

·科技风云·

宇宙学迎来“大数据时代”

近10年来,从射电波到 γ 射线来纵览整个宇宙——使用数字巡天和实时望远镜观测已经成为天文学家最常使用的工具,而且它所带来的改变也是翻天覆地的:观测的数据洪流,正等待科学家的精密梳理,以期得到宇宙学中最大谜题的答案。

始于2000年的斯隆数字巡天(SDSS),正是昭示着“大数据天文学”时代到来的里程碑。在此之前,天文学由“单个”天文学家或“某个”天文学家小组组成,其程序基本如下:先申请望远镜的使用时间,获得自己的数据,然后再对其进行分

析。1/3的夜空,获得了超过930000个星系和120000个类星体的信息,诸多的天文学家的研究工作已经不是自己去直接观测天空,而是对SDSS的公开数据进行分析,目前已经在此之上问世了2000多篇科学论文,其中最令人瞩目的成就当属“星流场”的发现——这一群恒星分布的跨度长达近1/4个天空,它们可能是小型星系被银河系吞噬时所产生的残骸。

随着观测技术的进步,“大数据天文学”的潮流会更加“汹涌而来”。2019年,当大口径全天巡视望远镜(LSST)在智利帕琼山上运行之时,它会将世界上最大的32亿像素数码相机对准天空,并以15s/帧,每晚拍摄2000次,最后生成30万亿字节数据的方式拍摄天区,而且准备持续10年——LSST将会成为“千万亿字节天文学”的核心。

天文学家则必须从如洪水般涌来的新数据中提炼出新的认知,这不仅关乎数据的数量,还关乎数据的质量和复杂性。适用软件要在几秒钟内处理数万亿字节的数据,找出其中的规律和异常,对关键信息进行可视化,甚至在这个过程中“自我学习”。此时,大数据技术显示出它在其他领域已经驾轻就熟的能力。

首先,大数据技术可以轻松地对数十亿个天体进行分类和整合。例如,这个天体是一颗恒星还是一个星系?如果它是一个星系,那它是旋涡星系还是椭

圆星系?如果是椭圆星系,它是圆的还是扁的?其次,针对大数据的算法不仅能更容易地发现规律,还能加速识别出异常现象。例如,之前搜寻到高红移类星体被认为要靠1百万分之一的“好运气”,而现在用大数据来筛选数十亿个天体,则会把这个“运气”百分数提高到不知多少个数量级。再者,在反映出宇宙整体的图像中,大数据技术也有效得多。例如,传统上天文学家只能利用分光仪来估计遥远星系的距离,但其中

对于习惯于行走在人类想象力的极限天文学家们,并不缺乏追求成功的勇气。

99%天体的光谱无法获得,而现在天文学家仅分析图像就能估计出天体的距离,极大地拓展了宇宙研究的三维空间。

不过,要使大数据技术充分施展才能,就必须要有足够多的“数据”。一群富有进取精神的天文学家开始讨论建立一个全球统一的虚拟天文台,就像互联网一样,把存放在全世界的各个地方的天文观测数据连接起来。现在,这个试验性的虚拟天文台——美国国家虚拟天文台正在不断建构之中,当然它也面临着不少问题,许多天文数据并不公开,而公开中的数据绝大部分并不兼容,设计一个完美的数据分析工具也绝非易事,但对于习惯于行走在人类想象力极限的天文学家们,并不缺乏追求成功的勇气(6月6日美国NASA网站)。

近日,有关“大数据天文学”的最新进展是,美国德克萨斯大学以Lewis Irvine教授为负责人的天文学小组研究出一种新的计算方法,估算出银河系可能拥有大约1亿颗可支持生命存在的行星,这是首次对银河系中可能存在比微生物更高级生命的地外世界的数量进行了量的估测。

首先,天文学家们对1000多颗行星的观测数据进行了深入研究,并考虑了行星密度、温度、形态、化学、年龄以及与中心恒星的距离等要素,计算出了各行星的“生命复杂性指数(BCI)”。结果表

明,大约有1%到2%的行星生命复杂性指数比“木卫二”要高,而木卫二是木星的一颗卫星,被认为是在太阳系中除地球外,最有可能孕育多种形式生命的星体。由于银河系中大约100亿颗恒星,因此“生命复杂性指数(BCI)”的计算结果意味着银河系可能存在着大约1亿颗类似的行星,距离我们最近、最可能存在生命的系外恒星系统之一“Gliese 581”,其2颗行星就可能存在复杂生命圈。复杂生命并不意味着智慧生命,但是比微生物更大更复杂的有机

体通常可以以多种生命形式存在,而地球距离“Gliese 581”恒星

系统大约20光年。看来,地球人并不孤独,只不过距离与拥有同样生命复杂性的外星生命很遥远罢了(美国6月9日《挑战》期刊)。

从浩淼宇宙转向现实世界,在中国近期浩浩荡荡的反腐风暴中,同样存在着一只“看不见的手”——大数据技术。在普通人眼中,金融行业中存在的都是“高智商”犯罪,那看得人眼花缭乱的数据、票证放在一起,怎么就能找到违法违规的线索?国家审计署金融审计司司长吕劲松回答:金融审计最核心最有效的手段,就依靠大数据,违规发放贷款、虚列支出、公款私存、多计支出、少计收入、使用虚假发票报销费用等等问题,普通人看着永远不重复的数据,利用大数据的追踪技术,却能找到同一骗术的共同标志,一个个问题追下去,异常的数据自然就会浮出水面,而一系列的大案、要案则现出了它们的冰山一角(6月20日中国科技网)。

不知不觉中,大数据技术已经从各个方面渗透到当今每一个学科和每一个社会职能领域,并带来巨大的价值,而从宇宙到每个人,更多的改变都在蓄势待发,让我们摒心静气,等待更大的奇迹出现。

文/杨书卷
(责任编辑 李娜)