

基于网格的计算化学过程可视化*

陈建华^{1,2}, 段启智^{1,2}, 金 钟¹, 刘 倩¹, 张宝花¹, 迟学斌¹

(1. 中国科学院计算机网络信息中心 超级计算中心, 北京 100190; 2. 中国科学院研究生院, 北京 100049)

摘要: 提出了一个基于网格的计算化学过程可视化解决方案, 实时地监控计算过程, 增加对计算过程的控制。网格提供了计算作业所需要的大量计算资源, 对计算过程的实时监控则增加了对计算过程的控制。为了达到实时的目的, 针对计算过程中的原子不变性, 提出了优化的三维可视化方法, 动态地设定检测的时间间隔, 并引入了断点续传的思想。

关键词: OPTView; Java3D; 计算过程可视化; 断点续传; 计算化学; 网格

中图分类号: TP311 **文献标志码:** A **文章编号:** 1001-3695(2012)02-0432-03

doi:10.3969/j.issn.1001-3695.2012.02.006

Grid-based interactive visualization of computational process on computational chemistry

CHEN Jian-hua^{1,2}, DUAN Qi-zhi^{1,2}, JIN Zhong¹, LIU Qian¹, ZHANG Bao-hua¹, CHI Xue-bin¹

(1. Super Computing Center, Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China; 2. Graduate University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: This paper presented a grid-based interactive visualization of computational process resolution to monitor the process in real time. The grid supplied large resources for the computation. And monitoring the process of computation helped to control the process. As atoms were never changed in the computation process, the paper gave an optimized 3D visualization method. It introduced dynamical interval setting and resume broken transfer in data retrieving from the grid.

Key words: OPTView; Java3D; computation process visualization; resume broken transfer; computational chemistry; grid

计算化学领域的作业计算量大、耗时长, 而且计算过程带有一定的猜测性和试探性, 有时候需要用户通过自己的经验来对计算过程进行取舍。目前, 计算化学在不同的学科分支和应用领域的应用已超过数千种^[1], 主流的计算软件如 Gaussian^[2]、NWChem^[3] 等, 分子结构可视化软件如 Gaussian view^[4]、JMol^[5] 等, 都已具备广泛的用户基础; 也有一些计算化学应用是基于网格来实现的, 如 GridChem^[6] 等, 这些应用都利用了网格丰富的计算资源, 为用户提供了统一便利的使用平台。

一般而言, 分子结构的模拟计算和三维可视化是分离的, 用户通常无法实时地观测到计算过程中分子结构的变化。然而, 用户在尝试性地给定计算初始参数时, 可能存在由于参数不合适而导致作业盲计算很长一段时间得不到很好的结果。尤其是对于一些利用网格进行分布式计算的大规模作业, 可能在计算过程完成一部分时就达到了用户期望的效果, 多余的计算只会造成时间和计算资源的浪费。因此, 对计算化学作业的实时监控, 即时地反馈计算过程中的分子信息, 不但能够有效地避免资源的浪费, 而且能够辅助用户更好地控制计算过程, 得到良好的计算结果。本文根据这一需求给出了一个基于网格的计算化学过程可视化的方案 (OPTView)。OPTView 系统

基于 Java 和 Java3D, 它集成了国家网格 (CNGrid)^[7], 提供了一个网格交互的平台, 并能实时地检测计算过程, 将计算过程中的结果进行实时可视化。OPTView 系统具有跨平台、实时、响应迅速和高可视化分辨率等优点。

本文从网格应用的角度出发, 对提交至网格的计算化学作业进行实时监控, 对不同计算阶段的分子结构进行三维可视化, 交互式地展示时变数据的实时变化情况。

1 设计与实现

1.1 整体设计

为了使系统低耦合、高内聚、高重用性, 同时屏蔽网格中间件的变化对系统造成的影响, 这里采用了如图 1 所示的设计。



图1 系统示意图

网络模块负责实现与网格的交互, 主要包括向网格提交作业、查询作业状态、下载计算过程文件、查询计算过程信息等。系统通过数据处理模块来实时地获取数据, 并实时地三维可视

收稿日期: 2011-07-06; 修回日期: 2011-08-11 基金项目: 国家“863”计划资助项目(2006AA01A119, 2009AA01A130)

作者简介: 陈建华(1986-), 男, 山东莱芜人, 硕士研究生, 主要研究方向为高性能计算与应用、网格计算等(cjh1123@gmail.com); 段启智(1986-), 男, 湖南益阳人, 硕士, 主要研究方向为高性能计算及应用、网格计算; 金钟(1974-), 男, 副研究员, 博士, 主要研究方向为高性能计算应用、高性能计算化学软件算法和应用技术等; 刘倩(1981-), 女, 博士, 主要研究方向为物理化学; 张宝花(1985-), 女, 硕士, 主要研究方向为计算化学、药物设计; 迟学斌(1963-), 男, 研究员, 博士, 主要研究方向为并行计算的应用与实现。

化计算过程。用户可以动态、实时地查看计算过程的状态信息,并根据这些信息对作业进行控制。

为了便于系统的维护和扩展,防止依赖传递和变化的扩展,系统采用了模块化设计,主要由六个模块组成,如图 2 所示。a)配置模块用来加载系统和用户的配置信息,它允许用户根据自己的偏好进行配置,使系统更加人性化;b)能量曲线图实时地显示计算过程中分子的能量趋势,用户可以动态地查看计算过程中每一个阶段的分子能量;c)3D 可视化模块提供了一个统一的接口用来构建 3D 分子,并且提供了分子的平移、放大、缩小、旋转等操作,用户可以实时地查看计算过程中分子的结构信息;d)图形化用户界面为用户提供了一个友好的可操作界面,将模型与视图分离;e)网格模块为系统的其他模块与网格的交互提供一个稳定、公共的接口,本文采用了 MVC 模型来实现网格模块。为了减小客户端的压力,以及加强对于跨平台和分布式的支持,本文采用了 RMI 的技术来实现客户和服务器的通信^[8]。数据处理模块一方面用来不断地从网格获取数据、处理数据,抽取出相应的信息;另一方面也可以获取本地保存的分子信息,为可视化提供数据准备。

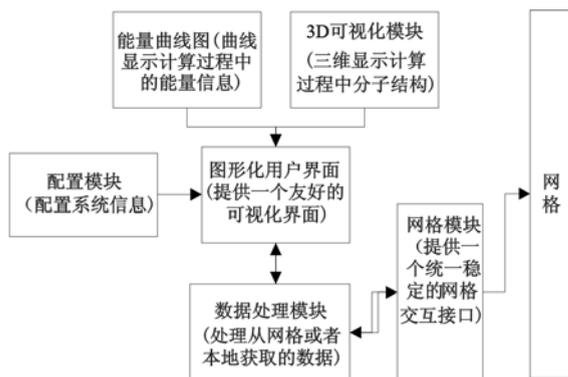


图2 系统模块

由于系统需要实时地监控网格中的计算化学作业,当计算过程进行到新的阶段时,需要动态地实时更新能量曲线图和分子 3D 图。一般来说,可以每隔一段时间定时地去查看数据是否有更新,如果发生更新再去更新能量曲线图和分子 3D 图。但是这就造成了在数据的两次更新之间,对数据的大量查看是无效的。尤其是当查看的时间间隔很短时,会对系统的效率造成很大的影响。为了解决这个问题,本文采用了观察者的设计模式^[9],如图 3 所示。

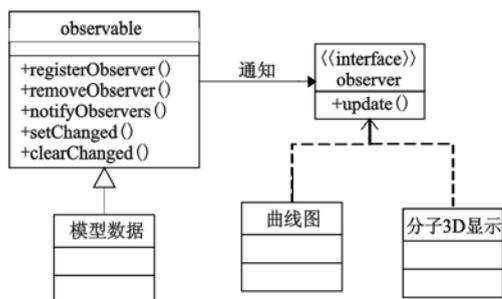


图3 观察者更新体系

系统中模型数据继承主题(observable 类),曲线图和分子的 3D 显示实现了观察者接口(observer),并且注册为模型数据的观察者。当从网格获取的数据发生改变时,作为主题的数据模型会发生改变,这时候它会通知注册在该主题上的观察者(曲线图和分子 3D 显示)进行更新。这样就达到了实时更新的效果,并且不会使系统增加额外的负担。

1.2 可视化原理

OPTView 系统中,分子的可视化采用了 Java3D 的技术来实现。Java3D 采用了场景图的方式,它是一个树型结构,由内容(content)和视图(view)两个分支组成。视图分支定义了视点等视图模型;内容分支是程序需要重点考虑的部分,用于描述场景所包含的所有图形对象以及针对这些对象的空间变换、光照、行为、背景等。为了方便描述,本文有如下的约定:BG 表示 Java3D 中的 BranchGroup, TG 表示 Java3D 中的 TransformGroup, TD 表示 Java3D 中的 Transform3D。

实时的可视化需要解决的主要问题有:当第一次显示时,如何快速地构建显示三维分子;当计算由一个阶段进入到另一个阶段时,如何快速地切换三维分子显示。

分子有多种显示方式,这里以最为常见的球棍模式为例进行阐述。在 OPTView 系统中,用球(sphere)来表示一个原子,用圆柱来表示两个原子之间的化学键^[10]。Java3D 提供了绘制三维球的接口,但是只能在原点绘制默认的较低分辨率的三维球体,需要通过 TG 和 TD 将球移动到指定的位置,这就会大大地降低可视化的速度。为此,本文在 OPTView 中提供了在给定的位置创建给定分辨率球体的方法。这就解决了实时可视化需要解决的第一个问题。

在一个计算化学的计算过程中,分子包含的原子类型和数量是不会变化的,不断变化的是每个原子的位置以及原子之间的化学键。对一般实现来说,当计算由一个阶段到达另一个阶段时,可以重新构建整个三维分子,但这样会造成空间的大量浪费,而且如果分子体系很大,重新构建三维分子的速度会非常缓慢。这里给出了自己优化的三维可视化方法。由于在整个计算过程中分子包含的原子是不会变化的,变化的只是它们的位置。因此完全可以将原子进行重用,只需不断变换它们在三维虚拟空间中的位置,并重新构建原子之间的化学键即可。

如图 4 所示的三维场景图,为了更好地重用和更高的灵活性,减少空间的使用,将三维场景图和分子分支图进行了分离。三维场景图提供了通用性的背景、光线以及一些分子的行为设置等。不同分子的三维可视化,只需要替换相应的分子分支图(molecule3D)即可,而不需要重新构建整个三维场景图。其中 behaviorTG 包含了分子具有的行为,contentBG 提供了替换分子分支图的能力。

如图 5 所示的分子分支图,moleculeBG 为分子分支图的根;moleculeTG 用来提供分子的整体放大缩小、平移等功能;moleculeBG2 用来包含分子结构中的原子分支(atom3D)以及原子之间的化学键(bondBG),它提供了化学键的替换能力;bondBG 包含了分子之间的化学键。为了提高效率、减少内存的使用,化学键可以共享一个 appearance。原子分支提供原子的平移和放大缩小的行为。在计算过程中原子的位置可能是不断变化的,就允许不断地对原子进行平移,而不需要在新的位置进行重新构建,这样就大大提高了可视化的效率。

随着计算过程的深入,当分子中原子的位置和原子之间的化学键不断变化时,只需要不断地平移原子的位置,同时删除旧的化学键、添加新的化学键即可。这样就避免了对整个分子的重新构建,提高了可视化的效率,缩短了系统响应的时间。同时,当需要可视化不同的分子时,只需要删除旧的分子分支图、添加新的分子分支图即可,也不需要重新设置背景、光线,以及分子所具有的行为。

根据表 3 所示的结果,可以定量地比较联网审计绩效各评价指标的重要程度。另外,三种方法计算出的各灰色综合关联度值比较如图 1 所示。由图 1 和表 3 可以看出:

a) 不论联网审计绩效的评价指标权重值采用何种方法确定,重要影响因素分析结果基本一致。

表 3 灰色综合关联度分析结果

评价指标	$\hat{\rho}_{ij}(RC/AHP)$	$\hat{\rho}_{ij}(RC)$	$\hat{\rho}_{ij}(AHP)$
C ₁	0.707	0.5034	0.603 6
C ₂	0.681 8	0.6199	0.722 4
C ₃	0.628 3	0.5142	0.564 2
C ₄	0.697 2	0.5185	0.598 7
C ₅	0.714	0.5158	0.650 8
C ₆	0.695 6	0.5449	0.675 5
C ₇	0.668 8	0.5042	0.584 5
C ₈	0.792 2	0.7424	0.646 3
C ₉	0.713 6	0.6733	0.607
C ₁₀	0.645 7	0.6102	0.573
C ₁₁	0.746 6	0.7076	0.623 5
C ₁₂	0.700 2	0.6601	0.600 3
C ₁₃	0.746 6	0.7076	0.623 5
C ₁₄	0.633 7	0.5999	0.566 9
C ₁₅	0.620 9	0.52	0.560 5
C ₁₆	0.792 2	0.7424	0.646 3
C ₁₇	0.838 6	0.5911	0.782 7
C ₁₈	0.625 3	0.581	0.562 7

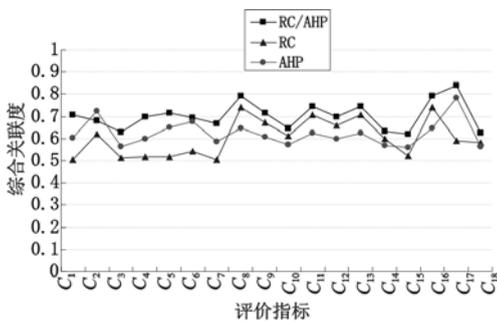


图 1 灰色综合关联度值计算结果比较

b) 综合灰色综合关联度值计算结果,联网审计的评价指标 C₈(审计效率的提高)、C₁₇(采集数据的合适性)、C₁₆(系统的友好性)是影响联网审计绩效的主要因素;而 C₁(硬件成本)、C₁₅(审计的频率)、C₁₈(系统升级频率)不是影响联网审计绩效的主要因素。也就是说,对于各个联网审计,经费投入多少不是影响联网审计绩效的主要因素,效益高、系统设计与开发质量高的联网审计系统才是最有效的系统。

4 结束语

IT 绩效审计是目前审计领域研究的一个前沿和热点问题,联网审计是 IT 审计研究的一个重要方向,研究联网审计的绩效评价问题具有重要意义。为了进一步明确影响联网审计绩效的有效因素,需要定量分析联网审计绩效各影响因素的重要程度。本文基于我国联网审计的特点,建立了联网审计绩效影响因素灰色关联分析模型,并对联网审计项目绩效评价的影响因素进行了定量分析,研究结果为我国今后实施联网审计项目提供了决策依据。

参考文献:

- [1] 陈伟, 张金城, QIU R. 计算机辅助审计技术(CAATs)研究综述[J]. 计算机科学, 2007, 34(10): 290-294.
- [2] GROOMER S M, MURTHY U S. Continuous auditing of database applications: an embedded audit module approach[J]. Journal of Information Systems, 1989, 3(2): 53-69.
- [3] ALEXANDER K, EPHRAIM F S, MIKLOS A V. Continuous online auditing: a program of research[J]. Journal of Information Systems, 1999, 13(2): 87-103.
- [4] REZAAE Z, SHARBATOGHILE A, ELAM R, et al. Continuous auditing: building automated auditing capability [J]. Auditing: A Journal of Practice and Theory, 2002, 21(1): 147-163.
- [5] CHEN Wei, ZHANG Jin-cheng, JIANG Yu-quan. One continuous auditing practice in China: data-oriented online auditing (DOOA) [C]//Proc of the 7th IFIP International Conference on e-Business, e-Services, and e-Society. Boston: Springer, 2007: 521-528.
- [6] 陈伟, QIU R, 刘思峰. 持续审计(CA)研究综述[J]. 小型微型计算机系统, 2008, 29(9): 1755-1760.
- [7] 国家“863”计划审计署课题组. 计算机审计数据采集与处理技术研究报告[R]. 北京: 清华大学出版社, 2006.
- [8] 陈伟, QIU R, 刘思峰. 一种基于数据匹配技术的审计证据获取方法[J]. 计算机科学, 2008, 35(8): 183-187, 194.
- [9] 陈伟. 一种基于等级法的联网审计绩效评价方法[J]. 计算机科学, 2010, 37(11): 111-116.
- [10] 陈伟. 一种基于 AHP 的联网审计绩效评价方法[J]. 审计与经济研究, 2011, 26(5): 47-52.
- [11] 刘思峰, 党耀国, 方志耕, 等. 灰色系统理论及其应用[M]. 5 版. 北京: 科学出版社, 2010.

(上接第 434 页)

参考文献:

- [1] Computer assistance to chemical research center [EB/OL]. [2010-10-06]. http://www.chem.ac.ru/Chemistry/Soft/.
- [2] Gaussian [EB/OL]. [2010-10-06]. http://www.gaussian.com/.
- [3] NWChem [EB/OL]. [2010-10-06]. http://www.nwchem-sw.org/index.php/Main_page.
- [4] ROY D, TODD K, JOHN M, et al. Gaussview, version 3.09 [R]. Shawnee Mission, KS: Semicem Inc, 2003.
- [5] Jmol: an open-source Java viewer for chemical structures in 3D [EB/OL]. [2010-10-06]. http://jmol.sourceforge.net/.
- [6] Computational chemistry grid [EB/OL]. [2010-10-06]. https://

www.gridchem.org/.

- [7] DEPEI Q. CNGrid: a test-bed for grid technologies in China [C]// Proc of the 10th IEEE International Workshop on Future Trends of Distributed Computing Systems. Washington DC: IEEE Computer Society, 2004: 135-139.
- [8] FREEMAN E, FREEMAN E, SIERRA K, et al. Head first design patterns [M]. [S. l.]: Oreily Media, 2007: 37-80.
- [9] SUN Yan-hua, SHEN Bin, LU Zhong-hua, et al. GridMol: a grid application for molecular modeling and visualization [J]. Journal of Computer-aided Molecular Design, 2008, 22(1): 119-129.
- [10] 欧阳刘彬, 孙衍华, 刘继风, 等. GridMol 系统中蛋白质可视化与建模的性能优化[J]. 计算机工程, 2009, 35(20): 242-245.