

### 机器学习方法在国际大型多波段巡天项目数据融合中的应用

近日，国家天文台张彦霞研究员、赵永恒研究员和北京大学吴学兵教授基于中国郭守敬望远镜（LAMOST）和美国斯隆数字巡天（SDSS）的光谱数据，结合欧洲航天局的 X 射线牛顿天文望远镜（XMM-Newton）数据、广域红外巡天探测器（WISE）数据及 SDSS 可见光测光数据，通过交叉证认方法获取了具有不同波段信息的多个已知样本。基于此，研究人员通过机器学习方法构建了适合不同波段样本的最优分类器，为 X 射线牛顿天文望远镜第九次发布数据（DR9）提供了分类预测和概率。该成果是天文信息技术在国际多个大型多波段巡天项目融合数据中的应用典范，为进一步开展 X 射线源的细致分析和研究工作具有重要的研究价值。该成果已在国际知名天文期刊英国《皇家天文学会月刊》（2021, MNRAS, 503, 5263）上发表。

二十世纪以来，随着各类地基和空间大型天文观测设备的研制和运行，天文观测覆盖了从射电、红外、可见光、紫外至 X 射线和  $\gamma$  射线的全电磁波段。天文学迎来了全新的发展阶段，即全波段-大样本-海量信息时期，这为探索各类天体和天文现象的物理本质提供了强有力的数据资源。从海量复杂的天文数据中挖掘信息和知识为天文学家提供了新的机遇和挑战。

欧洲航天局（ESA）于 1999 年发射的 X 射线牛顿天文望远镜（XMM-Newton），具有极高的光谱分辨能力，其目的是对宇宙天体进行高质量的 X 射线观测，为研究黑洞、中子星和脉冲星、星系团、活动星系核等高能现象的 X 射线发射源提供极具价值的海量数据资源，成为 ESA 历史上科学产出傲人的项目之一。2019 年 12 月，该项目对外发布了第九次高质量数据集（DR9），该批数据的观测日期从 2000 年 2 月至 2019 年 2 月，观测的独立源数量达到 55 万。



图1 左图为X射线牛顿天文望远镜，右图为郭守敬望远镜（LAMOST）。

交叉证认在多波段天文学研究中起着至关重要的作用,是多波段天文数据融合的关键技术和方法。通过对某一天体的多波段证认,可以获得该天体的不同波段信息,进而可以揭露更多的物理本质。近年来随着观测仪器和观测技术的大幅提升,天文数据的搜集能力大大加强,天文学步入大数据时代。

张彦霞等人首先利用 LAMOST 光谱数据和 SDSS 光谱数据,获得已知天体的分类信息,而后与 XMM-Newton 项目的 X 射线数据、WISE 近红外数据以及 SDSS 可见光数据交叉证认,获取了包含 7203 个星系、3558 颗恒星、21040 个类星体的已知样本,该样本仅占 X 射线牛顿天文望远镜第九次发布总观测数据的 5.78%,这就意味着 X 射线牛顿天文望远镜 DR9 数据集的绝大多数 X 射线观测数据仍为未知天体。

研究人员通过数据融合提取了已知样本分别在不同波段的特征信息。接着,利用机器学习方法针对不同波段的已知样本分别构建了不同的分类器,从中挑选出最优分类器对 X 射线牛顿天文望远镜 DR9 数据中的未知源给出了分类预测和概率。

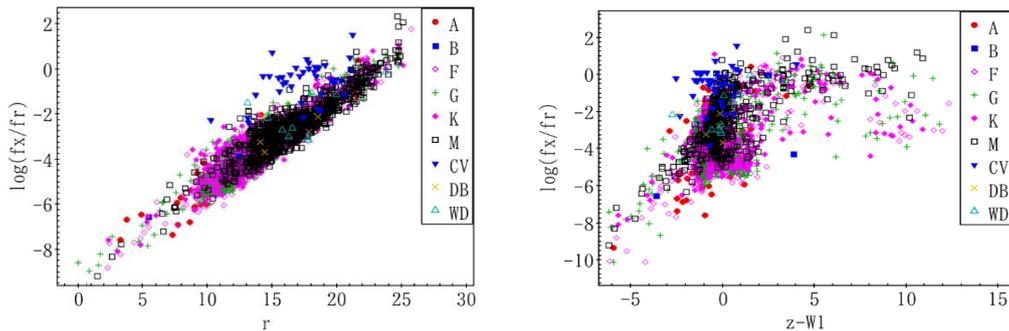


图2 不同光谱型恒星在二维参数空间中的分布

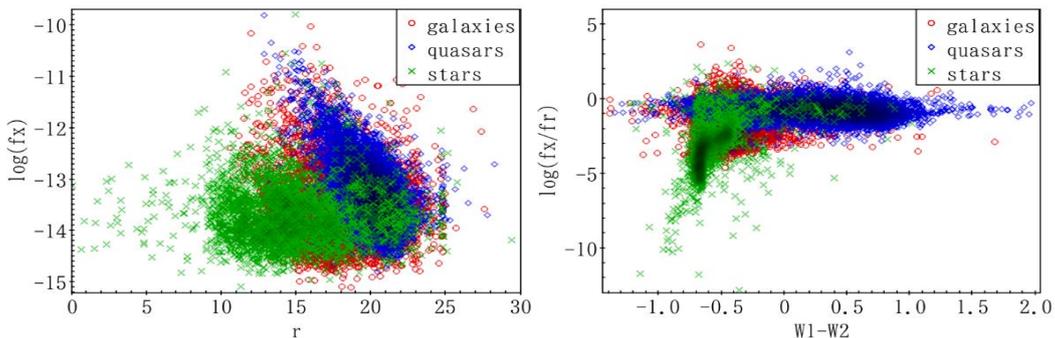


图3 星系、类星体和恒星在二维参数空间中的分布

该成果充分体现了机器学习方法在多个大型多波段巡天项目融合数据中的应用价值,该分类结果为进一步细致分析和研究 X 射线源具有非常重要的科学应用前景。

4月26日,中国科学院院网和《中国科学报》和对该成果进行了报道,引起大家关注。

论文链接: <https://academic.oup.com/mnras/article/503/4/5263/6232150?guestAccessKey=df807ff2-0560-4d35-8698-b417b3b74358>

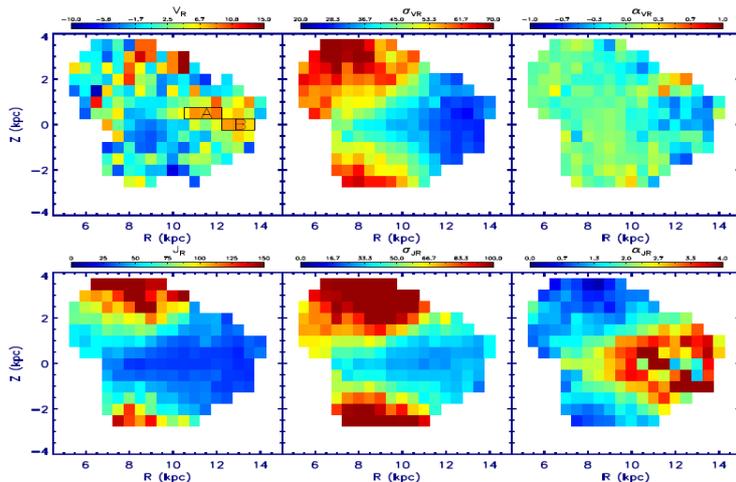
4d35-8698-b417b3b74358

## 研究人员基于 LAMOST-Gaia 数据对银盘恒星动力学演化的研究

近期，国家天文台武雅倩等人基于 LAMOST 和 Gaia 数据，计算了 LAMOST 数据中 64 万红巨星样本的运动学参数，分析了它们的年龄与运动学参数以及轨道参数的关系。并基于此对银河系薄盘和厚盘中恒星的金属丰度与运动学参数的关系给出了进一步验证与限制。该项研究成果已在国际知名天文期刊英国《皇家天文学会月刊》(2021, MNRAS, 501, 4917) 发表。

银盘化学动力学性质随时间的演化研究是天文学家长期以来的研究热点，对于理解星系盘的形成和演化历史具有重要意义。银河系盘结构被认为具有薄盘和厚盘两个组成部分。在化学丰度研究中，厚盘星比薄盘星的金属丰度  $[Fe/H]$  含量更贫， $[\alpha/Fe]$  含量更富一些，平均增丰程度超过 0.2 dex；在运动学研究中，厚盘星较薄盘星的速度弥散和轨道偏心率更大；在年龄的研究中，厚盘星形成于银河系早期，因而比薄盘星更年老。然而，我们对于银河系薄盘和厚盘的形成问题以及金属丰度梯度问题等还存在很大不确定性，造成这些不确定性的原因主要是银盘自身的演化，即不同年龄的恒星本身具有不同的空间结构。

2019 年，武雅倩等人从 LAMOST 数据中筛选出 64 万红巨星，并精确获取了它们的年龄和质量信息，为进一步理解星族合成和银河系演化历史提供了重要样本。最新成果中，武雅倩等人继续将这些红巨星样本与 Gaia DR2 的数据交叉，计算了它们的运动学参数，描绘出 LAMOST-Gaia 银盘恒星的三维速度、弥散、偏度等参数在年龄-金属丰度空间以及 R-Z 位置空间的变化，同时讨论了恒星的迁移对薄盘恒星的影响。最终发现在年龄-金属丰度空间，年轻的富金属恒星和年老的贫金属恒星的角动量分布区别很大。这些区别与化学空间区分薄盘和厚盘恒星是一致的。结果也证明



样本在 R-Z 空间的分布图，颜色分别表示径向速度和角动量空间的平均值（左图），弥散（中间图），偏度（右图），A 和 B 代表外盘的两个子结构。

通过计算分析这个红巨星样本的轨道参数，他们还发现样本中的厚盘星大部分出生于内盘，主要通过较高的离心运动到达目前的位置。薄盘恒星的轨道离心率比较低，它们的运动主要是靠径向迁移而完成。

## 观测运行部工作情况

4月，LAMOST 共观测了 56 个天区。理论观测时间为 270 小时，实际观测时间为 113.5 小时，占理论观测时间的 42.0%。受兴隆观测站天气原因\*影响，共 155.5 小时未能观测，占理论观测时间的 57.6%。

本月，望远镜仪器故障时间为 1 小时。

（天气原因\*：包括雨雪、大风、阴天、沙尘、多云等）

## 科学巡天部工作情况

- ✓ 更新和完善科学巡天的输入星表；
- ✓ 完成4月低分辨率和中分辨率2D光谱数据的处理和分析；
- ✓ 完成正式巡天日常观测计划的制定；4月份实际观测计划执行情况如下：M：7个，B：12个，V：5个，中分辨率：32个。共计40个。

（V为9m-14m亮天区；B为14m-16.8m较亮天区；M为16.8m-17.8m天区；F代表17.8m-18.5m天区。）

## 数据处理部工作情况

- ✓ 跟踪 LAMOST 用户使用数据情况和数据发布网站的使用情况；
- ✓ 解决和回馈用户提出的数据方面的问题；
- ✓ 完成DR7 v1.3版本数据的国内发布事宜；
- ✓ 完成4月份光谱数据的1D软件处理分析。

## 技术维护与发展部工作情况

主动光学、MA 机架跟踪电控系统自检和维护；MA、MB 子镜清洁及反射率测量；6 块金基紫外增强型反射镜清洗、清洗前后反射率测量；MA 镜罩轨道、镜室框架罩壳等日常清洁维护。MB 侧支撑中心孔因钢环及胶接工装加工，开展因钢胶接实验；望远镜焦面姿态光学复核及调整维护。

光谱仪日常维护，液氮灌注、像质自检维护、32 台 CCD 控制器自检、32 台 CCD 制冷温度及离子泵运行监测；高分辨率光谱仪光纤通光效率测试、自检测试及杜瓦抽真空维护；光谱仪中低色散观测模式切换、像质维护及观测；光纤定位闭环测试，问题单元排查、屏蔽并检查分析原因；光纤定位相机安装、稳定性测试；中色散定标灯支架制作与测试；制冷机组、除湿机等现场设施运行、检查、维护和日常安全巡视等工作；配合现场观测。



LAMOST 运行和发展中心

Center for Operation and Development of LAMOST Telescope