

# 基于质谱数据的肿瘤诊断 WEB 平台<sup>①</sup>

梁进科, 厉力华, 柯激情

(杭州电子科技大学 生命信息与仪器工程学院, 杭州 310018)

**摘 要:** 随着蛋白质质谱技术以及智能算法的发展, 使用智能算法分析临床质谱数据已在肿瘤早期诊断方面显示了良好的应用前景. 通过研究 MATLAB Builder JA Toolbox 和分析蛋白质质谱数据的智能算法, 结合临床蛋白质质谱数据分析的需求, 使用 JavaEE 技术, 设计和实现了基于质谱数据的肿瘤诊断 web 平台. 平台功能主要包括质谱数据可视化显示、智能检索和智能决策诊断三大功能. 同时还提供病人基本信息及肿瘤数据管理, 质谱生物特征位点设置及分析结果数据存储等附加功能.

**关键词:** 质谱数据分析; 可视化; 智能检索; 智能决策诊断; MATLAB; JavaEE

## Tumor Diagnosis WEB Platform Based on Mass Spectrum Data

LIANG Jin-Ke, LI Li-Hua, KE Ji-Qing

(College of Life Information Science & Instrument Engineering, Hangzhou Dianzi University, Hangzhou 310018, China)

**Abstract:** With the development of proteomics mass spectrometry and intelligent algorithm, using intelligent algorithm to analysis clinical mass spectrum data had showed a good application prospect in the tumor early diagnosis. Based on the study of MATLAB Builder JA Toolbox and intelligent algorithms which are used to analyze protein mass spectrum data, and combing with JavaEE technology, we design and achieve a tumor diagnosis web platform based on protein mass spectrum. The platform mainly achieves three big functions which are mass spectrum data visualization display, intelligent matching and intelligent diagnosis algorithm. While still providing patients' basic information and tumor data management, mass spectrometry biological characteristics site settings and the analysis results data storage such as additional functions.

**Key words:** mass spectrum data analysis; visualization display; intelligent information retrieval; intelligent decision-making diagnosis; MATLAB; JavaEE

## 1 引言

肿瘤是严重影响人类健康和生命的重大疾病, 肿瘤研究已从较早的组织细胞水平发展到分子水平. 疾病蛋白质组学主要研究寻找各种疾病的特异性标志蛋白质, 进而应用于临床诊断和药物开发等, 因此也常称为临床蛋白质组学. 随着蛋白质组学和质谱技术的发展, 临床质谱分析技术已在恶性肿瘤早期诊断等方面显示了良好的应用前景, 同时越来越多的智能算法被用来进行蛋白质质谱数据的分析和处理<sup>[1-4]</sup>. 因此, 如何处理和利用好临床样本的蛋白质质谱数据和基于

其所发现的质谱数据的生物标志物具有重要意义. 为此, 我们开发了一个完整的集病人信息管理与质谱数据智能分析为一体的肿瘤诊疗 Web 平台. 平台主要实现质谱数据可视化显示、质谱数据智能检索和质谱数据智能决策诊断三大功能. 同时还提供病人基本信息和质谱数据管理, 分析结果数据报表下载和肿瘤质谱特征位点设置等附加功能. 平台使医学工作者通过浏览器就可以对蛋白质质谱数据进行智能分析, 使用不同的智能算法为早期肿瘤诊断提供可靠的参考依据. 平台采用 B/S 模式, 用户不需要二次开发可以直接使

① 基金项目: 国家杰出青年科学基金(60788101); 国家自然科学基金(60801054, 60801055); 浙江省自然科学基金(Z2090299)

收稿时间: 2012-08-19; 收到修改稿时间: 2012-09-19

用, 减少对用户硬件的要求. 在服务器端可以快速高效的进行质谱数据的分析.

## 2 平台设计的总体方案

### 2.1 平台的主要功能

#### 2.1.1 蛋白质质谱图显示

不同质荷比的离子经质量分析器分开后, 到检测器被检测并记录下来, 经计算机处理后以质谱图的形式表示出来. 蛋白质质谱数据是高维数据, 其中峰值对病人的诊断具有重要意义. 在平台中我们利用 MATLAB Builder JA Toolbox 提供的 web-figure 实现质谱图的显示, web-figure 具有缩放、移动、3D 旋转功能<sup>[5-8]</sup>. 当进行放大显示时, 平台将会将显示区域存储为静态图, 方便使用者下载查看. 同时平台对具有意义的峰值进行提取.

#### 2.1.2 蛋白质质谱数据智能检索

智能检索使用 KNN 算法, 找到距离病人质谱数据最近的三个已知类型的质谱数据作为检索结果, 将检索结果与病人数据一同通过 web-figure 显示给医生, 同时将这三个检索数据的类别提供给医生. 平台提供基于全局位点和基于生物标志物位点两种检索方式.

#### 2.1.3 蛋白质质谱数据智能决策诊断

智能决策诊断主要使用智能分类算法对病人质谱数据在已知类型的病人质谱数据库中进行二分类并提供病人是结果类的概率. 平台主要提供的智能诊断算法主要稀疏表示分类和 SVM 分类<sup>[3]</sup>. 每种算法提供基于全部位点和基于生物标志物位点的决策诊断.

#### 2.1.4 设置质谱数据的生物标志物

绝大多数的蛋白对于癌症诊断来说是多余、不相关的, 其仅与少部分蛋白有关. 这些与疾病密切相关的少部分蛋白称为生物标志物. 对浙江省肿瘤医院提供的蛋白质质谱数据集, 我们实验室通过一些特征选择算法已经找到了相应的生物标志物<sup>[1]</sup>. 平台中默认使用我们找到的生物标志物, 同时提供给医生设置自己经验获得的质谱生物标志物的权限. 平台提供基于生物标志物的智能检索和智能决策诊断.

### 2.2 平台的技术思路

MATLAB 是由美国 MathWorks 公司发布的主要用于算法开发、数据可视化、数据分析以及数值计算的高级技术计算语言和交互式环境. MATLAB 提供了 Bioinformatics Toolbox, 其提供了质谱数据分析的基本方法. 在我们的研究过程中主要使用 MATLAB 编写

分析质谱数据的算法. MATLAB 提供了 MATLAB Builder JA Toolbox, 其可以将 M 函数编译成 Java 代码, 实现了 Java 对 MATLAB 算法的调用<sup>[7,8]</sup>. 使用 MATLAB Builder JA Toolbox, 我们可以将分析蛋白质质谱数据有关的 M 函数转换为 Java 类的方法, 结合 JavaEE 技术, 可以创建一个基于 WEB 的蛋白质质谱数据可视化和智能分析平台.

### 2.3 平台的体系结构

平台服务器端体系构造采用经典的 MVC 模型: 模型(Model), 视图(View) 和控制(Controller). Model 层实现系统中的业务逻辑, 通常可以用 JavaBean 或 EJB 来实现, 在平台中使用 JavaBean 实现. 平台的 Model 主要是 MATLAB 代码编译所得的 Java 类和与数据库操作相关的类. 平台中使用 MySQL 数据库, 操作数据库使用流行的 ORM 映射框架 Hibernate<sup>[9]</sup>. View 层用于与用户的交互, 通常用 JSP 和 Servlet 来实现, 本平台中 View 层主要功能是实现结果数据和可视化图像的显示及输入参数的获取. 在浏览器的显示使用 Javascript 的 ExtJS 框架可以很容易的实现桌面程序的效果<sup>[10]</sup>. Controller 层是 Model 与 View 之间沟通的桥梁, 平台中使用基本的 Servlet 和 Java 的 Ajax 框架 DWR 实现 Controller 层<sup>[11]</sup>, 其主要功能是从输入页面获得参数传递给 Model 层的 Java 类, 将 Model 层 Java 类的运算结果输出的输出界面进行数据库操作等.

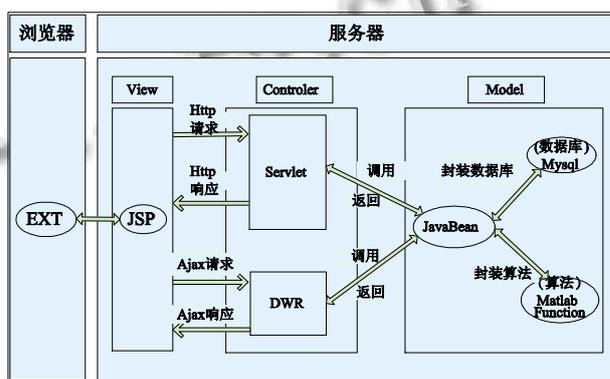


图 1 系统总体结构

## 3 平台的设计和实现

### 3.1 开发环境要求

操作系统: Windows Server 2003

开发语言: MATLAB 7.11.0, Java, Java Servlet Pages

开发工具: Eclipse IDE for Java EE Developers

其他要求: Java Runtime Environment 6 or higher,

Tomcat 6.0.32 or higher, MATLAB 7.9 or higher, MATLAB Toolboxes: MATLAB Compiler 4.11; Bioinformatics Toolbox; MATLAB Builder JA, MySQL5.0

### 3.2 M 函数编译为 Java 类的方法的操作

首先打开 MATLAB, 点击

Start->MATLAB->MATLAB Builder JA. 弹出 Deployment Project 窗口,新建一个 Deployment 工程,工程 Target 选择 Java Package, 存储路径为当前路径,确认后弹出一个新的窗口 Deployment Tool, 点击 Add Class 可以设置类的名称, 然后点击 add Files 添加要编译的 M 函数, 可以添加多个 m 函数, 然后点击菜单 Project->Build 开始编译 M 函数, 编译前点击菜单 Project->Package 可以修改包的名字, 被编译的 M 函数最后生成类的方法, 此方法与 M 函数同名.

编译完成后, 在存储的目录中会生成一个工程文件夹, 文件夹下有两个子文件夹: distrib 和 src, 在 src 和 distrib 下都有 M 函数编译生成的 jar 文件及其相关文件. M 函数编译成类方法的过程如图 2 所示.

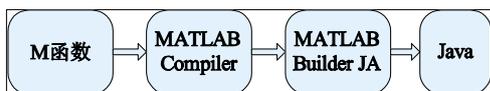


图 2 M 函数编译成 Java 的过程图

### 3.3 平台中使用 MATLAB 函数编译的 Java 类

#### 3.3.1 调用 M 函数编译生成的类方法

在 WEB 应用中使用 M 函数编译生成的 Java 文件的具体步骤如下:

(1) 将 MATLAB 提供的 javabuilder.jar 和 3.2 所编译成的 jar 文件放到 Web 工程的 WEB-INF/lib 下.

(2) 调用 M 函数编译生成的方法, 首先创建生成的类的对象, 使用对象调用与 M 函数同名的方法. 其对应规则如下

M 函数: [wb ByteArrayvalue Ys]  
=PlotMoreMZY(files,left,right,down,up,style)

Java 类的方法:Object [] result

=plotMZY.PlotMoreMZY(3,cellArray,other parameters.....);

plotMZY 为 M 函数编译生成的类创造的对象. PlotMoreMZY 是类的方法与 M 函数同名且对应. 方法的输出 Object[] result 对应 M 函数的多输出, 其次序与 M 函数的输出参数一一对应. 其第一输入参数决定了输出

数组的个数, 其值应大于等于零小于等于 M 函数的输出个数. 其他输入参数一一对应 M 函数的输入参数.

#### 3.3.2 方法输入参数的转变(Java2Matlab)

编译成的 java 类的方法的输入参数可以为 Java 类型的参数也可以为经过 MArray 子类转换成 MATLAB 内部矩阵类型. 当输入为 java 类型是原子数据类型或 String 时函数将其自动转化成 MATLAB 内部矩阵类型. 当输入 Java 类型为 Array, List, Map 等数据类型时需要将其转化成相应 MArray 子类. 如下代码所示为将 List 数据类型转化成 MCellArray 的过程.

```

List<String> files=new ArrayList<String>();
files.add("mspec01.csv");
int numberOfListElements = files.size();
int numberOfListColumns = 1;
MCellArray cellArray =new MCellArray
(numberOfListColumns,numberOfListElements);
Iterator<String> fileIterator=filesList.iterator();
for(int i=1;fileIterator.hasNext();i++){
String fileNameString=(String)fileIterator.next();
cellArray.set(i,fileNameString);
}
  
```

#### 3.3.3 方法输出参数的转换(Matlab2Java)

方法的输出 Object[] result 包含的是 MArray 的子类, 要转化成 Java 常用的类型需要进行如下所示的代码转化操作.

从 Object[] result 中获取 double 数值的示例代码  
MWNumericArray nu= (MWNumericArray)results[2];  
double change=nu.getDouble();

从 Object[] result 中获得 web-figure 的示例代码:

Object webfigure=MWJavaObjectRef

.unwrapJavaObjectRefs(results[0]);

WebFigure wb=(WebFigure)webfigure;

### 3.4 平台中使用 web-figure 显示图像

MATLAB Builder JA 提供了 web-figure 标签类, 可以将服务器生成的图线内嵌入 HTML 网页中从而实现传统 MATLAB Figure 的缩放、移动、功能. web-figure 的缩放功能不能细化坐标系数值的, 显然对于数量庞大的质谱数据显示是不够的, 在这里我们设计了一个导航图, 用户可以从图中选取要显示的局部, 然后经过服务器运算后将局部放大后显示到 web-figure 中. 这样坐标系数值也被放大和缩小.

在 M 函数中首先要使用 `f=figure('Position',[0,0,width,height])` 获得 figure 句柄, 然后开始画图, 画图完毕使用 `wb=webfigure(f)` 将句柄变成 WebFigure 类对象. 使用 M 函数 `returnByteArray=figToImStream('figHandle',f,'imageFormat','png','outputType','int8')` 可以获得此 figure 的静态图片的字节数组. 在 Java 中获得这个字节数组使用图像写入流可以生成此 figure 的静态图像.

要使用 web-figure 需将 MATLAB 提供的 web-figure.tld 放到 WEB-INF 下. 在 web.xml 下添加 WebFigure Servlet 标签.

```
<servlet>
    <servlet-name>WebFigures</servlet-name>
    <servlet-class>com.mathworks.toolbox.javabuilder.
        webfigures.WebFiguresServlet
</servlet-class>
</servlet>
<servlet-mapping>
    <servlet-name>WebFigures</servlet-name>
    <url-pattern>/WebFigures/*</url-pattern>
</servlet-mapping>
```

为了在 JSP 页面中显示 web-figure, 需要 JSP 页面与 Servlet 进行通信. 在 Servlet 中将 web-figure 对象可以被添加到 Session 或 Application 范围中, 但考虑到用户之间的并行性需求, 应将 web-figure 对象添加到与用户对应的 Session 中.

```
request.getSession().setAttribute("Vararg_Figure",wb);
request.getSession().setAttribute("Vararg_Figure_Binde",
new MWHHttpSessionBinder(wb));
```

同时需要在显示 web-figure 的 JSP 页面中引用 web-figure.tld, 具体语句为 `<% @taglib prefix = "wf" uri = "/WEB-INF/web-figure.tld" %>`, 然后在 JSP 中引用标签, 具体语句为 `<wf:web-figure= name="Vararg_Figure" scope = "session" root="web-figure" width="100%" height="100%" />`, 在网页的此位置会显示为 web-figure.

## 4 平台的主要功能和实现效果

### 4.1 病人管理

在图 3 的工具栏中添加、更新、删除按钮为病人信息管理功能.



图 3 病人信息管理界面

### 4.2 质谱数据显示

选择肿瘤类型和病人编号, 点击图 3 中的显示按钮, 平台经过对质谱数据进行降维、基线矫正、谱峰检测与质谱矫正等预处理后, 将其转化成质谱图显示出来. 图 4 中左下图为导航图, 在图中小框为选中的局部区域其被放大显示到 web-figure 中.

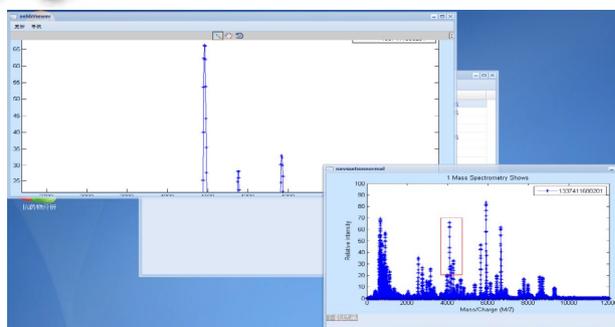


图 4 质谱数据全图显示

### 4.3 智能检索

图 5 中所示的左图其为病人的质谱图, 点击检索一按钮将获得的图 5 中的右图, 其代表数据库中与病人最相近的数据. 图中右上角标识出检索出病人的姓名及其肿瘤良恶. 点检索二、检索三按钮依次类推. 特征检索是在生物标志物上进行检索其使用与结果表达与检索相同.

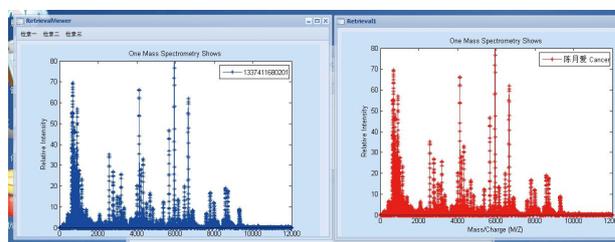


图 5 智能检索显示

### 4.4 智能决策诊断

使用稀疏决策对病人进行诊断, 结果显示算法是否成功运行、预测病人肿瘤良恶及其概率.

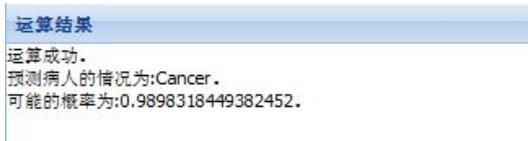


图 6 稀疏表示算法的智能决策诊断结果

#### 4.5 诊断报表的输出

对病人进行算法分析后将产生系统报表. 报表提供了病人的基本信息病人以及质谱数据在不同算法下的分析结果. 报表使用著名的 Java 类库 iText, 格式为 pdf<sup>[12]</sup>.

肿瘤蛋白质质谱数据诊断表				
编号:	1337411680201	姓名:	陈月爱	
性别:	女	年龄:	34	
肿瘤类型:	乳腺癌	算法:	稀疏表示	
预测类型:	Cancer	概率:	0.99	
肿瘤类型:	乳腺癌	算法:	特征峰值检测	
特征位点:	4700.0	6000.0	8330.0	/ /
敏感峰值:	9.0015475	13.277216	18.631419	/ /
肿瘤类型:	乳腺癌	算法:	检索	
病人姓名:	陈月爱	宋银凤	谢仁飞	钱金彩
病人类型:	Cancer	Cancer	Cancer	Cancer
病人姓名:	吴老桂	付美娟	洪秋娟	黄洁卿
病人类型:	Cancer	Cancer	Cancer	Control

图 7 生成的 PDF 报表

## 5 结语

本文设计的基于蛋白质质谱分析的肿瘤诊断 WEB 平台, 使用 MATLAB Java Builder 工具包将智能处理质谱数据的 M 函数编译成 Java 类的方法, 将开发方式统一成面向对象的模式, 同时使用 web-figure 显示质谱图, 完美的将 MATLAB 的强大数值计算和图形处理功能和 Java 在 Web 应用系统开发上的强大功能结合起来. 使用 web-figure 显示图像具有比单一图像结果更好的图像显示, 增强了系统的交互性. 平台涵盖了临床蛋白质质谱数据分析的常用功能, 同时还包含了一些我们自己在长期研究积累过程中提出或改进的算法. 今后, 我们将基于现有的基础, 为平台增加更多高效的智能决策诊断和智能检索的算法, 为临床早期肿瘤诊断提供更有力的算法模型支持.

## 参考文献

- 1 王尧佳, 祝磊, 韩斌, 厉力华, 郑智国, 牟瀚舟. 基于递归零空间线性判别分析算法的蛋白质质谱数据特征选择. 航天医学与医学工程, 2010, 23(5): 324-328.
- 2 Ke JQ, Zhu L, Han B, Dai Q, Wang YJ, Li LH. Sparse representation based feature selection for mass spectrometry data. Bioinformatics and Biomedicine Workshops (BIBMW), 2010 IEEE International Conference. 2010: 57-62.
- 3 柯激情. 基于稀疏表示的蛋白质质谱数据分析[硕士学位论文]. 杭州: 杭州电子科技大学, 2011.
- 4 曹源, 应晓敏, 王娜, 李爱玲, 王恒莎, 张学敏, 李伍举. BioSunMS: 基于质谱数据的疾病诊断与分型的软件系统. 第八届北京生命科学领域学术年会. 2008: 96-101.
- 5 Markiewicz T. Using MATLAB software with Tomcat server and Java platform for remote image analysis in pathology. Diagnostic Pathology, 2011(Suppl1): S18.
- 6 金万根, 高汝维, 王宇红. 利用 MATLAB Java Builder 开发 Web 远程仿真实验系统的新思路. 微计算机信息, 2011, 27(4): 204-206.
- 7 MATLAB Builder JA User's Guide. [2011-10-25]. [http://www.mathworks.cn/help/pdf\\_doc/javabuilder/javabuilder.pdf](http://www.mathworks.cn/help/pdf_doc/javabuilder/javabuilder.pdf).
- 8 MATLAB Application Deployment Web Example Guide. [2011-10-25]. [http://www.mathworks.cn/help/pdf\\_doc/javabuilder/example\\_guide.pdf](http://www.mathworks.cn/help/pdf_doc/javabuilder/example_guide.pdf).
- 9 艾里特, 等. 刘平利译. 精通 Hibernate. 北京: 机械工业出版社, 2009: 25-96.
- 10 徐会生, 康爱媛, 何启伟. 深入浅出 Ext JS. 第二版. 北京: 人民邮电出版社, 2010: 31-122, 341-345.
- 11 Zammetti FW. DWR 实战. 北京: 人民邮电出版社, 2009: 5-105.
- 12 Lowagie B. iText in Action Second Edition. Greenwich, Connecticut: Manning Publications Co., 2010: 105-203.