新疆天文台脉冲星数据检索

张海龙^{1,2*}, 王娜^{1,2}, 王杰¹, 聂俊^{1,2}, 袁建平^{1,2}, 王万琮¹

1. 中国科学院新疆天文台, 乌鲁木齐 830011;

2. 中国科学院射电天文重点实验室, 南京 210008;

*联系人, E-mail: zhanghailong@xao.ac.cn

收稿日期: 2018-12-13; 接受日期: 2019-01-16; 网络出版日期: 2019-06-11

国家重点基础研究发展计划(编号: 2015CB857100)、国家自然科学基金(编号: 11873082, U1531125, 11803080, 11503075)、国家重点研发计划(编号: 2018YFA0404704)和中国科学院青年创新促进会资助项目

摘要 新疆天文台南山观测站25 m射电望远镜(简称NSRT) 2002年L波段常温接收机制冷改造完成后, 脉冲星观测系统灵敏度达到了0.4 mJy, 实现了200多颗脉冲星的常年监测. NSRT通过近20年的观测产生了大量有价值的科学数据. 南山观测站脉冲星数据通过300 Mb/s专线传输到70 km外的新疆天文台本部数据中心进行长期存储, 通过南山观测站及新疆天文台本部的数据存储系统实现了脉冲星数据的远程异地容灾备份. 元数据提取、数据归档及发布由位于新疆天文台本部的数据发布系统完成, 截至2018年11月底已发布脉冲星观测数据87000多条, 最大文件1 GB, 脉冲星数据格式为PSRFITS. 脉冲星数据服务提供了锥形检索及多目标约束检索, 数据发布基于虚拟天文台标准并已将脉冲星数据服务注册到国际虚拟天文台联盟, 用户可通过标准虚拟天文台工具实现检索、访问与处理.

关键词 脉冲星, 元数据, 数据检索, 虚拟天文台

PACS: 47.27.-i, 47.27.Eq, 47.27.Nz, 47.40.Ki, 47.85.Gj

1 引言

脉冲星^[1]是大质量恒星演化到末期并经过核坍缩、超新星爆发后形成的产物, 具有体积小、密度大、高速旋转、强磁场等特性. 脉冲星的观测研究有助于揭示宇宙诞生及演化的奥秘, 对于推动天体物理学、粒子物理学和卫星导航等方面的发展具有非常重

要的科学意义, 并始终是天文学和天体物理学研究领域的热点. 自1967年Bell课题组^[2]用天线阵发现第一颗射电脉冲星以来, 其观测和理论研究都取得了飞速发展, 发现的脉冲星数目也在不断增加, 截至2018年12月已确认脉冲星总数达2659颗^[3].

新疆天文台南山观测站25 m射电望远镜, 经过改造后口径为26 m, 简称NSRT^[3], 是我国VLBI网的重要

1) <http://www.atnf.csiro.au/people/pulsar/psrcat/>

引用格式: 张海龙, 王娜, 王杰, 等. 新疆天文台脉冲星数据检索. 中国科学: 物理学 力学 天文学, 2019, 49: 099516

Zhang H L, Wang N, Wang J, et al. XAO pulsar data retrieval system (in Chinese). Sci Sin-Phys Mech Astron, 2019, 49: 099516, doi: 10.1360/SSPMA2018-00404

组成成员,是我国较早开展脉冲星观测研究的天文台站之一.经过二十几年的发展,已经成为全球脉冲星观测的重要基地.如何将观测到的脉冲星数据存储、归档、发布,从而实现脉冲星数据在全世界范围内共享是新疆天文台脉冲星观测研究过程中需要解决的问题.

2 脉冲星观测系统

1996年,NSRT开始尝试进行脉冲星实验观测,在327 MHz频率上观测取得成功,观测到8颗脉冲星,得到了4颗脉冲星的高质量数据和平均脉冲轮廓图,并观测到PSR B1133+16的反常模式.

1999年,在L波段(18 cm)建立了由双通道常温接收机、多通道滤波器、数字化器和降频转换器三部分组成的国内第一套256通道的脉冲星消色散观测系统^[4],该系统的中心频率为1540 MHz,消色散采用 $2 \times 128 \times 2.5$ MHz多通道模拟滤波器实现.2000年开始通过脉冲星到达时间观测系统对脉冲星进行监测,监测到流量密度大于4 mJy的强源有74颗.

2002年6月之前,NSRT使用L波段双通道常温接收机,消色散由模拟滤波器组(AFB)实现,AFB数据的格式是“Timer”^[5].2002年下半年,配备了L波段双极化制冷接收机,并投入了脉冲星观测,该系统的工作频率为1381.25–1701.25 MHz,中心频率为1540 MHz,极化器噪声温度小于8 K,接收机噪声温度小于10 K,观测灵敏度达到0.4 mJy.

2010年1月,与澳大利亚ATNF联合研制的DFB^[6]系统完成调试投入运行,更高的时间分辨率使NSRT能够监测280颗左右脉冲星,包括10颗毫秒脉冲星(MSP).DFB数据的格式为“PSRFITS”,记录的数据文件扩展名为*.rf,“psrchive”^[7]程序可以读取和分析数据.

3 脉冲星数据检索系统总体设计

NSRT脉冲星数据检索系统设计示意如图1所示,设计共包含三个部分:数据存储、数据检索和数据处理.数据存储部分涉及脉冲星观测数据获取与在线归

档;数据检索部分实现原始观测数据基于虚拟天文台^[8]标准的发布、检索与数据共享;数据处理部分主要功能是为用户提供计算资源,通过利用距离数据“最近”的计算资源提高数据的处理效率.

数据存储部分主要功能: NSRT数据获取及在线归档部分在南山观测站完成,脉冲星数据采集、处理及记录由望远镜接收系统完成.数据采集程序实时完成数据采集、数据预处理、脉冲星周期计算和周期叠加、数据存储.观测中典型的采样速率为1 ms,每次观测时间由脉冲星在该观测频段的流量强度决定,一般为4–16 min.数据经过预处理后写入暂存服务器中,经过确认有效后数据将实现永久归档.脉冲星数据通过速度为300 Mbps的专线实现南山观测站与新疆天文台本部存储系统间的相互同步,目前采用Rsync实现数据增量备份.

数据检索部分主要功能为脉冲星观测数据通过专线传输备份到新疆天文台本部后,数据发布系统将提取观测文件的元数据信息,并将元数据导入到相应的数据库中,在入库的同时建立相关索引,为数据发布做准备.新疆天文台采用GAVO^[9]作为数据发布框架实现元数据操作及基于虚拟天文台标准的数据发布.允许用户通过浏览器、标准虚拟天文台工具、各类脚本等访问数据检索平台.目前数据发布系统支持锥形检索与基于源名、观测日期、观测ID、终端等的多约束目标检索.用户可对检索输出的数据条数、输出字段、排序等内容进行自定义,目前脉冲星数据检索服务支持HTML, VOTable, FITS Table, CSV, tar等多种数据输出格式.

数据处理部分主要功能为提供计算资源. Taurus高性能计算系统^[2]、数据存储与数据发布服务器之间采用NFS方式实现数据互操作,可提供理论上56 Gb/s的数据交换速率.用户可通过高速网络将数据下载到Taurus高性能计算系统自身存储,并实现数据处理.用户处理后数据可根据需要进行归档、发布. Taurus集群系统由16个计算节点组成,单节点由2颗Intel E5-2692v2, Tesla K20M GPU, 64 G内存组成,整个系统共享一套高性能Lustre存储,容量为100 TB.经过Linpack^[10]测试其实际计算性能达到CPU理论值的93.5%, GPU理论值的80%,多节点写速度在650 MB/s左右^[11].

2) <http://taurus.xao.ac.cn/>

图 1 (网络版彩图) NSRT脉冲星数据检索示意图

Figure 1 (Color online) Overall design of NSRT pulsar data retrieval.

4 数据中心与脉冲星数据检索

新疆天文台数据中心³⁾界面如网络版补充材料图S1所示, 目前数据中心中已发布ADQL Query, PPMXL Catalog Cone Search, XAO DC Custom Uploading Crossmatcher, XAO Pulsar Data Query, XAO IDV Data Query, XAO SF Data Query等多项服务. 可检索的NSRT数据资源有脉冲星观测数据、活动星系核观测数据与分子谱线观测数据^[12].

4.1 Web页面检索

用户可通过输入地址⁴⁾的方式访问新疆天文台脉冲星数据检索页面, 或通过数据中心的“XAO Pulsar Data Query”链接打开检索页面. 截至2018年11月底已归档脉冲星数据87219条记录, 最大文件1 GB, 数据格式主要为PSRFITS (脉冲星数据专用文件格式, 基于FITS格式设计), 可通过PSRCHIVE完成后续的数据处理^[13].

点击网络版补充材料图S1中“XAO Pulsar Data Query”链接后面的“[i]”标志, 可以看到脉冲星数据库中具体表的字段相关信息、默认检索输出字段、VO-

Table输出字段信息等, 表1显示了NSRT脉冲星元数据表中的具体字段及相关描述. 点击“XAO Pulsar Data Query”链接后面的“[Q]”标志可打开检索页面, 检索页面的部分截图如网络版补充材料图S2所示.

检索界面介绍:

(1) “Position/Name”: 坐标位置信息, 主要用于输入脉冲星的赤经、赤纬信息, 允许输入SIMBAD识别的位置信息. 例如, RA~278.4, Dec~-3.6, M31等.

(2) “Search radius”: 搜索半径, 根据“Position/Name”中的信息输入搜索半径, 用于定义数据检索范围. 例如, SR~0.1, 单位为角分.

(3) “Target Object”: 观测源, 可直接选择或输入名称. 例如, Target Object~J2157+4017, 或~*J2157*等相关信息.

(4) “Date_obs”: 观测时间. 输入“>year-month-day”表示某年某月某日之后的数据, 例如, “<2018-09-20”表示2018年9月20日以前的数据; “year-month-day +/- num”表示某年某月某日前后num天的数据, 例如, “2018-09-20 +/- 3”为2018年9月17日-2018年9月23日6天内的数据.

3) <http://data.xao.ac.cn>

4) <http://data.xao.ac.cn/pul/pulsar/q/form>

表 1 脉冲星数据表字段信息

Table 1 Information of pulsar data table

表中字段名	检索显示对应名称	描述
Accref	product key	access key for the data
Backend	backend ID	backend ID
Beconfig	backend configuration file name	backend configuration file name
Date_obs	date_obs	observation date
Frontend	receiver ID	receiver ID
Object	target object	object observed
Obs_mode	observation mode	observation mode
Obsbw	observation bandwidth	observation bandwidth
Observer	observer	observer name
Obsfreq	observation frequency	observation frequency
Obsnchan	number of frequency channels	number of frequency channels
Pos	position/name	coordinates (as h m s, d m s or decimal degrees), or SIMBAD-resolvable object
Projid	project name	project name
Sr	sr	ROI radius in degrees

(5) “Observation frequency”: 观测频率, 观测频率是影响脉冲星观测的一项重要因素, 对脉冲星信号的色散效应、散射效应、频谱等均有一定影响, 可根据实际需要填写, 如用于脉冲星观测的L波段双极化制冷接收机的工作频率为1381.25–1701.25 MHz, 中心频率为1540 MHz, 可在此填写观测频率: >1456.250 MHz.

(6) “Output format”: 数据输出格式, 包含: HTML⁵⁾ (默认)、Text(fixed columns), tar, Text(with Tabs), JSON⁶⁾, VOTable⁷⁾, FITS⁸⁾ table, CSV共8种, 用户可根据科研需要适当选取检索结果输出类型.

不同类型的检索字段在检索框后由不同的标签进行了标注, 不同类型的数据标签代表不同的数据匹配模式.

“[?num. expr.]”标签表示数值表达式. 在输入框中允许输入Vizier-like表达式, 具体可参考链接http://data.xao.ac.cn/static/help_vizier.shtml#floats.

“[?char expr.]”标签表示字符串表达式. 输入框中支持通配符, 包括“?”、“*”, 分别代表任意一个字符和多个任意字符, “=”表示强制文字匹配, 字符串表达式

的其他运算符不区分大小写, 还可以实现字符串比较及相关匹配, 具体可参考链接http://data.xao.ac.cn/static/help_vizier.shtml#string.

“[?date expr.]”标签表示日期表达式. 日期必须以YYYY-MM-DD格式给出, 如“2004-01-02..2005-05-01”指定2004年01月02日至2005年05月01日之间的日期范围, 相应功能及用法可参考链接http://data.xao.ac.cn/static/help_vizier.shtml#date.

选择框–是下拉框或是其他点选框. 针对下拉框, 用户只能选择一个条目, 如Sort By. 在其他选择框中可以选择多个条目, 如Target Object, 用户可根据不同的需要进行选择.

以PSR J0332+5434 (PSR B0329+54)为例, 检索2018年11月11日以后观测频率大于1000 MHz, 观测带宽大于500 MHz的脉冲星到达时间观测数据, 共25条, 均以“*.rf”的格式存储, 以记录时间命名. 例如, 2018年11月12日的到达时间数据, 记录时间分别为: 2018-11-12T19:11:59Z, 2018-11-12T23:59:59Z, 数据命名分别为: “x181112_191314.rf, x181113_002927.rf”. 检索

5) <http://www.w3.org/MarkUp/>

6) <http://www.json.org/js.html>

7) <http://www.ivoa.net/documents/latest/VOT.html>

8) <http://fits.gsfc.nasa.gov/>

结果如网络版补充材料图S3所示, 点击预览图后面的“[Preview (pav-DFTp)]”可查看清晰脉冲轮廓图, 如图2所示.

对于大多数查询, 在查询页面底部附近有一个“Sort by”选择框, 允许用户设置排序和限制选项, 其中ASC代表升序, DESC代表降序, Limit to代表结果输出的具体行数, 默认值为100. 根据查询类型的不同, 选择要排序的列, 由于查询需要匹配大量的后台数据, 可能会导致结果响应速度显著减慢, 即便只返回100行记录, 也可能需要对数百万个项目进行排序.

4.2 VO工具检索

虚拟天文台(简称VO)注册中心提供了一种注册机制, 通过这种机制, VO应用程序可以发现和选择与特定科学问题相关的资源^[14]. IVOA在Registry Interfaces规范⁹⁾定义了相关的操作, 各天文数据中心可根据规范将数据服务注册到IVOA, 以便更多的天文研究人员发现相关的数据集, 新疆天文台已经完成了脉冲星数据服务的注册. 虚拟天文台提供了包括VOPlot, Mirage, OpenSkyQuery, Topcat, Aladin等在内的各种VO工具, 本文以Topcat检索为例.

(1) TOPCAT

TOPCAT^[15]是用于表格数据的一款交互式图形编辑器, 便于天文学家们分析天文数据. 由于使用了虚拟

天文台(VO)标准, 它可以顺利地与VO内部及外部的其他工具、服务系统和数据集协同工作. 在Starlink, AstroGrid, VOTech, AIDA, GAVO, GENIUS, DPAC等项目中得到了广泛应用.

(2) 利用TOPCAT检索

运行TOPCAT并通过VO菜单中的Cone Search搜索关键词“xaovo”, 可找到包括XAO Pulsar Data Query, PPMXL Catalog Cone Search等在内的4项新疆天文台已注册服务. 如网络版补充材料图S4所示, 本例选取XAO Pulsar Data Query服务, 自动生成Cone URL¹⁰⁾, 检索RA~10, Dec~20, Radius~180, Verbosity~3 (maximum), 即可检索到87155条脉冲星到达时间数据(图S4左半部分). 其中“J2000”表示为2000历元(以字母“B”开头的命名, 表示采用1950历元).

4.3 脉冲星数据可视化

通过TOPCAT可对检索到的脉冲星数据进行可视化操作: 打开TOPCAT, 在脉冲星检索结果界面选择“Send via SAMP”, 即可将数据传输到TOPCAT完成可视化操作. 使用TOPCAT Version 4.6-1, 检索条件为: 观测日期>2017-06-01、数据格式为“~*.rf”(2010年1月后DFB系统产生的脉冲星数据)的24284条结果进行可视化, 如网络版补充材料图S5所示, 可视化结果通过Spherical Plot在3D银河系坐标中标出.

图2 PSR J0332+5434脉冲轮廓图

Figure 2 Pulse profile of PSR J0332+5434.

9) <http://www.ivoa.net/documents/RegistryInterface/>

10) <http://data.xao.ac.cn/pul/pulsar/q/scs.xml?>

表 2 访问数据统计

Table 2 Access statistics

日期	访问次数	主机数
2016.11	11580	664
2016.12	15782	677
2017.01	13687	671
2017.02	10096	597
2017.03	8337	1005
2017.04	10493	736
2017.05	13318	796
2017.06	11881	823
2017.07	12519	928
2017.08	12873	889
2017.09	11093	767
2017.10	14107	839
2017.11	17616	868
2017.12	15381	1064
2018.01	18219	945
2018.02	18723	923
2018.03	23217	1293
2018.04	22769	1020
2018.05	26618	1857
2018.06	22235	3238
2018.07	12872	1372
2018.08	13012	1212
2018.09	21588	2303
2018.10	22829	2339
2018.11	23143	2218

4.4 脉冲星数据访问

2016年11月至2018年11月间脉冲星数据访问情况如表2所示, 月平均访问次数在1万次以上, 月平均访问主机数在500台以上.

5 结论

新疆天文台脉冲星数据归档与发布系统面向所有

科研人员使用, 实现了NSRT脉冲星观测数据的存储、备份、归档和发布, 实现了科研数据的全世界范围内共享. 数据发布以VO协议为基础, 实现了锥形检索及多目标约束检索服务. 用户可通过浏览器、各类脚本及标准VO工具实现数据的检索访问. 通过注册服务到IVOA, 使用户可以利用标准VO工具检索到新疆天文台脉冲星数据, 并实现简单的数据分析、处理与可视化操作.

致谢 感谢审稿人的宝贵意见, 数据中心测试与数据的预处理在新疆天文台Taurus高性能计算系统上完成.

补充材料

本文的补充材料见网络版<http://physcn.scichina.com>. 补充材料为作者提供的原始数据, 作者对其学术质量和内容负责.

参考文献

- 1 Manchester R N, Hobbs G B, Teoh A, et al. The Australia telescope national facility pulsar catalogue. *Astron J*, 2005, 129: 1993–2006
- 2 Hewish A, Bell S J, Pilkington J D H, et al. Observation of a rapidly pulsating radio source. *Nature*, 1968, 217: 709–713
- 3 Zhang H L, Zhu Y, Nie J, et al. Xinjiang Astronomical Observatory NSRT data storage system (in Chinese). *Astron Res Technol*, 2018, 15: 181–187 [张海龙, 朱艳, 聂俊, 等. 新疆天文台NSRT观测数据存储系统. *天文研究与技术*, 2018, 15: 181–187]
- 4 Wang N, Manchester R N, Zhang J, et al. Pulsar timing at Urumqi Astronomical Observatory: Observing system and results. *Mon Not R Astron Soc*, 2001, 328: 855–866
- 5 Wang N, Manchester R N, Zhang J, et al. Pulsar observations at urumqi observatory—present and future. *Acta Astron Sin*, 2003, 44: 207–214
- 6 Hampson G, Brown A, Vimiera C. A 1 GHz Pulsar Digital Filter Bank and RFI Mitigation System. Australia Telescope National Facility, 2008
- 7 Keith M J. Installation and use of pulsar search software. *Astron Res Technol*, 2012, 9: 219–228
- 8 Cui C Z, Xue Y J, Li J, et al. Virtual observatory, an e-science environment for astronomy (in Chinese). *Bull Chin Acad Sci*, 2013, 28: 511–518 [崔辰州, 薛艳杰, 李建, 等. 虚拟天文台——天文学研究的科研信息化环境. *中国科学院院刊*, 2013, 28: 511–518]
- 9 Demleitner M. DaCHS: Data Center Helper Suite. Astrophysics Source Code Library, 2018
- 10 Barrett R F, Chan T H F, D’Azevedo E F, et al. Complex version of high performance computing LINPACK benchmark (HPL). *Concurrency Computat-Pract Exper*, 2009, 41
- 11 Zhang H L, Ye X C, Wang J, et al. Taurus high performance computing system of Xinjiang Astronomical Observatory (in Chinese). *Astron Res Technol*, 2018, 15: 332–339 [张海龙, 冶鑫晨, 王杰, 等. 新疆天文台Taurus高性能计算系统. *天文研究与技术*, 2018, 15: 332–339]
- 12 Zhang H L, Wang J, Wang W Q, et al. Construction and application of the data center in Xinjiang astronomical observatory (in Chinese). *Astron Res Technol*, 2017, 14: 94–102 [张海龙, 王杰, 王万琼, 等. 新疆天文台数据中心建设与应用. *天文研究与技术*, 2017, 14: 94–102]
- 13 Hotan A W, van Straten W, Manchester R N. PSRCRIVE and PSRFITS: An open approach to radio pulsar data storage and analysis. *Publ Astron Soc Aust*, 2004, 21: 302–309
- 14 Demleitner M, Greene G, Le Sidaner P, et al. The virtual observatory registry. *Astron Computing*, 2014, 7-8: 101–107, arXiv: 1407.3083
- 15 Taylor M. TOPCAT: Desktop exploration of tabular data for astronomy and beyond. *Informatics*, 2017, 4: 18

XAO pulsar data retrieval system

ZHANG HaiLong^{1,2*}, WANG Na^{1,2}, WANG Jie¹, NIE Jun^{1,2},
YUAN JianPing^{1,2} & WANG WanQiong¹

¹ *Xinjiang Astronomical Observatory, Chinese Academy of Sciences, Urumqi 830011, China;*

² *Key Laboratory of Radio Astronomy, Chinese Academy of Sciences, Nanjing 210008, China*

Since the room temperature receiver was replaced by the L-band cryogenic receiver in 2002, the sensitivity of pulsar observation using NanShan 25-m Radio Telescope (NSRT), operated by Xinjiang Astronomical Observatory (XAO), has reached 0.4 mJy, allowing regular monitoring for more than 200 pulsars. NSRT has produced large amount of valuable scientific data through nearly 20 years of observations. The pulsar data is synchronized to XAO headquarters for long-term storage through the 300 Mb/s dedicated line, and remote disaster tolerance backup is realized through the data storage system of NanShan Station and XAO headquarters. Metadata extraction, data archiving and release are completed by the data server which located in XAO headquarters. By the end of November 2018, more than 87000 pieces of pulsar observation data have been released, with the data format of PSRFITS. Pulsar data service provides cone search and multi-objective constrained retrieval. Data release is based on Virtual Observatory (VO) standard and pulsar data query service has been registered to the International Virtual Observatory Alliance (IVOA). Users can search, access and process NSRT pulsar data through VO tools.

pulsar, metadata, query, virtual observatory

PACS: 47.27.-i, 47.27.Eq, 47.27.Nz, 47.40.Ki, 47.85.Gj

doi: [10.1360/SSPMA2018-00404](https://doi.org/10.1360/SSPMA2018-00404)