

虚拟天文台 ——新世纪的天文学革命^{*}

赵永恒 陈建生

(国家天文台 北京 100012)

摘要 介绍了虚拟天文台在天文学革命中的重要作用,着重评述了虚拟天文台的科学价值及其交叉学科的特性。虚拟天文台的出现将极大地推动天文学的发展,并对科学和技术等方面提出了新的挑战,同时也为中国天文学的发展带来了新的机遇。

关键词 天文学,信息科学,交叉学科



1 信息时代的天文学

400 年前伽利略发明了天文望远镜,使人类首次摆脱了用肉眼直接观测天空的历史,为从哥白尼开始的天文学革命提供了大

量的科学证据,因而后人把伽利略誉为“天空哥伦布”。历史悠久的天文学经过哥白尼、伽利略、开普勒和牛顿等人的发展,演变成了一门崭新的科学,同时也催生了现代科学技术。

150 年前,由于照相技术和光谱技术在天文观测中的应用,用人眼作为天文探测器的时代结束了,由此诞生了天文学的新分支——天体物理学,并发展成为现代天文学的主流。

50 多年前,在第二次世界大战中得到蓬勃发展的无线电技术使人类的视野跃出可见光的波段,从而诞生了射电天文学。其后不久,随着宇航时代的到来,天文观测不再局限于地面,由此空间天文学应运而生,人类对宇宙的观测扩展到了伽马射线、X 射线、紫外和红外波段。在过去的 100 年里,人类对天体的认识也从太阳系延伸到恒星、银河系、星系乃至整个宇宙,产生了恒星结构与演化和大爆炸宇宙学这两大理论体系,解决了恒星能源问题和化学元素的起源问题等。

从 10 多年前开始,天文学一直进行着革命性的变化,即开始步入信息时代的天文学。这一变化是由前所未有的技术进步所推动的,主要集中在以下几个方面:望远镜的设计和制造、大尺寸探测器阵列的开发、计算能力的指数增长以及互联网络的飞速发展。

望远镜技术的进步使得人类可以建造大型的空间天文台,为伽玛射线、X 射线、光学和红外天文

^{*} 修改稿收到日期:2001 年 12 月 21 日

的发展开辟了新的前景,同时也推动了新一代的大口径地面光学望远镜和射电望远镜的建造。在光学与近红外波段,已经有了高灵敏度高分辨率而尺寸不断增大的探测器阵列。伴随着这些技术进步,天文学家们正计划建造功能更强、口径更大的空间和地面望远镜,并将配备尺寸更大、像素更多的探测器。如同在计算机工业中反映计算能力随时间指数增长的摩尔定律一样,在过去 10 年中技术进步使得天文学的发展实际上也是遵循摩尔定律的^[1]。

随着众多先进的地面与空间天文设备的使用,将产生大量的观测数据。例如,目前哈勃空间望远镜每天大约产生 50 亿字节的数据,我国正在建造的 LAMOST 望远镜也将产生每天 30 亿字节的数据,而美国计划建造的大口径巡天望远镜将会达到每天 10 万亿字节的量级!

除了数据量的快速增长外,天文观测的方式也有了变化。这些先进的地面和空间天文台更多的是进行大规模的巡天观测,并产生质量均匀、标准统一的海量数据,通常是万亿字节的量级。这种天文观测模式的变化,不仅由于这些新设备能够快速获得观测数据,还由于计算机软硬件技术的发展能满足对数据进行快速的采集、处理及存档。

使天文学研究发生变化的另一个主要的技术进步是宽带高速网络,它允许在不同地点间进行海量数据的交换,使世界各地的天文学家都能够访问和使用这些数据,从而具有巨大的科学产出的潜在意义。

2 天文学的革命——虚拟天文台

数百年来,天文研究通常都是单个或几个天文学家进行少数天体的观测,加上大望远镜的观测时间又非常有限,因而无法研究那些需要大量数据来解决的天体物理问题。而在今后,技术上的进步将导致前所未有的天文数据流,造成所谓的“数据雪崩”现象。与以往不同的是,这些数据大部分是质量均匀的、且多波段并覆盖相当大的天区,其科学内容的丰富和深度是空前的,天文学家能利用这些数据进行各种科学问题的研究。

面对这些包含数亿个天体的多波段的海量数据,天文学家需要分析数据的工具,来进行诸如数据挖掘、先进的模式识别、大规模统计交叉关联、罕见天体以及时变的发现等工作。此外,有了这样的海量数据,在天文学历史上第一次可以将复杂的数值模拟与统计上完备的数据进行有意义的对比。而迅速发展的高速网络意味着全世界的天文学家都可以分享这些科学成果。

利用伽马射线巡天、X 射线巡天、紫外巡天、光学巡天、红外巡天和射电巡天所得到的观测数据,用适合的方法对数据进行统一规范的整理、归档,便可以构成一个全波段的数字虚拟天空;而根据用户要求而获得某个天区的各类数据,就仿佛是在使用一架虚拟的天文望远镜;如果再根据科学研究的要求而开发出功能强大的计算工具、统计分析工具和数据挖掘工具,这就相当于拥有了虚拟的各种探测设备。这样,由虚拟的数字天空、虚拟的望远镜和虚拟的探测设备所组成的机构便是一个独一无二的虚拟天文台。由此可见,虚拟天文台是在信息时代里天文学发展的必然产物。

新的巡天带来了巨大的科学发现的潜力,对这些巡天数据的联合使用,将涌现出全新的、无法预见的、意义重大的科学产出,这是一种仅靠单独使用其中某一部分数据所不能产生的新科学。人们越来越意识到,科学数据的获得、组织、分析和传播是持续而坚实地发展科学技术的基本要素。因此,投入一定的人力、财力、物力将所有符合特定规范的数据联合到虚拟天文台中,其科学意义是传统天文台所无法比拟的。

3 虚拟天文台的科学目标

目前,天文学家确定的虚拟天文台的主要科学目标是:

(1) 多观测参数空间的探索:把各个巡天数据统一到虚拟天文台中,将会有更广泛而复杂的应用。这些数据能提供全天十多个不同波段上的信息,在多维空间里展示整个天空的真实面貌。可以说,多种巡天数据在虚拟天文台完美结合,将会得到更加完善的真实的宇宙图像(多层次、大尺

度、系统性等等)。

(2) 稀有天体与新型天体的发现: 目前通过巡天来寻找稀有天体(如高红移类星体、褐矮星等)的项目正在蓬勃发展。假如某种有趣的天体或现象出现的概率是百万分之一或亿分之一, 那么就需要几百万或几亿个样本才有可能发现。这样, 在海量数据中进行彻底的宇宙探索来寻找稀有的未知类型天体具有更诱人的前景。因此, 虚拟天文台将会促进新的天文发现。

(3) 新兴的科学领域: 虚拟天文台对任何要求融合各类数据来研究天文现象的课题都具有重要的影响。虚拟天文台的出现促进了多波段天文学的发展, 不同波段的巡天数据的联合可以从更深层次来探索宇宙; 同时, 虚拟天文台推动了各种各样令人兴奋的科学探索, 如活动星系核和星系团的多层次研究、低表面亮度星系的形成和演化的研究、星系结构的研究等; 虚拟天文台的出现还促进了统计天文学的兴起, 如宇宙大尺度结构和银河系结构的定量分析、各种天体(特殊种类或特殊性质的恒星或星系、活动星系核、星系团等)完备样本的建立与研究, 等等。虚拟天文台的建立可以使天文学研究在数量和质量上大幅度提高。

(4) 数据挖掘技术: 从海量数据中发现稀有的天体或现象, 发现以前未知种类的天体或新的天文现象, 或者根据数据来区分不同类型的天体等, 都需要充分运用在信息科学中迅速发展的数据挖掘和知识发现技术。数据挖掘技术在虚拟天文台中的应用, 将使任何地方的天文学家在不依赖于大望远镜的情况下就可以做出一流的工作, 而这种研究方式完全不同于传统的天文学研究。运用数据挖掘技术可以有效地解决天文学中的“数据雪崩”问题, 这对天文学发展是至关重要的^[2]。

4 虚拟天文台的功能与特点

为了实现上述科学目标, 虚拟天文台应具有的主要功能是: (1) 将现有的多波段的大型数据库联合起来, 发展新的工具来实现对星表和图像数据进行查询; (2) 为建构未来大型天文数据库发展通用标准; (3) 为新的数据库提供兼容架构, 以从中得

到最大的科学回报; (4) 发展分析工具, 以在星表数据中进行天文发现和对联合数据作统计分析; (5) 发展在图像数据库中进行天体分类的工具; (6) 发展实现星表和图像数据库可视化的工具; (7) 为查询图像数据库、图像分析以及模式识别发展新方法; (8) 容纳数值模拟的结果, 并为观测数据与模拟结果的对比提供“工具箱”; (9) 与不断发展的数字图书馆和电子期刊进行链接。

与之相应的是, 在虚拟天文台的建设过程中将开展一系列的活动: (1) 建立可以使用数据流水线、存档和读取过程的公共系统, 来保证全世界的用户进行经济而高速的访问; (2) 开发一套通用的分布式的新型软件工具, 以进行海量数据的查询、关联、可视化和统计对比; (3) 协调建立高速的数据传输网络, 这对连接数据库、超级计算设备和广泛分布的用户是必需的; (4) 促进天文研究单位和其它学术机构之间高效率的合作, 用最少的基础设施费用做出更多的成果; (5) 与面临类似问题的其它学科科学家进行交流与合作; (6) 利用虚拟天文台独一无二的资源, 为科学普及和教育提供一个了解天文学和科学方法论的独特的窗口。

和传统概念的天文台不同, 虚拟天文台是以高速发展的网络技术和计算机科学为支撑的, 因而具有其本身的鲜明特点。首先, 虚拟天文台是快速发展的, 随着数据量的迅速增加、计算机网络的快速发展, 虚拟天文台的各种软硬件设施都要及时更新, 以保持其先进性; 第二, 虚拟天文台是分布式的, 无论其数据还是各种计算机软硬件资源都将分布在不同的国家和地区; 第三, 虽然在物理构成上具有分布性, 但在功能上必须是有机的统一体, 从而为用户提供整体一致的服务; 第四, 虚拟天文台是面向全球的, 它的资源将为全世界的天文学家所共享。此外, 拥有独一无二的科学资源的虚拟天文台必将承担起与它相称的社会责任, 为公众了解和利用科学提供服务。

5 虚拟天文台中的多学科研究

虚拟天文台所面临的技术挑战是十分明显的: 星表的大小是万亿字节, 而各类数据的总量将达到

千万亿字节;这些数据是非常复杂的,包含上亿甚至更多的天体,每个天体又有几十或几百个属性,这对于数据挖掘而言是一个至关重要的新问题;而对如此庞大的星表进行多变量相互关联将是一个大规模计算问题;如果还要对天体进行像素层次上的分析,计算的问题将更加突出。此外,还需要对元数据的表示与处理、大规模统计分析与相互关联以及分布式并行计算技术等进行多学科研究,来解决虚拟天文台所面临的前所未有的数据访问和计算问题。

整个天空的数字巡天的数据量约为 10 万亿字节,处理这样量级数据的工具和技术明显地需要发展新的信息技术来予以支持。当前,信息技术的研究与开发已在一些领域中展开,如大型数据库的统计分析与数据挖掘、分布式高性能计算网格、数据增强网格计算(数据网格)以及结构化数字信息管理(数字图书馆),这些研究在许多方面都和虚拟天文台所面临的问题有关,贯穿于这些科学分支中的信息技术和数据管理技术将会推动虚拟天文台的发展。

可以预计,虚拟天文台将在存储技术、信息管理、数据处理、分布和并行计算、高速网络、数据可视化和数据挖掘等各个领域突破现有技术的限制。这就要求天文界和信息科学界携起手来共同研究和开发虚拟天文台所需要的信息技术;要求与其它科学分支和超级计算中心进行合作,来开发元数据处理、数据处理和分布式计算的标准。而数据挖掘是一个需要天文学家、计算机科学家、数学家和软件专家一起合作才能解决的多学科问题。

大规模的数据以及用户与资源的广泛分布同样对通信网络的互联性提出了巨大挑战,虚拟天文台将是利用广域的高性能宽带网络来进行学术研究的一个富有创造性的例子。

6 中国天文学的发展机遇

1999 年美国天文学家首先提出了虚拟天文台的概念,很快就在国际天文学界引起了广泛的关注。美国国家科学院天文学及天体物理学发展规划委员会在题为“新千年的天文学和天体物理学”

的十年发展规划中把建立美国国家虚拟天文台作为优先推荐项目。美国在 2001 年财政预算报告中已单列了“国家虚拟天文台”的建设经费(作为美国自然科学基金会经费的增加部分)^[1];而欧洲的天文学家们正在紧锣密鼓地酝酿“欧洲虚拟天文台”的建设;在全世界,已有多家天文单位声称建立了虚拟天文台的原型。

我国天文学家对虚拟天文台也有很高的热情并积极建设。今年九月份国内天文学界召开了“虚拟天文台学术研讨会”,并成立了我国自己的“虚拟天文台科学工作组”,以研究和跟踪国际上相关领域的发展动向,确定“中国虚拟天文台”的建设目标,推动国内天文学界和相关学科对虚拟天文台的研究与发展。

虚拟天文台的出现为中国天文学提供了新的发展机遇。这是因为:(1)与在信息技术领域中互联网的建设相似,中国天文学界只有在虚拟天文台的建设中做出自己的贡献,才有可能充分利用虚拟天文台来获得科学上的发展;(2)目前我国天文学界业已开展一些小规模的巡天项目(如 BATC 巡天、密云米波射电巡天和超新星巡天等),而在未来,随着我国天文学界的重大科学工程项目如 LAMOST 望远镜、太阳空间望远镜(SST)和 500 米口径射电望远镜(FAST)等一批大型观测设备的建成,中国天文学将在国际天文学界拥有更多发言权,并将在未来的虚拟天文台中发挥重要的作用;同时,这些大型观测设备的科学产出也需要虚拟天文台这样的机构来配合;(3)虚拟天文台对于我国这样的发展中国家来说具有更深刻的意义,国力的限制使得我们不可能全面地发展各类型的大型天文观测设备,而由于国际上天文观测数据的开放性和虚拟天文台的建设与发展,我国的天文学家完全可以充分利用国际上最先进的设备所提供的高质量的数据来做出一流的科学研究工作;(4)由于虚拟天文台为天文学开辟出了一条崭新的研究道路,因此,必须在我国培养新一代的天文学家以适应时代发展的需要;(5)虚拟天文台设想的提出仅有三年的时间,可以说我们与其它国家处于同一起跑线,而我国飞速发展的信息技术和计算机网络技术已

提供了强有力的保障; (6) 虚拟天文台的出现为我国天文学家、计算机科学家、数学家和统计学家施展才华提供了新的机遇, 为我国多学科跨领域的交叉研究提供了新的舞台。

虚拟天文台是当代天文学发展的推进器、促进新发现的催化剂, 它将开辟出探索宇宙的新天地。在这历史性的机遇面前, 中国科学家们应该也一定能够为虚拟天文台的建设与发展做出自己应有的贡献。

参考文献

- 1 Brunner R, Djorgovski S, Szalay A. Virtual observatory of the future. Michigan: Astronomical Society of the Pacific, 2001, 343– 372.
- 2 Banaday A, Zaroubi S, Bartelmann M. Mining the sky. Berlin: Springer-Verlag, 2001, 1– 352.

Virtual Observatory ——New Revolution in Astronomy

Zhao Yongheng Chen Jiansheng

(National Astronomical Observatories, CAS, 100101 Beijing)

With the developments of the astronomical techniques and information sciences, it is the time of astronomy with an enormously increased flow of data. This paper describes the function of the Virtual Observatory for the revolution in astronomy, and its scientific discovery and technical challenges. The Virtual Observatory will provide new opportunities for Chinese astronomical community.

赵永恒 国家天文台研究员, 博士生导师。1964 年 2 月出生。1990 年获南京大学理学博士学位。国家“九五”重大科学工程 LAMOST 项目常务副总经理, 世界数据中心中国天文学科中心主任。主要从事活动天体的理论研究、高能天体的观测分析、多波段观测, 参加和主持过“攀登计划”项目、国家自然科学基金项目等研究工作。1997 年入选中国科学院“百人计划”, 1998 年获中国科学院第三届十大杰出青年称号, 1999 年获教育部科技进步奖三等奖。发表学术论文 50 余篇。