的L-D在第四轮中GF为0.939，ACC为81.7。在提高模型公平性的同时，模型仍然保持较高的ACC。表2中，使用EO为公平性度量，初始化阶段RNF方法的效果最好，但随着公平性的

提升，L-D 方法的表现最佳。

表 1 Adult 数据集上的 GF 效果(GF&ACC)

Table 1 GF effect on Adult dataset(GF&ACC)

| RNF | | EOR | | L-D | |
|---|---|---|---|---|---|
| **0.862** | 82.5 | 0.835 | 83.0 | 0.848 | **83.1** |
| 0.891 | 81.6 | 0.836 | 82.4 | **0.901** | **82.9** |
| 0.918 | 81.5 | 0.847 | 81.8 | **0.936** | **82.2** |
| 0.936 | 81.4 | 0.869 | 81.1 | **0.939** | **81.7** |

表 2 Adult 数据集上的 EO 效果(GF&ACC)

Table 2 EO effect on Adult dataset(GF&ACC)

| RNF | | EOR | | L-D | |
|---|---|---|---|---|---|
| **−0.096** | 82.3 | −0.099 | 83.4 | −0.115 | **83.5** |
| −0.074 | 82.1 | **−0.071** | 82.7 | −0.083 | **83.0** |
| **−0.043** | 81.9 | −0.049 | 82.1 | −0.057 | **82.7** |
| −0.036 | 80.6 | −0.034 | 80.8 | **−0.029** | **81.7** |

由表 3 和表 4 可以看到，在初始化阶段 RNF 和 EOR 方法有较好表现，而在后续实验中，随着公平性提高，L-D 方法的效果更加突出。在两个数据集中的实验均证明，本研究提出的拉格朗日对偶方法在公平性和准确性方面取得了最佳效果。

表 3 CelebA 数据集上的 GF 效果(GF&ACC)

Table 3 GF effect on CelebA dataset(GF&ACC)

| RNF | | EOR | | L-D | |
|---|---|---|---|---|---|
| 0.720 | **77.5** | **0.721** | 76.8 | 0.714 | 75.6 |
| 0.758 | **75.9** | 0.750 | 74.2 | **0.760** | 74.3 |
| 0.784 | 72.1 | 0.768 | 71.8 | **0.809** | **72.9** |
| 0.814 | 71.1 | 0.773 | 69.5 | **0.832** | **72.4** |

表 4 CelebA 数据集上的 EO 效果(GF&ACC)

Table 4 EO effect on CelebA dataset(GF&ACC)

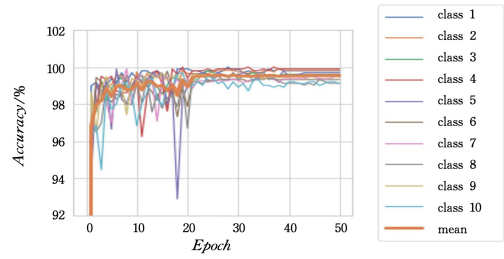
| RNF | | EOR | | L-D | |
|---|---|---|---|---|---|
| **−0.409** | 77.1 | −0.416 | 76.2 | −0.414 | **83.5** |
| **−0.380** | 73.8 | −0.403 | 75.5 | −0.378 | **83.0** |
| −0.376 | 77.9 | −0.378 | 73.6 | **−0.347** | **82.7** |
| −0.300 | 71.3 | −0.321 | 71.5 | **−0.293** | **81.3** |

### 4.2 隐私性分析

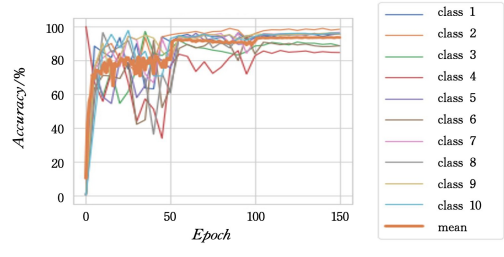
本实验构建 MNIST&colored 数据集，colored MNIST 数据集与 MNIST 数据集的类别相同，是彩色图像。将两个数据集混合，分别从两个数据集中随机挑选 5 个类别按照彩色图与灰度图 1∶9 的比例混合，另外挑选 5 个类别按照彩色图与灰度图 9∶1 的比例混合，将颜色设置为敏感属性，标记类别为 Class $x$，$x=(1,2,\cdots,10)$。将数据集样本扩充为 32×32 像素。

CIFAR-10 数据集包含 10 类彩色图像，标记为 Class $x$，$x=(1,2,\cdots,10)$。下文将分析 DP 对每个类别的影响。

图 2 给出了两个数据集在前期实验中，模型准确率的变化。由于引入噪声，训练过程中准确率的波动十分明显，在训练的早期阶段，各个类别的准确率波动很大，在学习率较小的训练后期阶段，仍存在类间的差异，但波动显著缩小。例如，对于 CIFAR-10 来说，各个类别波动趋于平稳，但 MNIST 数据集相比其他类别 Class 10 在训练后期仍有较大波动。因此，根据各个类别的准确率不同，需要分析 DP 对每个具体类别的影响。




(a) MNIST




(b) CIFAR10

图 2 实验数据集的准确率

Fig. 2 Accuracy of experimental datasets

在 MNIST 和 CIFAR-10 两个数据集的实验中隐私预算设置相同，$\varepsilon=k, k\in(2,4,6,8,10)$，$\delta=10^{-5}$。对比 IRDP-SGD 和 DP-SGD 两种方法的效果。在 MNIST 数据集中，先将样本扩充为 32×32 像素。如图 3 所示，本文提出的 IRDP-SGD 方法在不同的隐私预算下的效果均优于 DP-SGD，且随着隐私预算的增加，模型效果有显著提升。图 4 中，在 CIFAR-10 数据集下，IRDP-SGD 方法整体优于 DP-SGD，结论与上述相同。通过实验证明，IRDP-SGD 在隐私保护和模型效用上整体优于 DP-SGD 方法。

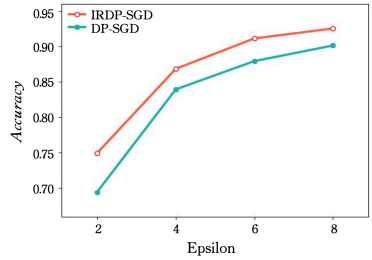



图 3 MNIST 数据集上 DP-SGD 和 IRDP-SGD 的对比

Fig. 3 Comparison of DP-SGD and IRDP-SGD on MNIST dataset

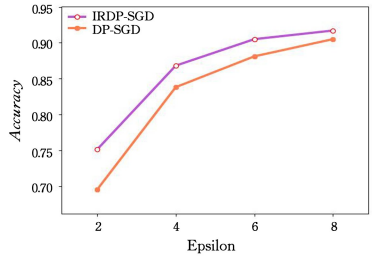



图 4 CIFAR-10 数据集上 DP-SGD 和 IRDP-SGD 的对比

Fig. 4 Comparison of DP-SGD and IRDP-SGD on CIFAR-10 dataset

### 4.3 小样本分析

小样本学习常用图像数据集为 Mini-ImageNet，其包含

100 类共 6000 张 84×84 彩色图。其中训练类别 64 个，验证类别 12 个以及测试类别 24 个。本实验中根据 Mini-ImageNet 构建相应数据集 Mini-ImageNet-COLOR。Mini-ImageNet-COLOR 中，将彩色图像转换为灰度图，通过混合彩色图和灰度图建立敏感属性。将数据集的 100 个类别分为两部分，第一部分的 50 个类别中彩色图与灰度图的比例为 9∶1，第二部分中彩色图与灰度图的比例为 1∶9。其他设置与 Mini-ImageNet 图像一致，实验中以色彩为敏感属性。

实验采用 5 way-5 shot 模型，实验结果表 5 所列。其中 $\varepsilon=k, k\in(2,4,6,8,10)$，$\delta=10^{-5}$ 代表隐私使用 DP-SGD 的训练方法，作为实验基准 $\delta=10^{-5}$。IRDP 表示不加入 L-D 的情况，隐私使用 IRDP-SGD 算法进行训练。L-D 为隐私使用 DP-SGD 进行训练。PF-LD 表示使用 L-D(IRDP-SGD)训练。SC-Reptile[32](SR)为一种安全协作小样本学习框架，用作本文方法的对比方法。

根据图 5 以及表 5 中的数据对比，本实验提出的 PF-LD 与其他实验方法相比，在 GF 以及 EO 方面都达到了领先水平。随着隐私预算的增加，DF-LD 方法受到的影响最小，在隐私和公平的平衡方面取得了良好的效果。在 ACC 的比较中，本文中的 DF-LD 方法与 IRDP 方法均有明显优势，在模型的准确性方面保持了良好的结果。根据实验对比，DF-LD 方法在隐私性、公平性和模型实用性方面达到了领先水平。

表 5 Mini-ImageNet-COLOR 数据集上 5 way-5 shot 模型的效果

Table 5 5 way-5 shot model effect on Mini-ImageNet-COLOR dataset

| | *GF* | *EO* | *ACC*/% |
|---|---|---|---|
| $\varepsilon=2$ | 0.396 | −0.720 | 60.8 |
| IRDP | 0.401 | −0.699 | **61.1** |
| L-D | 0.407 | −0.693 | 59.3 |
| PF-LD | **0.411** | **−0.675** | 61.0 |
| SR | 0.402 | −0.718 | 61.1 |
| $\varepsilon=4$ | 0.418 | −0.666 | 61.9 |
| IRDP | 0.422 | −0.643 | 61.9 |
| L-D | 0.428 | −0.625 | 60.5 |
| PF-LD | **0.438** | **−0.619** | **62.1** |
| SR | 0.420 | −0.731 | 61.7 |
| $\varepsilon=6$ | 0.447 | −0.594 | 62.3 |
| IRDP | 0.461 | −0.528 | **62.6** |
| L-D | 0.479 | −0.478 | 60.9 |
| PF-LD | **0.490** | **−0.462** | 62.3 |
| SR | 0.439 | −0.608 | 61.9 |
| $\varepsilon=8$ 0.483 | −0.565 | 62.8 | |
| IRDP | 0.506 | −0.475 | **63.9** |
| L-D | 0.525 | −0.441 | 61.8 |
| PF-LD | **0.548** | **−0.401** | 63.6 |
| SR | 0.477 | −0.592 | 62.8 |
| $\varepsilon=10$ | 0.498 | −0.511 | 63.9 |
| IRDP | 0.543 | −0.468 | **65.3** |
| L-D | 0.641 | −0.420 | 63.7 |
| PF-LD | **0.689** | **−0.361** | 65.2 |
| SR | 0.504 | −0.496 | 63.4 |

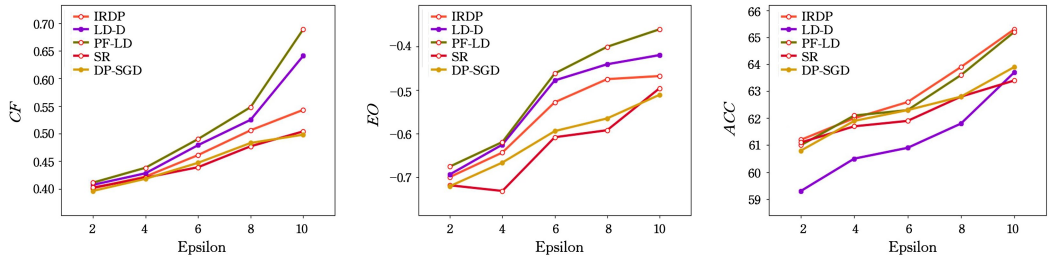



图 5 Mini-ImageNet-COLOR 数据集上 5 way-5 shot 模型效果中 EO/GF/ACC 标准下的对比

Fig. 5 Comparison of 5 way-5 shot model effects under EO/GF/ACC criteria on Mini-ImageNet-COLOR dataset

## 5 实验总结

在公平性分析、隐私性分析以及小样本分析实验中，对实验结果进行对比分析。将本文中的模型与现有模型进行进一步比较，证明本文模型在隐私性以及公平性方面具有显著优势，并且引入噪声后模型效用依然表现良好。通过对本文多轮实验的总结，得出以下结论。

1)IRDP 方法在准确率方面表现最佳，证明 IRDP 中提供了更加准确的个体隐私损失的计算。在本模型中，公平性度量标准 GF 和 EO 相比其他方法有所提高，证明 IRDP 对模型公平性有一定提升。

2)实验证明，PF-LD 方法对模型公平性的提升最佳，在准确率方面，由于考虑公平性与隐私性的平衡，准确率略低于 IRDP，但与 IPDP 方法接近。在确定隐私配置下，PF-LD 方法达到了模型隐私性、公平性以及效用的有效平衡。

3)在对比实验 SC-Reptile 中，SC-Reptile 模型在公平性和效用方面表现较差。探究其原因，可能是在 DP-SGD 方法中引入了噪声，使其对模型公平性产生负面影响。

**结束语** 本文提出了 PF-LD 方法，并在实验中证明了该方法在隐私性、公平性以及模型效用方面取得良好效果，达到三者的最佳平衡。该方法利用 Rényi 差分隐私提供更加精确的隐私损失计算，构建 Rényi 差分隐私过滤器，避免了因采用最大隐私损失计算给模型带来的负面影响。利用公平性度量构建约束条件，通过拉格朗日对偶方法寻求最强拉格朗日松弛，实现隐私性与公平性的平衡。

在隐私保护和公平性的研究中仍然面临许多挑战。本文方法在神经网络模型的隐私性和公平性保护方面效果明显，以本文方法为基础，将进一步寻求更加精确的隐私损失计算方法，提升神经网络模型的隐私性和公平性。

## 参 考 文 献

[1] CHANG H, SHOKRI R. On the privacy risks of algorithmic fairness[C]// 2021 IEEE European Symposium on Security and Privacy(EuroS&P). IEEE, 2021: 292-303.

[2] FARRAND T, MIRESHGHALLAH F, SINGH S, et al. Neither private nor fair: Impact of data imbalance on utility and fairness in differential privacy[C]// Proceedings of the 2020 Workshop

on Privacy-preserving Machine Learning in Practice. Association for Computing Machinery, 2020: 15-19.
[3] LIN Y, BAO L Y, LI Z M H, et al. Differential privacy protection over deep learning: An investigation of its impacted factors [J]. Computers & Security, 2020, 99: 102061.
[4] GARCIA V, BRUNA J. Few-shot learning with graph neural networks[J]. arXiv: 1711. 04043, 2017.
[5] MIRONOV I. Rényi differential privacy[C]// 2017 IEEE 30th Computer Security Foundations Symposium. IEEE, 2017: 263-275.
[6] DWORK C. Differential privacy [C]// Proceedings of the 33rd International Colloquium on Automata, Languages and Programming(ICALP), Part II. 2006: 1-12.
[7] ABADI M, CHU A, GOODFELLOW I, et al. Deep learning with differential privacy[C]// Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. ACM, 2016: 308-318.
[8] FELDMAN V, ZRNIC T. Individual privacy accounting via a Rényi filter [C] // Advances in neural information processing systems. New York: Curran Associates, Inc. 2021: 28080-28091.
[9] DU M N, YANG F, ZOU N, et al. Fairness in deep learning: a computational perspective[J]. IEEE Intelligent Systems, 2021, 4(36): 25-34.
[10] ZHANG T, ZHU T, LI J, et al. Fairness in semi-supervised learning: Unlabeled data help to reduce discrimination[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 34(4): 1763-1774.
[11] ZAFAR M B, VALERA I, ROGRIGUEZ M G, et al. Fairness constraints: Mechanisms for fair classification[C]// Artificial Intelligence and Statistics. PMLR, 2017: 962-970.
[12] LIANG Y, CHEN C, TIAN T, et al. Joint Adversarial Learning for Cross-domain Fair Classification [J]. arXiv: 2206. 03656, 2022.
[13] DU M, YANG F, ZOU N, et al. Fairness in deep learning: A computational perspective[J]. IEEE Intelligent Systems, 2020, 36(4): 25-34.
[14] PENG Y C, QIN X I, ZHANG L G, et al. Survey on Few-shot Learning Algorithms for Image Classification [J]. Computer Science, 2022, 49(5): 1-9.
[15] LU J Y, LING X H, LIU Q, et al. Meta-reinforcement Learning Algorithm Based on Automating Policy Entropy[J]. Computer Science, 2021, 48(6): 168-174.
[16] VINYALS O, BLUNDELL C, LILLICRAP T, et al. Matching networks for one shot learning[J]. Advances in Neural Information Processing Systems, 2016, 29: 3630-3638.
[17] TRIANTAFILLOU E, ZHU T, DUMOULIN V, et al. Meta-dataset: A dataset of datasets for learning to learn from few examples[J]. arXiv: 1903. 03096, 2019.
[18] TSIMPOUKELLI M, MENICK J L, CABI S, et al. Multimodal few-shot learning with frozen language models[J]. Advances in Neural Information Processing Systems, 2021, 34: 200-212.
[19] SNELL J, SWERSKY K, ZEMEL R. Prototypical networks for few-shot learning[J]. Advances in Neural Information Processing Systems, 2017, 30: 4077-4087.
[20] RAVI S, LAROCHELLE H. Optimization as a model for few-shot learning[C]// International Conference on Learning Representations. 2017.
[21] JAMAL M A, QI G J. Task-agnostic meta-learning for few-shot learning[C]// Conference on Computer Vision and Pattern Recognition. IEEE, 2019: 11719-11727.
[22] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[C]// International Conference on Machine Learning. PMLR, 2017: 1126-1135.
[23] ABBAS M, XIAO Q, CHEN L, et al. Sharp-MAML: Sharpness-Aware Model-Agnostic Meta Learning[C]// International Conference on Machine Learning. PMLR, 2022: 10-32.
[24] LI J, KHODAK M, CALDAS S, et al. Differentially private meta-learning[J]. arXiv: 1909. 05830, 2019.
[25] SLACK D, FRIEDLER S A, GIVENTAL E. Fairness warnings and Fair-MAML: learning fairly with minimal data[C]// Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. ACM, 2020: 200-209.
[26] ZHAO C, CHEN F, WANG Z, et al. A primal-dual subgradient approach for fair meta learning[C]// 2020 IEEE International Conference on Data Mining(ICDM). IEEE, 2020: 821-830.
[27] JAGIELSKI M, KEARNS M, MAO J M, et al. Differenially private fair learning [C] // International Conference on Machine Learning. PMLR, 2019: 3000-3008.
[28] FIORETTO F, MAK T W K, VAN HENTENRYCK P. Predicting ac optimal power flows: Combining deep learning and lagrangian dual methods[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2020: 630-637.
[29] TRAN C, FIORETTO F, VAN HENTENRYCK P. Differentially private and fair deep learning: A lagrangian dual approach [C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(11): 9932-9939.
[30] DU M, MUKHERJEE S, WANG G, et al. Fairness via representation neutralization[J]. Advances in Neural Information Processing Systems, 2021, 34: 12091-12103.
[31] BECHAVOD Y, LIGETT K. Penalizing unfairness in binary classification[J]. arXiv: 1707. 00044, 2017.
[32] XIE Y, WANG H, YU B, et al. Secure collaborative few-shot learning[J]. Knowledge-Based Systems, 2020, 203: 106157.

**WANG Jinghong**, born in 1967, Ph. D, professor, academic advisor, is a member of CCF(No. 58341S). Her main research interests include machine learning, data mining, and artificial intelligence.

**WANG Wei**, born in 1982, Ph. D, associate professor, academic advisor, is a member of CCF (No. 51382M). His main research interests include machine learning, knowledge representation, and virtual simulation.

(责任编辑:喻藜)