GECAM伽马暴数据分析算法和软件设计介绍

宋欣颖^{1*}, 熊少林^{1*}, 罗琦¹, 廖进元¹, 朱玥¹, 黄跃¹, 乔锐¹, 郭东亚¹, 陈维¹, 张凯^{1,2}, 蔡策¹, 肖硕¹, 彭文溪¹, 宋黎明¹, 郑世界¹, 王平¹, 李小波¹, 马想¹

1. 中国科学院高能物理研究所, 北京 100049;

2. 曲阜师范大学物理工程学院, 济宁 273100

*联系人, 宋欣颖, E-mail: songxy@ihep.ac.cn; 熊少林, E-mail: xiongsl@ihep.ac.cn

收稿日期: 2020-01-22; 接受日期: 2020-03-24; 网络出版日期: 2020-11-09

中国科学院战略性先导科技专项(编号: XDA15052700)和中国科学院前沿科学重点研究计划(编号: QYZDB-SSW-SLH012)资助项目

摘要 引力波暴高能电磁对应体全天监测器(Gravitational wave high-energy Electromagnetic Counterpart All-sky Monitor, GECAM)的物理数据分析工作至关重要, 所以科学用户需要一套能够涵盖所有伽马暴数据分析功能的软件, 且该软件具有满足用户友好、简单易用等特点. 本文描述了GECAM伽马暴数据分析软件的架构设计, 包括图形用户界面(Graphical User Interface, GUI)和命令行界面(Command Line Interface, CLI), 以及伽马暴数据分析的核心算法. 本文对核心算法的设计和分析流程进行了概述, 包括数据处理、本底分析、伽马暴定位、能谱拟合、上限计算、信号显著性计算、光变分析、外部触发事件关联概率计算, 以及对GECAM触发事件进行分类的朴素贝叶斯分类器, 为今后利用GECAM开展伽马暴数据分析工作提供了依据和参考.

关键词 引力波, 伽马暴, 数据分析, 软件设计

PACS: 04.80.Nn, 98.70.Rz, 07.05.Kf, 07.05.Hd

1 引言

引力波暴高能电磁对应体全天监测器(Gravitational wave high-energy Electromagnetic Counterpart All-sky Monitor, GECAM)项目是由中国科学院面向2020年前后引力波天文学研究的重要机遇而提出的专门探测引力波高能电磁对应体(简称引力波伽马暴)的机遇型空间科学探测项目.

GECAM由两颗微小卫星组成, 每颗卫星的有效

载荷包括伽马射线探测器(GRD)、荷电粒子探测器(CPD)和载荷处理器(EBOX), 利用25个GRD监测除了地球遮挡之外的全部天区. 两颗相同的GECAM卫星运行于相同轨道面内, 且轨道相位相反, 从而形成对全天的完整覆盖.

GECAM具有全天视场、高灵敏度、良好的定位精度和宽能段覆盖的特点, 具备在轨触发定位以及准实时上传探测结果、引导其他望远镜后随观测的能力. GECAM对引力波伽马暴以及其他全天随机发生

引用格式: 宋欣颖, 熊少林, 罗琦, 等. GECAM伽马暴数据分析算法和软件设计介绍. 中国科学: 物理学 力学 天文学, 2020, 50: 129511
Song X Y, Xiong S L, Luo Q, et al. Introduction to gamma-ray burst data analysis algorithm and software tools for GECAM (in Chinese). Sci Sin-Phys Mech Astron, 2020, 50: 129511, doi: [10.1360/SSPMA-2020-0020](https://doi.org/10.1360/SSPMA-2020-0020)

的高能暂现源的综合探测能力很大程度上优于国内外同类型的高能望远镜, 将是正在兴起的“多波段、多信使”天文学时代重要的高能望远镜之一. 除了跟引力波成协的伽马暴, GECAM还将探测跟快速射电暴和高能中微子成协的各种类型的伽马射线暂现源.

目前, 国际上已有许多针对伽马暴探测的空间望远镜配备的专门数据分析软件. 例如: 针对于费米伽马射线太空望远镜(Fermi, <https://fermi.gsfc.nasa.gov/>)上的伽马射线暴监测器(GBM)开发的RMFIT, 用于分析GBM触发数据^[1,2], 为用户处理GBM伽马暴数据提供了直观易用的工具.

GECAM数据分析软件是GECAM实现其科学目标的保证, 为此开发一款同样具有图形用户界面(Graphical User Interface, GUI)、易于使用的GECAM伽马暴数据分析软件是很有必要的.

2 软件需求

2.1 功能需求

本文所介绍的GECAM伽马暴数据分析软件属于GECAM科学应用系统科学研究支持分系统(具体介绍见文献[3]). 分系统的主要工作目标是: 结合载荷性能, 开展科学与观测模拟, 为在轨运行提供支撑; 根据有效载荷性能分析, 提出在轨标定方案; 完成数据分析软件开发, 提供科学用户支持; 建立科学值班系统; 建立管理GECAM主页, 提供综合信息服务平台, 提供数据产品和分析软件的发布功能, 发布研究进展等. 本软件是伽马暴数据分析和提供用户支持功能不可或缺的一部分.

因此, GECAM伽马暴数据分析软件需要涵盖科学用户在伽马暴数据分析过程中的绝大部分分析需求, 包括数据处理(区分每个探头的高低增益数据, 计算一段时间范围内的死时间, 计算时间/能道计数率)、光变曲线和能谱展示、本底分析、伽马暴定位、能量响应矩阵生成、计算总显著性、能谱拟合、光变分析、与外部触发事件关联性分析、在无伽马暴触发时针对外部触发的上限分析、对疑似伽马暴事件进行触发分类, 以便为后续的科学提供分析结果.

除此之外, 伽马暴数据分析软件需要具备与用户交互的功能, 既提供图形用户交互界面, 便于用户选

择分析参数或者输入指令; 又提供命令行界面(Command Line Interface, CLI), 便于用户单独调用某一个分析功能或者批量进行分析.

2.2 性能需求

本软件选择使用Python3.5以上版本对软件进行编写, 利用shell语言进行安装. 在推荐的硬件配置(内存: 8 G以上; 硬盘: 500 G以上)下, 软件满足性能需求: 同时处理两颗星(每颗星各有25个GRD)的数据, 数据载入速度为50 MB/s; 单个伽马暴响应矩阵生成时间 ≤ 10 s; 单个伽马暴能谱拟合时间, 对于每个伽马暴函数模型 < 20 s; 多探测器光变显示时间 < 2 s; 本底分析时间 < 200 s; 光变曲线与能谱切换显示时间间隔 < 1 s.

3 设计原理

本节描述GECAM伽马暴数据分析软件的设计原理. 本软件在功能上分为三个部分.

(1) GUI/CLI. GUI是与用户交互的图形界面, 用户从本界面上通过组件(按钮、文本输入框、图形操作界面)输入指令, 去调用数据处理模块, 或者调用核心算法, 两者返回需要显示的变量列表, 将其处理成给用户观看的图像或者文字, 以及分析结果. CLI是命令行界面, 用户安装本软件后, 在终端输入命令行和相应参数, 调用核心算法模块中的子模块或函数, 生成分析结果.

(2) 数据处理和共享变量模块. 接受GUI/CLI和核心算法的调用, 将用户选定的数据文件(FITS格式)进行处理, 生成便于软件分析的数据, 并将其定义为共享变量类, 在其他模块中引用, 达到共享变量的目的.

(3) 核心算法模块. 接受GUI/CLI的调用, 从数据共享变量模块读取数据, 分析之后返回分析结果至GUI/CLI或者共享变量模块, 供用户观看.

以下介绍了三个功能模块的设计细节, 且着重阐述了基于高能物理实验常用的伽马暴数据分析核心算法设计.

3.1 GUI/CLI设计

GUI界面设计主要使用多平台的工具包PyQt5(<https://pypi.org/project/PyQt5/>). 如图1所示, GUI主界面模块(GUI_mainframe_GECAM.py)中类GECAM_

图1 GECAM伽马暴数据分析软件GUI部分的主要架构, 包括各个组件及子界面

Figure 1 The main framework of GUI in GECAM GRB data analysis tools, including all widgets and sub-GUIs.

MainWindow(QMainWindow)由一个主要组件GECAM_MainWidget (QWidget), 即中心组件(Central-Widget)和菜单栏构成。

主要组件GECAM_MainWidget (QWidget)由布局管理、运行状态监视栏和左右面板构成。其中右面板包括一个图形展示界面, 继承自父类MPLwrapper (QWidget), 含有图形操作基类MPLbase (FigureCanvas)和Matplotlib库的工具栏Navigationtoolbar。左面板上罗列着许多功能组件(包括按钮和菜单栏等), 他们的布局由布局组件(QGroupBox, QVBoxLayout, QHBoxLayout, QGridLayout和QFormBoxLayout)管理, 当用户按下按钮或者菜单栏的下拉栏选项时, 依据Qt特有的信号槽机制, 会调用所连接的事件函数或者功能模块。如果该事件函数功能是发起一个新的子界面, 那么新的子界面便会弹出, 子界面的架构类似于GUI主界面, 具体的功能组件根据各自功能设计。

主界面左面板上的组件包括: 文件浏览选定按钮, 用户点击之后会弹出文件浏览选定子界面; 文件载入按钮, 用户点击之后, 本软件将会对载入的数据文件做处理, 使之生成时间和能道的二维数据, 方便软件进行分析; 光变曲线和能谱展示组件, 可以对某个探测器的能谱/光变切换展示, 对光变曲线重新分bin, 展

示不同时间/能量范围内的能谱和光变曲线; 本底分析范围选择按钮/文本输入框; 开启能谱拟合子界面按钮; 开启光变分析、上限计算、关联分析计算按钮。左面板所连接的子界面上组件的构成与该子界面的具体功能相关, 在此不再一一赘述。

右面板为图像展示组件和一个状态显示栏。图像展示组件的类基于Python第三方库Matplotlib (<https://matplotlib.org/>)中figure包中作图类以及包含图形用户界面组件matplotlib.widgets中的类和方法进行构造。通过用户点击图像, 实现选中探测器, 绘制本底/信号范围的功能。

GECAM数据分析软件主界面设计原型如图2所示。CLI界面的构造依赖于Python第三方库setuptools (<https://pypi.org/project/setuptools/>)。其中的entry_points语句对功能模块中的函数进行安装并命名。安装之后, 在终端键入命令行名称及参数, 可以实现命令行运行。

3.2 数据处理及共享变量模块设计

数据处理主要依赖Python的第三方库astropy (<https://www.astropy.org/>), 其包含天文学和天体物理学所需的关键功能和通用工具astropy.io。对于FITS格式

图2 GECAM伽马暴数据分析软件(图形用户)主界面原型

Figure 2 The prototype of main framework of GUI in GECAM GRB data analysis tools.

的输入文件, 提取时间和能道信息的二维数据, 修正死时间, 以便获得时间或者能道计数率, 做出对应探测器的光变曲线和计数能谱. 目前, 在推荐的硬件配置下, 测试中的数据处理速度达到约100 MB/s, 可以满足用户处理数据的速度需求.

为了用户与软件图形界面进行实时交互, 需要将一些变量在不同模块间达到实时共享. 因此, 构造一个类(Class globalVar), 定义需要共享的变量为类变量. 在其他模块中引用(import)该类, 可以达到实时共享变量的目的.

3.3 核心算法设计

GECAM伽马暴数据分析中核心算法包括本底分析、显著性计算、光变分析、伽马暴定位、响应矩阵生成、能谱拟合、上限计算、关联分析、二级数据生

成, 以及触发分类. 分析过程中, 各核心算法模块之间的流程关系如图3所示.

本软件的输入被定义为一级数据, 包含GECAM单星或双星每个探头探测到的伽马暴能道、时间及卫星姿态和工程辅助数据的FITS格式文件. 二级数据为经过本软件分析的物理结果, 包含本底形状信息、能谱形状、光变分析结果、信号显著性、与引力波事件的关联概率、与触发分类的类别和概率等信息的FITS格式文件.

用户进行伽马暴数据分析流程为: 通过本软件先将一级数据进行处理, 将时间和能道等信息变为本软件内部便于分析的数据格式(包含光变和能谱的二维数据信息), 然后进行本底分析, 生成净能谱数据和本底能谱数据, 用以计算信号显著性、光变分析, 并对伽马暴定位. 响应矩阵生成模块根据定位结果生成响应

图3 核心算法模块分析流程

Figure 3 Workflow for core algorithm.

矩阵, 用于能谱拟合. 关联分析模块根据定位结果计算本次伽马暴与引力波事件的关联概率. 如果用户没有在光变曲线上看到明显的伽马暴源信号, 能谱拟合的结果将用于上限计算. 所有计算模块的结果传入二级数据生成模块生成二级数据. 所有二级数据的结果将用于触发分类计算. 以上分析结果, 可根据用户需要, 通过GUI或者CLI调用相应模块进行运算或者显示.

对核心算法模块逐一介绍如下.

(1) 本底分析模块

本模块采用side bands方法估计本底. 用户通过GUI对伽马暴源信号左右的本底范围进行选择, 选出的本底范围作为side bands区域; 选用一定阶数的多项式为本底函数模型, 以最小二乘法拟合side bands区域的本底形状, 用来估计在计数能谱中源信号下的本底形状, 如图4所示. 算法对于每个能道的光变数据都进行本底分析, 从而给出源和本底在计数能谱上的形状.

(2) 显著性计算模块

算法思想如下: 假定每个探测器在给定信号区间内的本底数 b (由本底分析结果获得) 和给定信号区间内观测到的总事例数为 n_{obs} , 信号数为 $s = n_{\text{obs}} - b$. 根据 b 值计算源的统计显著性(Sigma) 和信号区内不含信号的概率 P 值^[4].

当 b 小于25时, b 的分布符合泊松分布规律. 将泊松分布累加为实验 P 值, P 值计算如下:

$$\begin{aligned}
 P &= P(n > n_{\text{obs}} \mid H_0) \\
 &= \sum_{n=n_{\text{obs}}}^{\infty} f(n; s=0, b) \\
 &= 1 - \sum_{n=0}^{n_{\text{obs}}-1} \frac{b^n}{n!} e^{-b},
 \end{aligned} \quad (1)$$

其中, H_0 为信号区内不含信号的假设. 由 $\text{Sigma} = s / \sqrt{b}$ 计算相应统计显著性.

当 b 大于25时, 我们认为期望值大的泊松分布近似为高斯分布, 以高斯分布计算无信号假设下本底数为 b 的概率, 即为 P 值. 由 P 值定义对应的统计显著性值($Sigma$)如下:

$$\int_{-Sigma}^{Sigma} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \equiv 1 - P. \quad (2)$$

P 值越小, 信号显著性越大, 即数据中存在伽马暴源信号的概率就越大. 在实际代码中, 由Python第三方库scipy中的模块stats计算 P 值和统计显著性.

(3) 光变分析模块

计算伽马暴持续时间及其统计误差. 将 S 记作随时间观察到伽马暴源信号事例数的积分. τ_f 记作从时间起始点到 S 达到总积分 S_{tot} 百分比为 f (%)的时刻.

$$\frac{f}{100} = \frac{\int_{\tau_s}^{\tau_f} (dS/dt) dt}{S_{tot}}. \quad (3)$$

由式(3)定义两种标志伽马暴持续时间的特征量 T_{90} 和

T_{50} 为 $T_{90}=\tau_{95}-\tau_5$, $T_{50}=\tau_{75}-\tau_{25}$, 文献[5]给出计算 T_{90} 和 T_{50} 统计误差的计算方法. 以 T_{90} 为例, 其统计误差决定于 τ_{95} 和 τ_5 , 这二者的误差又被流强积分 S_f 所决定, S_f 的统计误差分为两部分, 一部分为探测器计数带来的 $(dS_f)_{cnt}$, 另一部分为本底分析误差带来的 $(dS_f)_{fluc}$, 如式(4)和(5)所示:

$$(dS_f)_{cnt}^2 = \sum_{j=\tau_0}^{\tau_f} \left[\frac{\Delta C(t_j)}{\Delta t_j} \right] \Delta t_j, \quad (4)$$

$$(dS_f)_{fluc} = \sqrt{\left(\frac{\partial S_f}{\partial L_0} \right)^2 dL_0^2 + \left(\frac{\partial S_f}{\partial L_t} \right)^2 dL_t^2}. \quad (5)$$

$(dS_f)_{fluc}$ 由两部分独立误差组成, 分别是起始点 τ_0 和结束点 τ_t 的计数 $(\Delta C(t))$ 带来的误差, 这二者的误差都由本底的统计涨落决定, 由本底随时间的积分给出. 其中 dL_0 为 τ_0 所对应的时间bin内计数的泊松误差, dL_t 为 τ_0 到 τ_t 时间段内本底计数误差, 总误差由式(6)给出:

$$(dS_f)_{tot} = \sqrt{(dS_f)_{cnt}^2 + (dS_f)_{fluc}^2}. \quad (6)$$

除此之外, 光变分析模块还将给出给定时标下的流强峰值(Peak Flux), 即在给定的时标范围内对光变曲线积分, 找出伽马暴持续时间内流强的最大值.

(4) 定位模块

目前采用两种方法对伽马暴进行定位. 一种方法为用马尔可夫链采样法进行源位置和能谱的同时拟合[6], 适用于单星和双星的数据, 具体可见文献[7]; 另一种方法为交叉相关方法(Modified Cross-Correlation Function, MCCF)计算时间延迟定位, 适用于双星都有数据的情况, 算法在文献[8,9]中有详细叙述.

(5) 响应矩阵生成模块

响应矩阵包含了探测器对于伽马暴响应的具体信息, 包括探测器对于伽马暴光子的探测效率及能量分辨. GECAM载荷探测器的蒙特卡罗模拟数据给出全天方位角的响应矩阵库. 当用户通过GUI调用响应矩阵生成模块时, 算法从响应矩阵库中找出距用户给定方位角球面最近的三个方位角的响应矩阵, 利用插值法得到用户给定方位角的响应矩阵, 具体见文献[7].

(6) 能谱拟合模块

引入了NASA发行软件HEASoft中(<https://heasarc.gsfc.nasa.gov/heasoft/>)的PyXspec模块. 当用户开启能

图4 Side bands方法本底分析示意图. 绿色区域(BG)为选中的本底side bands, 橙色曲线(Curve for fitted BG)为多项式拟合得出的描述本底形状的曲线. 蓝色散点为光变数据点

Figure 4 Plot for BG fitting with side bands. The green regions denote the side bands of background selected. Orange curve represents the shape of background extracted from fitting with polynomial. Blue dots with error bars denote the data from light curve.

谱拟合功能后, 通过GUI选择软件提供的函数模型(包括Power-law, Band, Comptonized, Smoothly Broken Power Law共4种伽马暴能谱模型^[10], 以及黑体谱与这4种模型的叠加模型)并输入能谱拟合参数, 算法自动寻找响应矩阵, 并与函数模型卷积, 用二者卷积生成的分布去拟合伽马暴能谱的真实数据, 最终给出所选模型的参数值、伽马暴流强及统计误差。

(7) 上限计算模块

本模块针对于外部触发事件(例如引力波事件)未在GECAM监测数据上获得触发, 源信号不明显的情况. 根据该外部触发事件发生时间窗内(时间窗宽度由用户指定)的净计数能谱和本底形状进行能谱拟合, 然后根据能谱拟合结果(包括流强及统计误差)计算可能监测到的伽马暴流强在一定置信度下的上限^[4].

(8) 关联分析与关联分析模块

以外部触发事件是引力波事件为例. 由用户通过GUI输入本次伽马暴发生的时间 t_{GRB} , 引力波事件发生的时间 t_{GW} . 定义一个 p -value, 量化伽马暴事件与引力波事件的偶然符合性, 其含义是伽马暴事件与引力波事件无关的概率. p -value由两部分构成: 时间概率 P_{temporal} 和空间概率 P_{spatial} . 时间概率 P_{temporal} 由式(7)决定:

$$P_{\text{temporal}} = 2\Delta t_{\text{GRB-GW}} R_{\text{GECAM-GRB}}, \quad (7)$$

其中, $\Delta t_{\text{GRB-GW}} = t_{\text{GRB}} - t_{\text{GW}}$, 是伽马暴与引力波发生的时间差. $R_{\text{GECAM-GRB}}$ 是GECAM探测到伽马暴的时间概率.

为了计算 P_{spatial} , 参考文献[11]的算法, 定义信号统计量 $S = \sum_{i=1}^{N_{\text{pix}}} P_1 P_2$, P_1 和 P_2 是用户输入的伽马暴与引力波事件的定位概率. 用GECAM观测到所有与引力波无关的伽马暴事件作为样本, 将其定位概率密度在全天随机平移反转, 构造一个本底统计量 B , 利用显著性计算模块中计算 P 值的方法计算 P_{spatial} .

时间概率 P_{temporal} 和空间概率 P_{spatial} 是相互独立的, 因而 p -value= $P_{\text{temporal}} \times P_{\text{spatial}}$. 根据显著性计算模块的式(2)计算伽马暴与引力波事件相关联的统计显著性.

(9) 二级数据生成模块

将以上所有分析结果生成FITS格式的二级数据, 并上传至GECAM科学应用系统的触发暴发数据库, 便于用户进行下一步的研究.

(10) 触发分类与触发分类模块

当用户对触发数据进行了以上所有分析并得出了分析结果之后, 将已知类型的触发事件发生时的卫星位置、伽马暴源定位、光变曲线的形状、光变分析(持续时间)和能谱拟合的结果(计数能谱的硬度比)作为控制样本, 构造出一个朴素的贝叶斯分类器^[12]. 该分类器可作为软件插件, 区分触发的种类(粒子事件、伽马暴、地球伽马闪、太阳耀发等), 并计算触发事件属于每种触发的概率.

设触发种类(Class)为 c_i , $i=1, 2, 3, \dots, m$, m 为GECAM能够探测到的触发类型数目; 区分触发种类的特征量为向量 $\mathbf{f}=(f_1, f_2, f_3, \dots)$, 向量维数 n 为特征量个数.

GECAM观察到特征向量为 $\mathbf{f}=(f_1, f_2, f_3, \dots)$ 时, 触发事件属于种类 c_i 的概率记为 $P(c_i | \mathbf{f}=(f_1, f_2, f_3, \dots))$. 根据贝叶斯原理^[4], $P(c_i | \mathbf{f}=(f_1, f_2, f_3, \dots))$ 的计算方法如式(7)所示:

$$P(c_i | \mathbf{f}=(f_1, f_2, f_3, \dots)) = \frac{P(\mathbf{f} | c_i) \times P(c_i)}{P(\mathbf{f})}, \quad (8)$$

其中, $P(\mathbf{f} | c_i)$ 是触发种类为 c_i 时观察到特征量 $\mathbf{f}=(f_1, f_2, f_3, \dots)$ 的概率; $P(c_i)$ 为GECAM探测到触发种类 c_i 的概率; $P(\mathbf{f})$ 为特征量 $\mathbf{f}$ 被探测到的概率. 如果特征量分量 f_j 之间互相独立, 即符合朴素贝叶斯分类方法, 那么 $P(\mathbf{f} | c_i)$ 和 $P(\mathbf{f})$ 满足

$$P(\mathbf{f} | c_i) = \prod_j P(f_j | c_i), \quad (9)$$

$$P(\mathbf{f}) = \prod_j P(f_j). \quad (10)$$

$P(c_i)$ 和 $P(\mathbf{f})$ 都可以通过GECAM对已知触发种类观测数据的积累得到. 在积累了一定量的控制样本之后, $P(\mathbf{f} | c_i)$ 可以由每种分类里每个特征量的分布进行累积计数, 平滑或者拟合给出. 我们把这些已知触发种类的数据称作训练样本, 可以想见, 已知触发种类的观测数据越丰富, 统计量越大, 则这些概率密度函数的形状就越准确, 相对统计误差就越小, 我们所构造的GECAM触发分类器也就被训练得越充分. 对于一个未知触发类型的触发事件, 由式(7)可以得出每种触发分类下的概率, 概率最大者被作为伽马暴的触发类型.

需要特别指出的是, 在特征量的选取中, 并未要求特征量全部为概率密度函数分布连续的统计量, 推广

至包含离散分布的统计量, 如伯努利分布时, 分类器仍然适用. 分类器算法结构如图5所示.

4 总结

本文所展示的GECAM伽马暴数据分析软件的设计为用户提供了直观易用的GUI和CLI, 并将在之后的使用过程中采集用户的反馈和建议, 丰富和完善相关功能.

此外, 本文介绍的算法涵盖了绝大部分伽马暴数据分析的需求, 其中包括基于高能物理实验统计学方法开发的本底分析、信号显著性、上限分析算法; 也包括了基于伽马暴数据通用分析方法开发的光变分析算法; 为今后在GECAM上开展的伽马暴数据分析工作提供了依据和先例, 也将为后续研究提供准确而详实的分析结果. 除此之外, 基于贝叶斯原理开发出的朴素贝叶斯分类器被用于触发分类算法, 为天体物理同类分类问题提供了普遍的分类器雏形. 并且, 对于引力波暴电磁对应体——伽马暴的数据分析算法可用于扩展补充引力波多信使观测项目, 如硬X射线望远镜

图5 触发分类算法流程

Figure 5 The workflow for trigger classification algorithm.

镜(HXMT)^[13]及未来的增强型X射线时变与偏振空间天文台(eXTP)^[14].

参考文献

- Burgess J M, Preece R D, Baring M G, et al. Constraints on the synchrotron shock model for the Fermi GRB 090820A observed by gamma-ray burst monitor. *Astrophys J*, 2011, 741: 24
- Greiner J, Burgess J M, Savchenko V, et al. On the Fermi-GBM event 0.4 s after GW150914. *Astrophys J*, 2016, 827: L38
- Chen W, Song L M, Zheng S J, et al. Introduction of the scientific application system of GECAM (in Chinese). *Sci Sin-Phys Mech Astron*, 2020, 50: 129512 [陈维, 宋黎明, 郑世界, 等. GECAM卫星科学应用系统简介. 中国科学: 物理学 力学 天文学, 2020, 50: 129512]
- Zhu Y S. Probability and Statistics in Experimental Physics (in Chinese). Beijing: Science Press, 2006. 14: 429 [朱永生. 实验物理中的概率与统计. 北京: 科学出版社, 2006. 14: 429]
- Koshut T M, Paciesas W S, Kouveliotou C, et al. Systematic effects on duration measurements of gamma-ray bursts. *Astrophys J*, 1996, 463: 570
- Patil A, Huard D, Fonnesbeck C. PyMC: Bayesian stochastic modelling in Python. *J Stat Soft*, 2010, 35: 1
- Liao J-Y, Luo Q, Zhu Y, et al. The localization method of GECAM and simulation analysis (in Chinese). *Sci Sin-Phys Mech Astron*, 2020, 50: 129510 [廖进元, 罗琦, 朱玥, 等. GECAM定位方法及模拟分析. 中国科学: 物理学 力学 天文学, 2020, 50: 129510]
- Hurley K, Briggs M S, Kippen R M, et al. The ULYSSES supplement to the BATSE 3B catalog of cosmic gamma-ray bursts. *Astrophys J Suppl S*, 1999, 120: 399-408
- Li T P, Qu J L, Feng H, et al. Timescale analysis of spectral lags. *Chin J Astron Astrophys*, 2004, 4: 583-598
- Gruber D, Goldstein A, von Ahlefeld V W, et al. The Fermi GBM gamma-ray burst spectral catalog: Four years of data. *Astrophys J Suppl Ser*, 2014, 211: 12
- Abbott B P, Abbott R, Abbott T D, et al. Gravitational waves and gamma-rays from a binary neutron star merger: GW170817 and GRB 170817A. *Astrophys J*, 2017, 848: L13
- Zhang H. The optimality of Naive Bayes. In: Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference. Miami Beach, 2004

- 13 Li T P, et al. (The *Insight*-HXMT team). *Insight*-HXMT observations of the first binary neutron star merger GW170817. *Sci China-Phys Mech Astron*, 2018, 61: 031011
- 14 Zhang S N, Santangelo A, Feroci M, et al. The enhanced X-ray Timing and Polarimetry mission—eXTP. *Sci China-Phys Mech Astron*, 2019, 62: 029502

Introduction to gamma-ray burst data analysis algorithm and software tools for GECAM

SONG XinYing^{1*}, XIONG ShaoLin^{1*}, LUO Qi¹, LIAO JinYuan¹, ZHU Yue¹,
HUANG Yue¹, QIAO Rui¹, GUO DongYa¹, CHEN Wei¹, ZHANG Kai^{1,2}, CAI Ce¹,
XIAO Shuo¹, PENG WenXi¹, SONG LiMing¹, ZHENG ShiJie¹, WANG Ping¹,
LI XiaoBo¹ & MA Xiang¹

¹ *Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China;*
² *School of Physics and Physical Engineering, Qufu Normal University, Jining 273100, China*

The Gamma-Ray Burst (GRB) data analysis is of great importance for the research on Gravitational wave high-energy Electromagnetic Counterpart All-sky Monitor (GECAM); therefore, a user-friendly software tool that covers all the requirements on GRB data analysis has been developed for users. In this paper, the framework and architectural design of GRB data analysis tools software has been described. Meanwhile, details design including the graphical user interface and command line interface were introduced, as well as its installation. The core algorithms of the GRB data analysis were clarified and outlined, including background analysis, calculations on statistical significance, duration and peak flux in given time scale of light curve, fit for deposited energy spectrum, the calculation for the association of GRBs and external trigger event, and a trigger classifier based on Native Bayesian Classification, which offers a rationale and sample for the forward GRB data analysis in GECAM.

gravitational wave, gamma-ray bursts, data analysis, software design

PACS: 04.80.Nn, 98.70.Rz, 07.05.Kf, 07.05.Hd

doi: [10.1360/SSPMA-2020-0020](https://doi.org/10.1360/SSPMA-2020-0020)