

论文标题

徐文杰^{a,b*}, 第二作者^a, 第三作者^b

^a 中国科学院科技战略咨询研究院, 北京 100190

^b 中国科学院大学公共政策与管理学院, 北京 100049

摘要 在此撰写摘要。摘要应简要概括论文的主要贡献、方法和结果，通常在 150–250 字之间。

关键词: 关键词 1; 关键词 2; 关键词 3; 关键词 4

1 引言

在此撰写引言部分。介绍研究背景、问题描述和论文结构 [1]。

本模板支持中英文混排 (Chinese and English mixed typesetting)，数学公式和术语可以自然切换。

主要贡献包括：

- 贡献 1
- 贡献 2
- 贡献 3

2 相关工作

综述相关文献，将你的工作定位在现有研究中。

3 方法

3.1 问题建模

描述问题建模。关键公式应编号：

$$f(x) = \sum_{i=1}^n \alpha_i \cdot g_i(x) \tag{1}$$

定义 3.1 (术语名称). 在此给出定义。

定理 3.2 (定理名称). 在此陈述定理内容。

引理 3.3 (引理名称). 在此陈述引理内容。

3.2 算法

Algorithm 1 展示了主要算法流程。

4 实验

4.1 实验设置

描述数据集、基线方法和评价指标。

* 通讯作者: xuwenjie@example.edu.cn

Algorithm 1: 算法名称

Input: 输入参数
Output: 输出结果
1 步骤 1 描述
2 **for** 每次迭代 **do**
3 更新步骤
4 **return** 最终结果

表 1: 数据集统计

指标	数据集 A	数据集 B
样本数	XXX	XXX
特征数	XXX	XXX
类别数	XXX	XXX

表 2: 性能对比

方法	指标 1	指标 2	指标 3
基线 1	XX.X	XX.X	XX.X
基线 2	XX.X	XX.X	XX.X
本文方法	XX.X	XX.X	XX.X

4.2 实验结果

Table 2 展示了对比结果。结果表明本文方法优于所有基线方法。

5 结论

总结主要发现并讨论未来工作方向。

附录 A 补充证明

本节提供文中定理的详细证明过程。

A.1 定理 1 的详细证明

定理 A.1 (收敛性定理). 假设目标函数 $f(x)$ 满足 L -光滑条件, 且学习率 $\eta \leq \frac{1}{L}$, 则算法 I 保证在 T 次迭代后收敛:

$$\min_{t \in [T]} \|\nabla f(x_t)\|^2 \leq \frac{2(f(x_0) - f^*)}{\eta T} \quad (2)$$

证明. 根据 L -光滑性的定义, 对于任意 x, y 有:

$$f(y) \leq f(x) + \nabla f(x)^\top (y - x) + \frac{L}{2} \|y - x\|^2 \quad (3)$$

令 $y = x_{t+1} = x_t - \eta \nabla f(x_t)$, 代入上式得:

$$f(x_{t+1}) \leq f(x_t) - \eta \|\nabla f(x_t)\|^2 + \frac{L\eta^2}{2} \|\nabla f(x_t)\|^2 \quad (4)$$

$$= f(x_t) - \eta \left(1 - \frac{L\eta}{2}\right) \|\nabla f(x_t)\|^2 \quad (5)$$

当 $\eta \leq \frac{1}{L}$ 时, $1 - \frac{L\eta}{2} \geq \frac{1}{2}$, 因此:

$$f(x_{t+1}) \leq f(x_t) - \frac{\eta}{2} \|\nabla f(x_t)\|^2 \quad (6)$$

对 $t = 0, 1, \dots, T-1$ 求和:

$$f(x_T) \leq f(x_0) - \frac{\eta}{2} \sum_{t=0}^{T-1} \|\nabla f(x_t)\|^2 \quad (7)$$

整理并注意到 $f(x_T) \geq f^*$, 即得证。 \square

附录 B 补充实验

B.1 在不同数据集上的性能对比

Table 3 展示了本文方法在多个公开数据集上的详细性能对比。

表 3: 不同数据集上的性能对比 (准确率%)

方法	CIFAR-10	CIFAR-100	ImageNet	平均
ResNet-50	94.2	74.5	76.1	81.6
ResNet-101	95.1	77.3	78.5	83.6
ViT-B/16	96.2	81.4	80.2	85.9
本文方法	97.1	83.2	82.8	87.7

B.2 超参数敏感性分析

我们对关键超参数进行了敏感性分析。?? 展示了不同学习率和批量大小对模型性能的影响。

实验结果表明, 当学习率在 $[10^{-4}, 10^{-3}]$ 范围内, 批量大小在 $[32, 128]$ 范围内时, 模型性能相对稳定。

参考文献

- [1] First Author and Second Author. A sample paper title. *Journal Name*, 1(1):1–10, 2023.