

项目名称	DEEPMd-Kit深度势能计算软件性能优化
项目类型	创新训练项目
所属学院	计算机与大数据学院
学科一级门	计算机科学技术
学科二级类	计算机应用
项目期限	一年期 2022-05-01 至 2023-05-01

项目成员

学生姓名	学生学号	项目中的分工	成员类型
陈一凡	112001007	统筹全局，架构分析	主持人
姜前辰	032002513	算法分析，性能测试	成员
林子键	032004119	架构分析	成员
吴维峰	221900129	代码重构	成员
陈宇豪	032004108	论文撰写	成员

指导教师

姓名	所属学院
廖祥文	计算机与大数据学院
吴运兵	计算机与大数据学院

项目简介（200字以内，项目研究的目的和主要研究内容）

DeePMD-Kit是一个训练神经网络势能的代码包，旨在减少构建基于深度学习的原子间势能和力场模型和执行分子动力学所需的工作量，其荣获2020年戈登·贝尔奖的机器学习分子动力学应用奖项。其成功推动了分子动力学的跨越式发展，为解决分子模拟中的精度与效率难题带来了新的希望，我们将对其进行HPC方面的优化，在保证准确率的前提下，尽可能地减少训练模型所花费的时间和计算资源，将机器学习分子动力学潜力持续释放

负责人曾经参与科研的情况

(1)2022世界大学生超级计算机竞赛ASC，优化分子动力学模拟应用DeePMD-Kit

指导教师承担科研课题情况

廖祥文:

- [1]廖祥文，国家自然科学基金海峡联合基金项目，网络大数据协同建模与序列分析的理论和研究方法研究，课题号U1605251；
- [2]廖祥文，国家自然科学基金面上项目，基于多源语义表示学习的社交媒体文本属性情感分类研究，课题号61772135；
- [3]廖祥文，国家自然科学基金青年项目，融合用户社会影响力和用户个性化特征的社会媒介倾向性检索研究，课题号61300105；
- [4]廖祥文，陈国龙，福建省科技重大专项项目，面向公共安全的网络舆情感知云平台研制与产业化，课题号2013H6012；
- [5]陈国龙，廖祥文，教育部博士点基金联合资助项目，基于图模型的用户群体观点挖掘研究，课题号2012351410010；

[6]廖祥文, 陈国龙, 福州市科技计划项目, 面向政务系统的敏感情报识别, 课题号2012-G-113;

[7]廖祥文, 福建省自然科学基金项目,博客倾向性检索中的倾向性评分研究, 课题号2010J05133;

[8]廖祥文, 福州大学科技发展基金资助项目, 面向论坛的热点话题发现研究, 课题号2010-XQ-22;

[9]廖祥文, 福州大学科研启动项目, 面向博客的舆情分析, 课题号022224.

[10]廖祥文,孙健, 王斌。杨东波, 程学旗。在数字图书馆中所采用的检索系统和检索方法.专利号:ZL200610072075.6.

吴运兵

[1]国家自然科学基金面上项目, 61976054, 面向观点型机器阅读理解的关键技术研究, 2020/1-2023/12,在研, 参与(排名第三)

[2]福建省自然科学基金面上项目, 2017J01755, 知识图谱推理若干算法的研究与实现, 2017/4-2020/4, 在研, 主持

[3]福建省教育厅中青年教师教育科研项目, JAT160077, 知识图谱推理和学习的若干算法研究与实现, 2016/5-2018/5, 在研, 结题

[4]国家自然科学基金面上项目, 61772135, 基于多源语义表示学习的社交媒体文本属性情感分类研究, 2018/1-2018/12结题, 参与(排名第三)

[5]福建省教育厅A类项目, JA15082, 基于SDN的控制器部署方案与多路径路由算法的研究, 2015/5-2017/5, 已结题, 参与

[6]福建省自然科学基金面上项目, 2013J01230, 具有复杂结构的动态社交网络聚类方法研究, 2013/1-2015/12, 已结题, 参与

[7]福建省自然科学基金青年科技人才创新项目, 2012J05115, 云计算环境下高效视频传输的算法研究, 2012/1-2014/12, 已结题, 参与

指导教师对本项目的支持情况

作为指导教师, 现承担了包括国家基金海峡重点项目、福建省级重点项目等多个项目, 具备一定项目指导经验;同时, 指导老师对于本项目目前已经进行了多次资料调研, 已有积累, 能够为该项目提供科研及设备方面支持。

(一) 研究目的

分子模拟 (Molecular Simulation) 利用计算机以原子水平的分子模型来模拟分子结构与行为, 进而模拟分子体系的各种物理、化学性质的方法。它是在实验基础上, 通过基本原理, 构筑起一套模型和算法, 从而计算出合理的分子结构与分子行为。分子模拟不仅可以模拟分子的静态结构, 也可以模拟分子体系的动态行为。

分子动力学模拟是一套分子模拟方法, 通过对分子、原子在一定时间内运动状态的模拟, 从而以动态观点考察系统随时间演化的行为。通常, 分子、原子的轨迹是通过数值求解牛顿运动方程得到, 势能 (或其对笛卡尔坐标的一阶偏导数, 即力) 通常可以由分子间相互作用势能函数、分子力学力场、全始计算给出。对于考虑分子本身的量子效应的体系, 往往采用波包近似处理或采用量子力学的费恩曼路径积分表述方式处理。分子动力学模拟也常常被采用作为研究复杂体系热力学性质的采样方法。在分子体系的不同状态构成的系综中抽取样本, 从而计算体系的构型积分, 并以构型积分的结果为基础进一步计算体系的热力学量和其他宏观性质。

分子动力学模拟由Alder和 Wainwright于 1957年首先提出, 可以用于NPT, NVE, NVT等系综的计算, 与蒙特卡洛法相比在宏观性质计算上具有更高的准确度和有效性, 广泛应用于材料、能源、化工、生物、医学、激光、电子等众多领域。时下最热门的抗新冠病毒疗法和疫苗研究就是分子动力学的典型应用场景。

过去三十多年来, 分子动力学研究主要利用AIMD模拟方法, 分析原子和分子在固定时间段内如何运动和相互作用。但这种方法只能研究千、万数量级原子, 以及皮秒时间尺度的小型系统, 对复杂的化学反应、电化学电池、纳米晶体材料和辐射破坏之类的模拟已经很难有所突破。

以机器学习、深度学习为代表的人工智能成为分子动力学研究的突破口。而人工智能+科学研究+高性能计算 (AI+Science+HPC) 则意味着AI的复杂数据处理能力和Science的第一性原理与HPC的超强算力的结合。其代表性成果就是荣获2020年戈登·贝尔奖的机器学习分子动力学应用DeePMD-Kit。DeePMD-Kit基于深度学习表示多体势能的最新发展为解决分子模拟中的准确性与效率两难困境带来了新的希望。因此, 如何更高效地利用DeePMD-Kit进行精准的分子动力学模拟, 以及大规模高效地利用超算, 是分子动力学模拟需要重点研究的课题。

(二) 研究内容

由于分子系统通常由大量粒子组成, 因此无法通过分析确定此类复杂系统的性质, 分子动力学模拟通过使用数值方法绕过了这个问题。然而, 在数值计算中产生累积误差, 可以通过正确选择算法和参数来最小化, 但不能完全消除。而分子动力学模拟的的精度与效率也是人们一直奋斗的目标。所以我们团队计划, 利用DeePMD-Kit框架去进行三个最具代表性的物质 (Water, Copper and Al-Cu-Mg ternary alloy) 的深度势能

(Deep Potential) 训练，来模拟不同物质的分子动力学。通过超级计算平台对DeePMD-Kit进行优化，使用更短的时间，来进行分子动力学模拟。项目主要分为以下几个模块:数据分块与数据通信优化;并行分布式处理;单GPU优化加速;多GPU优化加速。

由于GPU的运算速度与效率远高于CPU，我们将优化重心放在GPU优化上。利用CUDA编程等方法进行并行处理，实现对DeePMD-Kit进行分子动力学模拟的优化。

(三) 国、内外研究现状和发展动态

分子动力学简史

- 1957年：基于刚球势的分子动力学法（Alder and Wainwright）
- 1964年：利用Lennard-Jone势函数法对液态氩性质的模拟（Rahman）
- 1971年：模拟具有分子团簇行为的水的性质（Rahman and Stillinger）
- 1977年：约束动力学方法（Rychaert, Ciccotti & Berendsen; vanGunsteren）
- 1980年：恒压条件下的动力学方法（Andersen法、Parrinello-Rahman法）
- 1983年：非平衡态动力学方法（Gillan and Dixon）
- 1984年:恒温条件下的动力学方法(Berendsen et al.)
- 1984年：恒温条件下的动力学方法（Nosé-Hoover法）
- 1985年：第一原理分子动力学法（Car-Parrinello法）
- 1991年：巨正则系综的分子动力学方法（Cagin and Pettit）
- 1993年：路径积分分子动力学

分子动力学模拟有着运动方程简单，从分子水平研究物理化学过程，宏观物理量作为结果输出等独特优势，近年来计算机科学迅猛发展，超级计算机的强大的计算能力以及多尺度计算的发展，在大空间采用宏观或介观方法模拟。因此分子动力学模拟越来越被各国科学家所重视。

(四) 创新点与项目特色

分子动力学模拟也存在不足：随着模拟原子数增多，计算量迅速增大，能模拟的空间尺度有限。但当今计算机科学的迅猛发展，计算能力越来越强的超级计算机的出现，多尺度计算的发展，使得越来越多的科研人员重视和青睐分子动力学模拟。

本实验在DeePMD-Kit的模型构建基础上，通过优化访存，调整进程数和线程数，数据分块等高性能优化的手段来进行DeePMD-kit的训练效率优化。

(五) 技术路线、拟解决的问题及预期成果

1.技术路线

本项目拟通过cmake平台联合编译cpp和cuda文件，通过GPU加速DeePMD-Kit分子动力学模拟计算。

2拟解决的问题

由于利用DeePMD-Kit进行分子动力学模拟时，训练模型的过程中计算量较大，耗费的时间长。要通过合理的资源调度分布到多个GPU上进行并行计算，减少训练模型所花费的时间和计算资源。

3.预期成果

- (1)通过调整代码架构，利用CUDA编程等并行方法对DeePMD-Kit实现GPU上的优化加速。
- (2)相关技术文档和研究论文。

(六) 项目研究进度安排

项目实施分以下几个阶段进行:

- 1.文献调研、方案设计(2022年05月~2022年07月)：通过多种渠道进行文献调研和信息检索，了解分子动力学模拟领域国内外发展现状，成果，主要技术信息。讨论并制定具体的研究方案，对模型进行概要设计
- 2.数据集的收集、修正和评估、整体代码架构的分析（2022年07月~2022年08月），从多个来源收集测试所需的数据集,并对数据集进行调整，修正，评估数据集的质量。对整体代码架构进行分析。
- 3.资源获取和部署（2022年08月~2022年09月）：调配实验所需资源，并完成环境的配置和程序的部署
- 4.代码架构的改写和性能分析（2022年09月~2023年02月）：根据预先设计的方案和概要设计对代码架构进行改写，利用CUDA编程和一些并行方法对DeePMD-Kit实现GPU上的优化加速,并进行性能分析和测试对比。
- 5.质量评估和项目总结（2023年02月~2023年05月）：对最终成果进行质量评估。撰写技术文档、研究论文以及其他相关材料。

(七) 已有基础

1.与本项目有关的研究积累和已取得的成绩

从2022年1月开始，我们团队就致力于研究基于DeePMD-kit的分子动力学模拟优化，我们阅读了许多相关的论文，了解了DeePMD-Kit相关原理以及算法，为了优化DeePMD-Kit,我们根据已有知识，尝试调整编译选项，重写代码，我们实验室团队在ASC2022的优化部分的DeePMD-Kit模拟挑战中取得了优异成绩，从全球400余支队伍中脱颖而出，成功晋级2022ASC世界大学生超级计算机竞赛总决赛。


2.已具备的条件，尚缺少的条件及解决方法

目前，已获得福建超算中心的GPU服务器支持和信息化办公室小组老师的支持，同时也获得上海超算中心的机时设备支持。本项目在ASC2022世界大学生超级计算机总决赛初赛取得优异成绩，已在理论及实践上得到世界权威比赛的认可。

尚缺少的是人员的理论知识不足以及显存更大的GPU卡，目前的解决方法是组织成员自学CUDA、MPI等并行编程以及优化策略，同时获得显存更大的专业卡的使用机时，需要我们和超算中心协商。

开支科目	预算经费(元)	主要用途	阶段下达经费计划（元）	
			前半阶段	后半阶段
1.业务费	5250	总预算	2625	2625
(1)计算、分析、测试费	2000	计算、分析测试需要消耗较高的硬件成本	1000	1000
(2)能源动力费	0	公共设施费用	0	0
(3)会议、差旅费	500	外出交流学习、访谈	250	250
(4)文献检索费	250	查阅有关资料	125	125
(5)论文出版费	2500	出版论文	1250	1250
2.仪器设备购置费	2000	购买有关设备	1000	1000
3.实验装置试制费	2000	购买需要的硬件和软件	1000	1000
4.材料费	250	购买所需图书，打印、复印相关资料	125	125
合计	9500 元			

附件

 附件 (../UpLoadFile/112001007709843.pdf)

查看项目进展记录 (../Admin/ViewZJZTJ.aspx?ItemNo=2060)	查看中期检查表 (../Admin/ViewZhongQi.aspx?ItemNo=2060&state=3)	查看项目成果 (../Admin/ViewJieDuanChengGuo.aspx?ItemNo=2060)	查看结题报告 (../Admin/ViewJieTiBiao.aspx?ItemNo=2060&Type=xx)
---	--	---	---

返回