

虚拟天文台

张彦霞 赵永恒

(中国科学院北京天文台 北京 100012)

摘要

评述了近几年来大天区面积巡天技术方面的最新进展、数据挖掘技术在天文学领域的重要应用,介绍了虚拟天文台的功能、结构及与之相关的课题。虚拟天文台的出现和发展,预示其在二十一世纪将具有广泛的应用前景,将对天文学发展起到巨大的推动作用,并在知识和技术等方面对天文学家提出了新挑战。

关键词:虚拟天文台—数字巡天—数据挖掘—聚类分析

1 引言

200 多年来,天文学研究的方式一般是分别由单个天文学家或几个天文小组对数量相对较少的天体进行研究。整个天文学界花费了大量的时间和精力以获取足够多的数据,然而这些数据对统计研究来说似嫌不够。而且由于使用大型观测设备的时间是有限的,致使许多需要大量高精度观测数据的天体物理问题得不到解决。

近十年来,各种技术正在经历史无前例的飞速发展,使得天文研究的方式方法发生了巨大的变化。望远镜口径的增大、大面积探测阵的发展、计算机计算能力的迅速提高和通讯网络的不断普及,都将给天文学带来革命性的变化。随着望远镜在设计和制造技术上的进步,大型地面和空间天文台的建立成为可能,从而促进了一系列天文学新分支的兴起,如:射线天文学、射线天文学和红外天文学。随着科学技术的进步,可以建造新一代大口径地面可见光和红外望远镜、以及毫米和厘米波段单天线阵和多元天线阵。更多的地面和空间天文设备、以及更大口径和更精密仪器的投入使用,将带来天文数据的飞速增长。例如:哈勃空间望远镜大约每天产生 5GB 的数据,筹建中的大口径巡天望远镜每天将产生 10TB 的数据。除了数据量的增长外,天文学家进行天文研究的方式也在相应地发生变化,尽管新的地面和空间观测站仍然将部分观测时间分配给单个观测者或单个观测计划,然而大部分时间将用于多波段的大面积天区的巡天计划。

随着几个大的巡天计划的实施以及大型天文观测设备的建造,天文学正在步入一个多波段、大天区数字巡天的新时代。许多地面天文台和空间天文台大天区的数字巡天计划正在酝酿或计划中,它们覆盖了较大的波长范围,从 X 射线、紫外、可见光、近红外到微波、射电^[1-2],这些巡天数据通常以 TB 计量,以后将成为天文数据的主要来源。粗略估计一下,若以 0.5 的分辨率,巡天则由 2×10^{12} 像素构成。对每个波段的巡天将有 4TB 数据,全部各波段的巡天结果将形成巨大的数据量,因此,Szalay 认为天文学正面临着一场“数据雪崩”^[2]。

改变天文研究特征的主要技术是宽带高速的信息网络传输技术。未来的网络传输速度将是相当快的,从而能从地面设备和空间设备中高效率地获取数据,并将数据高效率地传输到不同的场所。其科学潜力是不可估量的。

技术进步将会导致未来几年天文数据史无前例的增长,而且这些数据也各不相同,大多数数据是相关的巡天数据,覆盖多波段和主要天区。因此,其广度和深度是以往数据所不可比拟的,将为从事不同科学研究项目的科学家提供良好的研究素材。为了使用的方便,将数据系统列表化是非常重要的。另外,这也可以减少对以前观测过的源的重复观测。拥有百万甚至上亿个天体的多波段数据的数据库的存在,使得天文界可以获得这些数据并对其进行分析,同时也要求具备与之匹配的分析工具,并为数据挖掘提供用武之地,例如:复杂形态

的证认，大样本的交叉证认，发现稀有天体和一些天体的时间演化序列。具有如此大的数据量，在天文学历史上第一次可以将复杂的数字模拟与统计的多变量分析进行比较。高速与广泛分布的信息网络的快速发展，意味着可以使世界各国的天文学家获得这些数据。

近年来，计算机速度的飞速增长，存储技术的迅速提高，网络速度的加快，探测器效率的加强，使得多波段、 10^{12} 量级的数据联合在一起不再是异想天开。这些技术上的进步将完全改变大多数天文观测的方式，会对天文学本身产生重要影响。而且，人们逐渐意识到获得数据、组织数据、分析数据、传输数据，是科技持续增长的基本要素。这些因素决定有必要建立一个机构，使其能够充分地有效地实现这些技术，因此，这就需要建立类似虚拟天文台的机构，用以管理这些正在增长的天文数据的存储。很快人们就会意识到投入巨大的财力物力将所有的数据联合到虚拟天文台中，远比建立一个传统的天文台有意义得多。这样，天文学家可以很容易地迅速对特定天区进行观测，而不是坐等几个月或一年才可获得望远镜的观测时间。

未来几年里，在信息和技术发展的推动下，计算机速度还要增长，计算机内存量仍在加大而价格在下降，信息网络速度仍在加快，探测器和天文台的性能将越来越好，数据库容量会越来越大，我们将面临许多实质性的挑战。例如：怎样加工、记录原始数据；怎样通过现代计算机硬件和网络系统存储、合并、获取数据；怎样快速地、有效地探索及分析数据并将这些数据可视化，这是一个正在快速发展的领域，将促进天文学家和计算机科学家的合作。在天文发展的数字化时代，数据的快速增长能否导致我们对天文学的看法发生改变？我们是否要问一些有关宇宙的各种新问题并且用新的方法回答？怎样系统有效地探索这些丰富的信息？怎样从大量的数据中提取科学精华和知识宝藏？这都是虚拟天文台要解决的问题^[3]。由此可见，数据库的联合是大势所趋^[2]，虚拟天文台的建立是必然的^[4]。

在这种形势下，美国国家科学院在天文与天体物理发展的新十年展望中把国家虚拟天文台（简称 VO，美国称之为 NVO）列为第一优先发展项目。该项目在 1999 年夏天开始酝酿，立即引起天文学界的广泛重视，当年 11 月第一次会议成立了一个工作小组，不到半年时间，就吸引了将近 140 名来自世界各国的天文学家聚集在一起，讨论虚拟天文台的科学目标，技术上的挑战及其构造的计划。据估计很快就会通过美国国家科学基金会立项，近年内就会启动。因此，虚拟天文台的建立可以说是信息时代发展的产物。

2 地面和空间天文的迅速发展

巡天技术的发展使得天文学和空间科学正在发生着巨大的变化。在大量的覆盖多波段、几十 TB 的像素、测量了几十亿个探测源的成十上百个参数的数字巡天计划的推动下，天文学成为数据极为丰富的领域。这样，我们可以系统地分层次地做出宇宙的图形。数据挖掘（DM）是指从海量数据中获得有价值的信息或发现新知识的技术。例如：利用多波段星表和图像以及源的性质与环境来搜寻未知类型天体，或搜寻稀有天体（如高红移类星体和褐矮星）。其它还有用 DPOSS 的 JFN 三色图像搜寻低表面亮度星系和光学星系团，用 XMM/Megacam-VST/VIRMOS 寻找 $z>1$ 的类星体和星系团，使用 ROSAT 和 DSS 进行的 AERQS 类星体巡天和搜索年轻星及褐矮星等工作，使用 ROSAT 和射电寻找 Blazar 的 DXRBS 巡天等，数字化历史底片研究恒星的长期变化，多色图像巡天寻找高红移天体的 2dF 类星体巡天、CADIS 和 EIS 巡天等。丰富的数据能够推动科学研究在定量研究方面不断发展，例如：统计研究银河系和宇宙大尺度结构，发现稀有甚至完全新型的天体和现象。数据挖掘技术能使任何地方的科学家和学生在不依赖于大望远镜的情况下就可以做出一流的工作。

在现代天文学中系统的数据探索将发挥核心作用，数字化的数据库是获得数据的主要途径。随着大量的遍布各国的地面观测站（如 NOAO 和 NRAO 天文台）和空间观测站（如哈

勃空间望远镜)的飞速发展, 10^{12} 量级甚至大到 10^{15} 字节的数据组正在如潮水般的涌来, 发现了 10 亿或更多的天体。天文学正进一步向全波段发展, 在 5 年内要覆盖 13 个或更多波段。巡天项目在规模和数量上仍在持续增长。

目前, 在地面天文和空间天文方面已有一批巡天观测资料。例如:

红外: IRAS、2MASS、DENIS...

光学: SDSS、DPOSS、MAST、POSS、UKSTU、ESO 底片巡天、LCRS、EIS、CfAZCAT、2dF、GSC-II...

射电: FIRST、NVSS、WENCS、SUMSS、GB6、PMN、nC...

X 射线: HEASARC、ROSAT ...

微波背景辐射: COBE (以及将来的 MAP、Plank 卫星)

其中如: 2dF、SDSS、2MASS、DENIS、FIRST、Chandra 和 XMM 正在运行, 一些新设施如: VISTA、VST、UKIRT WF、SCUBA-2、SALT、ALMA、MAP 和 Plank 将在近几年内启动。随着巡天产生几十 TB 甚至几百 TB 的天文数据, 数据的存储、分析是摆在当前天文学家面前的紧迫任务。天文学家必须借助虚拟天文台才能将各个不同波段的数据统一成一个整体, 充分地、深入地进行研究和探索。

3 虚拟天文台的主要功能^[2]

虚拟天文台的主要功能是把分散的数据联合起来。各数据库中的数据必须保留和改进, 数据放在各个熟悉该数据的课题组那儿, 即数据分散在各地。各课题组应保留自己的数据, 提供表示各数据库的内容、文件和元数据。

各种各样的用户对数据的要求各不相同。大多数普通用户仅是粗糙地浏览, 用 WWW 界面作大量的十分简单的查询。对中心网站来说, 支持这些不太复杂的查询引擎是相当容易的。中等水平的用户要较详细地使用数据库。而面对高级用户时将遇到很多困难, 他们要对 10^{12} 量级的数据多次查询, 提取 10^9 量级的数据去进一步研究。大多数查询各具特色, 至少在一段时间内不会完全一样。通常 科学家开始在有限的数据库范围内探索天体的多波段特性, 逐渐转向更复杂的查询。一些重要的工作需要支持。例如: 在不同的数据库中浏览同一天区的天体, 探索其独立特性, 用一些条件限制在主要天区, 并以在天区上天体间的角分辨率为基础进行寻找, 与其它数据库的数据进行交叉认证, 创建个人数据组和新的数据库。

数据就其本身而言是多维的, 每一个天体可以通过流量、在天区的位置、尺度大小、红移等来表示。寻找特殊类型的天体 (如类星体), 须在 N 维空间中定义复杂的区域。需要考查空间关系, 如找最近的天体, 或找满足一定角距离的其他天体。如果满足标准的天体的数量太多, 那么快速地获取数据的文件就建立不起来。唯一的办法是与数据库本身连接, 直接将数据传送给分析工具。

利用下一代高速网络, 可以实现不同网站间的相互联系。高速的信息网络技术现实中仅有几项应用, 这将造成严重的网络负载。然而, 虚拟天文台将成为创新网络应用的主要范例。

联合机构的创建和维护包括几方面内容。首先要了解一些基本要求和定义适当的标准, 然后建立满足这样条件的工具。工具的创建需要通过加强计算机科学技术在原有基础上发展来实现。这远远超出天文学的范畴。现在的社会是高度的信息反馈化社会, 这些技术的发展仅有天文学家的参与是远远不够的, 必须与计算机科学家、统计学家、甚至来自信息工业的参与者合作。

4 虚拟天文台的科学目标

4.1 多观测参数空间的探索

虚拟天文台的主要科学目标之一是探索多波段巡天测量得到的源的多参数空间。这样的工作已经做过，每个源在多参数空间中被当作一个点或一个矢量。但是将各个巡天数据统一到虚拟天文台中，将会有更广泛的复杂应用。这些数据库将提供全天或几乎全天在 13 个不同波段的信息，在多维空间里展示全天的真面貌。每一个数据库的每个波段包含 10 亿个天体，多达好几个 TB。必然存在不同数据库间的重叠，同时在两个不同数据库间仍有许多未知的事件。一旦数据库完备，它们的联合将只允许一对一的查询。在不久的将来，可以设想用图像或像素、或用数据和图像来进行探索，而且可以在巡天中考虑上时间维数，来进行时间演化序列的探索。例如：探索纯数据组时发现稀有天体，充分利用数据的丰富信息（成百上千的测量参数），从较平常的类型中区别出有趣的天体，这包括最有可能发现以前未知的天体和现象。更详细的评述可参看 Djorgovski 等人的文章^[5]。

通常，一个完备的观测参数空间包含的量有：天体的坐标、速度或红移，有时也包括运动方式、一定波长范围内的流量（光谱、一组波长的图像）表面亮度、确定源的图像形态参数、一定时间范围内的能谱变化参数等等。任何巡天都受到自身的选择效应和测量极限的限制。例如：流量、表面亮度、角分辨率、光谱分辨率、样本大小等等，只能测得一部分观测参数。这样任何巡天数据提供的宇宙图像都是有限的，人们对宇宙的了解也只能是有限的。但是，多个巡天数据在虚拟天文台中的完美结合将克服这些限制，得到更加完善的真实的宇宙图像（层次化的、大范围的、概括性的等等）。

有时不同波段的巡天数据的结合会产生非同寻常的新发现，如类星体和强射电星系的发现，各种各样 X 射线源和极亮源的发现等等。对虚拟天文台而言，依以前知之甚少的观测参数空间为样本，多波段研究将会对其性质深入地了解。

最早系统地探索宇宙空间的设想应归功于 Zwicky^[6]，不幸的是由于当时观测技术的限制，他的想法未得以实现。但这个富有创意的思想在 1975 年 Harwit^[7]和 1986 年 Harwit 和 Hildebrand^[8]得以更深辟的讨论。他们认为开辟新的观测参数空间对探索新的发现是很有意义的。随着大量的巡天数据的网络共享、探索和分析技术的提高，人们正准备将这一想法付诸实施。

4.2 稀有天体与新类型天体的发现^[3, 5]

在巡天中寻找稀有的已知类型天体（如：高红移类星体、褐矮星）的项目正在蓬勃发展。稀有性往往是观测的选择效应造成的，就以褐矮星来说，在宇宙中是很普遍的，只是难以发现而已。已知类型天体的性质可能与巡天仪器的参数混在一起，如带通、流量等，在参数空间的特殊区域中，满足一定选择标准的天体才能被发现，例如寻找高红移类星体^[9-13]。类似的技术同样适用于寻找高红移星系^[14-15]和褐矮星^[16-20]。

当然，在多参数空间中发现更多稀奇古怪的未知的稀有天体或现象是更有意义的事情。例如：可以在大样本中寻找一些在小样本中不可能发现的罕见事件。假如某种有趣的天体出现的概率是百万分之一或一亿分之一，那么需要几百万或几亿个样本才有可能发现。稀有天体在一些参数空间中辨别不出来，但是在其它一些参数空间中却可以区分出来。在可见光和红外波段象恒星的未确定源，仅有的区别是宽波段的能谱分布。能谱分布是以不同带通间的流量比例为参量的。在色参数空间中进行搜寻，光度候选者还要进行光谱证认。1998 年 Djorgovski 等^[21]在 DPOSS 巡天中利用色参数空间发现了高红移类星体和 II 型类星体。在色参数空间中，正常恒星分布呈现香蕉形状，并形成一温度序列，而类星体具有不同于正常恒星的色，远离恒星区，从而以在色空间的不同位置就可以将类星体与恒星区分开，如图 1 所示^[3, 5]。然后再根据吸收线和发射线的特点就可以区分开高红移类星体和 II 型类星体。这两类类星体相当稀少，面密度小于每平方度 10^{-2} ，低于可靠的恒星与星系分类的界限。这样，为了统计性地探测一些有意义的样本，就需要大天区巡天并且依靠合适的选择方法。同

样的方法也适合于其它波段的低角分辨率的巡天，例如：在 IRAS 数据中区分恒星和星系，用射电指数把类星体从射电星系中辨别出来，用 X 射线的硬度比从样本中找出 AGN 等等。在可见光和近红外波段，用同样的方法可以找到各个红移的类星体的完备样本^[22-23]。同样，可以选出特殊谱类型的恒星用以探测银河系结构^[24]。如果确定星系的形态可以用恰当的方法参数化，则同样也可以在一定范围内的哈勃分类中将星系区分出来^[25]。

相比之下，在大的数据组中寻找稀有的未知类型天体将具有更加诱人的前景。这可以通过系统地寻找参数空间中的格格不入者来发现，即从统计学的角度显然不同于其它天体。SDSS 和 DPOSS 研究组在寻找高红移类星体时就利用这种方法发现了一些新类型的天体。在未探索的参数空间中系统地寻找远离者可以发现另外一些特殊天体，其中一些结果便是新天文现象的原型。如果这些新的天体或天文现象确实存在，而且可以在已有的数据中探测到，那么彻底的、大范围内无偏差的多波段的宇宙探索将可以发现它们。可见虚拟天文台可以促进新发现。

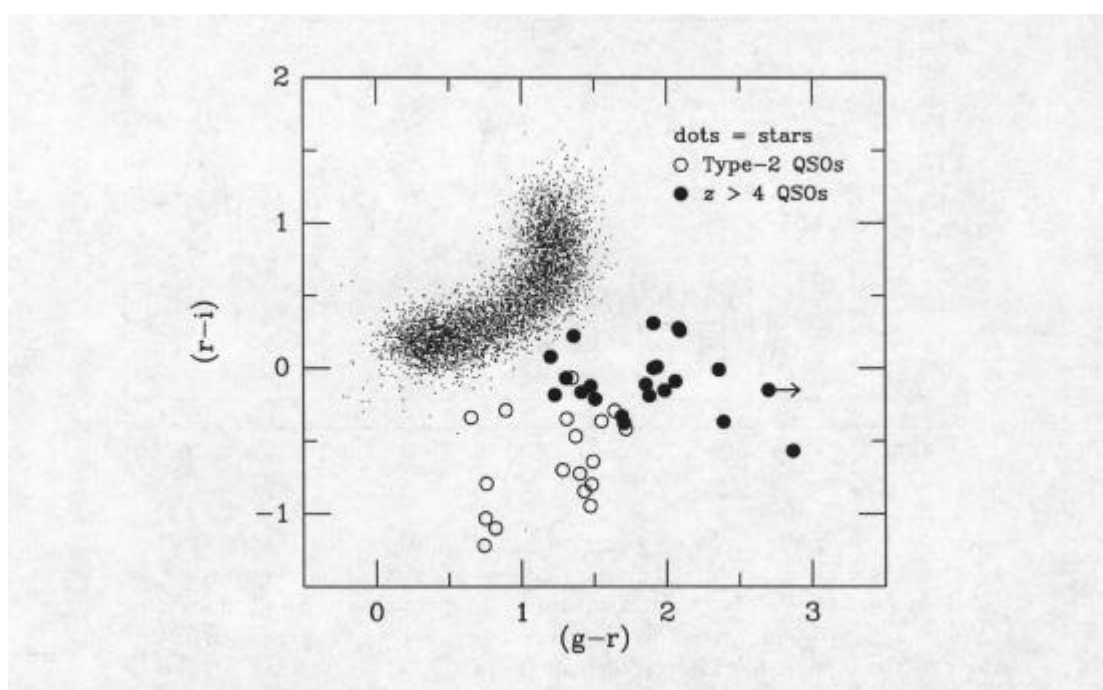


图 1 DPOSS 巡天得到的三色空间中天体的分布，其中点代表 r 约为 19 的正常恒星，实心圆圈代表 $z > 4$ 的类星体，空心圆圈代表 II 型类星体

4.3 边缘科学的兴起

尽管虚拟天文台对数据挖掘发现稀有天体具有较大的科学潜力，但更重要的是虚拟天文台对任何要求融合数据来研究天文现象的项目，都有着不可忽视的影响。虚拟天文台的出现促进了多波段天文学的发展，不同波段的巡天资料的联合可以从更深层次探索宇宙。典型的是对大多数工作最终要在光学波段证认，例如射电源或 X 射线源的的光学波段证认。对软件的基本要求是准确有效地从多波段交叉证认探测源。如果是一对一，那么工作就容易多了，但是要是一对多，那需要更复杂的算法和利用一些背景知识进行优化。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^[26-28]、星系结构的研究，这方面的研究对观测提出了挑战。这就需要充分发挥虚拟天文台的内在潜力。

虚拟天文台的出现也促进了统计天文学的兴起^[1, 3]，例如：宇宙大尺度结构和银河系结

构图像及其定量分析、各种天体（特殊种类或特殊性质的恒星或星系、AGN、星系团等）完备样本的建立与研究。如何系统地、统计地定量分析宇宙与将宇宙图像化及用于支持限制理论模型与理论思维的方法是天文学的热点。这再也不受来自小样本泊松误差的限制，但是了解巡天数据的系统误差和偏差将越来越重要。对这样的研究，样品的个数和大的视场是很重要的。从某种意义上讲，我们正在对天文学种类进行直接探索，通过新的数字巡天获得的信息将使之达到更高的准确度和更深层次的细节。

4.4 聚类分析及其引起的挑战

不管天体是已知的或未知的，数据被划分成各种不同类型的天体时，将遇到自动分类或聚类分析的问题。这是正在快速发展的数据挖掘（DM）和知识发现（KDD）领域的一部分。有关这方面的课题和方法可参看 Fayyad 等人的文章^[29]。虚拟天文台中典型的数据组具有一些性质：在 10^2 维中约有 10^9 数据矢量，数据组由若干种类的天体组成，如不同光谱类型的恒星、不同哈勃分类或形态的星系、类星体等等，这必须依靠强有力的分析技术支持。

聚类分析使用的技术有：

监督的分类方法，如：人工神经网络（ANN）或决策树。这种方法通常用于区分恒星与星系^[30-32]，在多参数空间中寻找具有预测特性的已知类型天体也可以用这种方法（如寻找高红移类星体）。

非监督的分类方法^[34-37]，如 EM(Expectation Maximisation), MCCV(Monte Carlo Cross Validation)。这些方法已用于确定数字巡天得到的星团数目，并将成为虚拟天文台分类工具的重要组成部分。

主分量分析方法（PCA）^[38-40]，用于多参数空间的降维，去掉一些无关或不重要的参量。

其它方法，如：最大似然法、非参数技术、信息瓶颈、小波、广义 Hough 变换、贝叶斯方法、独立分量分析方法（ICA）、Panalised likelihood 等。

一旦数据分成明显的种类，就可以对它们进行解释。问题是：某一类天体的性质间是否存在有趣的相关性？这些相关性可以反映新的天体物理知识，例如：恒星主序、星系基平面的相关性、Tully-Fisher 关系等，但同时这也使对聚类分析的统计解释复杂化。怎样才能证认出相关量并区分出无用的观测量？考虑到这些问题，在现实中盲目地运用聚类分析方法会产生误导或错误的结论。聚类分析方法必须足够有效地解决这些问题，而且分析结果必须有可靠的统计理论基础。应用在多参数空间中的数据可视化技术是聚类分析的另一重要组成部分，好的可视化技术对解释观测结果是很重要的，尤其在重叠的情况下。聚类分析有助于划分数据空间，从而找到不同寻常的天体，然后科学家证明并得出结论。另一重要课题是在处理虚拟天文台的大量巡天数据时提出的各种各样的问题，一些算法和模型需要相互作用或反复使用。

总之，应用这些方法发现新类型的天体、星团，并从结果中得出一些天体物理知识对虚拟天文台或其它应用是尤为重要的。看起来聚类分析作为从样本中选取有趣天体的方法，将完全不同于传统的天文学研究。

5 虚拟天文台的结构

当前主要任务是如何将这些高度分散的巡天数据联合起来推动知识发现，能胜任这项任务的正是虚拟天文台，其结构如图 2 所示^[1, 40]。Brunner 以数据服务的难易程度将研究项目分为两类：基础服务和高级服务^[40]。

基础服务中的数据已经作为商业数据的一部分，并且大多数数据库中心已使用。基础服务包括星表搜索引擎、系统元数据、数据关联、图像元数据引擎、图像数据库获得、查询

优化等。高级服务远超出平常的服务，要求对处理过的数据再加工才能应用，这样必须用特殊的软硬件才行。高级服务包括计算与数据网格、图像处理、统计分析、可视化、机器学习等。

Brunner 倡导的设计原则是不论在实际上还是应用上都要封装数据库，这样可以省去提供不同数据库间相互作用的努力（如：即插即用模型）。这种方法要求提供相应的方案给数据库，从而可以在当地的通信水平上充分发挥已有设备的作用。这种服务方式不仅减少了在通信上的投资，而且促进了提供分析工具者之间的协作。

要想使虚拟天文台立于不败之地，就必须在原有数据基础上吸收新的巡天资料和数据。这样，采用库文件转换数据和转换子库之间的元数据与元服务是必要的，这要求各分析工具能协调处理来自不同数据库的数据。所有的这些工作将在创建国家虚拟天文台或全球虚拟天文台后得以完成。拥有虚拟天文台，任何地方的科学家和学生将会如虎添翼地做出重要的开创性的工作。

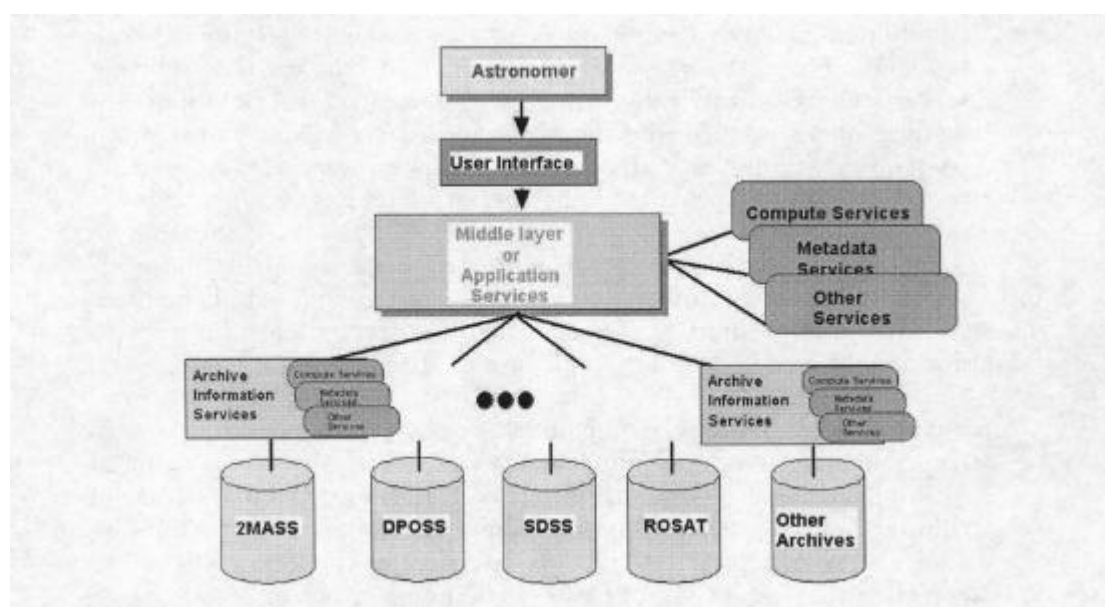


图 2 虚拟天文台结构示意图

6 结束语

综上所述，虚拟天文台是完全建筑于天文数据库和网络信息技术之上，需要强大的计算机与网络的软硬件支持，以及开发新的数据分析技术和知识发现工具。随着天文观测技术的飞速发展和天文数据处理技术的不断提高，虚拟天文台的出现使人观念一新，这对没有财力建造大型设备的发展中国家来说更具有深刻的意义。虚拟天文台的出现将为天文学家、计算机科学家、统计学家的精诚合作提供良好的机遇，开创一条系统地完整地探索宇宙的大道，并向我们展示未来天文的前景和面貌，使天文学家不用望远镜、足不出户就可以探索宇宙的美好愿望得以实现。同时，虚拟天文台的出现也向我们提出了严峻的考验，巡天技术的发展推动数据量的飞速增长，40000 平方度有 2×10^{12} 个像素，一个波段有 4 TB 数据，多波段有 10-100 TB，加上时间维，数据将达几个 PB，这就需要新的数据存储方法和新的分析工具以及强大的硬件支持，并需要培养下一代科学家以适应时代发展的需要。

参 考 文 献

- 1 Djorgovski S G, Brunner R J, astro-ph/0006043, 2000, in press
- 2 Szalay A S, Brunner R J, astro-ph/9812335
- 3 Djorgovski S G, Mahabal A A, Brunner R J, et al. astro-ph/0012453, 2000, in press.
- 4 Gilmore G, astro-ph/0011464, 2000, in press.
- 5 Djorgovski S G, Brunner R J, Mahabal A A, et al. astro-ph/0012489, 2000, in press.
- 6 Zwicky F, Morphological Astronomy, Berlin: Springer Verlag, 1957
- 7 Harwit M, QJRAS, 1975, 16: 378
- 8 Harwit M, Hildebrand R, Nature, 1986, 320: 724
- 9 Warren S, Hewitt P, Irwin M, et al. Nature, 1987, 325: 131
- 10 Irwin M, McMahon R, Hazard C, in The Space Distribution of Quasars, ed. D Crampton, ASPCS, 1991, 21: 117
- 11 Fan X, et al. Astron. J., 1999, 118: 1
- 12 Fan X, et al. Astron. J., 2000, 119: 1
- 13 Fan X, et al. Astron. J., 2000, 120: 1167
- 14 Steidel C, Adelberger K, Giavalisco M, et al. Ap. J., 1999, 519:
- 15 Dickinson M, et al. Ap. J., 2000, 531: 624
- 16 Kirkpatrick D, et al. Ap. J., 1999, 519: 802
- 17 Strauss M, et al. Ap. J., 522: L61
- 18 Burgasser A, et al. Ap. J., 2000, 120: 1100
- 19 Fan X, et al. Astron. J., 2000, 119: 928
- 20 Leggett S, et al. Ap. J., 2000, 536: L35
- 21 Djorgovski S G, Gal R R, Odewahn S C, et al. In Wide Field Surveys in Cosmology, eds. S Colombi et al. Gif sur Yvette:Eds. Frontieres, 1998
- 22 Wolf C, et al. Astron. Astrophys, 1999, 343: 399
- 23 Warren S, Hewitt P, Foltz C, M.N.R.A.S, 312: 827
- 24 Yanny B, et al. Ap. J., 2000, 540: 825
- 25 Odewahn S C, Windhorst R, Driver S, et al. Ap. J., 1996, 472: L13
- 26 Brunner R G, Djorgovski S G, Gal A A, et al. astro-ph/0010619, 2000, in press.
- 27 Brunner R G, astro-ph/0012361, 2000, in press.
- 28 Schombert J, astro-ph/0009080, 2000, in press
- 29 Fayyad U, Piatetsky-Shapiro G, Smyth P, et al. (eds.) Advances in Knowledge Discovery and Data Mining, Boston: AAAI/MIT Press
- 30 Weir N, Fayyad U, Djorgovski S G, et al. The SKICAT System for Processing and Analysing Digital Imaging Sky Surveys, Publ. Astron. Soc. Pacific, 1995, 107: 1243
- 31 Weir N, Fayyad U, Djorgovski S G, Automated Star/Galaxy Classification for Digitized POSS-II, Astron. J. 1995, 109: 2401
- 32 Fayyad U, Smyth P, Weir N, et al. Automated Analysis and Exploration of Image Databases: Results, Progress, and Challenges, J. Intel. Inf. Sys. 1995, 4: 7
- 33 Djorgovski S G, Carvalho R R, Odewahn S C, et al. astro-ph/9708218
- 34 Goebel J, Volk K, Walker H, et al. AA, 1989, 222: L5
- 35 De Carvalho R, Djorgovski S G, Weir N, et al. in Astronomical Data Analysis Software and Systems IV, eds. R Shaw et al. ASP Conference Series, 1995, 77: 272
- 36 Yoo J, Gray A, Roden J, et al. in Astronomical Data Analysis Software and Systems V, eds. G Jacoby, Barnes, ASP Conference Series, 1996, 101: 41

- 37 Brunner R J, Prince T, Good J, et al. astro-ph/0011222, 2000, in press
38 Adanti S, Battinelli P, Capuzzo-Dolcetta R, et al. 1994, Astron. Astrophys. Suppl. Ser. 108: 395
39 Connolly A J, Szalay A S, Bershadsky M A, et al. Astron. J. 1995, 110(3): 1071
40 Connolly A J, Szalay A S, Astron. J. 1999, 117: 2052

The Virtual Observatory

Zhang Yanxia Zhao Yongheng

(Beijing Astronomical Observatory, Chinese Academy of Sciences Beijing 100012)

Abstract

The new progress in the technology of the large area sky survey in recent years and the important roles which data-mining plays in astronomy are reviewed. Moreover, the functions and the architecture of the virtual observatory and the subjects related to it have been discussed. The existence and development of the virtual observatory will have a wide prospect of applications in the 21st century, will be a great push to astronomy and provide new challenges of knowledge and technology to astronomers.

Key words the virtual observatory —digital sky survey —data mining —clustering analysis