

脉冲星数据比对分析和可视化系统设计与实现*

张辉^{1,2,3} 王培⁴ 张蕾⁴ 许余云⁵ 岳友岭⁴ 潘之辰⁴
刘志杰^{1,3} 于徐红^{1,3} 游善平^{1,3} 姜家涛^{1,2,3} 谢晓尧^{1,3†}

(1 贵州师范大学贵州省信息与计算科学重点实验室 贵阳 550001)

(2 贵州师范大学数学科学学院 贵阳 550001)

(3 贵州师范大学FAST早期科学数据中心 贵阳 550001)

(4 中国科学院国家天文台 北京 100012)

(5 贵州水利水电职业技术学院管理工程分院 贵阳 551416)

摘要 脉冲星数据比对分析和可视化系统(PSRDB, URL: <http://www.psrdb.net/>), 由FAST (Five-hundred-meter Aperture Spherical Radio Telescope)早期科学数据中心团队为快速开展脉冲星候选体比对分析和数据管理研发. 通过前端数据提交页面, 接收和维护来自FAST及其他研究机构的候选体数据. 目前, PSRDB已收录自1967年人类发现第1颗脉冲星以来所有公开文献发表的2811颗脉冲星样本, 并采集了当前主要巡天项目尚未正式发表的源和候选体, 如FAST多科学目标同时扫描巡天(CRAFTS)候选体数据. 基于入库基础数据, 利用位置、周期、色散等参数进行比对分析, 辅助科研工作者在线检索匹配已知星表数据, 最后将检索匹配、比对分析结果生成图表供进一步分析. 目前, PSRDB已被应用于FAST脉冲星搜寻和候选体数据管理. 未来, PSRDB可在新源认证、后随观测、观测计划制定和原始数据处理流程设计等方面提供数据和工具支撑.

关键词 天文数据库, 脉冲星: 普通, 方法: 比对分析, 可视化, FAST

中图分类号: P111; 文献标识码: A

1 引言

自1967年8月英国天文学家Jocelyn Bell和其博士生导师Antony Hewish教授等人发现第1颗射电脉冲星CP 1919以来^[1], 脉冲星始终是射电天文学和天体物理学研究领域的热点之一. 脉冲星独特的物理性质, 如大质量、精确的自旋周期、强磁场和强引力

2020-06-15收到原稿, 2020-09-08收到修改稿

*国家自然科学基金项目(U1831131、U1631132、U1731238、11743002), 国家重点研发计划(2017YFA0402600), 中国科学院战略性先导科技专项(B类)(XDB23000000), 中国科学院天文大科学研究中心FAST重大成果培育项目(FAST[2019sr04]), 贵州省科技厅联合基金项目(黔科合LH字[2017]7338号)及贵州师范大学研究生创新基金项目(研创201528)资助

†xyx@gznu.edu.cn

等,使得它可被应用于开展深空探测、定位导航、引力波探测等,吸引着众多物理学家、天文学家和其他领域科学家竞相研究. 1974年美国著名物理学家R. A. Hulse和J. H. Taylor教授等人利用阿雷西博(Arecibo)射电望远镜首次发现了双星系统脉冲星(PSR B1913+16)^[2],从而发现了爱因斯坦相对论中关于引力波存在理论的间接证据^[3], Hulse和Taylor因此荣获1993年诺贝尔物理学奖. 迄今为止,公开文献资料发布的已知脉冲星样本已超过2811颗^[4].

为进一步研究这类天体个体和群体性物理特征、形成机制、演化进程和应用前景,射电望远镜的大量观测时间都投入到脉冲星搜寻项目中,以期探测发现新的样本或奇异样本,如脉冲星-黑洞(PSR-BH)双星系统、河外脉冲星、球状星团脉冲星、巨脉冲、带有伴星或者从未发现过的其他新类别脉冲星. 2016年9月位于贵州喀斯特偏远山区的五百米口径球面射电望远镜^[5](Five-hundred-meter Aperture Spherical Telescope, FAST)进入科学运行阶段,先后配置了超宽带接收机和L波段19波束接收机. 为充分利用FAST优异的探测能力,2018年李菂等人设计并实施了“FAST多科学目标同时扫描巡天”(The Commensal Radio Astronomy FAST Survey, CRAFTS)^[6],使用多个数字后端同时采集脉冲星、中性氢、分子谱线、暂现源、FRB (Fast Radio Burst)等多个科学目标观测数据^[7]. 当前,FAST已采集到观测数据容量超过1 PB,通过2 GB/s光纤专线网络传输至位于贵州师范大学校内的FAST早期科学数据中心进行统一存储,并通过联合研制的分布式并行计算加速系统及搜索数据库进行数据处理,成功获得超过500万个脉冲星候选体. 同时,FAST还多次探测到快速射电暴FRB121102重复爆发现象,并首次发现色散高达 $1812 \text{ pc}\cdot\text{cm}^{-3}$ 的新快速射电暴FRB181123^[8]. 根据刘鹏等人于2018年对FAST 19波束接收机脉冲星漂移扫描巡天探测能力模拟估算数据显示,FAST在赤纬 $-14^{\circ}12' - +65^{\circ}48'$ 天区范围可发现超过1600颗普通脉冲星和238颗毫秒脉冲星^[9]. 未来,FAST脉冲星巡天项目采集到的数据总量可达100 PB,经数据分析处理,预计可发现新脉冲星4000–5000颗^[10]. 另外,建设中的国际合作项目平方公里射电阵列(The Square Kilometre Array, SKA)相比于FAST,其综合灵敏度更高、探测天区更广、采集数据规模更大,伴随产生候选体和新脉冲星将更多,在SKA1-Mid和SKA1-Low两阶段预计可探测普通脉冲星约9000颗,毫秒脉冲星约1200颗^[11].

随着新的观测设备和数据分析处理方法应用于脉冲星科学研究,未来将可能发现更多新样本. 目前,国际上常用的脉冲星星表数据库有由澳大利亚国家天文台(Australia Telescope National Facility, ATNF)的Hobbs和Manchester等人开发的The ATNF Pulsar Catalogue (PSRCAT)、Green Bank North Celestial Cap (GBNCC)数据发布网站和The European Pulsar Network Pulse Profile Database (EPN)等. PSRCAT的星表数据库基于文本文件进行数据管理,它仅收录公开发表文献资料中的脉冲星参数和数据,不含候选体数据记录和尚未正式公开发表的源. GBNCC则通过自建网页发布其数据处理结果,包含186颗新发现的脉冲星,其中24颗为毫秒脉冲星,22颗为旋转射电暂现源(Rotating Radio Transient, RRAT). EPN则通过采集各类脉冲星轮廓图像为射电领域学者提供检索服务.

据文献调研,尚未有一套整合已知脉冲星、未发表源、候选体、巡天项目等数据并进行比对分析和可视化展示的综合专用数据库系统. 为更全面地采集已知脉冲星、

未发表源和候选体等数据, 为射电脉冲星科学研究提供更全面的数据记录和系统工具, 帮助FAST在脉冲星科学研究领域取得新的突破, 我们建立了面向已知脉冲星、未发表源和候选体数据采集, 集比对分析、数据管理和可视化为一体的专用数据库系统(The Pulsar Database, PSRDB, 网址: <http://www.psrdb.net/>). 目前, PSRDB已在FAST早期科学数据中心成功应用于FAST脉冲星数据分析处理项目.

2 数据库设计

2.1 设计目标与技术架构

2.1.1 设计目标

PSRDB为脉冲星数据提供规范统一的基础数据管理, 用于已知脉冲星及候选体的数据存储、比对分析和可视化, 为脉冲星科学研究中数据共享和匹配检索提供数据和平台支持, 具体设计目标如下:

(1)不同于ATNF或PSRCAT基于文本文件的数据管理方式, PSRDB采用结构化关系型数据库系统管理数据, 以便支撑不断新增脉冲星数据的在线管理需求, 解决文本格式存储数据容量有限、大量数据记录的维护和检索不便问题;

(2)收录FAST候选体数据及其他望远镜尚未正式公开发表的源以及1967年发现首颗脉冲星以来公开发表文献中已知脉冲星、主要巡天项目和望远镜设备数据, 提供数据检索、匹配分析、数据共享与简单统计应用分析, 为后续巡天观测计划制定、数据分析处理和候选体认证提供数据和工具支持;

(3)面向全球脉冲星科学研究组织和个人研究者, 提供开放的候选体数据存储管理, 建立脉冲星科学社区, 提供科研交流互动平台.

2.1.2 技术架构

PSRDB数据库系统采用Oracle公司开源数据库MySQL作为后端数据库¹, 应用服务部署于阿里巴巴公有云服务器(Elastic Compute Service, ECS)², 以保证系统的可用性、易扩展性和网络稳定性, 技术架构如图1. 脉冲星数据除了结构化的数据记录, 还包含大量的判别图像文件、描述文本等非结构化数据, 检索、比对分析和定位存在加载缓慢问题, 通过配置云对象存储(OSS)和内容分发网络(CDN)进行提速. PSRDB数据库系统以Spring Framework为核心框架, Spring MVC为模型视图控制框架, Mybatis作为数据持久层框架负责数据持久化, 同时引入Apache Shiro安全框架负责数据的安全和权限控制, 并选用Ehcache作为数据缓存框架对常用数据进行缓存, 使用Disruptor作为高性能并发框架用于实现网络的并发操作, 采用了Twitter Bootstrap作为前端页面展现框架, 可视化图表展示则采用开源Echarts³插件来实现, 从而使得前后端程序耦合性降低, 便于数据库后期持续扩展和维护, 未来计划开放部分数据接口(即开放应用编程接口, Open API).

¹<https://www.mysql.com/>

²<https://www.aliyun.com/>

³<https://echarts.apache.org/zh/index.html>

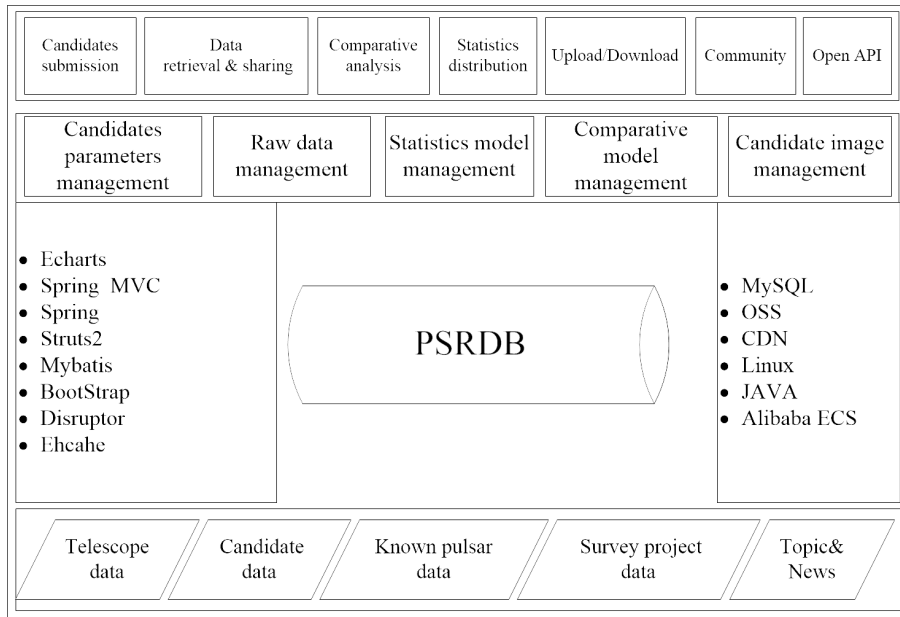


图 1 PSRDB数据库系统的技术架构图

Fig. 1 Technical architecture of the PSRDB

2.2 数据源

PSRDB汇集望远镜设备数据、已知脉冲星数据、未正式发表源、候选体数据、巡天项目数据等. 由FAST早期科学数据中心上传来自FAST脉冲星项目的所有候选体数据资源, 同时, PSRDB通过不定期查阅公开文献资料、主要巡天项目主页、相关数据库系统等方式进行基础数据采集和更新.

2.2.1 已知脉冲星

据ATNF脉冲星数据库数据显示, 1967年以来公开文献发表了2811颗脉冲星, 如表1, PSRDB通过检索ATNF数据库、查阅公开文献资料、巡天项目官网等多种方式采集了表1中已公开发表的脉冲星数据, 其中“misc”表示其他望远镜设备.

我们以两年为1个周期对1967—2020年世界主要望远镜设备(如FAST、Parkes、GBT、Arecibo等)脉冲星发现情况进行统计分析, 如图2, 其中“other”表示其他巡望远镜. 在最初20 yr里, 超过半数的脉冲星由Molonglo望远镜发现, 2000年以后Parkes望远镜成为发现脉冲星数量最多的探测设备, 一度超过其他望远镜发现数量的总和, 特别是Parkes开展的多波束脉冲星巡天(Parkes Multibeam Pulsar Survey, PMPS)^[12]直接贡献多达830颗新脉冲星. Fermi Gamma-ray Space Telescope (Fermi)⁴近年来在 高能(High-Energy, HE)脉冲星和毫秒脉冲星发现方面有突出贡献. Arecibo望远镜发现第1个孤立双星系统中质量较大的脉冲星PSR B1913+16, 观察结果与爱因斯坦广义相对论中有关引力波预言的理论计算结果非常接近, 从而间接证明了引力波的存在. 同时, Arecibo还发现了毫秒脉冲星PSR B1257+12, 该脉冲星周围有3颗行星, 其中2颗可能是

⁴<https://fermi.gsfc.nasa.gov/>

巨大的超级地球(Super-Earths), 另外1颗仅仅比月球稍大, 是到目前为止发现的最小的系外行星.

表 1 1967年以来主要望远镜设备脉冲星巡天探测发现脉冲星数量统计

Table 1 Statistics of pulsars discovered in pulsar surveys performed by the major telescopes since 1967

No.	Telescope	Number of surveys	Number of detected	Number of discovered
1	Arecibo	7	609	341
2	Fermi	2	123	123
3	Green Bank Telescope (GBT)	6	592	249
4	Molonglo	2	258	185
5	Giant Metrewave Radio Telescope (GMRT)	1	36	13
6	Effelsberg	1	17	14
7	Jodrell Bank Radio Telescope (Jodrell)	2	113	79
8	Parkes	9	2909	1381
9	Low Frequency Array (LOFAR)	1	369	53
10	FAST	1	2	2
11	misc	–	–	371
Total	–	32	5028	2811

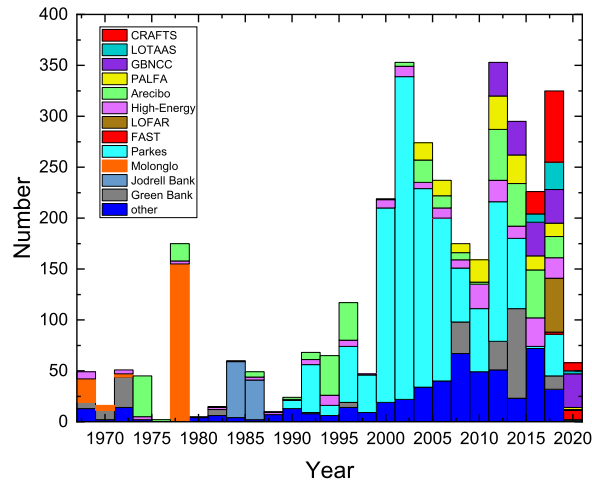


图 2 1967年以来世界主要望远镜设备脉冲星巡天探测发现脉冲星数量统计

Fig. 2 Statistics of pulsars discovered by the world's major telescopes and pulsar surveys since 1967

2.2.2 未发表源

PSRDB已采集尚未被ATNF脉冲星数据库收录的源522条, 分别来自FAST的多科学目标同时扫描巡天CRAFTS项目, Arecibo L波段馈源阵列脉冲星巡天(Pulsar Arecibo L-band Feed Array Survey, PALFA)^[13], 低频射电阵列LOFAR巡天(LOFAR Tied-Array All-sky Survey, LOTAAS)^[14], 绿堤射电望远镜GBT北半球脉冲星巡天(Greenbank Northern Celestial Cap Survey, GBNCC)^[15]和350 MHz频域巡天(GBT350)等主要巡天项目, 如表2. 这些巡天项目数据因元数据不一致, 阻碍了数据的交互、共享、汇总, 需要建立统一的数据标准, 利用数据仓库技术(Extract-Transform-Load, ETL)对源数据进行抽取、转换, 并输入到PSRDB以保证数据完整、正确及有效.

表 2 PSRDB数据库收录的来自主要望远镜尚未公开发表脉冲星统计
Table 2 Statistics of unpublished pulsars from the major telescopes contained in PSRDB

No.	Telescope	Survey	Known pulsar	Unpublished pulsar	Total
1	FAST	CRAFTS	11	130	141
2	Arecibo	PALFA	42	154	196
3	LOFAR	LOTAAS	35	39	74
4	GBT	GBNCC	18	164	182
		GBT350	0	35	35
Total	-	-	106	522	628

2.3 对比方法

2.3.1 比对流程分析

基于采集入库的脉冲星、未正式发表源和候选体数据等, 可对新产生的候选体进行在线比对和检索分析. 候选体比对分析基本流程为进入比对分析页面、填写比对参数(如赤经赤纬坐标、周期(P0)、色散值(DM)等)、选择比对方法、生成比对结果, 比对流程如图3. 具体流程操作为: 假设用户A拥有一颗新候选体C1, 比对流程第1步, 以C1名称作为参数进行在线检索, 若PSRDB候选体管理表“candidate”中未发现相同候选体, 则更新候选体表中参数如: 候选体识别码(candidate id)、坐标、周期、色散值、信噪比(s/n)、用户信息(user id)等; 第2步, 与PSRDB中已知脉冲星进行比对, 系统反馈结果列表, 若C1与某已知脉冲星参数吻合, 则认为发现了一颗已知脉冲星, 用户可选择标记C1状态为“Confirmed”, 并将该候选体保存至PSRDB; 若C1与PSRDB中已知脉冲星星表数据没有匹配上, 则进行候选体之间的比对, 如果未能匹配检索到已知候选体记录, 则将其状态更新为“unconfirmed”, 则它可能是新发现的脉冲星; 最后, 在上述比对过程中, 用户A可以自由选择是否将候选体相关数据保存至PSRDB, 并设置数据隐私权限和进行后期数据维护, PSRDB不会保留任何未经用户授权的候选体数据.

2.3.2 比对分析方法

脉冲星及候选体的比对分析是发现新脉冲星和进一步研究脉冲星性质特征的基础和必要过程, PSRDB提供4种不同比对方法进行在线比对分析.

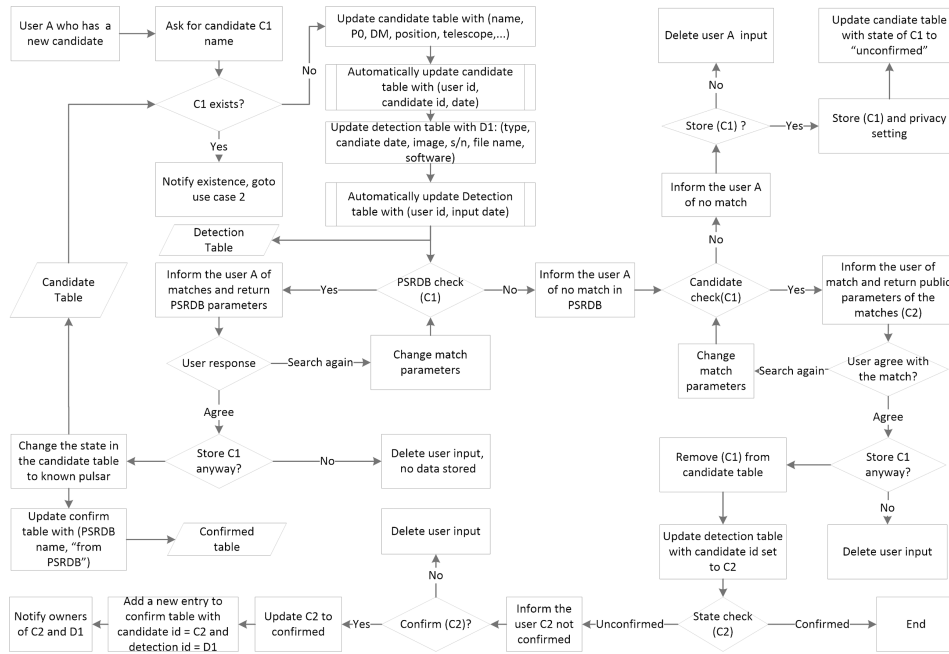


图 3 PSRDB候选体比对匹配分析流程示意图

Fig. 3 Flow chart of PSRDB candidate comparison and matching analysis

(1)基于脉冲星位置坐标的比对方法. 用户设置银经(gl)、银纬(gb)初值, 分别记作 gl_0 和 gb_0 (单位为 $^\circ$), 并以 (gl_0, gb_0) 为圆心坐标, 同时, 指定检索半径 r (单位为 $^\circ$)的大小, 匹配检索PSRDB数据库中包含在圆内的所有候选体或者已知脉冲星数据记录, 比对结果可以表格或者图形方式输出. 如(1)式:

$$(gl - gl_0)^2 + (gb - gb_0)^2 \leq r^2, \tag{1}$$

其中, 参数 gl 、 gb 分别表示脉冲星的银经和银经坐标, r 表示圆半径大小;

(2)基于DM和P0的比对方法. 用户通过给定色散(单位: $pc \cdot cm^{-3}$)和周期(单位: s)的区间范围, 结合逻辑判断(或(\parallel))、与($\&\&$)、非(!)等进行过滤, 检索PSRDB数据库中所有符合检索条件DM或者P0的样本数据, 如检索条件 $(P0 > 1 \ \&\& \ P0 < 6) \ \&\& \ (DM > 0 \ \&\& \ DM < 9)$, 并以表格或者图形方式直观输出比对结果;

(3)基于相似度(Similarity)的比对方法. 记 S 表示由脉冲星或者候选体部分参数(如表3)构成的集合, 如 $S = \{P0, DM, mass, gl, gb, ra, dec, s400, \dots\}$ 为脉冲星或者候选体属性参数集合, 其中 $mass$ 为质量、 ra 及 dec 分别为赤经和赤纬、 $s400$ 为流量密度. 利用其中部分参数构建特征向量, 检索匹配PSRDB数据库中给定范围的 n 维向量集合 $\{\vec{v}|\vec{v} = (x_1, x_2, \dots, x_n), x_i \in S, i = 1, 2, 3, \dots, n\}$. \vec{v}_0 表示用户输入待比对的候选体或者脉冲星特征向量, 特征向量各分量取值为用于比对分析中参与计算的脉冲星所对应各参数的数值. 相似度通过特征向量的夹角余弦值定义, 公式如下:

$$Similarity(\vec{v}, \vec{v}_0) = \frac{\sum_{i=1}^n (x_i y_i)}{\sqrt{\sum_{i=1}^n (x_i)^2} \sqrt{\sum_{i=1}^n (y_i)^2}}. \tag{2}$$

(4)基于距离的比对方法. 用户给定候选体 $\vec{v}_0 = (y_1, y_2, \dots, y_n)$, 通过计算其与数据库中已知脉冲星或者候选体 $\vec{v} = (x_1, x_2, \dots, x_n)$ 的 n 维欧式距离进行比对分析, 并给出比对分析结果列表, 如距离(3)式:

$$\text{dist}(\vec{v}, \vec{v}_0) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}, \quad (3)$$

其中, \vec{v} 和 \vec{v}_0 表示脉冲星的特征向量, $\vec{v} \in \{\vec{v} | \vec{v} = (x_1, x_2, \dots, x_n), x_i \in \mathcal{S}, i = 1, 2, \dots, n\}$.

表 3 候选体属性参数列表
Table 3 Parameter list of pulsar candidate

Parameter	Description	Parameter	Description
candidate_id	Candidate ID	user_id	User id
name	Candidate name	gl	Galactic longitude ($^{\circ}$)
osurvey	Surveys that detected the pulsar	gb	Galactic latitude ($^{\circ}$)
raj	Right ascension (J2000) (hh: mm: ss)	raj_d	Right ascension (J2000) ($^{\circ}$)
decj	Declination (J2000) (dd: mm: ss)	decjd	Declination (J2000) ($^{\circ}$)
P0	Barycentric period of the pulsar (s)	DM	Dispersion measure ($\text{pc}\cdot\text{cm}^{-3}$)
w10	Width of pulse at 10% (ms)	f0	Barycentric rotation frequency (Hz)
s400	Mean flux density at 400 MHz (mJy)	telescope	Survey telescope
tool	Data processing pipeline	type	Pulsar, RRATs
state	Unconfirmed and not checked, Unconfirmed but checked, Confirmed as know pulsar, Confirmed as new pulsar	project	Survey project that detected the candidate
Privacy	Keep public or private	date	Date of data generated

2.4 统计分析

为帮助脉冲星科学研究者理解样本点的分布特征, 也为巡天定标、新的巡天计划制定、脉冲星群体性和个性特征研究提供可量化的数据支撑, 对于样本点数值类型参数如周期、频率、色散量、质量等, 依据自定义检索范围和条件, PSRDB提供统计功能, 绘制频数分布直方图, 并计算参与统计样本的最大值、最小值、中位数、极差、均值、方差等统计量.

3 数据库功能介绍

3.1 已知脉冲星数据管理

PSRDB提供已知脉冲星检索匹配功能,如图4. PSRDB提供预定义脉冲星参数60多个,用户可根据不同科学目标在PSRDB检索比对页面勾选相关属性参数,输入相似度计算参数值和频数分布统计量,设置周期、色散、坐标和半径等进行检索,也可根据脉冲星名进行快速检索.

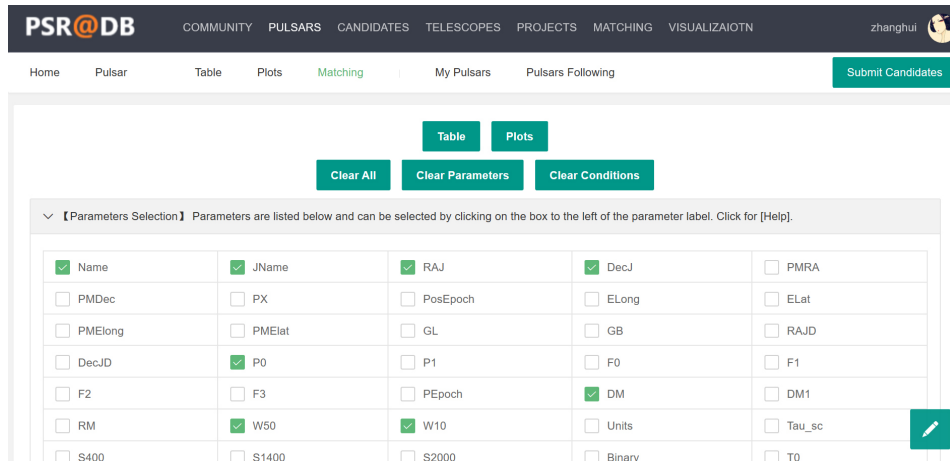


图 4 PSRDB数据库数据检索匹配页面

Fig. 4 Matching page of the PSRDB

检索结果以“TABLE”或“PLOT”形式输出。“TABLE”形式的数据展示如图5,通过比对页面【Logic Conditions】设置比对检索条件: $(P0 > 1 \ \&\& \ P0 < 6) \ \&\& \ (DM > 0 \ \&\& \ DM < 9)$, 共检索匹配到6颗符合条件的脉冲星. 如图6,应用(1)式和比对分析页面【Circular Boundary】功能,设置 $gl_0 = 10$ 、 $gb_0 = 10$ 、 $r = 5$, 共匹配到10颗符合条件的脉冲星. 匹配结果列表可按列属性进行2次升降排序,也可将数据在线打印或者以CSV、Excel和PDF等格式导出. 此外, PSRDB提供1-3维数据的图形可视化展示,可生成散点图、折线图、柱状图等,展示脉冲星的重要参数如坐标、色散值、周期之间的关联关系,并且所有的成图均可保存至本地. 如图7所示展示了目前为止所有已公开发表脉冲星(2811颗)在银河系中的空间分布,大部分脉冲星都集中在银道面上,这种现象与超新星遗迹的分布吻合,进而支持脉冲星是由超新星爆发所生成的观点.

用户可通过脉冲星详情页了解其更全面的各项参数,如果对重点关注的脉冲星想要进一步研究其进展, PSRDB还提供了个人收藏功能,通过点击详情页面“Following”进行收藏,方便下次预览和集中管理. PSRDB后台已知脉冲星数据管理页面如图8,提供脉冲星、候选体、未发表源、望远镜设备、巡天项目和科学社区等数据的增、删、改、查、排序和批量导入导出等功能.

No.	NAME	PSRJ	RAJ(hms)	DECJ(dms)	P0(s)	DM(cm ⁻³ p...	W50(ms)	W10(r)	OPERA...
1	B0809+74	J0814+7429	08:14:59.50	+74:29:05.70	1.29224144...	5.75066	41.4	89.5	Details
2	J1243+17	J1243+17	12:43:00	+17:52	1.2165	5	200		Details
3	B2327-20	J2330-2005	23:30:26.885	-20:05:29.63	1.64362218...	8.456	6.4	35	Details
4	J1503+2111	J1503+2111	15:03:54.60	+21:11:09.3	3.31400150...	3.2603	26	80	Details
5	B1133+16	J1136+1551	11:36:03.1198	+15:51:14.183	1.18791306...	4.84066	5.9	41.8	Details
6	J1336+33	J1336+33	13:36:00	+33:46	3.013	8.5			Details

图 5 已知脉冲星数据检索结果

Fig. 5 Search results of known pulsars

No.	PSRJ	RAJ(hms)	DECJ(dms)	RAJD(deg)	DECJD(deg)	P0(s)	DM(cm ⁻³ p...	W50(ms)	OPERA...
1	J1739-1313	17:39:57.821	-13:13:18.6	264.99092	-13.22183	1.21569761...	58.2	7.1	Details
2	J1732-1930	17:32:20.02...	-19:30:09	263.083	-19.5	0.48376999...	72.43	16	Details
3	J1720-1633	17:20:25.20...	-16:33:34	260.105	-16.56	1.56560114...	44.83	16.2	Details
4	J1711-1509	17:11:55.05...	-15:09:40	257.9794	-15.161	0.86880424...	59.88	12.2	Details
5	J1744-1134	17:44:29.40...	-11:34:54.6...	266.122524...	-11.58185591	0.00407454...	3.13695	0.137	Details
6	J1719-1438	17:19:10.07...	-14:38:00.942	259.7919705	-14.633595	0.00579015...	36.862	0.3	Details
7	J1748-1300	17:48:17.40	-13:00:52	267.0725	-13.014	0.39413334...	99.364	12.3	Details
8	J1744-1610	17:44:16.534	-16:10:35.8	266.06889	-16.1766	1.75720586...	66.67	36	Details
9	J1743-1351	17:43:37.62...	-13:51:38	265.9068	-13.861	0.40533694...	116.3	14.6	Details
10	J1731-1847	17:31:17.60...	-18:47:32.666	262.82337426	-18.7924072	0.00234455...	106.4711	0.13	Details

图 6 基于脉冲星银经(gl_0)、银纬(gb_0)和检索半径(r)的比对匹配结果

Fig. 6 Matching results based on the galactic longitude (gl_0), galactic latitude (gb_0) and radius (r) of a pulsar

3.2 候选体数据管理

3.2.1 属性参数

每颗候选体包含多个必要的属性,用于描述来源、名称、周期、色散值、坐标、探测望远镜、信号强度、发布时间、探测频率、项目编号等基础信息,如表3.考虑到软件兼容性、语义和使用习惯等问题,PSRDB数据库系统建设借鉴了ATNF的已知脉冲星参数空间,并增加了隐私(privacy)和状态属性(state),对候选体数据进行权限管理,并在后期运行维护中不断完善和扩充参数列表.

对于候选体的探测日志记录,PSRDB标记了候选体的处理软件、归属机构、诊断图、探测发现日期、候选体认证状态等关键信息,如表4所示,其中FFT、SP、FFA分别代表快速傅里叶变换、单脉冲搜索和快速傅里叶分析.

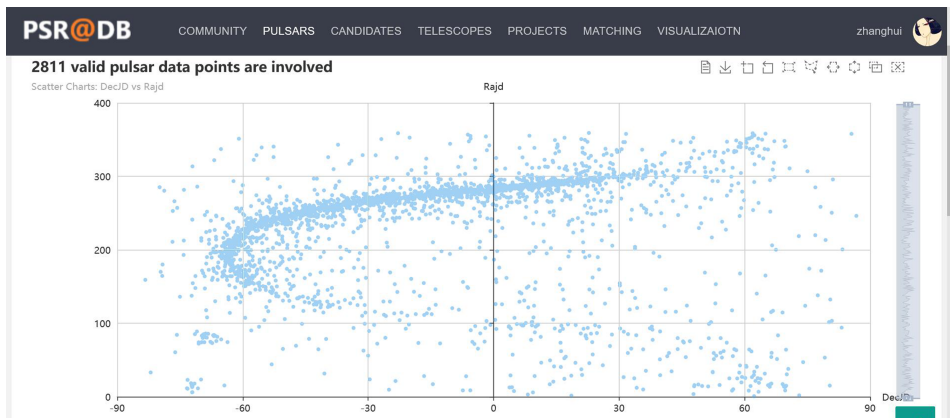


图 7 2811颗已知脉冲星在银河系中的空间分布

Fig. 7 Spatial distribution of 2811 known pulsars in the Milky way

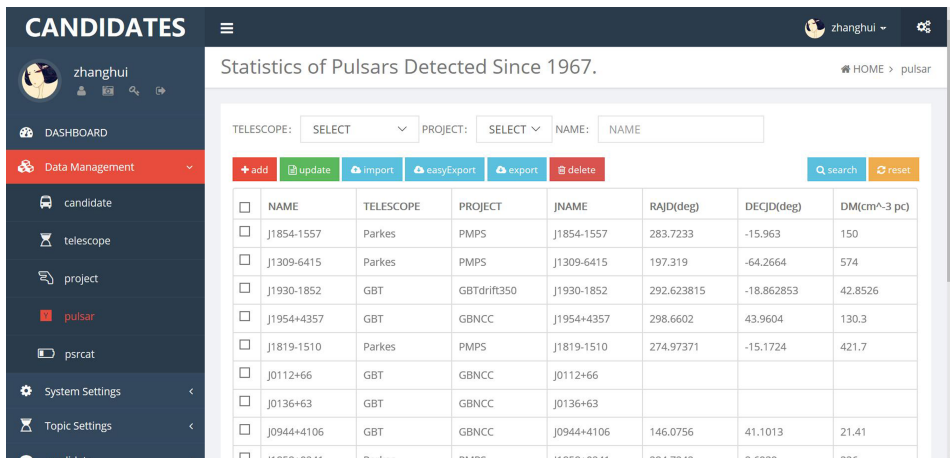


图 8 PSRDB已知脉冲星数据管理页面

Fig. 8 Known pulsar data management page of PSRDB

3.2.2 数据提交

PSRDB提供候选体数据提交功能, 以方便脉冲星科研团队成员及其他脉冲星研究组织在线提交候选体数据, 如图9. 在数据提交页面, 用户需填写观测设备、巡天项目、位置坐标、周期、色散等候选体关键属性数据, 设置隐私属性, 提交至PSRDB, 并可通过个人中心进行数据持续维护.

3.2.3 数据维护

候选体数据属于脉冲星科研团队的重要科研成果之一, 相关数据一般需进一步分析, 具有一定的保密、私有属性. PSRDB通过权限控制对用户权限进行管理, 只有同组用户才可以在线访问和分享数据, 未授权用户在前端“CANDIDATES”页面将无法查阅任何候选体相关信息. 候选体数据管理页面如图10, 提供数据的增、删、改、查、排序、批量导入和导出等功能. 当前, PSRDB已收录FAST巡天项目CRAFTS的优质候选体140多颗.

表 4 候选体数据处理及探测参数列表

Table 4 Parameters of candidate data processing and detection

Parameter	Description	Parameter	Description
detection_id	Detection ID	software	Software (string)
candidate_id	Candidate ID	sig	Signal to noise ratio (double)
user_id	User ID	institution	Institution (string)
date_obs	Date of observation	telescope	Telescope (string)
date_entry	Date of entry	success	Success (boolean)
detection_type	Detection type (FFT/SP/FFA ...)	img	Image (a link)
status	Candidate status	file_name	File name (string)

图 9 PSRDB候选体数据提交页面

Fig. 9 Candidate data submission page of PSRDB

3.3 科学社区

PSRDB除了面向射电脉冲星领域提供脉冲星、候选体等数据管理和分析外,还建立了脉冲星科学社区,为科研工作者提供互动交流平台,实现发起话题、问题回复等互动功能,如图11. 点击前端“COMMUNITY”导航进入社区页面,用户可发布和转载包括脉冲星、快速射电暴、RRAT等科研资讯,建立脉冲星搜索软、硬件方法等主题讨论区.

4 PSRDB在FAST中的应用

PSRDB在FAST脉冲星巡天项目过程中提供在线数据共享、检索、比对统计、验证等科研服务. FAST进入科学运行阶段以来,脉冲星科研团队已发现优质候选体140多

颗⁵, 其中, 张蕾等人公开发表了2颗, Cameron等人公开发表了11颗^[16-18], 如表5, 其中 RAJ2000和DECJ2000为定义在J2000坐标系下的赤经和赤纬, 部分脉冲星因需进一步开展科学研究尚未对外公开发表, 全部脉冲星数据均收录至PSRDB进行数据管理, 未授权用户无法查阅未公开发表数据.

NAME	STATUS	RA(hms)	DEC(dms)	DM(cm ⁻³ pc)	TELESCOPE	PROJECT
J0847+40	已审核	08:47:11	40:54:12	24.45	FAST	CRAFTS
J1919+2623	已审核	19:19:16	+26:23:55	96.4	FAST	CRAFTS
J1502+4654	已审核	15:02:19	46:54:08	27	FAST	CRAFTS
J2112+4029	已审核	21:12:34	+41:05:12	132	FAST	CRAFTS
J1942+3934	已审核	19:42:15	39:34:60	104.51	FAST	CRAFTS
J0344-0901	已审核	03:44:37.4574714	-09:01:02.28075	31	FAST	CRAFTS
J2025+2133	已审核	20:25:40	21:33:39	70.75	FAST	CRAFTS

图 10 候选体数据维护页面

Fig. 10 Data maintenance page of candidates

图 11 脉冲星科学社区

Fig. 11 Pulsar scientific community

⁵https://crafts.bao.ac.cn/local/ismwiki/index.php/Pulsar_Search_19-beam

表 5 11颗来自FAST巡天项目CRAFTS公开发表的脉冲星
Table 5 Eleven published pulsars from CRAFTS of FAST

No.	Name	P0 /ms	DM /($\text{pc}\cdot\text{cm}^{-3}$)	RAJ2000 (hh:mm:ss)	DECJ2000 (dd:mm:ss)	DIST_DM	PEPOCH (MJD)
1	J0021-0909	2316	25	00:21:46	-09:09:17.1	25.00	58068.0
2	J0803-0942	571.02	21.5	08:03:37	-09:37:44.2	0.84	58072.0
3	J0529-0715	689	80	05:29:01	-07:11:52.0	4.86	58046.0
4	J0344-0901	1226	31	03:44:37	-09:01:02	1.93	58066.0
5	J1851-0633	640	228	18:51:47	-06:34:14.1	6.21	57998.0
6	J1900-0134	1832	188	19:00:26	-01:34:38	5.00	57987.0
7	J1919+42621	651	96	19:19:16	+26:23:55	6.15	57988.0
8	J1926-0652	1612	85	19:26:35	-06:49:48.5	5.24	57996.0
9	J1931-0144	593	36	19:31:23	-01:32:54.3	1.33	57987.0
10	J1945+1211	4745	95	19:45:16	+12:11:37	3.92	57999.0
11	J2323+1214	3760	22	23:23:22	+12:08:36	2.30	57997.0

PSRDB通过比对分析页面【Frequency Distribution】功能设置组数5, DM范围0-1000 $\text{pc}\cdot\text{cm}^{-3}$, 对140颗FAST优质候选体的色散进行频数分布统计分析, 如图12. 数据显示有高达123颗优质候选体色散值集中在200 $\text{pc}\cdot\text{cm}^{-3}$ 以内, 这对下一阶段FAST脉冲星数据处理中制定消色散方案有重要参考意义.

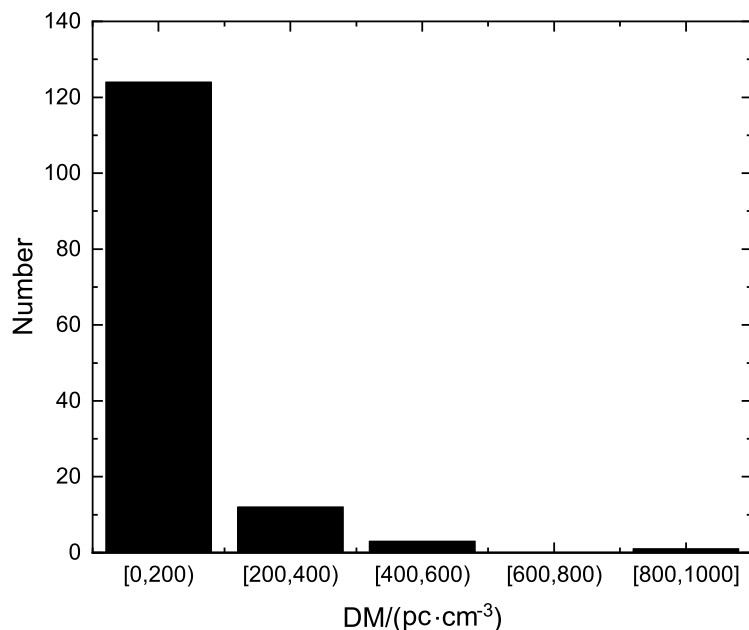


图 12 来自FAST的140颗优质候选体色散频数分布情况统计

Fig. 12 Statistics for the DM distribution of 140 new candidates from FAST

5 总结与展望

PSRDB已采集1967年自人类发现第1颗脉冲星以来所有公开发表的脉冲星2811颗,同时收录了包括CRAFTS、GBNCC、PALFA、LOTTAS等巡天项目尚未公开发表的源. PSRDB提供脉冲星和候选体检索匹配、验证、比对分析等功能,并实现了检索匹配数据的可视化展示. 目前, PSRDB已应用于FAST脉冲星数据管理,收录了140颗新优质脉冲星.

PSRDB为FAST下一步开展脉冲星搜寻、巡天计划制定、新源验证和数据处理等科研工作提供数据和工具支撑,同时,也可以为个人和其他科研机构提供候选体数据管理、新脉冲星比对和已知脉冲星检索匹配服务. 下一步我们将持续更新PSRDB数据记录,完善巡天项目、望远镜设备参数和补充已有脉冲星和候选体的部分缺失数据,设计更为丰富的统计分析模型和工具,扩展数据可视化功能.

致谢 感谢中国科学院国家天文台李菂教授, 澳大利亚联邦科学与工业研究组织George Hobbs教授、王晨教授和代实博士以及来自Parkes望远镜的数据支持, 特别感谢贵阳学院张正东博士、网易工程师张翔、软件独立开发者王存见对本文的建设性建议和帮助.

参考文献

- [1] Hewish A, Bell S J, Pilkington J D H, et al. *Nature*, 1968, 217: 709
- [2] Hulse R A, Taylor J H. *ApJ*, 1975, 195: L51
- [3] Weisberg J M, Nice D J, Taylor J H. *ApJ*, 2010, 722: 1030
- [4] Manchester R N, Hobbs G B, Teoh A, et al. *AJ*, 2005, 129: 1993
- [5] Li D, Pan Z C. *RaSc*, 2016, 51: 1060
- [6] Li D, Wang P, Qian L, et al. *IMMag*, 2018, 19: 112
- [7] Li D, Dickey J M, Liu S. *RAA*, 2019, 19: 016
- [8] Zhu W W, Li D, Luo R, et al. *ApJ*, 2020, 895: L6
- [9] 刘鹏, 王培, 李菂, 等. *天文学进展*, 2018, 36: 173
- [10] Li D, Nan R D, Pan Z C. *Proceedings of the International Astronomical Union*, 2012, 8: 325
- [11] Levin L, Armour W, Baffa C, et al. *Proceedings of the International Astronomical Union*, 2018, 13: 171
- [12] Manchester R N, Lyne A G, Camilo F, et al. *MNRAS*, 2001, 328: 17
- [13] Deneva J S, Cordes J M, McLaughlin M A, et al. *ApJ*, 2009, 703: 2259
- [14] Coenen T, Van Leeuwen J, Hessels J W T, et al. *A&A*, 2014, 570: A60
- [15] Stovall K, Lynch R S, Ransom S M, et al. *ApJ*, 2014, 791: 67
- [16] Qian L, Pan Z C, Li D, et al. *SCPMA*, 2019, 62: 959508
- [17] Zhang L, Li D, Hobbs G, et al. *ApJ*, 2019, 877: 55
- [18] Cameron A D, Li D, Hobbs G, et al. *MNRAS*, 2020, 495: 3515

Design and Implementation for Comparative Analysis and Visualization System of Pulsar Data

ZHANG Hui^{1,2,3} WANG Pei⁴ ZHANG Lei⁴ XU Yu-yun⁵ YUE You-ling⁴
PAN Zhi-chen⁴ LIU Zhi-jie^{1,3} YU Xu-hong^{1,3} YOU Shan-ping^{1,3}
JIANG Jia-tao^{1,2,3} XIE Xiao-yao^{1,3}

(1 *Guizhou Key Laboratory of Information and Computing Science, Guizhou Normal University, Guiyang 550001*)

(2 *School of Mathematical Sciences, Guizhou Normal University, Guiyang 550001*)

(3 *FAST Early Science Data Center, Guizhou Normal University, Guiyang 550001*)

(4 *National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012*)

(5 *Institute of Management Engineering, Guizhou Vocational and Technical College of Water Resources and Hydropower, Guiyang 551416*)

ABSTRACT The PulSaR DataBase (PSRDB, freely available at <http://www.psrdb.net/>), designed and implemented by the joint team from Guizhou Normal University (GZNU) and the Five-hundred-meter Aperture Spherical radio Telescope (FAST), is a highly dedicated, multipurpose and open-access database for the storage, visualization and comparative analysis of pulsar and its candidate data. Users can upload their candidate data through a unified user interface, and manage or share the data online. PSRDB encompasses records for more than 2811 known pulsars, gathered from pulsar literature and online database, published since 1967. PSRDB also includes new sources from major pulsar surveys, such as the Commensal Radio Astronomy FAST Survey (CRAFTS), which have not been officially published. In addition to its comprehensive database-derived, literature-derived data, PSRDB also supports a common text search function among the data and interactive visualization of the curated pulsar distributions with the most commonly used pulsar parameters and their uncertainties based on a web user interface. The PSRDB offers parameters and several comparative analysis methods, online real-time matching and retrieval of known pulsars, and multi-dimensional quantitative statistical analysis can be carried out in a predefined parameter space. Finally, results of the matching and comparative analysis can be visualized and displayed dynamically. The PSRDB can also provide candidate data management for individuals and institutes in the pulsar science community, and provide data support in new source confirmation observations, follow-up observation, new observation plan design and raw data processing, etc. Currently, PSRDB has been applied to the data processing and candidate management of FAST in FAST Early Science Data Center. PSRDB has been strenuously designed to address the broad demands of pulsar scientists, physicist, astrophysicist and members of the pulsar science community.

Key words astronomical data bases, pulsars: general, methods: comparative analysis, visualization, FAST