



谷典典

北京大学 计算机学院

研究方向：机器学习系统、分布式机器学习、集群管理、大模型训练

(+86) 130-2002-2611 (电话/微信)

gudiandian1998@pku.edu.cn (邮箱)

<https://gudiandian.github.io> (主页)

教育背景

北京大学 计算机学院

软件研究所 导师：刘譞哲 长聘副教授 指导教师：金鑫 长聘副教授

2020 - 2025 (预期)

计算机软件与理论 博士研究生

北京大学 信息科学技术学院

计算机科学与技术系

2016 - 2020

计算机科学与技术 理学学士

参与项目

工程项目：基于 Kubernetes 的 MLOps 平台——北京大学空间技术与系统研究中心 ([平台 demo](#))

参与平台的开发和部署。该平台可以自动完成算力用量评估、运行环境配置、调度和部署，为开发者简化了深度学习任务的部署流程。目前平台已在北京大学计算中心内部部署测试。平台集成了包括 ElasticFlow 在内的多个学术界领先的深度学习训练调度算法和优化方法，后续还将为实验室更多技术提供落地验证平台。

科研项目：面向长序列训练的大模型训练框架——上海人工智能实验室实习期间完成

将大模型训练的并行维度解耦，重新设计并整合划分空间，在该空间下自动化模拟并搜索得到不超出 GPU 内存限制的高效并行方案；通过通信算子和计算的“选择性交叠”进一步充分利用闲置的计算资源，提升训练效率；通过通信内存池和反碎片化技术进一步提升 GPU 内存利用率。

科研项目：服务器无感知的弹性分布式深度学习模型训练系统——微软亚洲研究院实习期间完成

为开发者抽象了一组“服务器无感知” (serverless) 的 API，屏蔽大量资源配置细节，从而大幅度降低了模型训练任务的开发复杂度；通过调度算法和作业放置算法保证尽可能多的模型训练任务可以在预期截止时间内完成，为开发者提供性能保障。

科研项目：面向深度学习模型训练的低碳集群调度系统

对深度学习训练的能耗建模，利用能耗模型和数据中心电网的实时碳密度数据，对集群中深度学习模型训练作业使用的资源进行调整，从而在集群碳排放量不超过预算的同时优化深度学习训练作业的性能。

论文发表

1. (CCF-A) *ElasticFlow: An Elastic Serverless Training Platform for Distributed Deep Learning* [ASPLOS'23]
Diandian Gu, Yihao Zhao, Yinmin Zhong, Yifan Xiong, Zhenhua Han, Peng Cheng, Fan Yang, Gang Huang, Xin Jin, Xuanzhe Liu
2. (CCF-A) *Rise of Distributed Deep Learning Training in the Big Model Era: From a Software Engineering Perspective* (导师一作，本人二作) [TOSEM'23]
Xuanzhe Liu, Diandian Gu, Zhenpeng Chen, Jinfeng Wen, Zili Zhang, Yun Ma, Haoyu. Wang, Xin Jin
3. (CCF-A) *基于元算子的深度学习框架缺陷检测方法* [计算机学报'22]
谷典典, 石屹宁, 刘譞哲, 吴格, 姜海鸥, 赵耀帅, 马鄂
4. (CCF-A) *Promal: Precise Window Transition Graphs for Android via Synergy of Program Analysis and Machine Learning* [ICSE'22]
Changlin Liu, Hanlin Wang, Tianming Liu, Diandian Gu, Yun Ma, Haoyu Wang, Xusheng Xiao
5. (CCF-B) *Roaming Through the Castle Tunnels: An Empirical Analysis of Inter-app Navigation of Android Apps* [TWEB'20]
Yun Ma, Ziniu Hu, Diandian Gu, Li Zhou, Qiaozhu Mei, Gang Huang, Xuanzhe Liu



6. (CCF-B) *LinkRadar: Assist the Analysis of Inter-app Page Links via Transfer Learning* [CIKM'19]
Diandian Gu, Ziniu Hu, Shangchen Du, Yun Ma
7. (CCF-A 在投) *GreenFlow: A Carbon Efficient Scheduler for Deep Learning Workloads* [TPDS'24]
Diandian Gu, Yihao Zhao, Peng Sun, Xin Jin, Xuanzhe Liu
8. (CCF-A 在投) *InternEvo: Efficient Long-sequence Large Language Model Training via Hybrid Parallelism and Redundant Sharding* [ATC'24]
Qiaoling Chen*, Diandian Gu*, Guoteng Wang, Xun Chen, YingTong Xiong, Ting Huang, Qinghao Hu, Xin Jin, Yonggang Wen, Tianwei Zhang, Peng Sun
9. (CCF-A 在投) *基于云边协同的深度学习作业调度方法* [软件学报'24]
谷典典, 金鑫, 刘讚哲



实习经历

上海人工智能实验室 AI 训练与计算部门 模型训练研发实习生 导师：孙鹏	2023.07 - 2024.01
微软亚洲研究院 系统与网络研究组 全职研究实习生 导师：熊一帆	2021.04 - 2022.03
谷歌（北京） 工程实习生 导师：张蕾	2018.07 - 2018.09



获奖情况

北京大学计算机学院斯伦贝谢奖学金	2023.12
北京大学优秀科研奖	2022.12
北京大学永赢基金奖学金	2021.12
北京市普通高等学校优秀毕业生	2020.05
北京大学优秀毕业生	2020.05
北京大学本科生科研优秀项目	2020.05
北京大学信息科学技术学院华为奖学金	2019.12
北京大学三好学生	2023.12、2021.12、2018.12、2017.12
北京大学休斯顿校友会奖学金	2018.12
国家奖学金	2017.12



学生工作

防疫委员，北京大学计算机学院	2022.05-2023.03
常任班长，北京大学计算机学院软件研究所 20 级（46 人）	2020.09 至今
带班辅导员，北京大学信息科学技术学院 2020 级 8 班（63 人）	2020.09 至今
团支书，北京大学信息科学技术学院 2016 级 4 班（65 人）	2017.09-2020.07
课程助教，软件工程	2020 春、2020 秋
课程助教，分布式机器学习：理论与系统	2021 秋



专业能力

- ☑ 编程语言：C、C++、Python、Java、JavaScript、LaTeX、Markdown
- ☑ 英语能力：CET-4 665 CET-6 645 GRE: 151+169+3.5
- ☑ 知识/技能：深度学习系统、机器学习； CUDA 编程； Kubernetes、Docker； Git、Github 开源协作