

张博涵

brandin_zhang@163.com | [Github链接](#) | [Blog链接](#)

北京海淀 | 2023级本科生 | 中国农业大学 (985/双一流A类)



教育背景

中国农业大学 | 人工智能专业

2024 - 至今

GPA: 3.91/4.00

专业排名: 1 / 29

中国农业大学 | 电子信息大类培养

2023 - 2024

GPA: 3.93/4.00

大类排名: 1 / 241

科研经历

学术论文发表

- (CCF-B 录用、第一作者) **Multi-label Twin Support Vector Machine with Probabilistic Output and Its Safe Screening Rule**. Bohan Zhang, Xincheng Lv, Chang Xu, Yitian X
 - 领域背景:** 多标签排序中的大规模核方法长期被视为高精度基线,但在实践中往往受制于高代价的网格化超参搜索与高复杂度 $O(n^3)$ 训练;同时,传统基于几何距离的排序/阈值判决对数据分布与尺度高度敏感,导致鲁棒性不足、性能难以充分发挥。
 - 贡献:** 一方面,我深入挖掘了模型稀疏结构,基于凸分析与变分不等式,设计了一种严格凸的训练与递归调参加速算法,在多数据集上实现稳定 $4\times+$ 加速,并给出不损伤性能的严格数学证明;另一方面,我进一步提出了改进PMLTSVM,以“双超平面+ Sigmoid 校准”的概率化机制替代多标签 TSVM 中缺乏理论支撑的阈值判决,从而获得更稳健的标签预测与排序表现。
 - 本人负责:** 全部数学推导与代码实现,论文绘图,论文写作。
- (CVPR' 2026 Findings 录用、第一作者) **Learning through Creation: A Hash-Free Framework for On-the-Fly Category Discovery**. Bohan Zhang, Weidong Tang, Zhixiang Chi, Yi Jin, Zhenbo Li, Yang Wang, Yanan Wu
 - 领域背景:** On-the-Fly Category Discovery (OCD) 旨在在推理阶段一边识别已知类别,一边从不断到来的数据流中发现新出现的未知类别,离线训练阶段只使用已知类别的监督信号。现有方法要么依赖固定的标签监督,要么借助扩散模型的数据增强来强化骨干网络,但它们都没有显式地把“测试时真正需要的发现任务”纳入训练目标。从根本讲,让一个仅在有限标注数据上优化得到的模型,在推理时去完成一种性质上不同的“发现”目标,是不合理的;这种训练—测试目标的不一致,造成了离线学习阶段与在线发现阶段之间明显的优化错位。
 - 贡献:** 我们提出 Learning through Creation (LTC): 一个完全基于特征、且不依赖哈希的框架,将“新类别意识”直接注入离线学习过程。其核心是一种轻量级的在线伪未知样本生成器,通过核能量最小化与熵最大化 (MKEE) 驱动。不同于以往在训练前一次性生成合成样本的做法,我们的生成器会随模型动态共同演化,以几乎可以忽略的开销按需合成伪新类实例。这些样本再通过带自适应阈值的双重最大间隔目标被纳入训练,从而借助显式“创造”强化模型对未知区域的刻画与检测能力。我们在七个基准上的大量实验表明, LTC 稳定优于现有方法,在全类准确率上带来 1.5% 到 13.1% 的提升。
 - 本人负责:** 主导全部数学推导与代码实现、论文写作。

荣誉奖项

奖学金: 国家奖学金、学业一等奖学金等。

数理竞赛: 全国部分地区大学生物理竞赛一等奖、

全国大学生数学竞赛一等奖(北京赛区)

北京市大学生数学竞赛一等奖(甲类)

全国中学生物理竞赛省级一等奖等等。

Kaggle全球算法竞赛银牌 计算机视觉方向 排名42/1428

项目经历

百度-中国农业大学联培基地第三期菁英班

学习图论, 动态规划等基础算法知识, 在企业导师指导下通过课程夯实机器学习, 计算机视觉基础。

复现大语言模型开源项目: MiniMind.

逐行研究并实验了源码, 系统梳理了RoPE, GQA-attention架构等基础组件的底层实现, 探究了Pretrain SFT LoRA DPO等基本技术的使用, 整理为技术博客两篇: <https://www.cnblogs.com/zbohan/p/19102925>。

