

基于 Web 的色质数据解析平台设计与开发

孔波¹, 邹有², 卢红兵¹, 杨华武¹, 度苏行¹

(1. 湖南中烟工业有限责任公司技术研究中心, 湖南长沙 410007;
2. 中南大学信息与网络中心高性能计算中心, 湖南长沙 410083)

摘要:色质联用仪数据分析软件系统是处理和分析色质联用仪样品数据的主要工具,是色质联用分析技术不可或缺的重要组成部分。目前分析色质联用仪产生的样品数据,还需要安装特定的商用仪器软件,这些商业分析软件通常与仪器捆绑销售,使用的经济成本高,其便利性和共享性也非常欠缺。为了更加准确、方便、快捷地分析色质联用仪产生的样品数据,提供高效、灵活、低成本的数据解析服务,基于 Node.js 的 Express 应用框架,设计开发了基于 Web 的色质数据解析平台。该平台可以实现色质数据的可视化、谱峰检测以及质谱检索等功能。为了提高数据处理效率,采用 GPU 加速技术对多任务并发的运算进行加速。该平台使用便捷、数据处理高效,易于部署。实验结果表明,该平台在保证解析结果准确性的同时,能够极大地提高色质数据解析效率。

关键词:色质解析;质谱检索;CUDA;Node.js

中图分类号:TP312

文献标识码:A

文章编号:1673-629X(2019)12-0198-07

doi:10.3969/j.issn.1673-629X.2019.12.036

Design and Development of GC/MS Data Analysis Platform Based on Web

KONG Bo¹, ZOU You², LU Hong-bing¹, YANG Hua-wu¹, TUO Su-xing¹

(1. Technology Research & Development Center, Hunan Branch of China Tobacco Industry Co., Ltd., Changsha 410007, China;
2. High Performance Computing Center, Information and Network Center of Central South University, Changsha 410083, China)

Abstract: Data analysis software system of GC-MS is the main tool to process and analyze sample data of GC-MS and is an indispensable and important part of GC-MS analysis technology. At present, it is necessary to install specific commercial instrument software to analyze sample data produced by GC-MS. This commercial analysis software is usually sold in bundles with instruments, which are of high economic cost and lack of convenience and sharing. In order to analyze the sample data produced by GC-MS more accurately, conveniently and quickly, and provide efficient, flexible and low-cost data analysis service, we design and develop a Web-based color data analysis platform based on Express application framework of Node.js. The platform can realize the visualization of color data, peak detection and mass spectrometry retrieval. In order to improve the efficiency of data processing, we use GPU acceleration technology to speed up multi-task concurrent operation. The platform is convenient to use, efficient in data processing and easy to deploy. The experiment shows that the platform not only ensures the accuracy of the analytical results, but also greatly improves the resolution efficiency of GC-MS data.

Key words: GC/MS analysis; spectral identification; CUDA; Node.js.

1 概述

色谱-质谱联用仪^[1]是色谱分离技术和质谱分析技术紧密结合的产物,在环境监测、食品安全、生命科学等领域有着广泛的应用。色谱-质谱联用仪主要是

由硬件和配套的分析软件两部分组成,其工作流程是通过一系列的物理、化学反应,将输入的待测物质转化成为一系列的二进制色质数据,并交由仪器自带的分析软件进行分析,最后得出分析结果。随着计算机技

收稿日期:2018-12-13

修回日期:2019-04-16

网络出版时间:2019-09-24

基金项目:湖南省教育科学“十三五”规划课题(XJK17BXX011)

作者简介:孔波(1984-),男,博士,研究方向为化学计量学、烟草化学。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190924.1534.002.html>

术的快速发展,新的技术和设备不断出现,不同设备通常配备着不同的数据分析软件,这些软件的性能和使用方法上存在较大的差异,也为用户使用这些软件产生了一定的障碍。

色谱-质谱联用仪主要有两种形式:气相色谱-质谱联用(GC-MS)^[2]和液相色谱-质谱联用(LC-MS)^[3],这两种方法互为补充,适用于不同性质化合物的分析。

气质联用技术是最早商品化的联用仪器,适用于小分子、易挥发、热稳定、能气化的化合物,其质谱仪器的电离源一般采用电子轰击电离(EI)源。液质联用技术主要适用于大分子、难挥发、热不稳定、高沸点化合物的分析。相较于液质联用技术,色质联用技术出现较早,1957年,Holmas和Morrell第一次实现了气相色谱仪与质谱仪(GC-MS)连接使用^[4]。之后,随着气相色谱技术和质谱技术的不断进步,以及色质联用接口技术的发展,这项分析技术因其高通量、高分辨率的特点已经成为药品分析^[5]、环保监测^[6]、农药残留监测^[7],以及兴奋剂监测^[8]等领域的重要监测手段。

目前,随着生命科学的发展,待测样品更加复杂,分子范围更大,高效色谱-质谱联用技术的研发愈受到重视。近年来,在国内外研究人员的共同努力下,高效色谱-质谱联用技术研究取得了一些进步,但是仍然存在一些问题。这些问题主要表现在如下两个方面。

首先,硬件设备对实验条件要求较高,分析软件庞大、笨重,且与硬件设备捆绑使用。由于仪器设备对运行环境有较高的要求,必须在专门的实验室环境下运行,不便于移动。另外,实验人员必须在仪器周围,全程参与、监督整个分析过程。即使实验人员将生成的色质数据进行暂时保存,在后期的分析过程中还需要安装庞大、复杂的专用分析软件,而这些专用软件都是与相应硬件设备捆绑销售,利润极高,极大地增加了数据分析成本,在这一系列复杂软件中有部分软件还对使用范围作了限制,这给数据分析应用带来了巨大的不便,严重影响了该领域的发展。

其次,仪器分析软件存在多种缺陷,在某些应用场景中无法得到准确的结果,而仪器的封闭性强,更好的解决方案无法载入。仪器分析软件存在多种缺陷,在诸如重叠峰解析等一些应用场景中无法得到准确的结果。另外,即使研究人员对于这些缺陷有比较好的解决方案,也会因仪器设备和软件开放性较差而无法实施。分析软件的缺陷和封闭性严重影响了该领域的应用和发展。

虽然不同色谱-质谱联用仪配备的数据分析软件各不相同,但是它们实现的功能却大致相似。比如,曲

线平滑与去噪、线性-非线性拟合、模式识别、数据管理及其常见的运算模式等。当前这些设备配备的软件在功能上存在低水平的冗余,在数据分析层面这些软件却没有较好的兼容性。综上所述,设计和开发一款准确性高、开放性好、灵活便利的色质数据分析平台具有重要的现实意义。

目前,国外研究人员已经开发了一些用于色质数据解析的软件或软件包,例如:由Biller和Biemann提出的“点比较”算法广泛应用于纯净质谱图的自动提取^[9]。Colby在Biller算法的基础上进行了延伸,将同时达到最大峰的质荷比归为同一物质,以获得准确的保留时间信息^[10]。斯坦福大学的Dromey等提出了“峰型比较”的退卷积算法,被广泛用于数据噪声去除^[11]。尽管这些软件都得到了很好的应用,但是使用前还是需要对其进行安装和配置,通用性受到了一定程度的限制。基于Web平台的解析系统可以更加方便地为用户提供服务,其无需安装和配置,点击浏览器,输入网址就可以进行相应操作。目前,基于Web平台的解析系统较少,比较典型的有AIST^[12]和SDBS^[13],这两个系统是运行在Web上免费提供质谱检索功能的检索系统。

目前,国内色谱-质谱联用仪的绝大多数市场份额都由进口设备占领,而这些进口设备中,又以美国安捷伦(Agilent)公司和日本岛津(shimadzu)株式会社生产的色谱质谱联用仪为主。该研究以烟草行业使用较多的安捷伦气相色谱-质谱联用仪输出的色质数据为主要输入数据,开发一款基于Web的色质数据解析平台。平台所接受的数据输入适用于安捷伦仪器产生的MS格式数据,平台所涉及的算法适用于通用的气相色质数据分析。国内研究人员也已经开发了一些解析用色质数据的软件,例如:北京科技大学的陈海勇等研发了用于TVS-2200红外热像仪图像数据解析的分析软件^[14];南京理工大学的董海波等研发了用于VECTOR32型FTIR光谱仪图像数据解析的软件^[15];中国农业大学的王一鸣等研发了近红外光谱品质检测软件^[16]。虽然在国内外研究人员的努力下,数据解析软件的研发取得了长足的发展,但是仍然不能满足现实需求。

Express是一个简洁而灵活,基于node.js的Web应用框架,提供了一系列强大特性有助于创建各种Web应用和丰富的HTTP工具。使用Express可以快速地搭建一个完整功能的网站。Express框架核心特征包括:(1)可以设置中间件来响应HTTP请求,中间件就是处理HTTP请求的函数,它最大的特点就是,一个中间件处理完,再传递给下一个中间件。应用程序在运行过程中会调用一系列的中间件,每个中间件可

以从应用程序实例接收三个参数,依次为 request 对象(代表 HTTP 请求)、response 对象(代表 HTTP 回应)、next 回调函数(代表下一个中间件)。每个中间件都可以对 HTTP 请求(request 对象)进行加工,并且决定是否调用 next 方法,将 request 对象再传给下一个中间件;(2)定义了路由表用于执行不同的 HTTP 请求动作;(3)可以通过向模板传递参数来动态渲染 HTML 页面。

文中基于 Node.js 的 Express 应用框架,设计开发了基于 Web 的色谱数据解析平台。该平台可以实现色谱数据的可视化、谱峰检测以及质谱检索等功能。为了提高数据处理效率,采用 GPU 技术对多任务并发的运算进行加速。该平台使用便捷、数据处理高效,易于部署。

2 气相色谱-质谱联用技术原理描述

气相色谱(GC)作为一项成熟的分析技术可连接多种检测器,在火焰离子化检测器(FID)质谱检测器中最为常见。质谱检测器可以通过标准谱图库提供化合物的结构信息,并具有高灵敏度等优势。气相色谱-质谱(GC-MS)联用技术是鉴定挥发性复杂混合物中化合物的首先方法,其分析过程是通过将挥发性有机混合物样品经由微注射器进入气相色谱的进样口,经过色谱柱、电离源、质量分析器并最终达到检测器,从而被检测出转化成质谱信号输出到计算机存储处理。化合物在给定的色谱分析条件下会在某一时间段内流出色谱柱,并在某一时间点达到流出浓度的峰值,这一时间点被称为该化合物的保留时间。多种化合物的混合物根据其成分,保留时间的不同依次通过色谱柱。在电离源中,色谱的流出物被电离成碎片离子,碎片离子进入质谱分析部件,首先被电离成带电离子,然后进入质量分析器,不同质荷比(m/z)的离子将在不同的位置、时间和轨道进行分离,得到按质荷比大小排列的质谱,从而被物质分析器检测出。一个色谱检测的实例如图 1~图 3 所示。

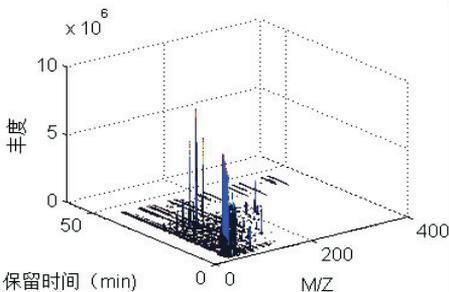


图 1 色谱数据矩阵的三维视图

(x 轴表示质荷比, y 轴表示保留时间, z 轴表示离子丰度(或相对丰度))

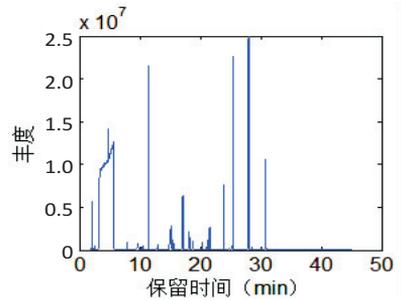


图 2 将图 1 中数据矩阵按保留时间累加得到的总离子流图

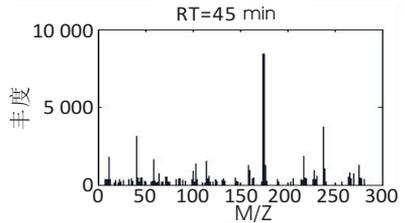


图 3 将图 1 中数据矩阵保留时间为 49.24 min 的数据得到的该时刻质谱图

2.1 色谱数据格式

不同厂商生产的色谱-质谱联用仪,其输出的色谱数据格式各有不同。例如:安捷伦使用的是 MS 格式、岛津使用的是 CDF 格式、Finnigan 使用的是 RAW 格式、Waters 使用的是 DAT 格式等^[17]。这些文件包含了实验数据的色谱和质谱信息,并且以独有的二进制格式存储,使用二进制存储可以有效地减少文件占用的磁盘空间,但是同时也给第三方使用带来了一定的不便。例如,安捷伦使用的 MS 文件存储结构如表 1 所示。

表 1 MS 文件存储结构

位置(字节)	说明
0-3	MS 文件标识
4-23	测量类型说明
24-148	文件信息
149-178	操作员信息
179-208	测量时间
209-259	仪器说明
260-263	色谱数据的起始地址
264-277	保留
278-281	色谱点数
282-285	测试起始点
286-289	测试终止点
290-765	保留
766-文件末尾	色谱数据

2.2 色谱数据解析

色谱数据解析主要包括谱峰检测和质谱检索两大关键技术,谱峰检测是在色谱中确认峰的位置,并计算

其大小。质谱检索则是将未知组分的质谱与质谱数据库数据进行匹配度计算,从而实现定性。

3 基于 Web 的色质数据解析平台设计与实现

目前,Web 服务技术已从基础设施的构建与概念推广阶段向大规模应用阶段快速发展。B/S (Browser/Server) 结构即浏览器和服务器结构,它是随着 Internet 技术发展而兴起,对 C/S (Client/Server) 结构进行优化或者改进的结构。在 B/S 结构下,用户工作界面通过浏览器实现,不同身份,不同地点,不同介入方式(比如 LAN, WAN)的用户都可以通过浏览器共同访问和操作用户工作界面,能够更好地实现资源共享和互联互通。另外, B/S 结构的开发和维护成本相对较低。B/S 结构系统能够一次性开发到位,功能添加和修改只需在服务器端进行,大大减小了系统维护成本。随着 JAVA 这样的高级语言的发展,基于 B/S 结构的平台系统能够很好地满足用户跨平台使用的需求。

3.1 平台框架

Node.js 是基于 Chrome V8 引擎的 JavaScript 运行环境,由于其轻量而高效,被广泛应用于 Web 和移动应用开发之中,近年来随着桌面应用的发展, Node.js 由于其灵活性已被广泛应用在桌面应用的开发之中。色质数据解析过程中存在大量文件读写和数值计算操作,而 Node.js 的非阻塞式 I/O 模型可以有效地满足这种情形下的任务并发。基于上述原因,该项目组选用 Node.js 的 Express 应用框架进行系统开发。图 4 展示了平台的基本架构。客户端(Web 端)主要承担数据输入、数据展示、用户交互、结果展示、结果输出等功能,服务端则承担数据运算功能,包括文件解析、谱峰检测和检索等。

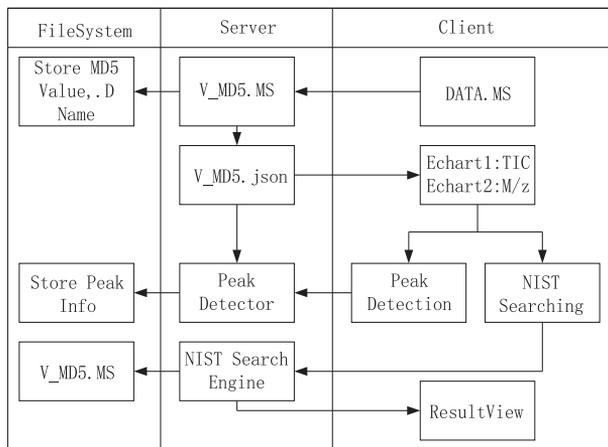


图 4 基于 Web 的色质数据解析平台架构

图 4 中的箭头走向揭示了色质数据从输入平台至生成解析结果的一般流程。其具体解析过程可以描述

为以下 4 个步骤:(1) 客户端输入 MS 格式数据后,服务端对其内容进行解析,得到色质数据矩阵。随后,服务端以数据文件的 MD5 校验码为文件名,以 JSON 格式进行存储。此过程中采用 MD5 校验码为文件名,主要是考虑到 MS 数据通过网络传输到服务端需要一定的时间,如果客户端上传的内容在服务端已经存在,则可直接调用历史数据,避免重复传输耗费不必要消耗的资源;(2) 服务端返回色质数据矩阵用于客户端的可视化展示。客户端采用百度公司开发的 JS 可视化插件 Echarts 进行可视化展示,该插件功能强大且完全免费,完全可以满足平台的数据可视化需求;(3) 客户端发出谱峰检测请求,服务端调用谱峰检测器进行色质数据的谱峰检测,并保存检测结果;(4) 客户端发出质谱检索请求,服务端调用质谱检索引擎对谱峰位置进行检索,保存检索结果,并返回客户端用于查看。

3.2 平台实现

从 3.1 节所述的数据解析流程可以分析出系统主要功能模块,包括:MS 数据解析、谱峰检测器、质谱检索引擎三大部分。下面将对三大核心部分的详细设计与实现分别进行描述。

3.2.1 数据解析

根据表 1 提供的 MS 文件存储结构,采用 C++ 开发 MS 文件解析模块。解析 MS 文件有两个关键点:(1) MS 数据采用 Little endian 模式存储,即低位字节存放在内存的低地址端,按字节读取多字节数据(如 long, int 等)时,需采用相应的模式进行转换;(2) MS 数据中不同信息的数据存储所占的字节数略有不同,如表 2 展示了主要数据信息的单个数据占用字节,只有按字节读取正确的字节数,并转换成对应的数据类型,才能得到正确的色质数据。

表 2 MS 文件存储结构

标识	所占字节数	数据类型
保留时间	4 字节	float
扫描点 ID	2 字节	int
总离子流	4 字节	float
质谱数据	2 字节	float

3.2.2 谱峰检测器

谱峰检测在信号处理领域应用非常普遍。在色谱、光谱数据分析的诸多相关研究之中也有涉及。文中采用连续的小波变换,将峰检测问题转换成小波系数矩阵的脊线检测问题,以此定位谱峰位置。连续小波变换是对信号缩放和平移的基函数乘积的积分,其数学表述如式 1 和式 2 所示。

$$C(a, b) = \int_{-\infty}^{+\infty} s(t) \psi_{a,b}(t) dt \quad (1)$$

$$\psi_{a,b}(t) = \frac{1}{a^{1/2}} \psi\left(\frac{t-b}{a}\right), a \in R^+, b \in R \quad (2)$$

其中, $s(t)$ 为信号函数; a 为尺度因子; b 为平移因子; $\Psi(t)$ 为墨西哥帽小波基函数; $\Psi_{a,b}(t)$ 为经缩放和平移的小波基函数; C 为小波系数矩阵。

谱峰检测算法的基本流程可以描述如下:

(1) 对原始数据进行小波变换, 得到不同尺度下的小波系数;

(2) 连接不同尺度下小波系数的局部最大值, 得到脊线;

(3) 通过脊线确定峰的位置;

(4) 过滤信噪比低的假峰。

该平台基于 Armadillo C++ 科学计算库实现, 图 5 展示了真实数据下, 使用上述算法进行谱峰检测得到的最终检测结果。其中 x 轴表示保留时间, 单位是分钟; y 轴表示数据的丰度。

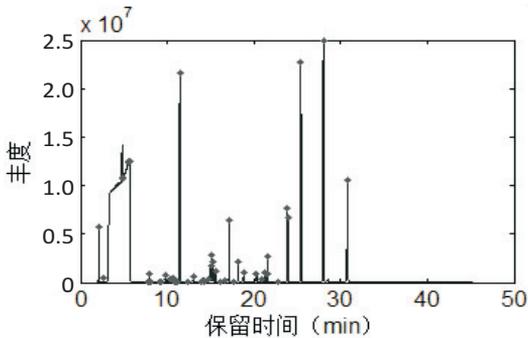


图 5 使用连续小波变换进行谱峰检测的结果

3.3 质谱检索引擎

在一些包含复杂成分的色质数据中, 经过谱峰检测后会得到许多待检索的未知质谱, 每一个未知质谱都需与标准质谱库中的质谱计算匹配度, 以此来决定其对应的定性结果。目前使用较多的标准质谱数据库有 Wiley 数据库和 NIST 数据库, 两者均为商业数据库。以 NIST08 版数据库为例, 该库的主库收录了 212 961 种物质的标准质谱。假设未知质谱数量为 n , 质谱库标准质谱数量为 m , 则需进行 mn 次匹配度计算, 以烟草行业焦油数据为例, 一个数据的匹配度计算量就达数以千万计。与桌面端应用不同的是, Web 应用还需考虑多任务并发情形下的计算效率, 当有 p 个并发检索任务时, 匹配度计算量为 $m \sum (n_1 + n_2 + \dots + n_p)$ 。为加速匹配度计算, 该项目组提出了一系列预检索算法, 先采用简单策略快速过滤质谱库中明显不是候选结果的质谱, 以此来减少后续匹配度计算的次数, 这些策略对提高检索速度都起到了非常重要的作用。但面对 Web 模式下的多任务并发带来的大量未知质谱检索, 仍然难以在可观的时间内完成。为此, 文中利用 GPU 在并行计算中的优势, 设计并实现了基于

CUDA (compute unified device architecture) 的快速检索方法。CUDA 是 NVIDIA 公司针对其 GPU 产品生产的一套可以使用类 C 语言进行通用计算的开发环境和软件体系, 通过它可以调用 GPU 来完成协同计算。目前许多领域的计算问题都通过 GPU 加速得到了改观。GPU 是一种非常适合于单指令多数数据流 (SIMD) 模式计算的硬件, 针对质谱检索问题, 由于每个未知质谱的检索过程完全独立且完全相同, 因此可以很好地应用到 GPU 上进行加速运算。此外, CUDA 环境专门提供了一套 GPU 版本的线性代数库 (CUBLAS), 可以非常方便高效地实现向量运算的并行化。

另外, 为使计算速度更快, 本平台也采用一种基于强峰过滤的预检索方法。该方法首先对谱库进行筛选, 得到候选质谱, 然后采用夹角余弦公式进行未知质谱与候选质谱的匹配度计算。基于强峰的预检索方法为: 将未知物谱图中最强峰与库中标准谱图第一至第八强峰进行比较, 判断是否存在能匹配的质谱峰; 然后将未知物谱图中第二强峰与库中标准谱图第一至第九强峰进行比较, 判断是否存在能匹配的质谱峰; 以此类推, 直到未知物谱图中第八强峰与库中标准谱图第一至第十六强峰进行比较; 如果都有能匹配的质谱峰, 则入选为候选质谱, 否则抛弃。使用夹角余弦进行匹配度计算时, 为减小未知质谱中杂峰带来的影响, 仅采用与候选质谱相同出峰位置的谱峰参与计算。在计算过程中, 质谱库数据首先从主机内存复制到 GPU 端的全局内存 (Global Memory) 之中, 以便每个线程都能访问该数据。随后, 系统会给每一个未知质谱分配一个线程 (Thread) 用于预检测和匹配度计算。CUBLAS 库中用于匹配度计算的 `cublasSdot` 内积函数描述如下。

```
//计算大小为 arrSize 的向量 a、b 的夹角余弦匹配度
float MatchFactor(float * a, float * b, int arrSize) { float * d_a; float * d_b;
//为设备端数据分配内存空间
cudaMalloc((void **) &d_a, sizeof(float) * arrSize);
cudaMalloc((void **) &d_b, sizeof(float) * arrSize);
cudaMemcpy(d_a, a, sizeof(float) * arrSize, cudaMemcpyHostToDevice);
cudaMemcpy(d_b, b, sizeof(float) * arrSize, cudaMemcpyHostToDevice);
cublasCreate(&handle_cos);
float * result = (float *) malloc(sizeof(float));
float * a_result = (float *) malloc(sizeof(float));
float * b_result = (float *) malloc(sizeof(float));
//调用 cublas 库进行内积运算
cublasSdot(handle_cos, channel, d_a, 1, d_b, 1, result);
cublasSdot(handle_cos, channel, d_a, 1, d_a, 1, a_result);
cublasSdot(handle_cos, channel, d_b, 1, d_b, 1, b_result);
cublasDestroy(handle_cos);
```

```
return * result/(sqrt( * b_result) * sqrt( * a_result));
}
```

4 运行测试与分析

表 3 描述了平台开发环境和服务器端的软、硬件情况。实验中所采用的客户端和服务端为同一计算机。其中,CPU 的型号是 i7,主频为 3.4 GHz,内存大小为 8 GB,GPU 的型号是 NVIDIA GTX1060,显卡显存为 6 GB,CUDA 版本是 8.0,CUDA 单元数为 1 280。

表 3 平台开发环境

项目内容	描述信息
操作系统	Windows 10 x86_64
Node.js 版本	8.9.0
Express 版本	4.15.2
Echarts 版本	3.8.5
CPU	Intel® Core™ i7-6700 @3.4 GHz
内存大小	8 GB
GPU 型号	NVIDIA GTX1060
CUDA 单元数	1 280
显存大小	6 GB
CUDA 版本	8.0
浏览器	Chrome 66.0.3359.181

客户端和服务端都采用 Windows 10 操作系统。应用程序的运行需要 Java 环境支撑(JDK1.8 以上),系统数据存储 MS 文件存储结构中,通过 C++开发 MS 文件解析模块,实现 MS 文件解析。

4.1 检测结果

实验采用安捷伦 7890A-5975C 气质谱仪对某烟叶产地的烟丝在 600℃ 下的裂解产物进行检测,并将输出的色质数据输入该平台进行检测,实验中的部分检测结果见表 4。

表 4 检测结果

保留时间/min	CAS 号	600 ℃
2.561	000534-22-5	4.17
2.561	000534-22-5	4.17
2.561	000534-22-5	4.17
5.138	000109-97-7	—
15.199	000100-71-0	—
15.334	000497-23-4	0.44
16.853	001551-16-2	—
25.975	005256-65-5	1.37
29.492	021835-00-7	—
29.573	005682-69-9	0.48
30.119	000823-82-5	0.24

续表 4

保留时间/min	CAS 号	600 ℃
30.134	000090-05-1	1.54
31.166	000496-76-4	—
35.698	054814-64-1	—
37.055	1000098-14-8	1.67
40.299	000488-17-5	—
47.219	000097-54-1	0.51
48.846	000487-19-4	3.34
48.993	000091-55-4	0.27
50.65	004854-85-7	—
53.989	1000314-35-7	0.34

4.2 检测结果分析

实验采用安捷伦 7890A-5975C 气质谱仪对某烟叶产地的烟丝在 600℃ 下的裂解产物进行检测,检测结果如表 4 所示。从表 4 可以看出,该平台可以较清晰地展现出谱峰检测的结果。为了验证平台解释色质数据的准确性,将该平台产生的解析结果与安捷伦 7890A-5975C 气质谱仪产生的最后分析结果进行了比较,比较结果如表 5 所示。其中第一列为文中方法所产生的保留时间;第二列为文中方法产生的解析结果利用 NIST05 检索软件在标准库中的检索结果;第三列为安捷伦气质谱仪所产生的保留时间;第四列是安捷伦气质谱仪所产生的解析结果利用 NIST05 检索软件在标准库中的检索结果。从表 5 给出的实验结果可以看出,文中方法和安捷伦气相色谱仪两种方法产生的解析结果利用 NIST05 检索软件在标准库中的检索结果具有较好的一致性,这也说明该项目组开发的解析平台准确可靠。

表 5 检测结果准确性验证

文中保留时间/min	文中 NIST 检索值	安捷伦保留时间/min	安捷伦 NIST 检索值
2.561	918	2.559	914
2.561	935	2.526	927
2.561	916	2.532	902
5.138	921	5.113	940
15.199	934	15.165	915
15.334	927	15.328	918
16.853	934	16.878	915
25.975	917	25.916	928
29.492	904	29.424	913
29.573	943	29.532	926
30.119	924	30.140	929
30.134	931	30.058	927

续表 5

文中保留时间/min	文中 NIST 检索值	安捷伦保留时间/min	安捷伦 NIST 检索值
31.166	945	31.094	932
35.698	947	35.192	948
37.055	934	37.063	936
40.299	946	40.215	948
47.219	953	47.221	959
48.846	976	48.866	978
48.993	984	48.909	987
50.65	987	50.66	992
53.989	992	53.917	998

5 结束语

色谱-质谱联用仪对实验条件要求较高,分析软件庞大、笨重,且与硬件设备捆绑使用。由于仪器设备对运行环境有较高的要求,必须在专门的实验室环境下运行,不便于移动。另外,实验人员必须在仪器周围,全程参与、监督整个分析过程。即使实验人员将生成的色质数据进行暂时保存,在后期的分析过程中还需要安装庞大、复杂的专用分析软件,而这些专用软件都是与相应硬件设备捆绑销售,利润极高,极大地增加了数据分析成本,在这一系列复杂软件中有部分软件还对使用范围作了限制,这给数据分析应用带来了巨大的不便,严重影响了该领域的发展。其次,仪器分析软件存在多种缺陷,在某些应用场景中无法得到准确的结果,而仪器的封闭性强,更好的解决方案无法载入。分析软件的缺陷和封闭性严重影响了该领域的应用和发展。文中基于 Node.js 的 Express 应用框架,设计开发了一款基于 Web 的色质数据解析平台,实现色质数据可视化、谱峰检测以及质谱检索功能,并采用 GPU 加速技术解决多任务并发的运算速度问题。该平台便捷、高效,部署和使用都十分简单。实验结果表明,该平台在保证解析结果准确性的同时,能有效地提高工作效率。

参考文献:

[1] 刘凤萍,曹晨忠,吴湘江,等. 用分子拓扑指数预测饱和酯

类化合物的气相色谱保留指数[J]. 计算机与应用化学, 2008, 25(6): 711-716.

- [2] 陈 艳. 多氯代二苯并呋喃(PCDFs)在不同色柱上气相色谱保留行为的 QSRR 研究[J]. 计算机与应用化学, 2007, 24(7): 968-972.
- [3] 杨辉华,任洪军,李灵巧,等. 基于 Sector/Sphere 的气相色谱质谱联用多样本并行对齐算法[J]. 计算机应用, 2013, 33(1): 215-218.
- [4] 高丙坤,岳茂兴,李兆云. 基于 C# 的饱和烃气相色谱含水指数的计算[J]. 计算机技术与发展, 2012, 22(1): 205-208.
- [5] 杨 娜,杨 林. 多氯联苯分子空间坐标与气相色谱相对保留时间的 QSPR 研究[J]. 计算机与应用化学, 2012, 29(2): 219-222.
- [6] 金 浩. 新连接性指数与脂肪醇气相色谱保留指数的关系[J]. 计算机与应用化学, 2011, 28(2): 229-233.
- [7] 张 莉. 面向气相色谱分析的温度控制系统研究[J]. 微机计算机信息, 2010, 26(34): 53-54.
- [8] 赵 君. 多种兴奋剂及其代谢产物超高效液相色谱质谱联用检测方法研究[D]. 北京:北京化工大学, 2017.
- [9] 俞 光. 基于主成分分析方法的气相色谱-四极质谱联用仪中纯净谱图提取研究[C]//第二十七届中国(天津)2013IT、网络、信息技术、电子、仪器仪表创新学术会议论文集. 天津:天津市电子学会, 2013: 6.
- [10] 罗荣模,王泳涛. 色谱工作站中谱峰识别的设计与实现[J]. 中国计量学院学报, 2003, 14(1): 24-28.
- [11] 蒋学慧. 色谱-质谱联用仪数据处理关键技术的研究[D]. 天津:天津大学, 2013.
- [12] 刘志广,王雅彬,张永策. 基于 ActionScript 程序设计的气相色谱分析过程模拟与网络虚拟仪器的建立[J]. 计算机与应用化学, 2006, 23(12): 1323-1326.
- [13] 李艳艳,孙多永,朱仲良,等. 基于气相色谱-主成分分析的红木分类识别方法研究[J]. 计算机与应用化学, 2010, 27(2): 237-240.
- [14] 扈 庆. 分析仪器数据格式及质谱检索系统的研究与应用[D]. 长春:吉林大学, 2006.
- [15] 董海波,张红雨,王乃岩,等. 红外谱图图形文件转换方法[J]. 计算机与应用化学, 2000, 17(5): 475-476.
- [16] 祝诗平. 近红外光谱品质检测方法研究[D]. 北京:中国农业大学, 2003.
- [17] 宋 爽. 气相色谱-质谱联用仪的纯净谱图提取与检索算法的研究[D]. 天津:天津大学, 2012.