深圳大学考试答题纸

(以论文、报告等形式考核专用)

二○ 一八 ～二○ 一九 学年度第 二 学期

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 课程编号 |  | | 课序号 | |  | | 课程名称 | | 数学建模 | | 主讲教师 |  | | 评分 |  |
| 学 号 | 2016150182  2016150180  2016150111 | | | 姓名 | | 邓文丰  伍梓豪  蒋超杰 | | 专业年级 | | 计算机与软件学院 2016级 计软5班 | | | | | |
|  | | | | | | | | | | | | | | | |
| 教师评语： | | | | | | | | | | | | | | | |
| 题目： | | 拉曼光谱成分分析问题 | | | | | | | | | | |  | | |

拉曼光谱成分分析问题

**摘要**

本文通过对混合物的拉曼光谱的运算时间和准确度等指标分析论证该拉曼光谱仪用于用于检测的合理性，根据提供的二十种纯净物的信息建立一个纯净物光谱的特征峰库；对混合物光谱的特征峰和建立的特征峰库做峰值匹配为混合物做定性分析；通过线性回归拟合的方法计算出每一种成分对于这种混合物的系数判断成分和含量；通过前面建立的模型估计剩下几种未知物质的组成成分和含量

针对问题一，通过对给定的数据找出一个区间里面的最高峰值作为特征峰的点，对这些特征峰作为依据作为后面的匹配，对二十种纯净物都是这么处理，就可以得到一个所有纯净物的特征峰库。

针对问题二，用纯净物图像对混合物进行匹配，如果对于一个纯净物的所有特征峰波数对应的位置这个混合物都出现了峰值，就认为匹配成功，即这个混合物里面包括这个纯净物。

针对问题三，对于问题二里面的计算出来有可能属于混合物成分的纯净物，对他们光谱和混合物光谱进行线性回归拟合，拟合结果里面系数比较高的部分就认为他们在混合物里面浓度占比比较高。并且可以通过这些成分之间的系数比值可以计算出各组分之间的比值推断出浓度的具体值。

针对问题四，利用上面建立出来的模型对未知成分的混合物进行判断，并分别评价模型的定性分析和定量分析是否可靠。

**关键词**：区间最大值，特征峰匹配，线性回归拟合

一、问题重述

**1 . 1** 引言

拉曼光谱(Raman spectra)，是一种散射光谱。拉曼光谱分析法是基于印度科学家C.V.拉曼(Raman)所发现的拉曼散射效应，对与入射光频率不同的散射光谱进行分析以得到分子振动、转动方面信息，并应用于分子结构研究的一种分析方法。

拉曼光谱仪主要适用于科研院所、高等院校物理和化学实验室、生物及医学领域等光学方面，研究物质成分的判定与确认；还可以应用于刑侦及珠宝行业进行毒品的检测及宝石的鉴定。

该仪器以其结构简单、操作简便、测量快速高效准确，以低波数测量能力著称；采用共焦光路设计以获得更高分辨率，可对样品表面进行um 级的微区检测，也可用此进行显微影像测量。

拉曼光谱技术的优越性：具有快速、简单、可重复、且更重等优越性，要的是无损伤的定性定量分析，它无需样品准备，样品可直接通过光纤探头或者通过玻璃、石英、和光纤测量。此外

1 由于水的拉曼散射很微弱，拉曼光谱是研究水溶液中的生物样品和化学化合物的理想工具。

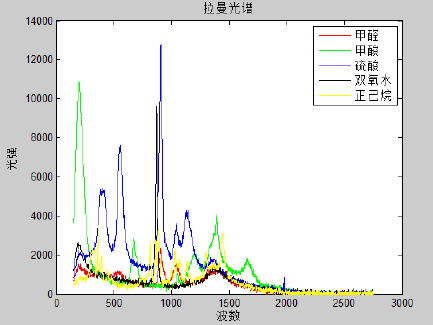
2 拉曼一次可以同时覆盖50-4000 波数的区间，可对有机物及无机物进行分析。

3 拉曼光谱谱峰清晰尖锐，更适合定量研究、数据库搜索、以及运用差异分析进行定性研究。在化学结构分析中，独立的拉曼区间的强度可以和功能集团的数量相关。

4 因为激光束的直径在它的聚焦部位通常只有0.2-2 毫米，常规拉曼光谱只需要少量的样品就可以得到。这是拉曼光谱相对常规红外光谱一个很大的优势。而且，拉曼显微镜物镜可将激光束进一步聚焦至20 微米甚至更小，可分析更小面积的样品。

5 共振拉曼效应可以用来有选择性地增强大生物分子特个发色基团的振动，这些发色基团的拉曼光强能被选择性地增强1000 到10000 倍。

拉曼散射光谱（如下图）可以很典型地来反应一种物品的特性，不同物品的拉曼散射光谱会呈现出不同的特征峰的波数和强度。因此，拉曼散射光谱可以准确地刻画生物样品和化学化合物的特性。



**1 . 2** 问题的提出

现某公司开发了一批拉曼光谱仪，对吡啶、丙酮、二氯甲烷、环己烷、甲醇、甲醛、甲酸、硫酸、三氯甲烷、双氧水、四氢呋喃、硝酸、乙醇、乙腈、乙醚、乙酸乙酯、正己烷、苯巴比妥、硫酸安非他明、盐酸麻黄碱20 种试剂标准品和15 种混合物品（这些混合物品都是由20 种试剂标准品中的2-4 种标准品混合而成的，其中10 种混合物品的成分已知，5 种混合物品的成分未知）进行了测试（拉曼谱图数据详见数据包）。

1. 请构建拉曼光谱分析的模型框架，并通过试验分析验证10 种已知混合物品的成分和含量，以此来验证所构建的模型的准确性。通过模型计算的时间和准确度等指标来论证该拉曼光谱仪用于海关、机场等场所安全检查的可行性。

2. 通过你们的模型判断5 种未知混合物品的组成成分和含量。

**1 . 3** 问题相关数据

* 测试数据在 data 文件夹中
* 代码部分放在 src 文件夹中
* 数据文本都是以如下格式存储，其中列向量 a 表示该样品拉曼散射光谱的波数，列向量 b 该样品拉曼散射光谱的强度（参见拉曼光谱图）；列向量 c, 列向量 d, 列向量 e, 列向量 f 暂无特殊意义。

**1 . 4** 要解决的问题

**二、模型的假设**

* 标准溶液的拉曼光谱具有代表性；
* 混合物之间不会发生相互反应；
* 混合物光谱和纯净物标准光谱在同一条件下测量；

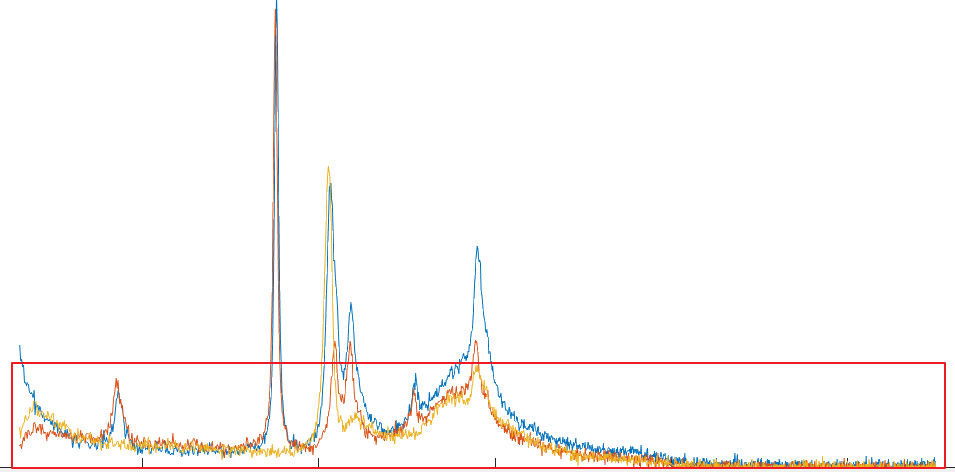
|  |  |  |
| --- | --- | --- |
|  | 三、符号说明 |  |
| 符号 |  | 意义 |
| N1 |  | 试剂的最高值区间范围 |
| Q1 |  | 试剂样本阈值 |
| N2 |  | 混合物的最高值区间范围 |
| Q2 |  | 混合物样本阈值 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

四、问题分析

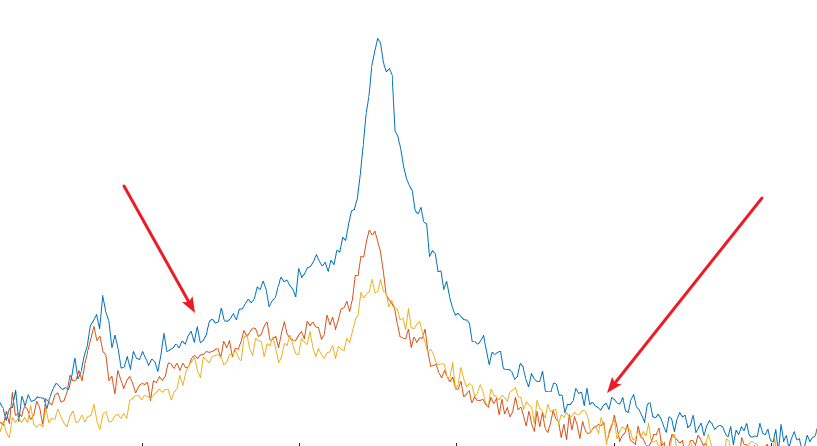
**4 . 1** 问题一分析

观察数据发现混合物和它的成分之间存在特征峰出现位置波数会很接近，所以可以采用峰值匹配的方法对这些混合物进行定量分析。第一个问题里面，我们做的是收集纯净物拉曼光谱的特征值。

为了收集到这些特征峰的信息，我们采用了一种求一个大小一定区间的大小的最大值的方法。



首先需要处理的是上图红色部分对于选取特征峰来说属于干扰信息，首先我们需要确定一个阈值，然后把图像的基准线抬高到这个阈值的位置，可以排除比较多的干扰。



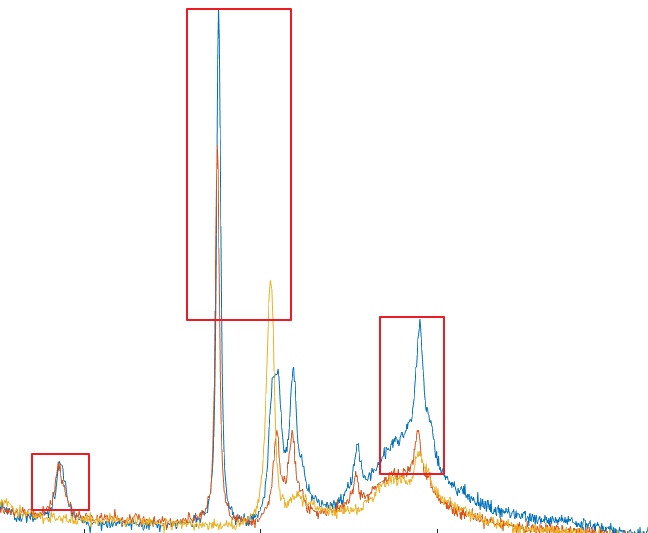
去除光强强度比较底的干扰后，还需要去除上图这种锯齿状的干扰。这里采用一种求区间最大值的方法确定一个N值，遍历图像里面的每一个点如果这个点的前面N个点和后面N个点的光强强度都小于该点的光强强度，说明这个点是区间里面的一个最大值。

我们认为这种最大值就是后面做峰值匹配需要的特征峰数据，需要注意的是在后面的匹配中使用纯净物去匹配混合物，所以在这一步对混合物光谱峰值和纯净物光谱峰值使用不同的筛选参数

|  |  |  |
| --- | --- | --- |
|  | 纯净物 | 混合物 |
| 最低阈值 | 2000 | 1000 |
| 判断区间大小 | 30 | 10 |

**4 . 2** 问题二分析

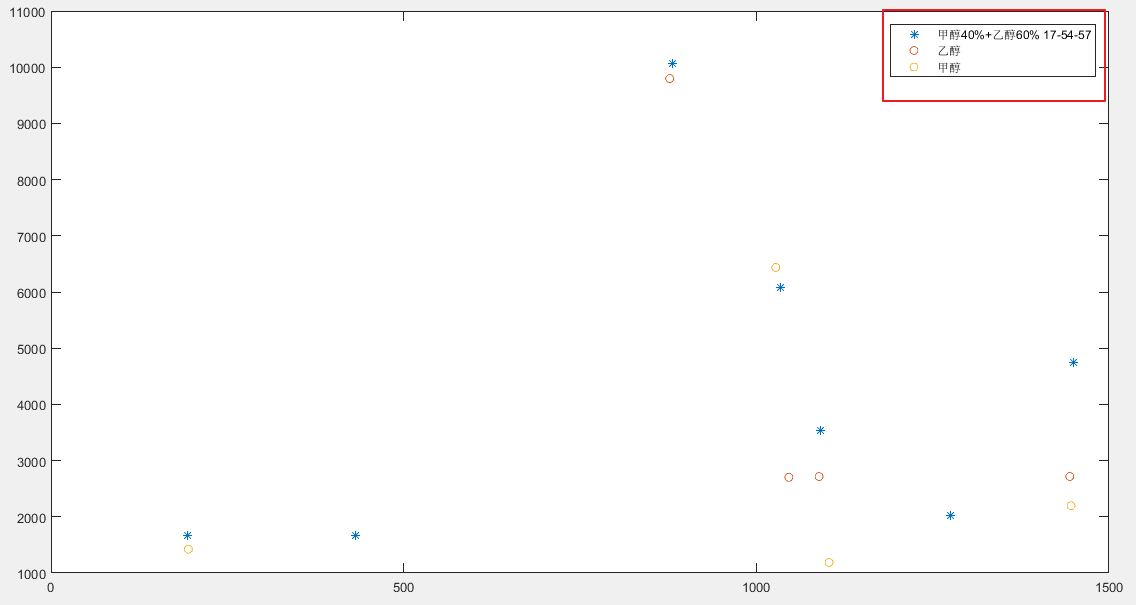
通过问题一计算出来的纯净物特征峰的数据进行匹配，观察发现混合物和里面成分的特征峰会有对应关系。在波数比较接近的位置出现特征峰，如下图乙醇和甲醇的混合物里面，混合物的拉曼光谱会在甲醇或者乙醇出现特征峰值的位置出现特征峰。所以我们利用这个特性对一个混合物的成分进行定性分析：



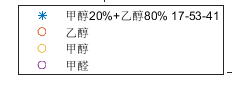
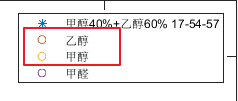
具体的匹配方法是使用问题一中用到的方法找出特征峰的所有峰值，然后使用纯净物的特征峰数据去匹配这个混合物的数据，如果纯净物数据出现特征峰的时候这个混合物也出现了特征峰，就认为匹配成功了。

对于下面这个例子，红色点表示乙醇的特征峰，黄色点表示甲醇的特征峰，蓝色点表示混合物的特征峰，匹配的时候如果混合物的特征峰和纯净物的特征峰波数差的绝对值相差小于15就认为这两个特征峰重叠。

运行这种匹配可以准确匹配出来一个混合物里面的两种成分。



对于这种定性分析方法我们运行十组已经知道结果的数据来评价判断是不是准确，结果是给定已知的十种混合溶液都可以在可能是成分的溶液里面都可以发现真实的溶液成分。这些成分可以作为后面定性分析和定量分析的依据。



**4 . 3** 问题三分析

光谱的强度信号可以认为是多个已知纯净物的叠加效果，所以我们使用线性拟合的方法把上面问题二中筛选出来的纯净物光谱图进行合成成为混合物的光谱信号。对于这种多个多个信号合成的问题，我们这里使用了建立线性回归关系的方法进行拟合。

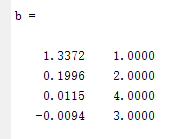
对于求出来的回归系数其中的最大者就是混合物里面含量最大的成分，回归系数的比例就是混合物浓度的比例。回归方程：

**Y = Ki \* Xi (i = 1 , 2 ... n) n是该种混合物有可能的成分个数**

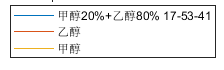
下面开始处理数据：

观察发现混合物数组的长度和纯净物的数组长度不一致，为了顺利回归拟合需要把这两个数组统一到同一个维度。所以使用插值的方法进行处理，处理完之后两个数组在同一个维度。Interpl(x , y , F)插值公式对比较短的Y变量进行插值。

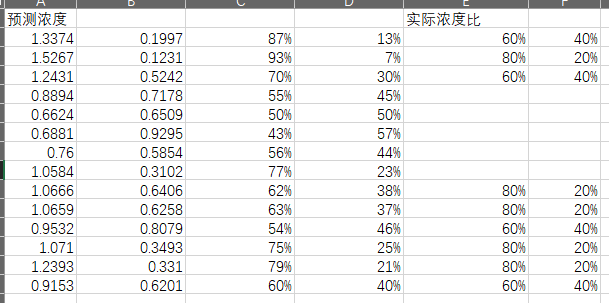
然后使用regress()函数进行线性拟合，通过前面计算出来可能的X拟合出来Y，并返回系数b，通过分析系数b可以的出来成分之间的关系。可以看到每一种成分里面对应的b有不同的值，b越大的说明成分含量越高。所以利用这种关系进行成分判断和含量分析。



可以看到几组分析出来的数据，定性分析的准确度比较高，给定的十种已知成分的混合物里面判断完全正确。



定量分析：通过输出出来的系数比例，我们可以对一直含量的十种物质进行分析，可以看到大部分比较贴近，大的误差在百分十左右。



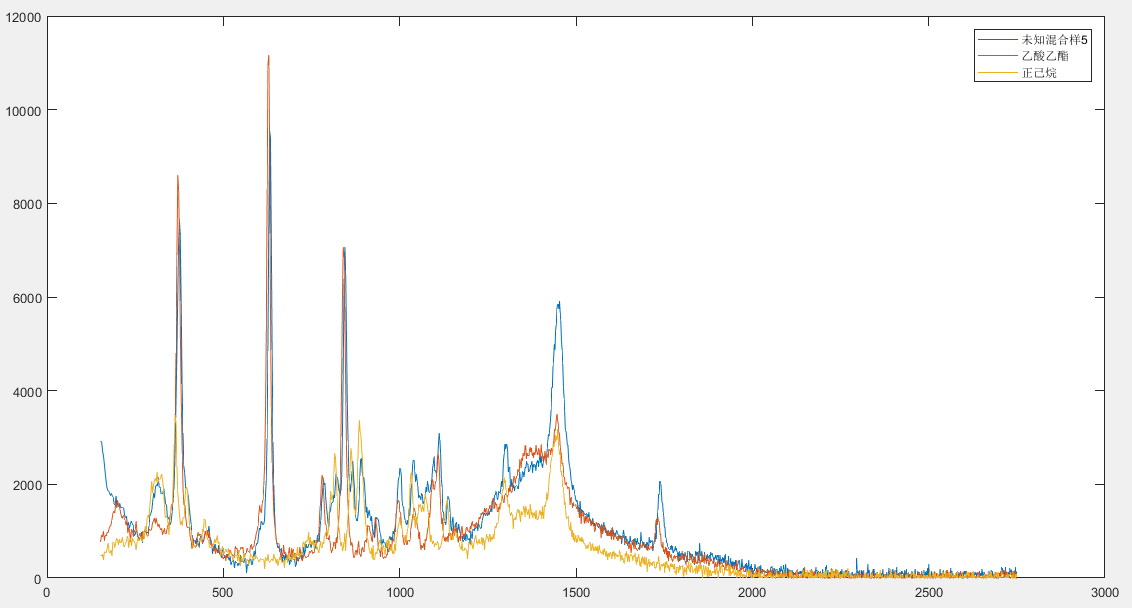
**4 . 4** 问题四分析

使用上面建立的模型对未知的物种混合物进行成分分析和含量分析

**分析结果：**

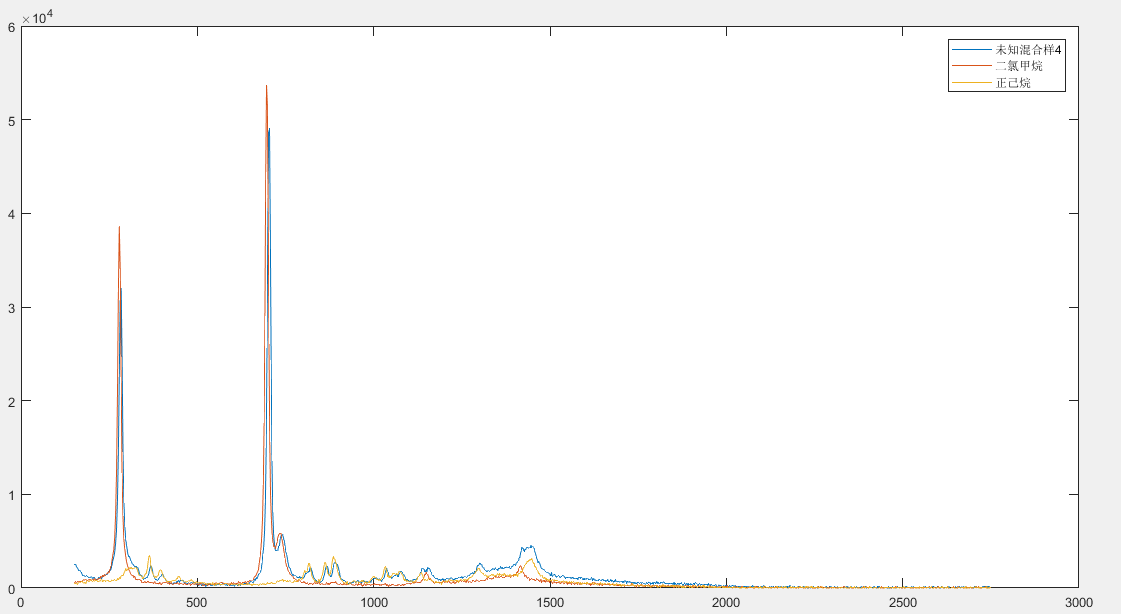
未知混合物一：系数比值为1.05:0.31，所以认为是77%乙酸乙酯和22%正己烷





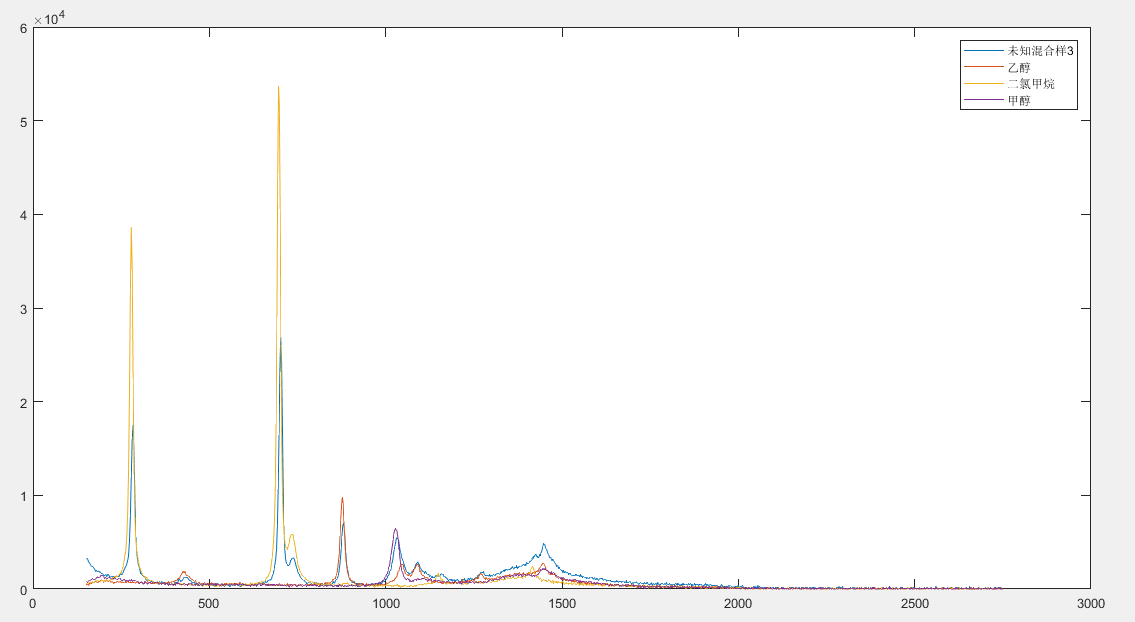
未知混合物二：系数比值为0.76:0.58，所以这是56%的二氯甲烷和44%的正己烷





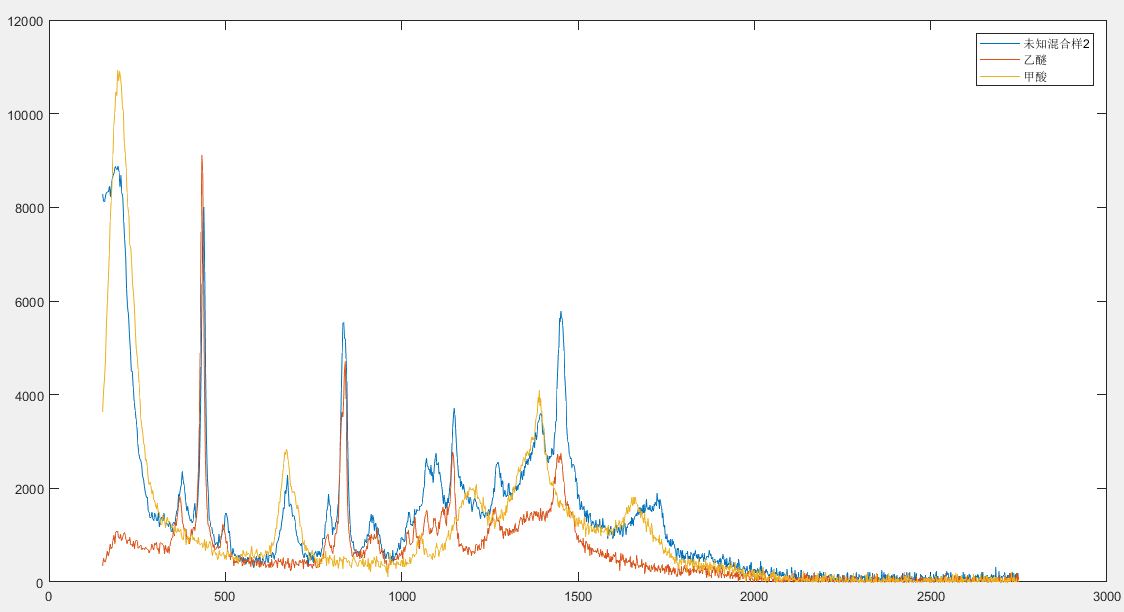
未知混合物三：系数比值为0.68:0.92，所以这是38%乙醇和22%的二氯甲烷和38%的甲醇





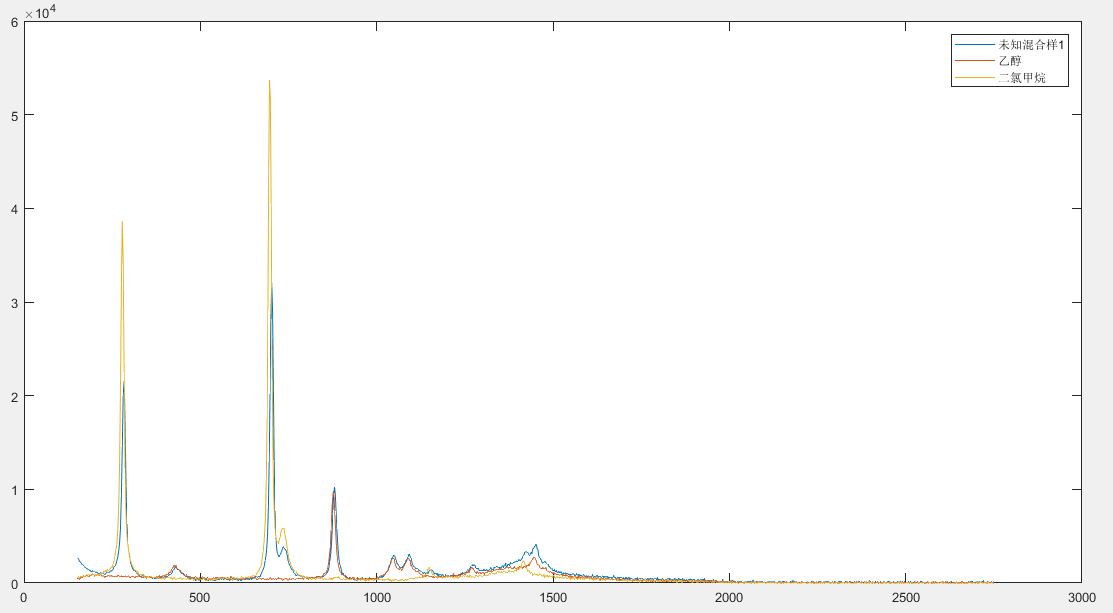
未知混合物四：系数比例为0.66：0.65，所以认为这是一般甲酸和一般乙醚





未知混合物五：系数比例是0.88:0.71，所以这是55%的乙醇和45%的二氯甲烷





五、模型评价

本文建立的这种模型在定性分析方面有很高的准确率，对于成分一直的对象没有出现错误的判断，这也是拉曼光谱的特性，可以明显区分出不同物质之间的特点。定量分析方面由于有部分物质的特征峰并不并不明显已经拟合的方法本身就存在误差，所以会出现百分之十的误差，但是这个误差大小对于问题中提到的海关、机场等场所安全检查是完全可以接受的。

六、参考文献

[1] 作者，全国大学生数学建模竞赛论文模板

附录 **A MATLAB** 源程序

见附件