

密级 _____



中国科学院大学
University of Chinese Academy of Sciences

博士学位论文

自动搜寻特殊及稀少天体光谱的相关问题研究

作者姓名: 韦鹏

指导教师: 赵永恒 研究员 罗阿理 研究员

中国科学院国家天文台

学位类别: 博士

学科专业: 天文技术与方法

培养单位: 中国科学院国家天文台

2014年5月

Typeset by L^AT_EX 2_ε at March 5, 2014

With package C^ASthesis v0.2 of C_TE_X.ORG

Research on automatic search for special and rare Celestial
Spectra

By

Peng Wei

A Dissertation Submitted to
University of Chinese Academy of Sciences
In partial fulfillment of the requirement
For the degree of
Doctor of Astronomical Technology and Method

National Astronomical Observatories
Chinese Academy of Sciences

March, 2014

摘 要

本文是中国科学院学位论文的 L^AT_EX 模板。除了介绍 L^AT_EX 文档类 CASthesis 的用法外，本文还是一个简要的学位论文写作指南。

关键词： 中科院，学位论文，L^AT_EX 模板

Abstract

This paper is a thesis template of Chinese Academy of Sciences. Besides that the usage of the \LaTeX document class `CASthesis`, a brief guideline for writing the thesis is also included.

Keywords: Chinese Academy of Sciences (CAS), Thesis, \LaTeX Template

目 录

| | |
|------------------------|-----|
| 摘要 | i |
| Abstract | iii |
| 目录 | v |
| 第一章 绪论 | 1 |
| 1.1 引言 | 1 |
| 1.2 国内外巡天项目介绍 | 1 |
| 1.2.1 光学及红外巡天项目 | 2 |
| 1.2.2 射电巡天项目 | 6 |
| 1.2.3 其他波段巡天项目 | 7 |
| 1.2.4 国内巡天项目 | 9 |
| 1.2.5 LAMOST巡天 | 10 |
| 1.3 数据挖掘在天文中的应用 | 13 |
| 1.3.1 分类 | 13 |
| 1.3.2 回归分析 | 13 |
| 1.3.3 聚类 | 13 |
| 1.3.4 关联规则 | 13 |
| 1.3.5 离群数据挖掘 | 13 |
| 1.3.6 特征选取与转变 | 15 |
| 1.4 本文研究背景 | 16 |
| 1.5 本文的工作介绍和篇章安排 | 17 |
| 参考文献 | 19 |
| 发表文章目录 | 31 |

| | |
|----------|----|
| 简历 | 33 |
| 致谢 | 35 |

表 格

插图

第一章 绪论

1.1 引言

每当我们抬头仰望星空的时候，在感叹宇宙之无穷之余，也充满了对于未知星空的探索的欲望。对未知的探索也推动了天文学的发展，顾炎武就曾在《日知录》[6]中有“三代以上，人人皆知天文：七月流火，农夫之辞也；三星在户，妇人之语也；月离于毕，戍卒之作也；龙尾伏辰，儿童之谣也”之说。

天文学是研究宇宙空间天体、宇宙的结构和发展的学科。“遂古之初，谁传道之？上下未形，何由考之？日月安属，星陈安列？”，屈原在《天问》[7]对未知的星空提出了一系列问题。天问学的研究内容包括天体的构造、性质和运动规律等。宇宙是如何形成和演化的？宇宙中数百亿个星系是如何分布和演化的？银河系包含有千亿颗恒星，它又是如何形成和演化的？天文学是一门古老的科学，自有人类文明史以来，天文学就有重要的地位。在整个人类文明发展史中，这些基本而又深奥的问题始终激励着人们对自然界进行不断地探索。主要通过观测天体发射到地球的辐射，发现并测量它们的位置、探索它们的运动规律、研究它们的物理性质、化学组成、内部结构、能量来源及其演化规律。

天文学是一门特殊的基础学科，它的研究对象是天体，因此该领域的研究工作的开展特别依赖于观测。自从1609年伽利略首先将一架30倍的望远镜指向天空以来，人类对于银河系乃至更高一级的宇宙的认识不断深入，这同时也促使着人类在天文领域的技术手段一直不断的进步。尤其是近十几年来，天文望远镜和终端仪器设备的技术得到了快速发展。自从上世纪90年代开始，“多目标光纤光谱仪技术”日趋成熟，它可实现多根光纤将视场中的多个待测天体的光引入到光谱仪同时进行观测，使得望远镜一次可获得几百至数千个源的光谱数据，显著的提高了观测效率，再加上大型CCD在天文观测设备上的广泛使用，大大提高了望远镜终端设备的收集能力。观测仪器设备及数据收集能力的大幅提高，使得我们迈入了天文观测数据的“雪崩”时代。

1.2 国内外巡天项目介绍

随着望远镜中观测设备及观测能力的不断尽力，尤其是在CCD技术不断发

展之后，越来越多的望远镜及观测设备投入到巡天观测中。巡天项目使用的望远镜从地面设备到太空，观测波段几乎覆盖了整个电磁波波段。本小节将对次进行简要介绍：

1.2.1 光学及红外巡天项目

从伽利略将望远镜指向太空到20世纪的400多年间，几乎所有的天文设备都是光学波段的观测。光学及红外波段的观测形式主要分为成像测光观测及光谱观测两种，而相应的巡天项目一般也分为成像测光巡天和光谱巡天两种。下面对这些巡天项目进行简单描述：

- 斯隆数字化巡天（Sloan Digital Sky Survey，缩写为SDSS）是使用位于新墨西哥州阿帕契点天文台的2.5米口径望远镜进行的红移巡天项目。该项目开始于2000年，以阿尔弗雷德·斯隆的名字命名，计划观测252006年，斯隆数字化巡天进入了名为SDSS-II的新阶段，进一步探索银河系的结构和组成，而斯隆超新星巡天计划搜寻Ia型超新星爆发，以测量宇宙学尺度上的距离。2008年10月31日，SDSS-II发布了最后一次数据。斯隆数字化巡天第三期工程SDSS-III已经于2008年7月启动，将持续至2014年。斯隆数字化巡天使用口径为2.5米的宽视场望远镜，测光系统配以分别位于u、g、r、i、z波段的五个滤镜对天体进行拍摄。这些照片经过处理之后生成天体的列表，包含被观测天体的各种参数，比如它们是点状的还是延展的，如果是后者，则该天体有可能是一个星系，以及它们在CCD上的亮度，这与其在不同波段的星等有关。另外，天文学家们还选出一些目标来进行光谱观测。目标的位置用钻孔的方式记录在铝板上，小孔的后面接有光纤，将目标天体的光引入摄谱仪。望远镜每次可以同时拍摄640个天体的光谱，每晚大约需要6到9块铝板对天体进行定位，
- 2度视场星系红移巡天(Two-degree-Field Galaxy Redshift Survey)，或2dF、2dFGRS是天文学在1997年至2002年4月11日之间使用AAO天文台的3.9米AAO望远镜进行的红移巡天观测。史蒂夫马杜克斯和约翰皮科克是这个计划的主持人，巡天测量的数据在2003年6月30日出版。在勘测的这一局部宇宙的部分，确定了大规模的结构。截至2007年1月，它是仅次于2000年开始的斯隆数字巡天之下，规模第二大的巡天观测。2度视场星系红移巡天计划的名称来源是因为所使用的仪器每次测量的面积大约

是2平方度，针对南天和北天的银极，总共测量了1,500平方度[2]。观测区域的选择则是依据先前的APM 星系巡天的结果（史蒂夫马杜克斯和约翰皮科克也参与此工作）[2]，在赤经上涵盖了75度宽的两个扇状区，在纬度上的涵盖是银北极7.5度的范围，银南极15度的范围，在银南极还测量了数百个独立的2平方度区域。（参见这张图中，那些黑色的小圈代表勘测的区域，红色的栅格是早先APM星系巡天调查的区域）。总计，在五年之中，使用了272个观测的夜晚。在光度学上测量了382,323个目标，在244,591个天体的光谱中有232,155个是星系（221,414有良好的品质），12,311个是恒星，还有125个类星体[3]。

- RAVE is an observational program in the field of Near Field Cosmology which exploits our position inside the Milky Way to study its formation and evolution in greater detail than is possible for other galaxies. RAVE focuses on obtaining stellar radial velocities to study the motions of stars in the Milky Way's thin and thick disk and stellar halo. The vast majority of stars in our Galaxy has no velocity measurements, and particularly no time-consuming radial velocity measurements. RAVE utilizes fiber optics to perform multiple, simultaneous spectroscopy on up to 150 stars in a single observation. In this way it can obtain a representative sample of the nearby stars in our Galaxy which are all around, over a wide area of the sky. For the majority of RAVE stars proper motion data is also available. Combining the proper motions with distances derived using the photometric parallax method, the transverse velocity of these stars can be calculated. With the observed radial velocities the full 6D phase space information for the stars is then obtainable. RAVE is complementary to the SDSS project's SEGUE program as it a southern hemisphere, wide-field, intermediate-depth, intermediate spectral resolution survey with limited wavelength coverage. Conversely, SEGUE is a northern hemisphere survey with selected angular coverage, deep exposures, low-resolution but large wavelength coverage. Most of the stars observed by RAVE are between 1500 and 13000 light years from the sun and so RAVE surveys halfway to the Galactic center and halfway to the edge of the Galaxy's disk.

- 2微米全天巡天 (Two Micron All-Sky Survey) (2MASS) 的工作开始于1997年, 完成于2001年。使用的两架望远镜, 分别位于北半球美国亚利桑那州的霍普金斯山和南半球智利沙罗陀洛洛, 以确保能观测到全部的天空。这是迄今最雄心勃勃的巡天计划, 经过整理的最后数据已经在2003年公布。全部的天空都使用红外线的2微米邻近的3个波段: J ($1.25 \mu\text{m}$), H ($1.65 \mu\text{m}$), 和Ks ($2.17 \mu\text{m}$) 完成扫描的工作。这次巡天的目的包括: 发现在隐匿带内的星系, 这是在可见光的范围内被我们的银河系遮蔽的天空。第一次对棕矮星的搜寻。对低质量恒星的大规模搜寻, 这些恒星在我们的银河系和其他的星系中都是很普遍的。建立所有被检出的恒星和星系的目录。
- The UKIRT Infrared Deep Sky Survey or UKIDSS is an astronomical survey conducted using the WFCAM wide field camera on the United Kingdom Infrared Telescope on Mauna Kea in Hawaii. Survey observations were commenced in 2005. UKIDSS consists of five surveys covering a range of areas and depths, using various combinations of five near-infrared filters. UKIDSS as a large-scale near infrared survey follows 2MASS and anticipates the VISTA telescope in the Southern hemisphere. It aims to cover 7500 square degrees of the Northern sky. Four particular areas of investigation for UKIDSS are: "the coolest and nearest brown dwarfs, high-redshift dusty starburst galaxies, elliptical galaxies and galaxy clusters at redshifts $1 < z < 2$, and the highest-redshift quasars, at $z = 7$ ".
- 广域红外线巡天探测卫星 (Wide-field Infrared Survey Explorer, WISE) 是NASA的红外线空间望远镜, 于2009年12月14日发射[1][2][3]。WISE搭载口径40厘米的红外线望远镜, 以3至25微米的波长, 六个月的时间进行巡天。WISE的红外线侦测器比之前的红外线巡天太空望远镜, 如IRAS、AKARI、COBE灵敏一千倍以上[4]。一般预期WISE一天可以发现数十颗小行星[5]。WISE预定将拍摄全天99%的影像, 且同一区域影像至少将拍摄八幅以增加精确度。WISE将位于526公里高的太阳同步轨道并至少运行10个月。预估WISE将拍摄约150万幅影像, 平均每11秒拍摄1幅[6]。每幅影像的视野是47角分。每个区域将被观测过至少10次[7]。WISE的影像将拍摄太阳系、银河系以及宇宙深处的影像。在这些影像中将可增

进我们对小行星、棕矮星和主要辐射红外线的星系的认识。WISE同时也是用来取代1999年3月发射失败的广角红外线探测器[8]。WISE将以红外线进行巡天。它的侦测器阵列在120、160、650、2600微央斯基(μJy)和3.3、4.7、12和23微米有5个标准差的极限[10]。因此WISE比1983年发射的IRAS在12和23微米波长灵敏10万倍；比1990年代的COBE在3.3和4.7微米波长灵敏50万倍[10]。第一波段：3.4微米，观察恒星和星系的宽波段。第二波段：4.6微米，观测棕矮星等亚恒星天体内部热能的热辐射。第三波段：12微米，观测小行星。第四波段：22微米，观测恒星形成的尘埃（温度约70—100 K）。

- Pan-STARRS 泛星计划（Panoramic Survey Telescope And Rapid Response System, Pan-STARRS，直译为全景巡天望远镜和快速回应系统）是一个正在进行中的巡天计划；该计划将对全天空天体进行天文测量和光度测定。该计划将比较同一天区不同时间的变化以期能发现彗星、小行星、变星等天体；尤其是有撞击地球威胁性的近地天体。泛星计划将建立一个所有在夏威夷能观测到，视星等最暗可达24等的天体数据库，总共可观测全天四分之三的区域。泛星计划要使用四个口径1.8米的望远镜组成阵列，这四个望远镜可能位于夏威夷毛纳基山天文台或海勒卡拉火山。计划的四座望远镜，PS4，将每次观测同一天区，观测影像将由电脑比对去除因为芯片缺陷造成的像素坏点或宇宙射线影响，最后将四个望远镜收集的光线汇集，可产生相当于口径3.6米的望远镜分辨率。
- 盖亚任务（Gaia）是欧洲空间局（ESA）的空间望远镜[4]。本任务的目的是要绘制一个包含约10亿颗或银河系1%恒星的三维星图[5][6][7]。作为依巴谷卫星的后继任务，盖亚任务是ESA 在2000年以后的远期科学任务。盖亚任务在约5年的任务中将可观测到视星等最暗为20等的天体。它的目标包含：确认10亿颗恒星的位置、距离和每年自行运动量。对视星等15等恒星的精确度为20 μas ，20等则为200 μas 。侦测数万个太阳系外行星系统[8]。能够发现轨道在地球和太阳之间的阿波希利型小行星。这个区域对地面望远镜相当难以观测，因为该区域几乎只在白昼时才会出现在天球[9]。侦测最多50万个类星体。阿尔伯特·爱因斯坦广义相对论的更精确实验。盖亚任务的资料将可建立在银河系内甚至之外的极高分辨率三维星图，并且绘出恒星的运动方向以了解银河系的形成与演化。分光

光度法量测可以得知被观测恒星的物理性质，以确认其光度、有效温度、表面重力和组成元素。这种大规模的恒星观测将提供许多基本观测资料以解决关于银河系形成、结构与演化等多种重要问题。大量的类星体、星系、太阳系外行星和太阳系天体也将同时被观测。

- 30米望远镜（Thirty Meter Telescope, 简称TMT）是一座由美国、加拿大、日本、中国、巴西、印度等国[5]参与建造的地面大型光学望远镜。和E-ELT类似：其主镜是一块由492块六边形镜面拼接所组成的分割式主镜，30米/98英尺的主镜直径仅次于E-ELT的42米；配备有自适应光学系统；能观测红外线等。在波长大于0.8微米的光的范围内，自适应光学系统使它的图像清晰度比哈勃太空望远镜高10倍；巨大的主镜使它的观测清晰度比现行的大型地面光学望远镜高10至100倍。预计将进行关于暗能量，暗物质等进一步的研究、研究星系在过去130亿年的聚合和发展、研究超质量黑洞和星系之间的联系、进行行星和恒星结构的研究、进行太阳系、行星大气的化学研究和气象研究、在太阳系外的星球上搜寻生命等科研任务。
- 大型综合巡天望远镜（Large Synoptic Survey Telescope, LSST）是一个计划中的广视野巡天反射望远镜，将每三天拍摄全天一次。LSST预计在2010年开始动工，并在2015年启用[5]。LSST在所有大口径光学望远镜（主镜直径8米级）中设计相当特别；其视野高达直径3.5度（9.6平方度）。相比较之下，在地球所见太阳和月球的视直径是0.5度（0.2平方度）。再加上LSST的巨大口径使它收集光线能力极强，光展量（Etendue）高达319 m²degree²[3]。LSST的科学目标特别是：观测深空中弱重力透镜以侦测暗能量和暗物质。寻找太阳系中的小天体，尤其是近地小行星和古柏带天体。侦测光学瞬变现象，尤其是新星和超新星。观测银河系。

1.2.2 射电巡天项目

射电望远镜是主要接收天体射电波段辐射的望远镜。射电望远镜的外形差别很大，有固定在地面的单一口径的球面射电望远镜，有能够全方位转动的类似卫星接收天线的射电望远镜，有射电望远镜阵列，还有金属杆制成的射电望远镜。当前正在进行和规划中的射电巡天项目有：

- 平方千米阵 (Square Kilometre Array, 缩写为SKA) 是计划中的下一代巨型射电望远镜阵, 工作在0.10 - 30GHz的波段, 有效接收面积可以达到大约1平方公里, 灵敏度将比目前世界上最大的射电望远镜还要高50倍。平方千米阵将由上千台天线组成, 其中有一半天线位于中央直径5公里的区域内, 另有四分之一的天线散布在周围150公里的区域内, 其余的分布在大约3000公里的范围内, 呈螺旋形排列。平方千米阵预计能够探测到宇宙大爆炸之后第一代恒星和星系形成时发出的电磁波、揭示磁场在恒星和星系演化过程中的作用、探测暗能量产生的种种效应, 甚至有人希望能够接受到地外智慧生命发出的无线电信号。
- 阿雷西博天文台 (Arecibo Observatory) 位于波多黎各的阿雷西沃山谷中, 是世界上最大的单面口径电波望远镜, 直径达305米, 后扩建为350米, 由史丹佛国际研究中心、国家科学基金会与康奈尔大学管理。阿雷西博望远镜是固定望远镜, 不能转动, 只能通过改变天线馈源的位置扫描天空中的一个带状区域。非常著名的SETI@home项目就是该望远镜的一个重要研究课题。
- 甚大天线阵 (Very Large Array, 缩写为VLA) 是由27台25米口径的天线组成的射电望远镜阵列, 位于美国新墨西哥州的圣阿古斯丁平原上, 海拔2124米, 是世界上最大的综合孔径射电望远镜。甚大天线阵每个天线重230吨, 架设在铁轨上, 可以移动, 所有天线呈Y形排列, 每臂长21千米, 组合成的最长基线可达36千米。甚大天线阵隶属于美国国家射电天文台 (NRAO), 于1981年建成, 工作于6个波段, 最高分辨率可以达到0.05角秒, 与地面大型光学望远镜的分辨率相当。天文学家使用甚大天线阵做出了一系列重要的发现, 例如发现银河系内的微类星体、遥远星系周围的爱因斯坦环、伽玛射线暴的射电波段对应体等等。

1.2.3 其他波段巡天项目

由于地球大气的作用, 地基望远镜只能观测光学、射电及近红外等非常小的窗口。而对于其他波段的观测只能通过发射空间天文卫星进行观测。主要的空间天文卫星巡天项目有:

- 紫外紫外线天文学是研究天体紫外线辐射的天文学分支学科; 观测电磁

波波长大约在100到3200埃之间。天体的紫外线光谱可用来了解星际介质的化学成分、密度以及温度；以及高温年轻恒星的温度与组成。星系演化的讯息也可从紫外线观测得知。以紫外线观测天体的结果会与光学观测有很大的差异。许多在光学观测上相对温度较低的恒星在紫外线观测时却显示是高温天体，尤其是在演化阶段早期或晚期恒星。远紫外分光探测器（Far Ultraviolet Spectroscopic Explorer，缩写为FUSE）是约翰霍普金斯大学应用物理实验室为美国宇航局研制的一颗紫外天文卫星，是美国宇航局“起源计划”的一部分。加拿大和法国也参与了卫星的研制。该卫星于1999年6月24日在卡纳维拉尔角用德尔塔 II 型火箭发射升空，运行在高度为773千米的近地轨道上，倾角约25度，周期为100分钟。远紫外分光探测器工作在电磁波谱中波长为90-120纳米的紫外波段，主要科学目标包括研究宇宙大爆炸初期的氦合成、宇宙中各种化学元素的丰度、星系的化学演化、星际介质等。星系演化探测器（Galaxy Evolution Explorer，缩写为GALEX）是美国宇航局2003年发射的一颗紫外天文卫星，主要目的是观测星系，特别是那些包含大量年轻恒星、辐射出强烈紫外线的星系，研究它们的形成和演化机制。

- X射线X射线天文学是以天体的X射线辐射为主要研究手段的天文学分支。X射线天文学中常以电子伏特（eV）表示光子的能量，观测对象为0.1keV到100keV的X射线。其中又将0.1keV-10keV的X射线称为软X射线，10keV-100keV称为硬X射线。由于X射线属于电磁波谱的高能端，因此X射线天文学与伽玛射线天文学同称为高能天体物理学。宇宙中辐射X射线的天体包括X射线双星、脉冲星、伽玛射线暴、超新星遗迹、活动星系核、太阳活动区，以及星系团周围的高温气体等等。1970年12月，美国在肯尼亚发射了人类历史上第一颗X射线天文卫星——乌呼鲁卫星。该卫星在轨期间进行了系统的X射线巡天，确定了339个X射线源，发现了许多银河系中的X射线双星、来自遥远星系团的X射线，以及第一个黑洞候选天体——天鹅座X-1。截至2006年，正在工作的X射线天文卫星有欧洲的XMM-牛顿卫星、美国的罗西X射线时变探测器、钱德拉X射线天文台、日本的朱雀卫星。此外，欧洲的国际伽玛射线天体物理实验室（INTEGRAL）、美国用于观测伽玛射线暴的雨燕卫星（Swift）、日本用于观测太阳的日出卫星也安装有X射线观测设备。

- 伽马射线天文学是指以伽马射线研究宇宙的天文学分支。伽马射线是可穿透整个宇宙的电磁波中最高能量的波段，也是电磁波谱中波长最短的部分。伽马射线可由太空中的超新星、正电子湮灭、黑洞形成、甚至是放射衰变产生。近年主要的太空伽玛射线天文台是国际伽玛射线天体物理实验室（INTEGRAL）和费米伽玛射线空间望远镜。

1.2.4 国内巡天项目

受到望远镜及观测仪器的限制，国内的巡天项目起步较晚，但也陆陆续续的开展了大量极具影响的巡天项目。

- BATC大视场多色巡天是1992年由陈建生院士发起并组建的大视场、大样本的高精度测光和低分辨光谱的巡天计划。BATC所采用的望远镜是60/90厘米施密特望远镜。它配备4096x4096像元的CCD和15个中等带宽滤光片和多种窄带滤光片（覆盖从3000埃到10000埃的波段）。这就组成了在国际上极富创新特色的大视场、高精度、低分辨率的大样本天体光谱巡天系统。
- 中德6cm银道面巡天项目利用中科院新疆天文台南山站25米直径射电天文望远镜，在6厘米波段（频率为5GHz）开展银河系射电偏振巡天，并和德国100米直径射电望远镜在21厘米波段进行的银道面射电巡天结果进行比对研究。巡天项目共观测约4500小时，覆盖银道面银经10度到230度、银纬正负5度之间共2200平方度的天区，成为迄今为止全球利用陆基望远镜开展的银道面偏振巡天中观测频率最高的大范围巡天项目。
- “宇宙第一缕曙光”探测项目（即21CMA）。现在，天文学家们已能观察到距今130亿年的星光，但再往前追溯却遇到了障碍，除了宇宙大爆炸时留下的一些辐射信号外，我们观察不到任何光信号，因此这一时间段被称为宇宙的“黑暗时代”。由于宇宙中第一代天体形成于“黑暗时代”，因此探寻这一时段宇宙的奥秘是所有天文学家梦寐以求的事。“宇宙第一缕曙光”探测项目则为中国天文学家揭开谜底提供一个机会。
- 500米口径球面射电望远镜（英语：Five hundred meter Aperture Spherical Telescope，简称FAST）是中国一座在建的射电望远镜。FAST位于贵州省平塘县克度镇大窝凼洼地，利用喀斯特洼地的地势而建。FAST已

于2008年12月26日开工，预计2016年9月完工。其科学目标包括1. 巡视宇宙中的中性氢2. 观测脉冲星——研究极端状态下的物质结构与物理规律3. 主导国际甚长基线干涉测量网——获得天体超精细结构4. 探测星际分子——研究恒星形成与演化及探索太空生命起源5. 搜索星际通讯信号——寻找地外文明

- 深度uvby β 巡天是中国科学院国家天文台即将展开的一个在Stromgren-Crawford(SC)测光系统下的全天巡天计划。SC系统可以为我们带来可以和高分辨率光谱相媲美的恒星参数。该巡天项目计划使用新疆南山的1米望远镜进行全天巡天，获取大量科学数据，并展开研究。
- 硬X射线调制望远镜（Hard X-ray Modulation Telescope）是一台已知计划中世界最高灵敏度和最好空间分辨本领的空间硬X射线望远镜，它将实现空间硬X射线高分辨巡天，发现大批高能天体和天体高能辐射新现象，并对黑洞，中子星等重要天体进行高灵敏度定向观测，推进人类对极端条件下高能天体物理动力学、粒子加速和辐射过程的认识。

1.2.5 LAMOST巡天

郭守敬望远镜（LAMOST，大天区面积多目标光纤光谱天文望远镜）是一架横卧南北方向的特殊的中星仪式反射施密特望远镜。它是由反射施密特改正板MA（大小为5.72米 \times 4.40米，24块对角线长1.1米，厚度为25毫米的六角形平面子镜组成）、球面主镜MB（大小为6.67米 \times 6.05米，37块对角线长为1.1米，厚度为75毫米的六角形球面子镜组成）和焦面构成。球面主镜及焦面固定在地基上，反射施密特改正板作为定天镜跟踪天体的运动，望远镜在天体经过中天前后时进行观测。天体的光经MA反射到MB，再经MB反射后成像在焦面上。焦面上放置的光纤，将天体的光分别传输到光谱仪的狭缝上，然后通过光谱仪后的CCD探测器同时获得大量天体的光谱。LAMOST应用薄镜面主动光学加拼接镜面主动光学技术，在曝光1.5小时内可以观测到暗达20.5等的天体，使其成为大口径兼大视场光学望远镜的世界之最。同时，采用并行可控的光纤定位技术，在5度视场，直径为1.75米的焦面上放置4000根光纤，同时获得4000个天体的光谱，使其成为世界上光谱获取率最高的望远镜。

LAMOST安放在中国科学院国家天文台兴隆观测站，该站地处燕山主峰南麓，位于河北省兴隆县连营寨（东经7小时50分，北纬40度23分），海拔960米。

LAMOST在大规模光学光谱观测和大视场天文学研究方面，居于国际领先的地位。仪器主要性能指标：

- MA尺寸：5.72米×4.4米
- MB尺寸：6.67米×6.05米
- 有效通光口径：4米
- 视场角直径：5度
- 焦面线直径：1.75米
- 焦距：20米
- 光纤数：4000根
- 光谱覆盖范围：370-900纳米
- 光谱分辨率：1/0.25纳米
- 观测天区：赤纬-10° 到+90°

作为国家重大科学工程，LAMOST于2001年9月开工，2008年10月落成，2009年4月通过了中国科学院组织的工艺鉴定和建安、财务、设备、档案四个专业组的验收，并于2009年6月4日圆满通过了国家发展和改革委员会主持的国家验收。LAMOST的建成，突破了天文望远镜大视场与大口径难以兼得难题，成为目前国际上口径最大的大视场望远镜，是我国光学望远镜研制的又一里程碑，显著提高了我国在大视场多目标光纤光谱观测设备领域的自主创新能力。

LAMOST的建成使我国在大规模天文光谱观测研究工作跃居国际领先地位，为我国在天文学和天体物理学许多研究领域取得重大科研成果奠定了基础。为了充分发挥LAMOST的威力，获得最大的科学回报，天文学家们结合望远镜的功能和特点为它制定了一系列的观测计划，设计了三大核心研究课题：河外星系的观测，银河系结构和演化以及多波段目标证认三个方面。

1. 河外星系的观测。LAMOST首个科学目标是研究宇宙和星系，一个是星系红移巡天，另一是通过获取的数据进一步研究星系的物理特性。对近

千万个星系、类星体等河外天体的光谱观测,将在宇宙学模型、宇宙大尺度结构、星系形成和演化等研究上做出重大贡献。星系物理是目前国际天文界相当热门的话题,宇宙的诞生、星系的形成以及恒星和银河系结构等前沿问题都建立在对星系物理的研究基础之上。研究宇宙大尺度结构依赖于星系红移巡天的工作。获取星系的光谱就能得到星系的红移,有了红移就有可以知道它的距离,有了距离就有了三维分布,这样就可以了解整个宇宙空间的结构。同时可以研究包括星系的形成、演化在内的宇宙大尺度结构和星系物理。这是一个环环相扣的工程,而获取星系的光谱则是最基础的一环。LAMOST的目标是观测1000万个星系、100万个类星体、外加1000万颗恒星的光谱。LAMOST建成后,由于要比SDSS计划所观测的星系和类星体的数目多十倍,由此可以预计,LAMOST将会以更高精度的方式来确定宇宙的组成和结构,从而使人类对暗能量和暗物质有更加深刻的认识。

2. 银河系结构和演化。LAMOST第二个科学目标是研究恒星和银河系的结构特征,对大量恒星的光谱巡天将在银河系结构与演化及恒星物理的研究上做出重大贡献。主要瞄准更暗的恒星,观测数目更多一些,这样可以更多了解银河系更远处的恒星的分布和运动情况,弄清银河系结构。因为LAMOST能够做大量恒星的样本,所以可以尽量选更多、更暗的星来做大范围的研究。恒星是众多星系的重要组成部分。通过一颗恒星的光谱,天文学家可以分析出其密度、温度等物理条件,可以分析出其元素构成和含量等化学组成,还可以测量出其运动速度和运行轨迹等。研究了不同种类的恒星的分布,可以研究出银河系的结构和银河系的形成。
3. 多波段目标证认。LAMOST的第三个科学目标是多波段证认,结合红外、射电、X射线、伽马射线巡天的大量天体的光谱观测将在各类天体多波段交叉证认上做出重大贡献。因为光谱理论充分,经验也多,这也造就了其它手段搜集到的有关天体的资料最终还是要通过光谱来确认。作为光谱获取率最高的天文望远镜,LAMOST对光学天文学的意义是不言而喻的。而多波段证认本身也是LAMOST的三大课题之一,通过与其它波段巡天望远镜,如X射线和望远镜相结合,它在许多天文学前沿问题的解决上都能起到相当大的作用。

1.3 数据挖掘在天文中的应用

数据挖掘（Data Mining, DM）是目前人工智能和数据库领域研究的热点问题，所谓数据挖掘是指从数据库的大量数据中揭示出隐含的、先前未知的并有潜在价值的信息的非平凡过程。数据挖掘是一种决策支持过程，它主要基于人工智能、机器学习、模式识别、统计学、数据库、可视化技术等，高度自动化地分析企业的数据库，做出归纳性的推理，从中挖掘出潜在的模式，帮助决策者调整市场策略，减少风险，做出正确的决策。

知识发现过程由以下三个阶段组成：（1）数据准备，（2）数据挖掘，（3）结果表达和解释。数据挖掘可以与用户或知识库交互。数据挖掘是通过分析每个数据，从大量数据中寻找其规律的技术，主要有数据准备、规律寻找和规律表示3个步骤。数据准备是从相关的数据源中选取所需的数据并整合成用于数据挖掘的数据集；规律寻找是用某种方法将数据集所含的规律找出来；规律表示是尽可能以用户可理解的方式（如可视化）将找出的规律表示出来。

按照方法的不同，本文把数据挖掘在天文学中应用分为以下几类：

1.3.1 分类

1.3.2 回归分析

1.3.3 聚类

1.3.4 关联规则

1.3.5 离群数据挖掘

巡天观测的数据集除了有大量的样本可以做上述几类研究外，不可避免的会包含一些数据对象，这部分数据对象与数据的一般行为或模型不一致。我们称这些数据为离群数据，对这些数据的挖掘称为离群数据挖掘。离群数据挖掘被广泛应用于如欺诈检测、医疗处理、公共安全、工业损毁检测、入侵检测等方面。离群数据检测方法很多，主要可分为以下两类：

- 监督、半监督及无监督方法

如果可以通过专家标记正常和离群数据的事例，则可以使用这些数据来建立离群数据检测的模型。所使用的方法可以划分为：

1. 监督方法。监督方法针对数据的正常性和异常性建模。领域专家考察并标记基础数据的一个样本。然后，离群点检测可以用分类问题来建模，任务是学习一个可以识别离群数据的分类器。
2. 无监督方法。无监督的离群数据检测方法假设正常数据在某种程度上是“聚类的”。也就是说，该类方法预计正常数据遵循远比离群数据频繁的模式。正常数据不必落到一个组群，具有高度的相似类，而是可以形成多个组群，每个组群具有不同的特征。同时，离群数据将是远离正常数据的群组。
3. 半监督方法。半监督离群数据检测方法可以看做半监督学习方法的应用。当有一部分被标记的正常数据时，可以使用它们与邻近的对象一起，训练一个正常对象的模型。然后，使用这个正常对象的模型来检测离群数据，不符合这个正常对象模型的对象被分类为离群数据。

- 统计方法、基于近邻型的方法和聚类的方法

离群数据检测方法对离群数据与其余数据作为假设。根据所做假设，可以把离群数据挖掘、的方法分类三类：

1. 统计学方法。统计学方法（又称为基于模型的方法）对数据的正常性作出假设。先假设正常的数据对象有一个统计（随机）模型产生，而不遵循该模型的数据即是离群数据。按照学习模型的不同，基于统计学的离群数据挖掘方法可以划分为两个主要类型：参数方法和非参数类型。
2. 基于近邻性的方法。基于近邻性的方法假定一个对象是离群数据，则它在特征空间中的最近邻也远离它，即该对象与它的最近邻之间的邻近性明显的偏离于数据集中其他对象与它们近邻之间的邻近性。常用的基于近邻性的离群数据挖掘方法包括基于距离和基于密度的离群数据挖掘。
3. 基于聚类的方法。基于聚类的离群数据挖掘算法假定正常数据属于大的、稠密的簇，而离群点属于小或者稀疏的簇，或者不属于任何簇。

天文学的一个重要目的是发现异常、稀少甚至于位置类型的天体目标，而离群数据挖掘作为一种重要的手段可以很好的从巡天数据中有效的找出其中特征异常而且数量相对较少的数据。除了前面介绍的利用分类的方法搜寻一些特定类型的观测目标以外，还有一些工作利用其他离群数据挖掘的方法进行特殊天体的搜寻：

- 文献[8]中对SDSS DR7中红移从 $Z=0.6$ 到 4.3 的类星体，利用自组织映射的方法投影到二维空间中，然后选取密度较低区域的光谱进行人眼查看，进而从中搜寻出异常的类星体光谱。他们搜寻出来的异常类星体光谱主要包括宽吸收线、红端连续谱异常及弱发射线三种类型。
- 屠良平等人[9]在搜寻超新星时，其中主要一步是对搜索范围进行约减，他们使用的方法是在超新星特征空间中利用局部孤立性因子算法（LOF）将搜寻范围约减到前10%的离群数据[10]。
- 文献[11]提出了一种新的离群数据挖掘方法-MCLOF方法，通过给每一条光谱标记一个数值（离群因子）来作为异常及稀少光谱的比较，进而快速、自动并有效的挖掘其中的离群数据。然后将该方法应用于SDSS DR8中接近50万条恒星光谱中，按照信噪比大于3、离群因子大于0.85的标准共筛选出37033条离群数据。离群数据中包括发射线恒星、双星、碳星及连续谱异常的恒星等。

1.3.6 特征选取与转变

通常情况下，巡天中的观测目标都包含了若干属性如光谱中不同波长处的流量、多波段的星等、位置信息、运动学特征等，但对于某些具体的问题来说这些属性并不都是必须的。同时大量的属性导致了高维空间中低密度甚至空洞的属性空间。这使得对数据的归纳及有用结果的产生变的非常困难，因此需要对数据进行降维，从而在保持尽量多信息的情况下使用较少的属性。

常见的线性降维方法包括主成分分析(Principal Component Analysis, PCA)[12]、多维缩放（Multi-dimensional scaling，MDS）[13]、因子分析[14]、投影追踪[15]、线性判别分析[16]、局部保留投影[17]、独立分量分析[18]等。常见的非线性降维方法包括自组织映射网络[19]、核主成分分析(Kernel Principal Component Analysis, KPCA)[20]、等距映射算法[21]、局部线性嵌入[22]等。

PCA作为一种传统而且有用的数据特征提取方法，被大量应用于天文数据的数据挖掘中。

- 文献[23-25]在SDSS及LAMOST数据中搜寻激变变星候选体时，先对SDSS中已知的激变变星光谱进行PCA分析，将观测光谱投影到此特征空间中，然后利用分类算法对所有观测光谱进行二分类，进而搜寻其中的候选体。
- 文献[26]中将PCA方法应用到近10万条SDSS的恒星光谱数据中。按照g-r颜色0.02星等的间隔，将所有光谱分成65个不同分分组，对每个分组进行PCA分析。他们发现前4个特征光谱就可以非常好的来描述观测光谱。同时还讨论特征系数和恒星大气参数之间的关系。
- 文献[27]中在利用实测光谱进行恒星参数估计时，利用PCA方法对实测光谱进行降维，以此为特征训练非线性回归模型来求恒星参数。
- 文献[28]中将PCA应用于简单星族的样本中来筛选一些对年龄敏感的方法。

此外，LLE[29, 30]、IsoMap[31]、核熵成分分析[32]等方法也被应用天文数据的降维中。

1.4 本文研究背景

特殊、未知、稀少天体在天文学中的研究中有着极其重要的意义[56]，如星系中的超新星[57, 58]、银河系中的激变变星[59-65]及贫金属星[47, 50]等。这类天体是比较的稀少的，需要在大样本的数据进行寻找。LAMOST的巡天计划的结果将产生大量的数据，通过人工或者交互式的方法显然是不可行的，因此需要研究相关的算法来从其中发现特殊、未知天体的光谱。

对于特殊未知天体的发现的方法，主要分为两类：有指导的学习和无指导的学习。有指导的学习即分类[66]，通过事先给出有分类的两类或多类光谱训练一个分类器，对于新的光谱通过该分类器给出一个分类。无指导的学习主要包括聚类和离群点挖掘两种，聚类通过数量化关系来描述不同光谱之间的相似度，相似的光谱聚集在一起形成簇，而不在任何簇中的光谱为奇异光谱，离群点挖掘是指大量样本寻找离群数据的方法来寻找奇异光谱。

1.5 本文的工作介绍和篇章安排

本研究的工作主要分为两部分：（1）设计并实现适用于LAMOST光谱的恒星大气参数测量系统，包括光谱预处理、参数测量等模块，其中参数测量模块支持扩展性，现已集成SSPP[16-18]、UlySS [19]等软件包，并加入PLS方法,随着研究的不断进行，会有更多的方法集成进来。软件系统采用Python，结合GTK实现图形用户界面，运用多线程编程计算实现对海量光谱的快速批量处理。（2）研究可用于LAMOST光谱中发现特殊天体的数据挖掘的算法，主要包括有指导和无指导两类，其中前者主要发现一些已知的特殊天体，而后者主要是发现一些未知的特殊天体。研究的方法主要包括随机森林算法及遗传算法等等。

参考文献

- [1] Detlev Koester, Jonathan Girven, Boris T Gänsicke, and P Dufour. Cool dz white dwarfs in the sdss. *arXiv preprint arXiv:1105.0268*, 2011.
- [2] Daniel J Eisenstein, James Liebert, Hugh C Harris, SJ Kleinman, Atsuko Nitta, Nicole Silvestri, Scott A Anderson, JC Barentine, Howard J Brewington, J Brinkmann, et al. A catalog of spectroscopically confirmed white dwarfs from the sloan digital sky survey data release 4. *ApJS*, 167(1):40, 2006.
- [3] Yue Wu, A-Li Luo, Hai-Ning Li, Jian-Rong Shi, Philippe Prugniel, Yan-Chun Liang, Yong-Heng Zhao, Jian-Nan Zhang, Zhong-Rui Bai, Peng Wei, et al. Automatic determination of stellar atmospheric parameters and construction of stellar spectral templates of the Guoshoujing Telescope (LAMOST). *RAA*, 11(8):924, 2011.
- [4] Adam S Bolton, David J Schlegel, Éric Aubourg, Stephen Bailey, Vaishali Bhardwaj, Joel R Brownstein, Scott Burles, Yan-Mei Chen, Kyle Dawson, Daniel J Eisenstein, et al. Spectral Classification and Redshift Measurement for the SDSS-III Baryon Oscillation Spectroscopic Survey. *AJ*, 144(5):144, 2012.
- [5] John J Bochanski, Andrew A West, Suzanne L Hawley, and Kevin R Covey. Low-mass dwarf template spectra from the Sloan digital sky survey. *AJ*, 133(2):531, 2007.
- [6] 顾炎武, 周苏平, and 陈国庆. 日知录. 北方妇女儿童出版社, 2001.
- [7] 屈原 and 陆元炽. 天问. 北京出版社, 1987.
- [8] H Meusinger, P Schalldach, R-D Scholz, M Newholm, A de Hoon, B Kaminsky, et al. Unusual quasars from the sloan digital sky survey selected by means of kohonen self-organising maps. *AA*, 541:77, 2012.

- [9] LiangPing Tu, ALi Luo, FuChao Wu, and YongHeng Zhao. A method of searching for supernova candidates from massive galaxy spectra. *Science China Physics, Mechanics and Astronomy*, 53(10):1928–1938, 2010.
- [10] 屠良平, 罗阿理, 吴福朝, and 赵永恒. 海量星系光谱中的超新星候选范围自动约减. *光谱学与光谱分析*, (12):3420–3423, 2009.
- [11] Peng Wei, Ali Luo, Yinbi Li, Jingchang Pan, Liangping Tu, Bin Jiang, Xiao Kong, Zhixin Shi, Zhenping Yi, Fengfei Wang, et al. Mining unusual and rare stellar spectra from large spectroscopic survey data sets using the outlier-detection method. *MNRAS*, 431(2):1800–1811, 2013.
- [12] Ian T Jolliffe. *Principal component analysis*. Springer verlag, 2002.
- [13] Joseph B Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, 1964.
- [14] Louis Leon Thurstone. *Multiple factor analysis*. 1947.
- [15] Peter J Huber. Projection pursuit. *The annals of Statistics*, pages 435–475, 1985.
- [16] Alan Julian Izenman. *Linear Discriminant Analysis*. Springer, 2008.
- [17] Xiaofei He and Partha Niyogi. Locality preserving projections. In *NIPS*, volume 16, pages 234–241, 2003.
- [18] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent component analysis*, volume 46. John Wiley & Sons, 2004.
- [19] Teuvo Kohonen. *Self-organizing maps*, volume 30. Springer, 2001.
- [20] Bernhard Scholkopf, Alexander Smola, and Klaus-Robert Müller. Kernel principal component analysis. In *Advances in kernel methods-support vector learning*. Citeseer, 1999.

-
- [21] Joshua B Tenenbaum, Vin De Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.
- [22] Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.
- [23] 姜斌, 罗阿理, and 赵永恒. 基于随机森林的激变变星候选体的数据挖掘. *光谱学与光谱分析*, 32(2):510, 2012.
- [24] 姜斌, 罗阿理, and 赵永恒. 海量光谱中激变变星候选体的数据挖掘. *光谱学与光谱分析*, 31(8):2278–2282, 2011.
- [25] Bin Jiang, Ali Luo, Yongheng Zhao, and Peng Wei. Data mining for cataclysmic variables in the Large Sky Area Multi-Object Fibre Spectroscopic Telescope archive. *MNRAS*, 430(2):986–995, 2013.
- [26] Rosalie C McGurk, Amy E Kimball, and Željko Ivezić. Principal Component Analysis of Sloan Digital Sky Survey Stellar Spectra. *AJ*, 139(3):1261, 2010.
- [27] P Re Fiorentin, CAL Bailer-Jones, YS Lee, TC Beers, T Sivarani, R Wilhelm, C Allende Prieto, and JE Norris. Estimation of stellar atmospheric parameters from sdss/segue spectra. *arXiv preprint astro-ph/0703309*, 2007.
- [28] Xu Kong and FZ Cheng. Principal component analysis for spectral indices of stellar populations. *Monthly Notices of the Royal Astronomical Society*, 323(4):1035–1040, 2001.
- [29] Bu Yude, Pan Jingchang, Jiang Bin, and Wei Peng. Stellar spectral subclass classification based on locally linear embedding. *Publications of the Astronomical Society of Japan*, 65:81, 2013.
- [30] Scott F Daniel, Andrew Connolly, Jeff Schneider, Jake Vanderplas, and Liang Xiong. Classification of stellar spectra with local linear embedding. *The Astronomical Journal*, 142(6):203, 2011.

- [31] 卜育德, 潘景昌, and 陈福强. 基于isomap 算法的恒星光谱离群点挖掘. 光谱学与光谱分析, (1), 2014.
- [32] Yong De Hu, Jing Chang Pan, and Xin Tan. High-dimensional data dimension reduction based on keca. *Applied Mechanics and Materials*, 303:1101–1104, 2013.
- [33] Donald G York, J Adelman, John E Anderson Jr, Scott F Anderson, James Annis, Neta A Bahcall, JA Bakken, Robert Barkhouser, Steven Bastian, Eileen Berman, et al. The sloan digital sky survey: Technical summary. *AJ*, 120(3):1579, 2000.
- [34] Paula Szkody, Scott F Anderson, Marcel Agüeros, Ricardo Covarrubias, Misty Bentz, Suzanne Hawley, Bruce Margon, Wolfgang Voges, Arne Henden, Gillian R Knapp, et al. Cataclysmic variables from the sloan digital sky survey. i. the first results. *AJ*, 123(1):430, 2002.
- [35] Paula Szkody, Oliver Fraser, Nicole Silvestri, Arne Henden, Scott F Anderson, James Frith, Brandon Lawton, Ethan Owens, Sean Raymond, Gary Schmidt, et al. Cataclysmic variables from the sloan digital sky survey. ii. the second year. *AJ*, 126(3):1499, 2003.
- [36] Paula Szkody, Arne Henden, Oliver Fraser, Nicole Silvestri, John Bochanski, Michael A Wolfe, Marcel Agüeros, Brian Warner, Patrick Woudt, Jonica Trampusch, et al. Cataclysmic variables from the sloan digital sky survey. iii. the third year. *AJ*, 128(4):1882, 2004.
- [37] Paula Szkody, Arne Henden, Oliver J Fraser, Nicole M Silvestri, Gary D Schmidt, John J Bochanski, Michael A Wolfe, Marcel Agüeros, Scott F Anderson, Lee Mannikko, et al. Cataclysmic variables from sloan digital sky survey. iv. the fourth year (2003). *AJ*, 129(5):2386, 2005.
- [38] Paula Szkody, Arne Henden, Marcel Agüeros, Scott F Anderson, John J Bochanski, Gillian R Knapp, Lee Mannikko, Anjum Mukadam, Nicole M Silvestri, Gary D Schmidt, et al. Cataclysmic variables from sloan digital sky survey. v. the fifth year (2004). *AJ*, 131(2):973, 2006.

-
- [39] Paula Szkody, Arne Henden, Lee Mannikko, Anjum Mukadam, Gary D Schmidt, John J Bochanski, Marcel Agüeros, Scott F Anderson, Nicole M Silvestri, William E Dahab, et al. Cataclysmic variables from sloan digital sky survey. vi. the sixth year (2005). *AJ*, 134(1):185, 2007.
- [40] Paula Szkody, Scott F Anderson, Keira Brooks, Boris T Gänsicke, Martin Kronberg, Thomas Riecken, Nicholas P Ross, Gary D Schmidt, Donald P Schneider, Marcel A Agüeros, et al. Cataclysmic variables from the sloan digital sky survey. viii. the final year (2007–2008). *AJ*, 142(6):181, 2011.
- [41] A Rebassa-Mansergas, A Nebot Gómez-Morán, Matthias R Schreiber, Boris T Gänsicke, A Schwöpe, J Gallardo, and Detlev Koester. Post-common envelope binaries from sdss–xiv. the dr7 white dwarf–main-sequence binary catalogue. *MNRAS*, 419(1):806–816, 2012.
- [42] Yinbi Li, Ali Luo, Gang Zhao, Youjun Lu, Juanjuan Ren, and Fang Zuo. Metal-poor hypervelocity star candidates from the sloan digital sky survey. *ApJL*, 744(2):L24, 2012.
- [43] Xiang-Qun Cui, Yong-Heng Zhao, Yao-Quan Chu, Guo-Ping Li, Qi Li, Li-Ping Zhang, Hong-Jun Su, Zheng-Qiu Yao, Ya-Nan Wang, Xiao-Zheng Xing, et al. The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST). *RAA*, 12(9):1197, 2012.
- [44] Gang Zhao, Yong-Heng Zhao, Yao-Quan Chu, Yi-Peng Jing, and Li-Cai Deng. LAMOST spectral survey—An overview. *RAA*, 12(7):723, 2012.
- [45] A-Li Luo, Hao-Tong Zhang, Yong-Heng Zhao, Gang Zhao, Xiang-Qun Cui, Guo-Ping Li, Yao-Quan Chu, Jian-Rong Shi, Gang Wang, Jian-Nan Zhang, et al. Data release of the LAMOST pilot survey. *RAA*, 12(9):1243, 2012.
- [46] Li-Cai Deng, Heidi Jo Newberg, Chao Liu, Jeffrey L Carlin, Timothy C Beers, Li Chen, Yu-Qin Chen, Norbert Christlieb, Carl J Grillmair, Puragra Guhathakurta, et al. LAMOST Experiment for Galactic Understanding and Exploration (LEGUE)—The survey’s science plan. *RAA*, 12(7):735, 2012.

- [47] Hai-Ning Li, Gang Zhao, Norbert Christlieb, A-Li Luo, Jing-Kun Zhao, Yong-Heng Zhao, Jian-Jun Chen, and Zhong-Rui Bai. Test observations that search for metal-poor stars with the guoshoujing telescope (lamost). *RAA*, 10(8):753, 2010.
- [48] Xue-Bing Wu, Zhao-Yu Chen, Zhen-Dong Jia, Wen-Wen Zuo, Yong-Heng Zhao, A-Li Luo, Zhong-Rui Bai, Jian-Jun Chen, Hao-Tong Zhang, Hong-Liang Yan, et al. A very bright ($i=16.44$) quasar in the 'redshift desert' discovered by the guoshoujing telescope (lamost). 10(8):737, 2010.
- [49] Juanjuan Ren, Ali Luo, Yinbi Li, Peng Wei, Jingkun Zhao, Yongheng Zhao, Yihan Song, and Gang Zhao. White-dwarf-main-sequence binaries identified from the lamost pilot survey. *AJ*, 146(4):82, 2013.
- [50] Y Huang, X-W Liu, H-B Yuan, M-S Xiang, Z-Y Huo, Y-H Hou, G Jin, Y Zhang, and X-L Zhou. Hst and lamost discover a dual active galactic nucleus in j0038+ 4128. *arXiv preprint arXiv:1402.2740*, 2014.
- [51] Liang-Ping Tu, A-Li Luo, Fu-Chao Wu, Chao Wu, and Yong-Heng Zhao. New supernova candidates from the SDSS-DR7 spectral survey. *RAA*, 9(6):635, 2009.
- [52] Nanbo Peng, Yanxia Zhang, Yongheng Zhao, and Xue-bing Wu. Selecting quasar candidates using a support vector machine classification system. *MNRAS*, 425(4):2599–2609, 2012.
- [53] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. Lof: identifying density-based local outliers. In *ACM Sigmod Record*, volume 29, pages 93–104. ACM, 2000.
- [54] Hiroaki Aihara, Carlos Allende Prieto, Deokkeun An, Scott F Anderson, Éric Aubourg, Eduardo Balbinot, Timothy C Beers, Andreas A Berlind, Steven J Bickerton, Dmitry Bizyaev, et al. The eighth data release of the Sloan Digital Sky Survey: first data from SDSS-III. *ApJS*, 193(2):29, 2011.

-
- [55] Karl Glazebrook, Alison R Offer, and Kathryn Deeley. Automatic redshift determination by use of principal component analysis. I. Fundamentals. *ApJ*, 492(1):98, 1998.
- [56] VV Shimansky, DN Nurtdinova, NV Borisov, and OI Spiridonova. Is wd 1437-008 a cataclysmic variable? *Astrophysical Bulletin*, 66(4):442–448, 2011.
- [57] AH Batten. Binary and multiple star systems, 1973.
- [58] A Rebassa-Mansergas, C Agurto-Gangas, MR Schreiber, BT Gänsicke, and D Koester. White dwarf main-sequence binaries from sdss dr 8: unveiling the cool white dwarf population. *MNRAS*, 433(4):3398–3410, 2013.
- [59] George P McCook and Edward M Sion. A catalog of spectroscopically identified white dwarfs. *ApJS*, 121(1):1, 1999.
- [60] D Koester and S Knist. New dq white dwarfs in the sloan digital sky survey dr4: confirmation of two sequences. *A&A*, 454:951–956, 2006.
- [61] A-Li Luo and Yong-Heng Zhao. Steps Towards a Fully Automated Classification and Redshift-measurement Pipeline for LAMOST Spectra. I. Continuum level and wavelength estimation for galaxies. *Chinese Journal of Astronomy and Astrophysics*, 1(6):563, 2001.
- [62] A-Li Luo, Yan-Xia Zhang, and Yong-Heng Zhao. Design and implementation of the spectra reduction and analysis software for LAMOST telescope. In *Astronomical Telescopes and Instrumentation*, pages 756–764. International Society for Optics and Photonics, 2004.
- [63] A-Li Luo, Yue Wu, Jingkun Zhao, and Gang Zhao. Automated stellar spectral analysis software for survey spectra. In *Proc. of SPIE Vol*, volume 7019, pages 701935–1, 2008.
- [64] FF Wang, AL Luo, and YH Zhao. Calibration of LAMOST spectral analysis. In *SPIE Astronomical Telescopes and Instrumentation: Observational Fron-*

- tiers of Astronomy for the New Decade*, pages 774031–774031. International Society for Optics and Photonics, 2010.
- [65] J Falcón-Barroso, P Sánchez-Blázquez, A Vazdekis, E Ricciardelli, N Cardiel, AJ Cenarro, J Gorgas, and RF Peletier. An updated MILES stellar library and stellar population models. *A&A*, 532:A95, 2011.
- [66] F. Wang, H. Zhang, A. Luo, Z. Bai, B. Du, and Y. Zhao. The effects of spectrograph slit modes on the accuracy of stellar radial velocity measurement and atmospheric parameter estimation. *Science in China G: Physics and Astronomy*, 56:1833–1839, October 2013.
- [67] X-W Liu, H-B Yuan, Z-Y Huo, L-C Deng, J-L Hou, Y-H Zhao, G Zhao, J-R Shi, A-L Luo, M-S Xiang, et al. LSS-GAC-A LAMOST Spectroscopic Survey of the Galactic Anti-center. *arXiv preprint arXiv:1306.5376*, 2013.
- [68] Zhongrui Bai. LAMOST 2D pipeline. *Proceedings of the International Astronomical Union*, 8(S295):189–190, 2012.
- [69] Young Sun Lee, Timothy C Beers, Thirupathi Sivarani, Carlos Allende Prieto, Lars Koesterke, Ronald Wilhelm, Paola Re Fiorentin, Coryn AL Bailer-Jones, John E Norris, Constance M Rockosi, et al. The segue stellar parameter pipeline. i. description and comparison of individual methods. *AJ*, 136(5):2022, 2008.
- [70] Željko Ivezić, Branimir Sesar, Mario Jurić, Nicholas Bond, Julianne Dalcanton, Constance M Rockosi, Brian Yanny, Heidi J Newberg, Timothy C Beers, Carlos Allende Prieto, et al. The Milky Way Tomography with SDSS. II. Stellar Metallicity. *ApJ*, 684(1):287, 2008.
- [71] Zhenping Yi, Ali Luo, Yihan Song, Jingkun Zhao, Zhixin Shi, Peng Wei, Juanjuan Ren, Fengfei Wang, Xiao Kong, Yinbi Li, Peng Du, Wen Hou, Yanxin Guo, Shuo Zhang, Yongheng Zhao, Shiwei Sun, Jingchang Pan, Liyun Zhang, Andrew A. West, and Haibo Yuan. M Dwarf Catalog of the LAMOST Pilot Survey. *AJ*, 147(2):33, 2014.

-
- [72] JK Zhao, AL Luo, TD Oswalt, and G Zhao. 70 DA White Dwarfs Identified in LAMOST Pilot Survey. *AJ*, 145(6):169, 2013.
- [73] Yue-Yang Zhang, Li-Cai Deng, Chao Liu, Sebastien Lépine, Heidi Jo Newberg, Jeffrey L Carlin, Kenneth Carrell, Fan Yang, Shuang Gao, Yan Xu, et al. DA White Dwarfs Observed in the LAMOST Pilot Survey. *AJ*, 146(2):34, 2013.
- [74] Hans-Peter Kriegel, Peer Kröger, Erich Schubert, and Arthur Zimek. LoOP: local outlier probabilities. In *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 1649–1652. ACM, 2009.
- [75] CA Whitney. Principal components analysis of spectral data. I-Methodology for spectral classification. *Astronomy and Astrophysics Supplement Series*, 51:443–461, 1983.
- [76] Coryn AL Bailer-Jones, Mike Irwin, and Ted Von Hippel. Automated classification of stellar spectra—II. Two-dimensional classification with neural networks and principal components analysis. *MNRAS*, 298(2):361–377, 1998.
- [77] CW Yip, AJ Connolly, DE Vanden Berk, Z Ma, JA Frieman, M SubbaRao, AS Szalay, GT Richards, PB Hall, DP Schneider, et al. Spectral classification of quasars in the Sloan Digital Sky Survey: Eigenspectra, redshift, and luminosity effects. *AJ*, 128(6):2603, 2004.
- [78] J Sánchez Almeida and C Allende Prieto. Automated unsupervised classification of the Sloan Digital Sky Survey stellar spectra using k-means clustering. *ApJ*, 763(1):50, 2013.
- [79] Harinder P Singh, Ravi K Gulati, and Ranjan Gupta. Stellar spectral classification using principal component analysis and artificial neural networks. *MNRAS*, 295(2):312–318, 1998.
- [80] Anthony C Danks and Michel Dennefeld. An Atlas of Southern MK Standards From 5800 to 10,200 Å. *PASP*, pages 382–396, 1994.

- [81] A-Li Luo et al. Data release one of the LAMOST survey. (9):in preparation, 2013.
- [82] JianMin Si, ALi Luo, YinBi Li, JianNan Zhang, Peng Wei, YiHong Wu, FuChao Wu, and YongHeng Zhao. Search for carbon stars and DZ white dwarfs in SDSS spectra survey through machine learning. *Science China Physics, Mechanics and Astronomy*, 57(1):176–186, 2014.
- [83] Christopher P Ahn, Rachael Alexandroff, Carlos Allende Prieto, Scott F Anderson, Timothy Anderton, Brett H Andrews, Éric Aubourg, Stephen Bailey, Eduardo Balbinot, Rory Barnes, et al. The Ninth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Baryon Oscillation Spectroscopic Survey. *ApJS*, 203(2):21, 2012.
- [84] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. LOF: identifying density-based local outliers. In *ACM Sigmod Record*, volume 29, pages 93–104. ACM, 2000.
- [85] Jeffrey L Carlin, Sébastien Lépine, Heidi Jo Newberg, Li-Cai Deng, Timothy C Beers, Yu-Qin Chen, Norbert Christlieb, Xiao-Ting Fu, Shuang Gao, Carl J Grillmair, et al. An algorithm for preferential selection of spectroscopic targets in LEGUE. *RAA*, 12(7):755, 2012.
- [86] Jeffrey L Carlin, HJ Newberg, TC Beers, YS Lee, B Yanny, L Deng, H Zhang, G Zhao, Y Zhang, A Luo, et al. Comparison of Derived Stellar Parameters for LAMOST and SEGUE Spectra. In *Bulletin of the American Astronomical Society*, volume 43, page 33413, 2011.
- [87] Li Chen, Jin-Liang Hou, Jin-Cheng Yu, Chao Liu, Li-Cai Deng, Heidi Jo Newberg, Jeffrey L Carlin, Fan Yang, Yue-Yang Zhang, Shi-Yin Shen, et al. The LEGUE disk targets for LAMOST’s pilot survey. *RAA*, 12(7):805, 2012.
- [88] Daniel J Eisenstein, David H Weinberg, Eric Agol, Hiroaki Aihara, Carlos Allende Prieto, Scott F Anderson, James A Arns, Éric Aubourg, Stephen

- Bailey, Eduardo Balbinot, et al. Sdss-iii: Massive spectroscopic surveys of the distant universe, the milky way, and extra-solar planetary systems. *AJ*, 142(3):72, 2011.
- [89] Fan Yang, Jeffrey L Carlin, Chao Liu, Yue-Yang Zhang, Shuang Gao, Yan Xu, Li-Cai Deng, Heidi Jo Newberg, Sébastien Lépine, Jin-Liang Hou, et al. The LEGUE input catalog for dark night observing in the LAMOST pilot survey. *RAA*, 12(7):781, 2012.
- [90] Yue-Yang Zhang, Jeffrey L Carlin, Fan Yang, Chao Liu, Li-Cai Deng, Heidi Jo Newberg, Hao-Tong Zhang, Sébastien Lépine, Yan Xu, Shuang Gao, et al. The LEGUE high latitude bright survey design for the LAMOST pilot survey. *RAA*, 12(7):792, 2012.
- [91] 范晔 et al. 后汉书, 1965.

发表文章目录

- [1] *Mining unusual and rare stellar spectra from large spectroscopic survey data sets using the outlier-detection method*, Peng WEI, ALi LUO et al., MNRAS(SCI,IF2012=5.5), 2013, 431, 2
- [2] *On the Construction of New Stellar Classification Templates Library for LAMOST Spectra Analysis Pipeline*, Peng WEI, ALi LUO et al., AJ (SCI,IF2012=4.9),in press
- [3] *Outlier detection in LAMOST I: method description and first results from DR1*, Peng WEI, ALi LUO et al., in preparation for ApJ(SCI,IF2012=6.7)
- [4] *M dwarf catalog of LAMOST pilot survey*, Zhenping YI, ALi LUO et al. (6/20), AJ(SCI,IF2012=4.9),2014,147,33
- [5] *Search for carbon stars and DZ white dwarfs in SDSS spectra survey*, Jianmin SI, ALi LUO et al. (5/7), Science China Physics, Mechanics and Astronomy(SCI,IF2012=1.169),2014,57,176
- [6] *Data mining for cataclysmic variables in the Large Sky Area Multi-Object Fibre Spectroscopic Telescope archive*, Bin JIANG, ALi LUO, Yongheng ZHAO & Peng WEI (4/4), MNRAS(SCI,IF2012=5.5), 2013, 430, 2
- [7] *White-dwarf-Main-sequence Binaries Identified from the LAMOST Pilot Survey*, Juanjuan REN, ALi LUO et al. (4/8), AJ(SCI,IF2012=4.9), 2013, 146, 4
- [8] *Stellar Spectral Subclass Classification Based on Locally Linear Embedding*, Yude BU, Jingchang PAN, Bin JIANG, Peng WEI et al. (4/4), PASJ(SCI,IF2012=2.4), 2013, 65, 81

- [9] *Spectral Feature Extraction Based on the DCPCA Method*, Yude BU, Jingchang PAN, Bin JIANG, Fuqiang, CHEN, Peng WEI et al. (5/5), PASA(SCI,IF2012=3.1), 2013, 30, 24
- [10] *Automatic determination of stellar atmospheric parameters and construction of stellar spectral templates of the Guoshoujing Telescope (LAMOST)*, Yue WU, ALi LUO et al. (10/13), RAA(SCI,IF2012=1.348), 2011, 11, 8
- [11] *GA-Optimized Random Forest Classification for High Dimensional Data*, Jingchang Pan, Peng WEI et al. (2/5), ICIC Express Letters(EI), 2011, 5, 5
- [12] *Outlier Detection of Celestial Spectra Based on Modified GA*, Jingchang Pan, Peng WEI et al. (2/5), ICIC Express Letters(EI), 2011, 5, 1
- [13] 恒星光谱参数自动测量中不同模板匹配度量方法比较,刘杰等(3/5), 光谱学与光谱分析(EI,SCI,IF2012=0.3), 2013, 2, 558
- [14] 基于光谱相似度的恒星大气参数自动测量方法,刘杰等(4/5), 光谱学与光谱分析(EI,SCI,IF2012=0.3), 2012, 12, 3402
- [15] 一种新的活动星系光谱红移自动测量方法,屠良平等(4/6), 光谱学与光谱分析(EI,SCI,IF2012=0.3), 2013, 10, 2858
- [16] 基于统计窗的恒星连续谱自动拟合方法, 潘景昌等(3/6), 光谱学与光谱分析(EI,SCI,IF2012=0.3), 2013, 8, 2260
- [17] 自动识别发射线恒星光谱的新方法, 潘景昌等(3/5), 光谱学与光谱分析(EI,SCI,IF2012=0.3), 2012, 6, 1689

简 历

基本情况

韦鹏，男，山东临沂人，1987 年 1 月出生，未婚，中国科学院国家天文台在读博士研究生。

教育状况

2004 年 9 月至 2008 年 7 月，山东大学(威海)信息工程学院，本科，专业：计算机科学与技术。

2008 年 9 月至 2011 年 7 月，山东大学(威海)机电与信息工程学院，硕士研究生，专业：计算机应用技术。

2004 年 9 月至 2008 年 7 月，中国科学院国家天文台，博士研究生，专业：天文技术与方法。

工作经历

无。

研究兴趣

数据挖掘与模式识别，贝叶斯理论及应用。

联系方式

E-mail: weipeng@lamost.org

致 谢

从我进入大学到博士毕业，我用掉了整整十年。

十年前，我连我们那个小城我都没出过，知道的听到的还是那么的有限，可那时的我总是喜欢幻想以后，喜欢和朋友讨论如果我上了大学会干什么；

十年前，从早到晚总有做不完的习题，喜欢带着讲义去办公室直接找老师讨论问题，喜欢偷偷走过某班的门前看一眼自己喜欢的姑娘；

十年前，自己的脾气十分暴躁，没有多少朋友；

十年年，语数外理化生，每一科都不是自己的短板，自己会写出优美的语句来赞美美好的生活，熟练各种三角函数和立体几何，记住的单词曾一度超过6000，能够仔细分析小车运动的每一个过程，最擅长类似猜谜一样的化学题，对于各种遗传概率都信手拈来；

十年间，我从我们那个小城到了近海的另外一个小城，然后有一天走近了帝都。

十年后，每次同学聚会，都会和几个好朋友一起回忆大家做过的那些事情，所谓青春好像已经成为了一段美好的记忆；

十年后，每天从早到晚都坐在电脑前面，看看paper，写写代码，然后和导师汇报讨论进展；

十年后，有一天忽然发现当初喜欢的姑娘已经嫁做人妇，好多同学都已经为人父为人母；

十年后，自己学会了忍耐，学会了控制的情绪，朋友还不是特别多，但是知心的朋友不少；

十年后，我总是说，我学的是计算机，干的是天文数据处理；

十年，我的身高从170cm长到了175cm，体重也涨到了令人恐怖的一个数字。

十年前，我迈入大学校门的时候我刚17周岁，十年后，将要结束学生生涯的时候我已经27周岁。

从93年进入小学开始自己的学生生涯，到现在已经是整整21个年头，其中有10个年头在离开家乡的外地。一路走来，慢慢的自己在成长，从一个不懂事的小孩，慢慢的成长为一名青年。回首这些年，尤其是在中国科学院国家攻读博士学位这几年，想感谢的人特别多。

感谢我的博士生导师赵永恒研究员。赵老师九十年代从国外回来之后一直从事LAMOST的相关研究及管理工作，可以说赵老师为LAMOST奉献了自己的青春年华。虽然赵老师对我的直接指导不多，但是他的为人和对科学研究的态度深深的影响了我。赵老师对学生非常关心，每次开会或者年底中期答辩或开题答辩的时候，赵老师总是非常认真的在听，而且总会提出非常好的意见和建议。2012年在韶关参加会议时，赵老师对我的工作进行评价的场景，现在仍然历历在目，从他的话语中我听到了赵老师对这个工作的肯定，也是从那时候开始，我更坚定了从事现在这个方向研究的决心。赵老师的头上慢慢的出现了几根白发，但赵老师仍然在继续努力着。不敢奢求成为像赵老师这样的人，但是我们一直也在为此而努力着。

感谢我的博士生导师罗阿理研究员。

感谢我的硕士生导师潘景昌教授。

感谢姜斌老师

感谢屠良平老师

感谢张健楠老师

感谢李荫碧师姐

感谢杜薇师姐和王凤飞师兄

感谢吴悦师姐

感谢杜老师、艾老师和马老师

感谢LAMOST的所有工作人员和同学。

感谢山东大学威海机电与信息工程学院的老师

感谢山东大学威海计算机应用技术研究所的所有老师及同学

感谢我的父母寒心茹苦抚养我长大，特别感谢已经在逝去的父亲。