

ATM 交易状态特征检测与异常分析

2017 年“深圳杯”数学建模夏令营论文

研究背景

ATM 是 Automatic Teller Machine 的缩写，意为自动柜员机。银行业内通常称之为自助机具或自动存取款机，因大部分用于取款又称自动取款机。它是一种高度精密的机电一体化智能装置，利用磁性代码卡或智能卡实现金融交易的自助服务来代替银行柜面人员的工作。作为银行储蓄或信用卡终端设备，持卡人可通过 ATM 使用信用卡或储蓄卡，根据密码办理自动取款、查询余额、转账划拨，还可进行现金存款、存折补登、更改密码、手机充值、支票存款（国内暂无）等业务。

随着现代的自动化程度提高、网络和金融电子化的迅速发展，无人监管的 ATM 机覆盖日趋变广，为众多用户带来了便捷，但也由于 ATM 机庞大的数量和广泛的分布，对 ATM 机的维修、监管也成为了银行的一大难题。为了使 ATM 机能够正常的为顾客提供即时服务，银行需要有合理的手段即时侦测到 ATM 的异常并采取维修措施，防止由于故障影响了原本应有的高效便捷的交易。

ATM 机常见的故障包括 1、分行侧网络传输节点故障，前端交易无法上传请求，导致业务量陡降；2、分行侧参数数据变更或者配置错误，数据中心后端处理失败率增加，影响交易成功率指标；3、数据中心后端处理系统异常（如操作系统 CPU 负荷过大）引起交易处理缓慢，影响交易响应时间指标；4、数据中心后端处理系统应用进程异常，导致交易失败或响应缓慢。

针对当前 ATM 机存在的问题，国内外的研究学者已经进行了很多研究，如用结合 UML 与 Petri 网对自动取款机进行建模^[1]。或者将当前火热的神经网络算法与数据挖掘应用到 ATM 机交易数据的分析中^[2]。云计算系统和自适应的检测方法大大减小了故障出现的概率和检测的开销^[3]。新兴的数学工具如小波变换也为金融交易时间序列中异常现象的检测提供了良好的方法^[4]。

但在当前研究过程中，仍存在很多问题。比如在故障的聚类分析上，过多的继承与运行时的多态使得程序的运行效率很低，聚类的标准局限于圆形和球形的簇使得此方法的应用面比较窄，并不适用于本次建模的数据分类。在故障实时的滤波分析上，有关二阶向量差分方程数字滤波器的稳定性还有待讨论并且实际应用性与容错性问题尚未解决。在云系统故障检测上，客户请求数量波动的特殊考虑与具体异常故障的分析还不够明确。

针对已有的技术和等待解决的问题，在方法和技术不断更新的现在，正迫切需要一套能够有效利用现有工具即时对 ATM 机的异常状态进行报警的方案，以便银行能及时解决存在的问题。

摘要

ATM 机交易故障问题是银行交易过程中的一大重症，严重影响了交易的高效性与便捷性。交易量，交易时间，响应时间，成功率等的交易状态量之间相对独立而又相互关联的错综复杂的关系，又加深了问题处理与剖析的难度。本文旨在对 ATM 机交易故障的问题建立先定性分析，后定量建模的逐步优化，层层深入的多维度综合模型，并提出合理的建议与规划。

针对交易状态特征参数的选择提取分析问题，本文先对题给数据进行简单分析并根据数据及题目应用性特点选择出描述 ATM 状态的参数，且通过图像分析与假设检验得出了几个参数在节假日与非节假日的区别。并针对交易量与时间的实时函数关系进行了拟合求参。本文创新地改用梯度下降法对各交易参数的分布进行拟合，可得出拟合的函数表达式，从而较为精确的得到各具体参数值，通过图像与语言文字的宏观分析，并结合参数的定量细致研究，从而建立了良好的交易状态特征参数体系，并且为后续的异常方案检测提供了好的方向。

针对异常数据分析的问题，本文采用差分滤波分析和小波分析两种方法，建立了两个异常数据分析模型，分别侧重运算的高效性与报警的时效性。差分滤波模型即基于异常数据在频域上具有高频分量的先验条件，对离散数据进行差分滤波，并与原始数据进行比对，得到每一时刻数据对应的异常因子。小波分析模型即对数据同时在时域和变换域展开，通过定量分析，得到每一时刻的异常状态。

针对报警系统建立的问题，本文提出动态阈值模型，以解决交易量大时，异常数据难以表达的问题。从 ATM 机相对稳定这一基础结论出发，根据报警次数与失败量的相关性设计出并计算出合理的阈值函数。对于异常显著的数据，利用异常分析中滤波得到的参考数据构建阈值函数，并从报警系统的稳定性入手，进行求参。本文还根据报警严重性，建立了多级报警模型，能够精准的反应现实中 ATM 机交易的异常状态的同时，还能反应异常的类别。并且本文通过数值实验验证了算法的有效性，模型的精确性，可靠性以及稳定性。

针对由于交易系统异常导致的系统滞后性波及问题，本文提出了基于动态排队论、波及效应以及满意度函数的模型。通过泊松分析的数学推导及数据验证，为后文建立顾客到 ATM 机前交易的动态排队论打下铺垫。并通过动态排队论初步建立了波及问题的概念化的表示，并进一步通过波及效应模型，再结合相关性参数分析，得出具体的量化处理与经验公式，旨在对报警系统的预警性与影响前瞻性作进一步的参考与完善。再结合模型拓展，将波及问题的本质归结于满意度函数的研究与分析，建立了人性化的综合模型，为波及问题的解决提供方向。

关键词： 梯度下降法 滤波器 波及效应 动态排队 动态阈值

目录

研究背景.....	1
摘要.....	2
一、问题重述.....	1
二、问题分析.....	2
三、模型假设.....	3
四、模型的建立与求解.....	4
4.1 模型一：梯度下降法拟合求参模型.....	8
4.1.1 问题分析.....	8
4.1.2 模型建立与求解.....	8
4.1.3 模型结论.....	12
4.2 模型二：假设检验研究节假日与其他日期的参数差异.....	13
4.2.1 问题分析.....	13
4.2.2 模型建立与求解.....	13
4.2.3 模型结论.....	14
4.3 模型三：滤波法异常分析模型.....	15
4.3.1 问题分析.....	15
4.3.2 模型的建立与求解：.....	15
4.3.3 模型检验：.....	22
4.4 模型四：ATM 交易系统故障的波及影响模型：.....	24
4.4.1 问题分析.....	24
4.4.2 模型的建立.....	24
4.4.3 模型的求解.....	28
4.4.5 模型的结论.....	30
4.4.6 模型的评价.....	30
4.5 模型五：报警系统的建立.....	31
4.5.1.1 问题分析.....	31
4.5.1.2 模型的建立与求解.....	31
4.5.1.3 模型的评价.....	32
4.5.2.1 优化模型分析.....	32
4.5.2.2 模型的建立与求解.....	33
4.5.2.3 模型的检验.....	40
4.5.2.4 模型的评价.....	42
五、需要的进一步背景资料和数据.....	43
六、模型总结.....	44
七、模型评价.....	45
八、参考文献.....	46
九、附录.....	47

一、问题重述

某商业银行的 ATM 应用系统包括前端和后端两个部分。前端是部署在银行营业部和各自助服务点的 ATM 机（系统），后端是总行数据中心的处理系统。前端的主要功能是和客户直接交互，采集客户请求信息，然后通过网络传输到后端，再进行数据和账务处理。持卡人从前端设备提交查询或转账或取现等业务请求，到后台处理完毕，并将处理结果返回到前端，通知持卡人业务处理最终状态，我们称这样完整的一个流程为一笔交易。商业银行总行数据中心监控系统为了实时掌握全行的业务状态，每分钟对各分行的交易信息进行汇总统计。汇总信息包括业务量、交易成功率、交易响应时间三个指标，各指标解释如下：

- 1、业务量：每分钟总共发生的交易总笔数；
- 2、交易成功率：每分钟交易成功笔数和业务量的比率；
- 3、交易响应时间：一分钟内每笔交易在后端处理的平均耗时(单位：毫秒)。

交易数据分布存在以下特征：工作日和非工作日的交易量存在差别；一天内，交易量也存在业务低谷时间段和正常业务时间段。当无交易发生时，交易成功率和交易响应时间指标为空。

商业银行总行数据中心监控系统通过对每家分行的汇总统计信息做数据分析，来捕捉整个前端和后端整体应用系统运行情况以及及时发现异常或故障。常见的故障场景包括但不限于如下情形：

- 1、分行侧网络传输节点故障，前端交易无法上送请求，导致业务量陡降；
- 2、分行侧参数数据变更或者配置错误，数据中心后端处理失败率增加，影响交易成功率指标；
- 3、数据中心后端处理系统异常（如操作系统 CPU 负荷过大）引起交易处理缓慢，影响交易响应时间指标；
- 4、数据中心后端处理系统应用进程异常，导致交易失败或响应缓慢。

附件是某商业银行 ATM 应用系统某分行的交易统计数据。假设附件的数据对应的交易状态无异常。你的任务是：

- （1）选择、提取和分析 ATM 交易状态的特征参数；
- （2）设计一套交易状态异常检测方案，在对该交易系统的应用可用性异常情况下能做到及时报警，同时尽量减少虚警误报；
- （3）设想可增加采集的数据。基于扩展数据，你能如何提升任务（1）（2）中你达到的目标？

二、问题分析

ATM 机交易故障问题是银行交易过程中的一大重症，严重影响了交易的高效性与便捷性。ATM 机交易量，交易时间，响应时间，成功率共同反映了交易状态。交易数据特征的提取，异常数据的分析判断，报警阈值的设立，以及异常状态的波及影响对报警系统的建立起到重要的作用。在建立报警系统的过程中，除了要充分利用交易数据的内在信息，也要考虑到 ATM 机交易的现实客观规律，从中得到先验条件，设立更合乎实际的模型，最大可能的减少误报、漏报现象。

交易数据特征的提取，是报警系统建立的根基。需要找到数据与数据之间的相关性，数据变化的周期性，并对数据分布建立合理的数学模型，提取出有意义的数据来，寻找数据的特征规律，为异常数据分析做准备。

异常数据分析的重点在于找到正常的参考数据的并与之作比对，由于数据本身都是实时动态的，所以参考数据也应是实时动态的。我们需要通过算法找到一个能够表达数据异常状态的异常因子，这个异常因子应是接近静态的，且能对异常有着敏感的表达。

报警阈值的建立则是在异常分析的基础上，为异常因子设立一个合理的阈值，这个阈值的设立应满足客观规律，比如报警次数与交易失败量有着较强的相关性等等。在设立阈值时也需要同时考虑到报警次数的合理性以及报警系统的稳定性，尽可能让阈值附近的数据少之又少。

交易系统异常导致的系统滞后性波及问题涉及到动态排队论、波及效应以及满意度问题。需要通过数学推导及数据验证，建立顾客到 ATM 机前交易的动态排队论。做出波及问题的概念化的表示，并进一步通过波及效应模型，再结合相关性参数分析，得出具体的量化处理与经验公式，旨在对报警系统的预警性与影响前瞻性作进一步的参考与完善。再结合模型拓展，将波及问题的本质归结于满意度函数的研究与分析，建立了人性化的综合模型，为波及问题的解决方向。

针对模型扩展与模型优化问题，本文根据已建立的模型和 ATM 的工作原理进行分析，确定要增加的数据和对已建立模型的进一步改进方案，从而得到更符合实际情况、精确度更高的报警系统。

三、模型假设

1. 所有的 ATM 机都是相同的，不考虑设置位点，使用寿命等对 ATM 机交易状态的影响。
2. 在所获得的数据中，每次银行处理故障的行为对参数变化的影响都相同。
3. 网络传输节点故障、数据中心后端处理系统异常等异常情况相互间独立。
4. 暂时不考虑除交易发生时间、响应时间、成功率之外的因素对系统状态是否异常的影响。
5. ATM 机相对稳定。
6. 报警分布与交易失败量有一定的相关性。
7. 不在无交易进行的时段对系统状态进行评价。
8. 交易量较小的时候，数据波动的可能性较大，异常的判断标准应更宽松。

四、模型的建立与求解

首先根据已有数据，分别画出每日交易量随时间变化分布图、每月里每日交易总量分布图、每日成功率随时间变化分布图、每日里响应时间随时间变化图、每个月的响应时间频数分布图等。

如下：

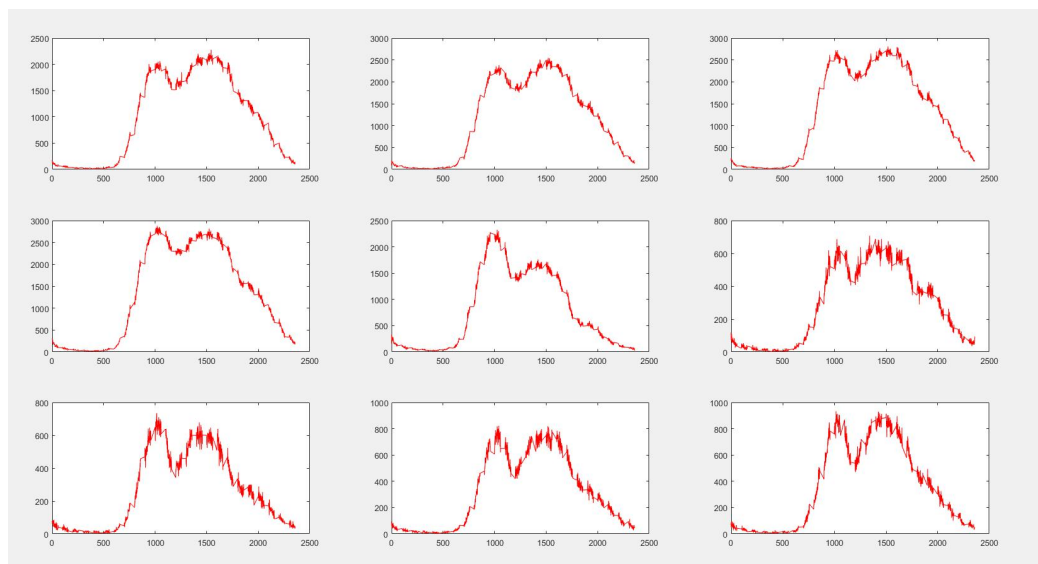


图 1 每日交易量随时间变化分布图---以一月为例

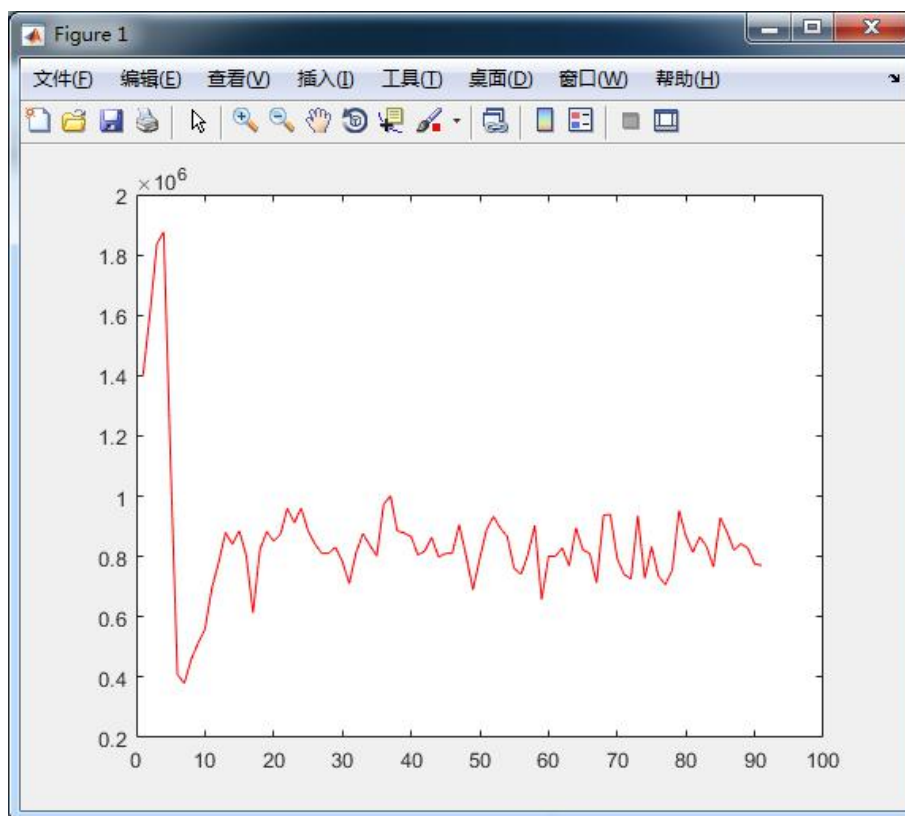


图 2 四个月的每日交易总量分布图

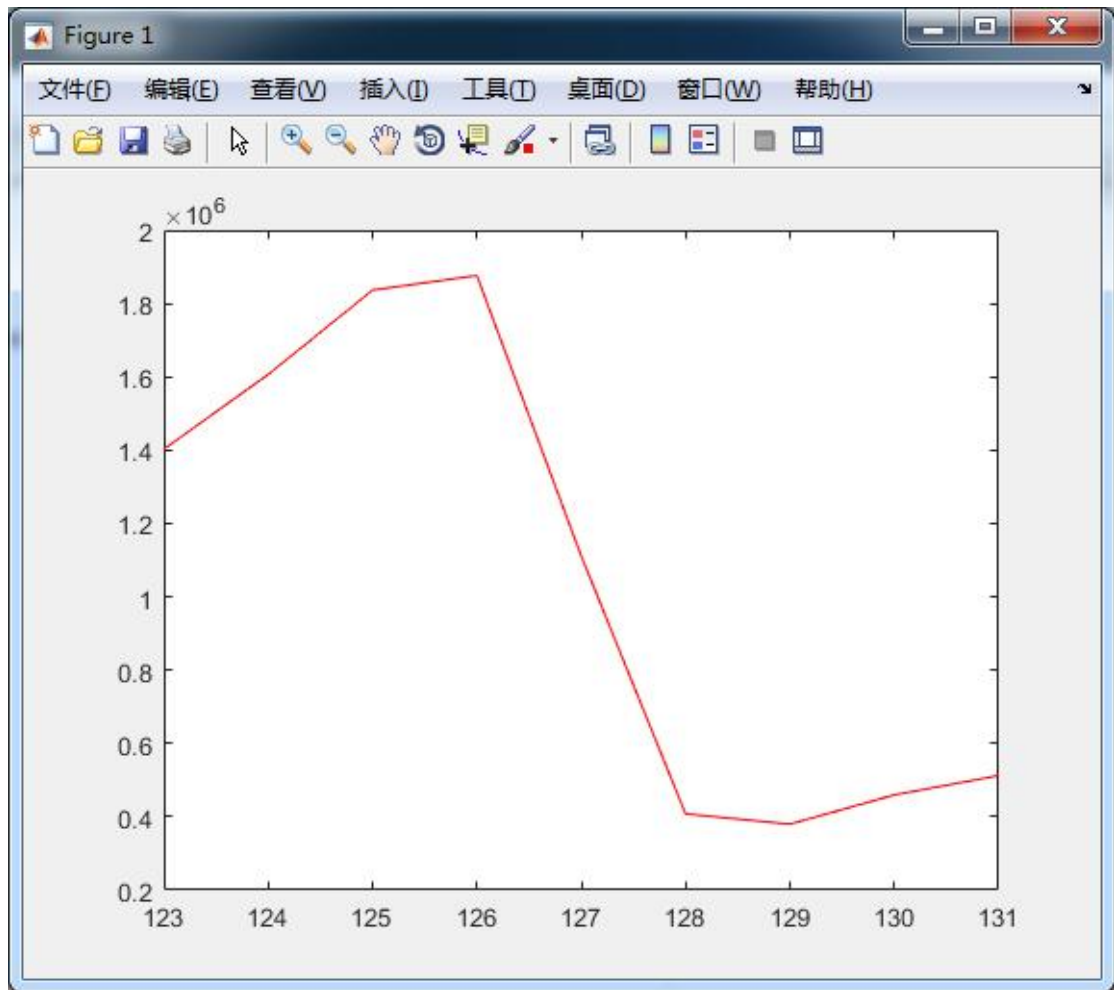


图 3 一月里每日交易总量分布图

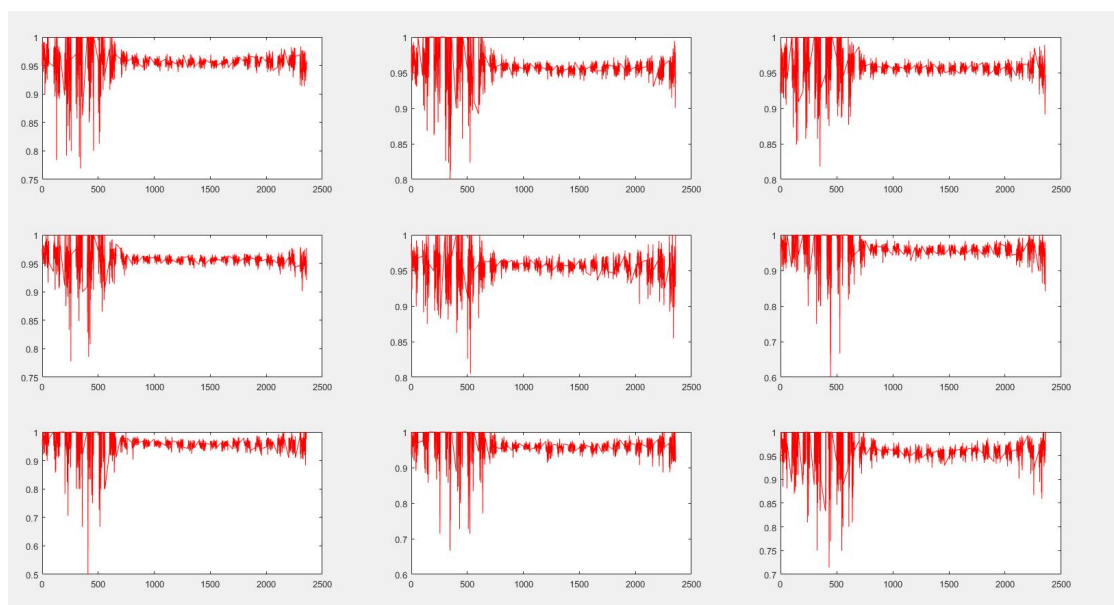


图 4 每日成功率随时间变化分布图---以一月为例

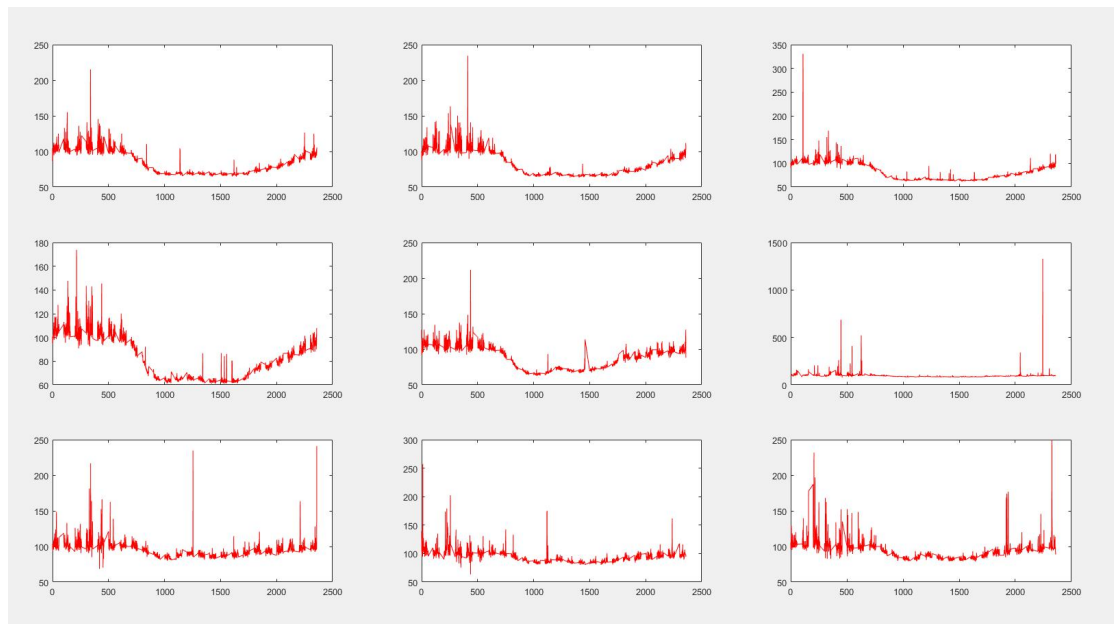


图 5 每日里响应时间随时间变化分布图

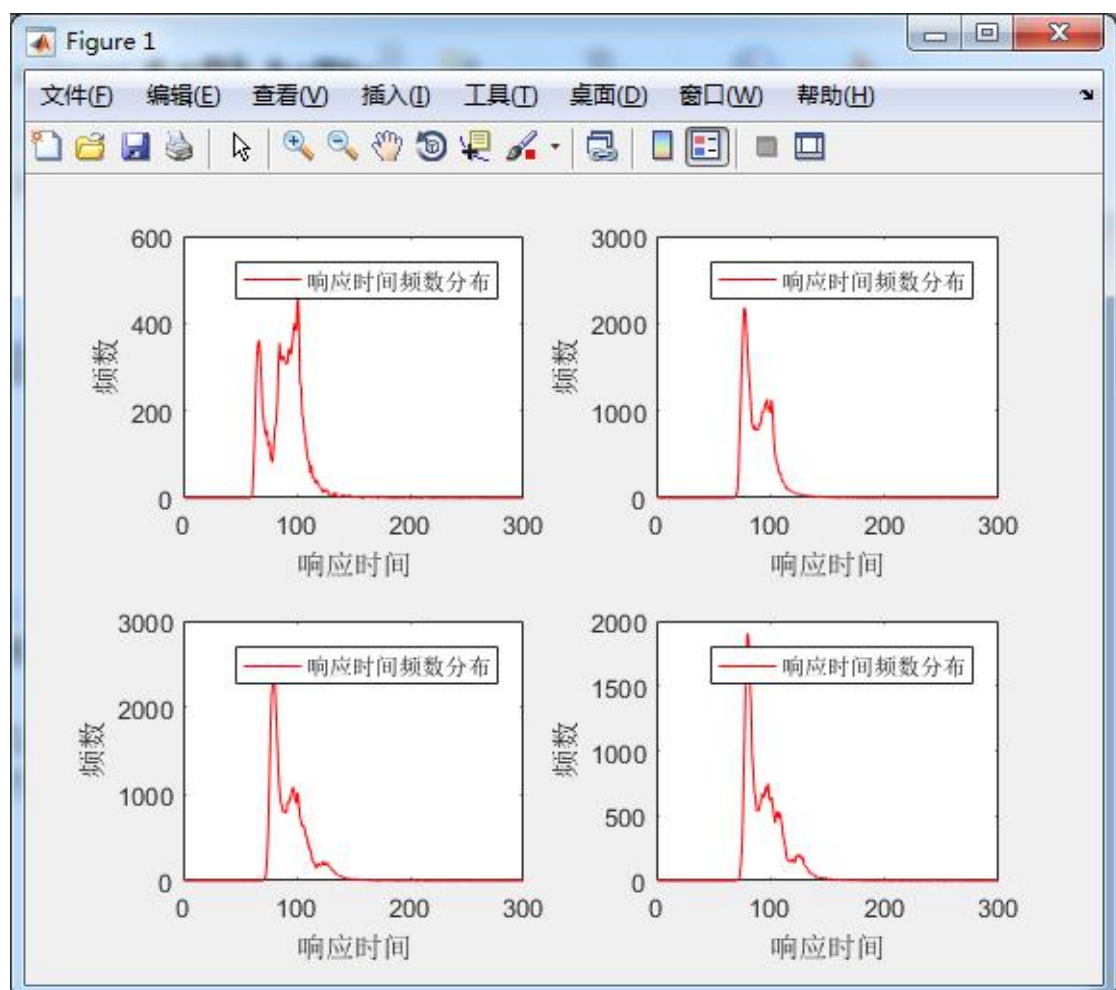


图 6 每个月的响应时间频数分布图

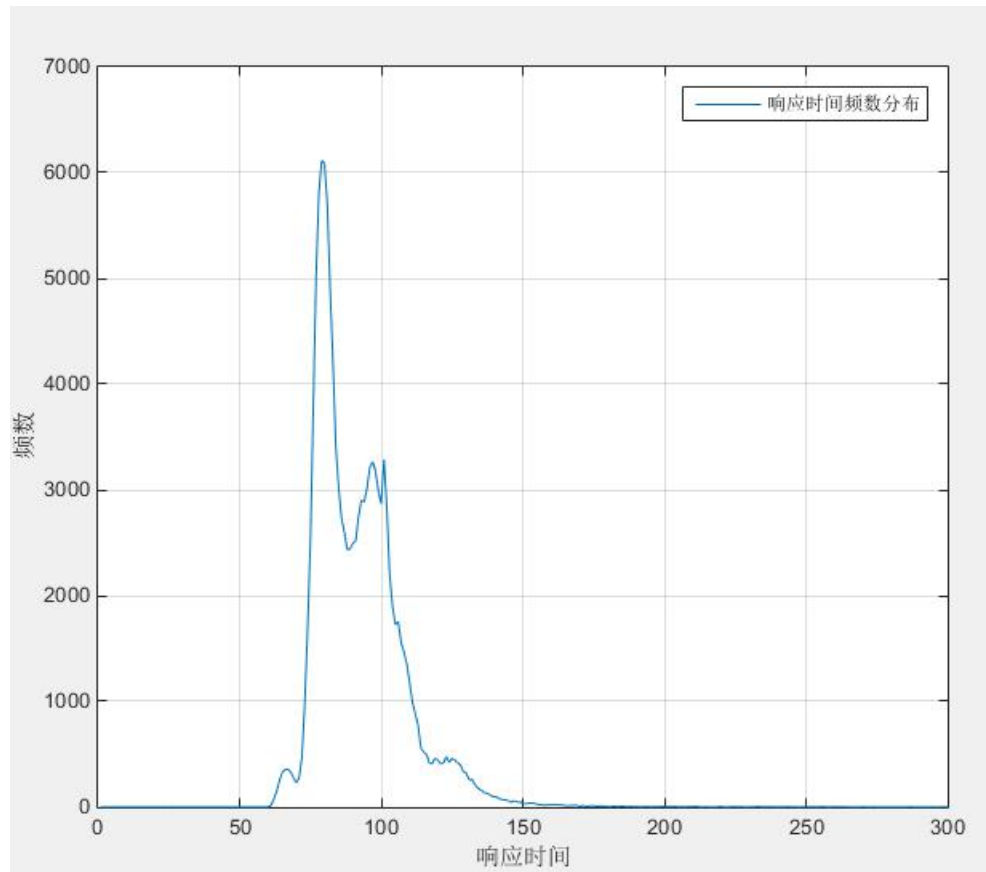


图 7 综合响应时间频数分布

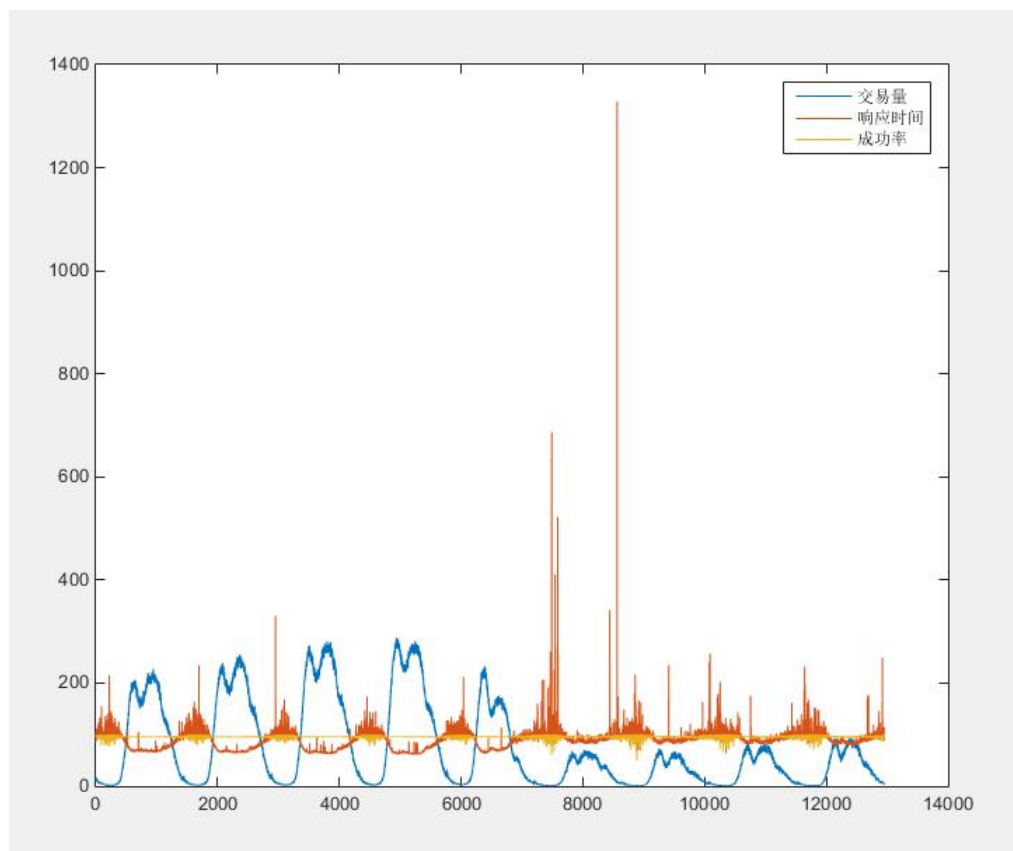


图 8 一月份交易数据，单位已调整

基于已有图形，可得结论有：

- (1) 每日交易量随时间的变化在早晚高峰期满足两类不同的正态分布；
- (2) 交易总量在 1 月 28 号、29 号附近发生显著波动，分析可知，该情况由春节因素引起，春节前交易量较大，而在春节当日及之后几天交易量较低，符合常理；
- (3) 交易量较大时，响应时间较小且稳定，成功率较大且稳定。交易时间小可能是由于 ATM 机交易系统存在固有响应时间，固有响应时间不随交易量的变化而变化，所以当交易量大时，平均固有响应时间会下降，所以响应时间会下降。
- (4) 响应时间频数分布并不是类正太分布，而是一种多峰分布，可见 ATM 机存在多种交易状态。
- (5) 每日成功率在凌晨六点之前波动较大，之后趋于稳定。
- (6) 每日里响应时间的波动非常大，但由响应时间的频数分布图可知：正常的响应时间应该约在 55-130 范围内；

4.1 模型一：梯度下降法拟合求参模型

4.1.1 问题分析

由每日交易量随时间的变化图可知，交易量在早晚各有一个高峰。交易量的主要特征体现在早晚两个高峰中，对交易量分布的分析也需要从这早晚两个高峰入手。考虑到现实意义和模型的可操作性，本文认为交易量在早晚高峰期分别满足满足两种不同的正态分布，而总的交易分布，应是这两个正态分布的合分布。

对于这种多峰分布的复杂分布，在求解参数时，为了使拟合分布与实际数据有着很好的贴合度，本文构建表达拟合分布与实际数据差别的损失函数，并利用梯度下降算法使损失函数达到最小值，从而精确的拟合实际数据。

4.1.2 模型建立与求解

在进行拟合时，需要一个机制去评估拟合的是否比较好，所以说需要对做出的函数进行评估，一般这个函数称为损失函数（loss function）本文记为 J 函数。

设 q 为真实交易量值， h 为拟合估计值

则令

$$J(\theta) = \sum_i [q_{(i)} - h(\theta_i)]^2$$

如何调整 θ 以使得 $J(\theta)$ 取得最小值有很多方法，本文采用梯度下降法 (Gradient Descent)

梯度下降法是按下面的流程进行的：

- 1) 首先对 θ 赋合理的初值。
- 2) 改变 θ 的值，使得 $J(\theta)$ 按梯度下降的方向进行减少。

为了更清楚，给出下面的图：

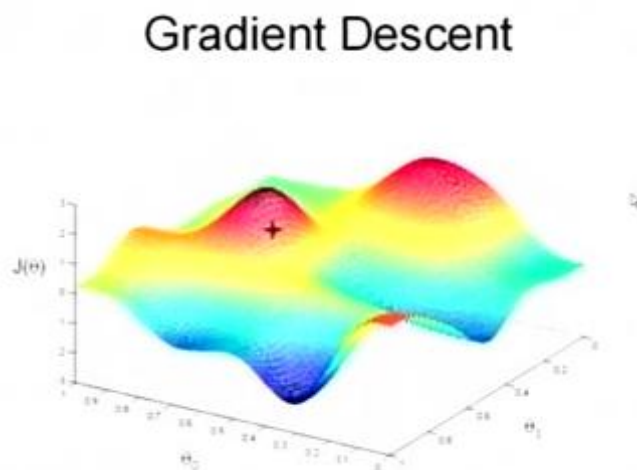


图 9

这是一个表示参数 θ 与误差函数 $J(\theta)$ 的关系图，红色的部分是表示 $J(\theta)$ 有着比较高的取值，需要的是能够让 $J(\theta)$ 的值尽量低。也就是深蓝色的部分。

在上面提到梯度下降法的第一步是给 θ 给一个初值，假设随机给的初值是在图上的十字点。

然后将 θ 按照梯度下降的方向进行调整，就会使得 $J(\theta)$ 往更低的方向进行变化，如图所示，算法的结束将是在 θ 下降到无法继续下降为止。

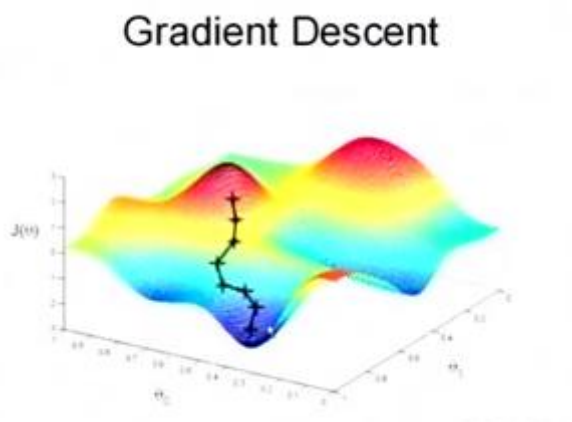


图 10

简单来说，梯度下降法的流程如下：

```

for i=1:n
     $\theta_{i+1} = \theta_i - \alpha \cdot \text{grad } J(\theta_i)$ 
end

```

模型求解：

因为每日交易量随时间变化的分布为一个双峰的正态分布，所以本文设拟合函数

$$h = Ae^{-\frac{(t-\mu_1)^2}{\sigma_1^2}} + Be^{-\frac{(t-\mu_2)^2}{\sigma_2^2}}$$

误差函数

$$J_{(\theta)} = \sum_i [q_{(i)} - h_{(\theta)}]^2$$

$$\theta = [A \ B \ \mu_1 \ \mu_2 \ \sigma_1 \ \sigma_2]$$

式中：

A -----每日第一个交易高峰的峰值

B -----每日第二个交易高峰的峰值

μ_1 -----第一个正态分布的期望

μ_2 -----第二个正态分布的期望

σ_1^2 -----第一个正态分布的方差

σ_2^2 -----第二个正态分布的方差

下面将误差函数对各参数进行求导：

$$\frac{\partial J}{\partial A} = \sum_i -2[q_{(i)} - h_{(\theta)}]e^{-\frac{(t-\mu_1)^2}{\sigma_1^2}}$$

$$\frac{\partial J}{\partial B} = \sum_i -2[q_{(i)} - h_{(\theta)}]e^{-\frac{(t-\mu_2)^2}{\sigma_2^2}}$$

$$\frac{\partial J}{\partial \mu_1} = \sum_i 4[q_{(i)} - h_{(\theta)}] A e^{-\frac{(t-\mu_1)^2}{\sigma_1^2}} \frac{(t-\mu_1)}{\sigma_1^2}$$

$$\frac{\partial J}{\partial \mu_2} = \sum_i 4[q_{(i)} - h_{(\theta)}] B e^{-\frac{(t-\mu_2)^2}{\sigma_2^2}} \frac{(t-\mu_2)}{\sigma_2^2}$$

$$\frac{\partial J}{\partial \sigma_1^2} = \sum_i 4[q_{(i)} - h_{(\theta)}] A e^{-\frac{(t-\mu_1)^2}{\sigma_1^2}} \frac{(t-\mu_1)^2}{\sigma_1^4}$$

$$\frac{\partial J}{\partial \sigma_2^2} = \sum_i 4[q_{(i)} - h_{(\theta)}] B e^{-\frac{(t-\mu_2)^2}{\sigma_2^2}} \frac{(t-\mu_2)^2}{\sigma_2^4}$$

因此得到 $grad J(\theta)$ 即可带入进行求解。

4.1.3 模型结论

在进行拟合求解之后，本文得到如下结果：

以 1 月 25 号为例：

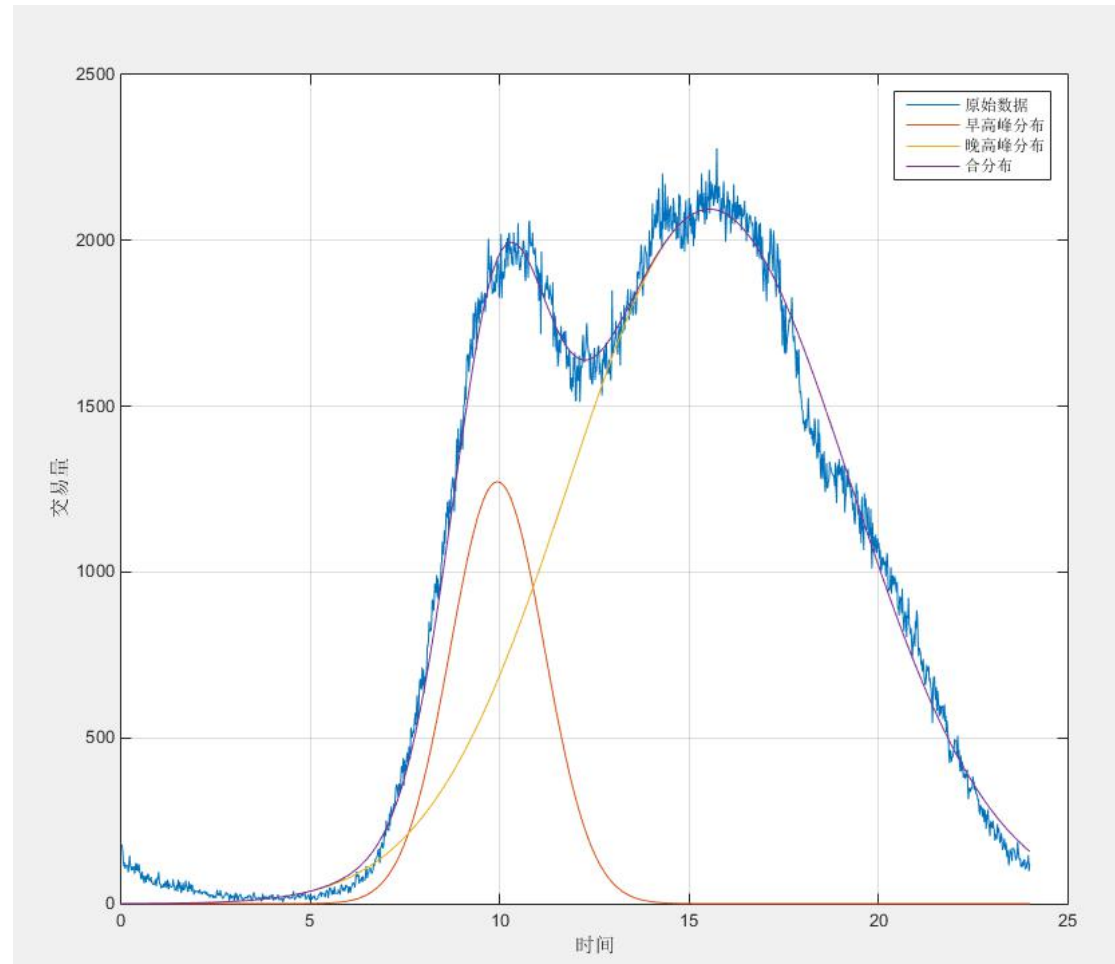


图 11 1 月 25 号交易量变化

表 1 1 月 25 号交易量分布拟合参数

每日交易量两个高峰时刻	t_1	9:58	t_2	15:33
每日交易早晚高峰的交易量	A	298345.283	B	1555071.917
每日交易早晚高峰的交易时间方差	σ_1^2	5652.732	σ_2^2	55252.175
每日交易量总和	m	1836442		
每日交易高峰时间段成功率的期望	E	95.73%		

表 2 1 月 28 号（春节）交易量分布拟合参数

每日交易量两个高峰时刻	t_1	10:14	t_2	15:13
每日交易早晚高峰的交易量	A	45058.736	B	363438.733
每日交易早晚高峰的交易时间方差	σ_1^2	3479.648	σ_2^2	56121.431
每日交易量总和	m	459496		
每日交易高峰时间段成功率的期望	E	96.05%		

可见受春节因素影响，日交易量受影响较大，但每日交易量的高峰时刻基本不受影响。

4.2 模型二：假设检验研究节假日与其他日期的参数差异

4.2.1 问题分析

由题目中的数据可直观得出 1 月 23 日到 27 日的交易量由于春节的影响而较高，二月则没有特殊节日，交易量较稳定。题目中也指出假日与平时的交易状况存在不同，会影响 ATM 机交易的相关参数。由于节假日期间和普通工作日期间人们的日常活动大不相同，因此会出现交易次数的快速升高和下降，影响已有的交易量、成功率和响应时间这些参数。现检验我们对节假日期间与普通工作日的交易量、成功率、响应时间存在差别的猜想。

4.2.2 模型建立与求解

样本容量足够大的时候，各个样本的期望值之间的大小关系可以反应对应参数的差别和大小关系。现在分别检验春节前、春节期间与工作日的数据指标关系。由于两总体方差未知，可用 T 假设对总体期望假设验证，对于本例由于样本数量较多，T 统计量趋于 U 统计量，因此采用 U 假设检验。

欲检验假设 $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 \neq \mu_2$

$$U = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

式中：

\bar{X} -----样本一的均值

\bar{Y} -----样本二的均值

S_1^2 -----样本 1 的方差

S_2^2 -----样本 2 的方差

n_1 -----样本 1 的容量

n_2 -----样本 2 的容量

表 3 春节前、春节对应时刻统计量 U 统计量的数值

交易量	春节前	45.4097
	春节	14.7667
成功率	春节前	-1.4741
	春节	-0.5685
响应时间	春节前	-12.5395
	春节	-1.3583

选取 $\alpha=0.05$ ，查表可得 $u_{1-\frac{\alpha}{2}}=1.96$ 。当 $|u|>u_{1-\frac{\alpha}{2}}$ 时即可认为落在拒绝域内，拒绝原假设 H_0 。

4.2.3 模型结论

由检验结果可知：

1. 对于交易量，U 值远远大于 $u_{1-\frac{\alpha}{2}}$ ，在节假日如春节期间与平时的工作日存在显著差异
2. 对于成功率，节假日与工作日的期望可以接受原假设，认为不存在显著差异。
3. 对于响应时间，春节前的响应时间与工作日的响应时间差异较大而春节期间的响应时间与工作日的响应时间差异较小。

4.3 模型三：滤波法异常分析模型

4.3.1 问题分析

通过之前对交易数据的分析，可以发现交易数据存在着波动性变化，而异常数据实则是在原有的波动数据上的一个冲激。在对数据进行异常分析时，要充分考虑到数据的动态性。利用冲激在频谱上有很多的高频成分，而数据本身波动性的频谱成分主要集中在低频部分的先验条件，建立数字低通滤波器，滤出掉数据的高频成分，即可得到保留数据原有波动性变化且不含冲激的参考数据。所得的参考数据代表着每时刻交易数据的应用的正常状态。拿原始交易数据和参考数据进行比对即可得到异常数据的时刻。

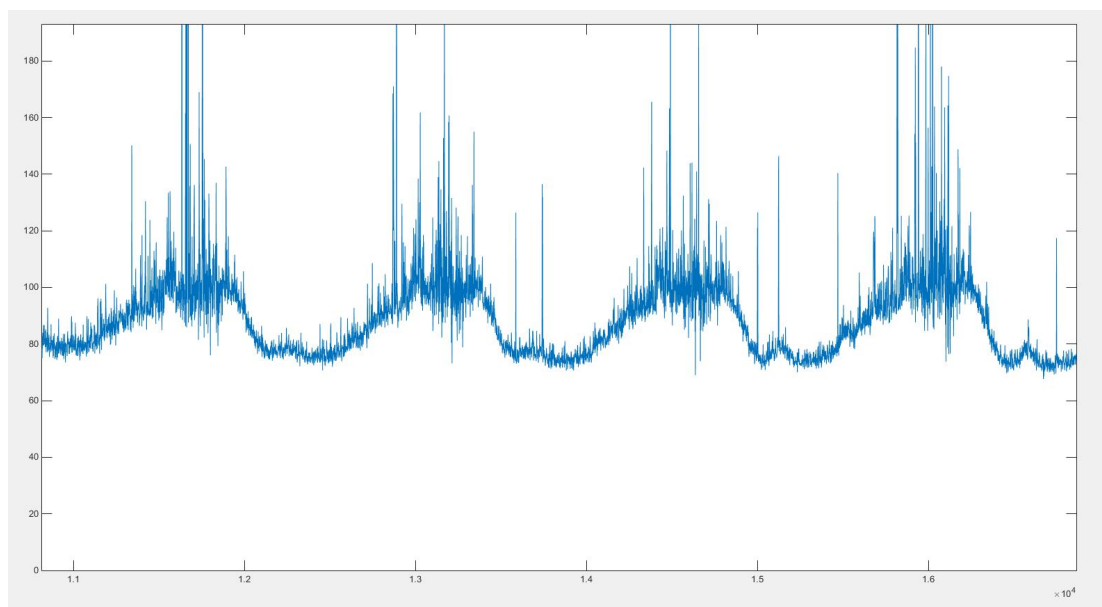


图 12 某时间段响应时间的波动性

除了运用数字低通滤波器对数据进行异常分析外，也可以运用小波变换将交易数据直接在时域和变换域同时展开。小波变化中的变换域对冲激十分敏感，并能同时知道冲激发生的时刻，根据这个特征也可以找到异常数据。

4.3.2 模型的建立与求解：

数字滤波分析：

一位数字滤波器的时域表达式为：

$$\sum_k a_k y(n-k) = \sum_r b_r x(n-r)$$

Z 域表达式:

$$H(z) = \frac{Z[\sum_k a_k y(n-k)]}{Z[\sum_r b_r x(n-r)]} = \frac{\sum_k a_k z^{-k}}{\sum_r b_r z^{-r}}$$

考虑到滤波器的计算简便性和低通性，这里选择常用低通数字滤波器：

$$y(k) = Qx(k) + (1-Q)y(k-1)$$

其中 Q 为滤波系数，取值为 0 到 1，Q 值越大通频带越大，滤波效果越差，但数据变化率大时不会产生惰性失真，Q 值越小低通性越好，滤波效果也就越好，但容易产生惰性失真。

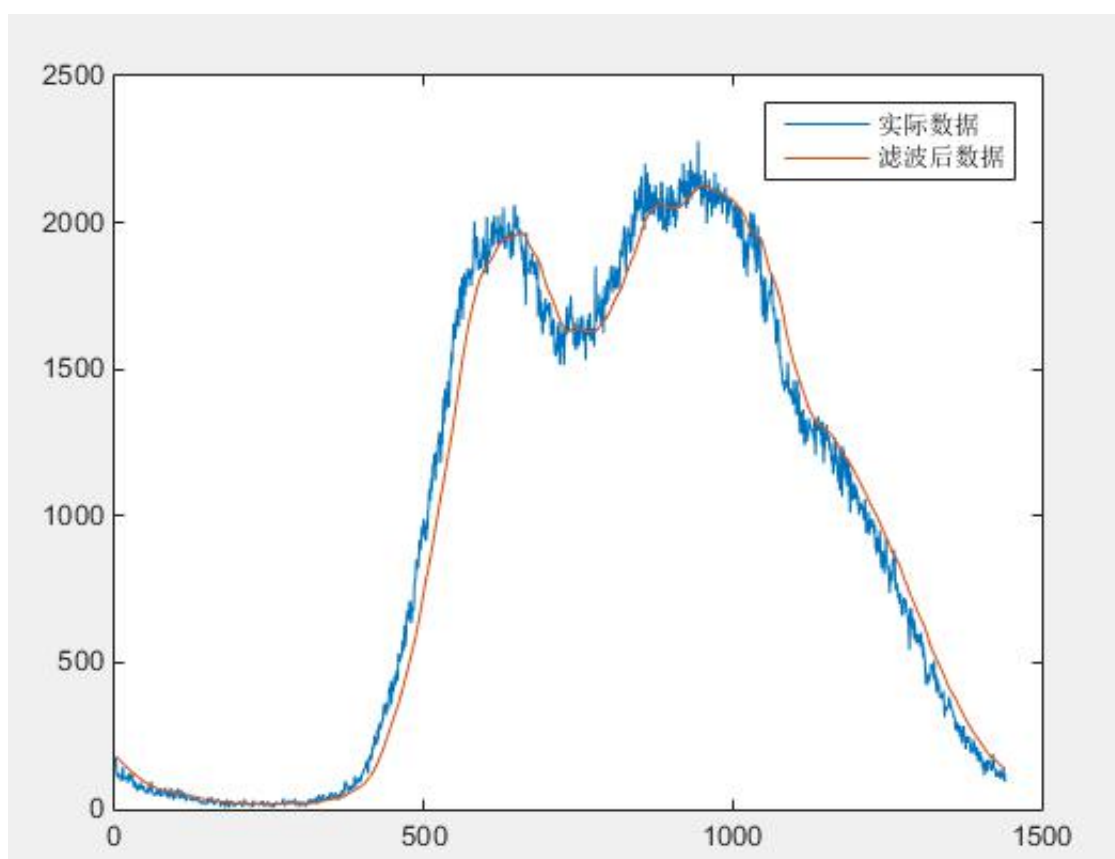


图 13 Q 值过小时产生的惰性失真

针对响应时间滤波，经过多次实践本文取 $Q=0.2$ ，既可以保证滤波效果，又不会导致惰性失真。其具体滤波模型时域表达式:

$$rc(k) = 0.2r(k) + 0.8rc(k-1)$$

其中:

rc -----响应时间的参考数据

Z 域表达式:

$$H(z) = \frac{0.2}{1 - 0.8z^{-1}}$$

令 $z = e^{-j\omega}$ 则得到频域表达式：

$$H(\omega) = \frac{0.2}{1 - 0.8e^{j\omega}}$$

一个周期的频谱图如下：

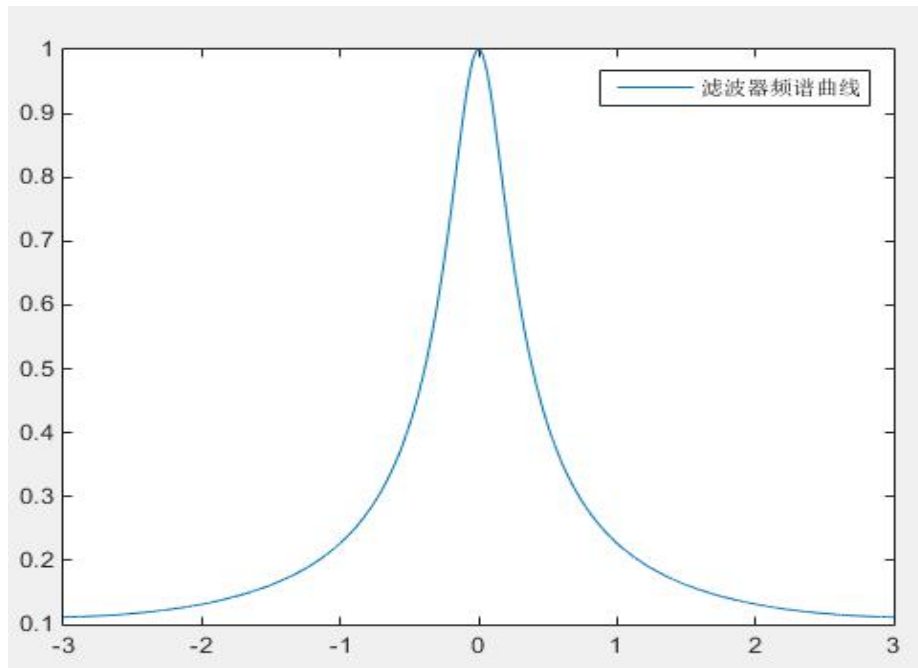


图 14

可见该滤波器具有低通功能

将交易响应时间通过该滤波器即可得到滤波后的参考数据：

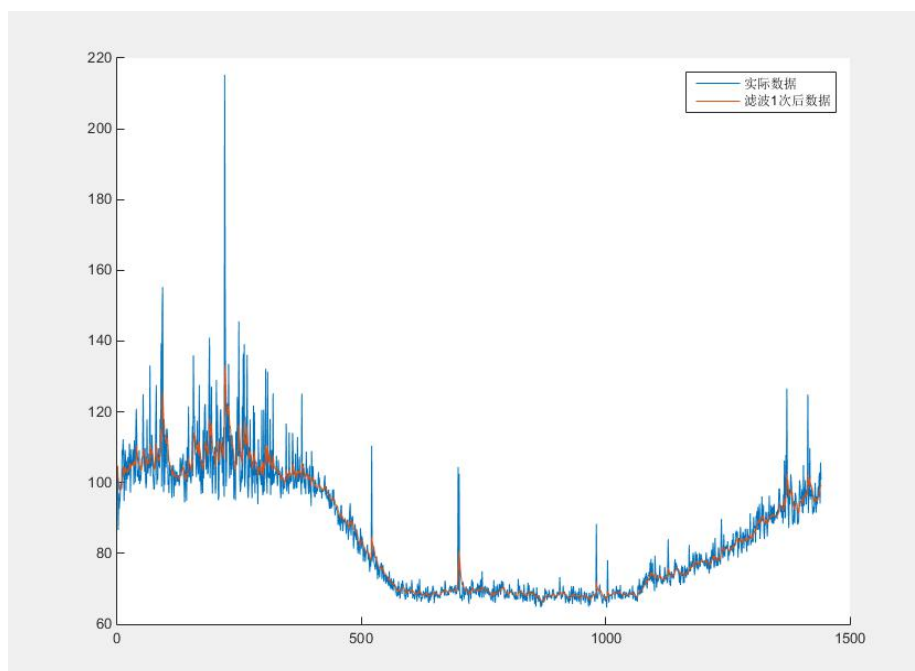


图 15 1 月 23 日响应时间滤波一次

由图可以看出滤波后在保留原有数据动态特这的情况下一定程度的减少了异常冲激，但幅度仍然不够，如果通过减小滤波系数 Q 来增大对滤波效果的话又会造成惰性失真，所以本文采用对此滤波的方法来提高滤波效果，即将滤波一次后的参考数据再次通过滤波器得到效果更好的参考数据，这样一来增强了滤波效果，二来不会产生惰性失真，因为滤波一次后的参考数据的动态变化率不会比原始数据的大，故不可能发生惰性失真。

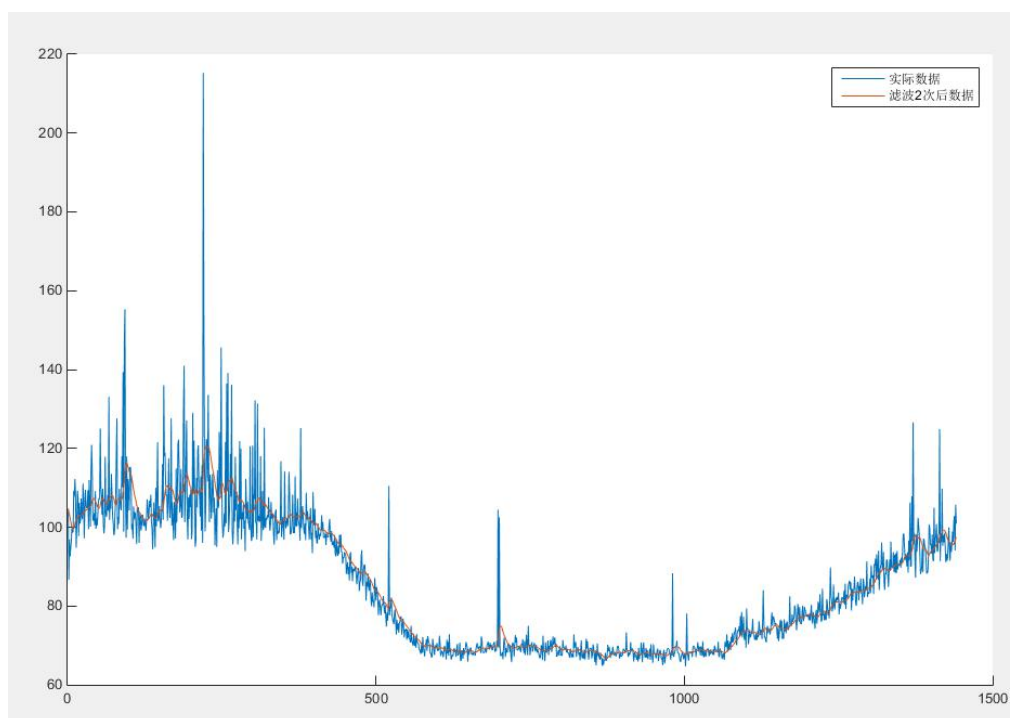


图 16 1 月 23 日响应时间滤波两次

可见滤波两次后，得到的参考数据较为理想。

得到参考数据之后将实际数据与参考数据想出就可以排除原油数据的动态因素（静态化），得到每一时刻数据的异常状态，我们定义这个值为响应时间异常因子，符号为 α_{tr} 。

$$\alpha_r = \frac{r}{rc}$$

其中：

α_r -----响应时间异常因子

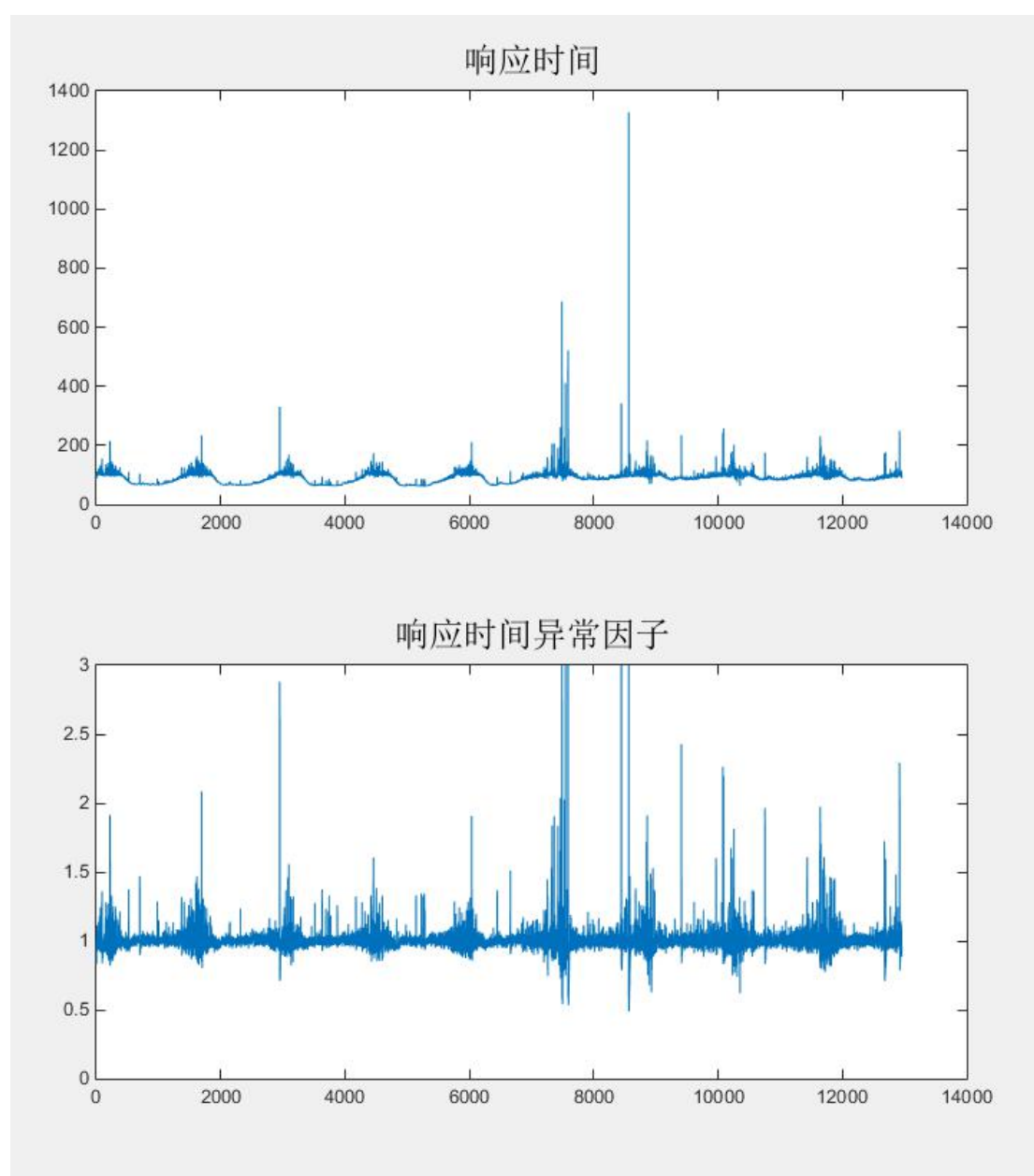


图 17 1 月份响应时间和响应时间异常因子

除了可以对响应时间运用数字滤波法提取异常因子外，对失败量（f），终端总响应时间（tr）都可提取相应的失败量异常因子 α_f 、终端总响应时间异常因子 α_{tr} 。

下图则是一月份终端总响应时间和终端总响应时间异常因子：

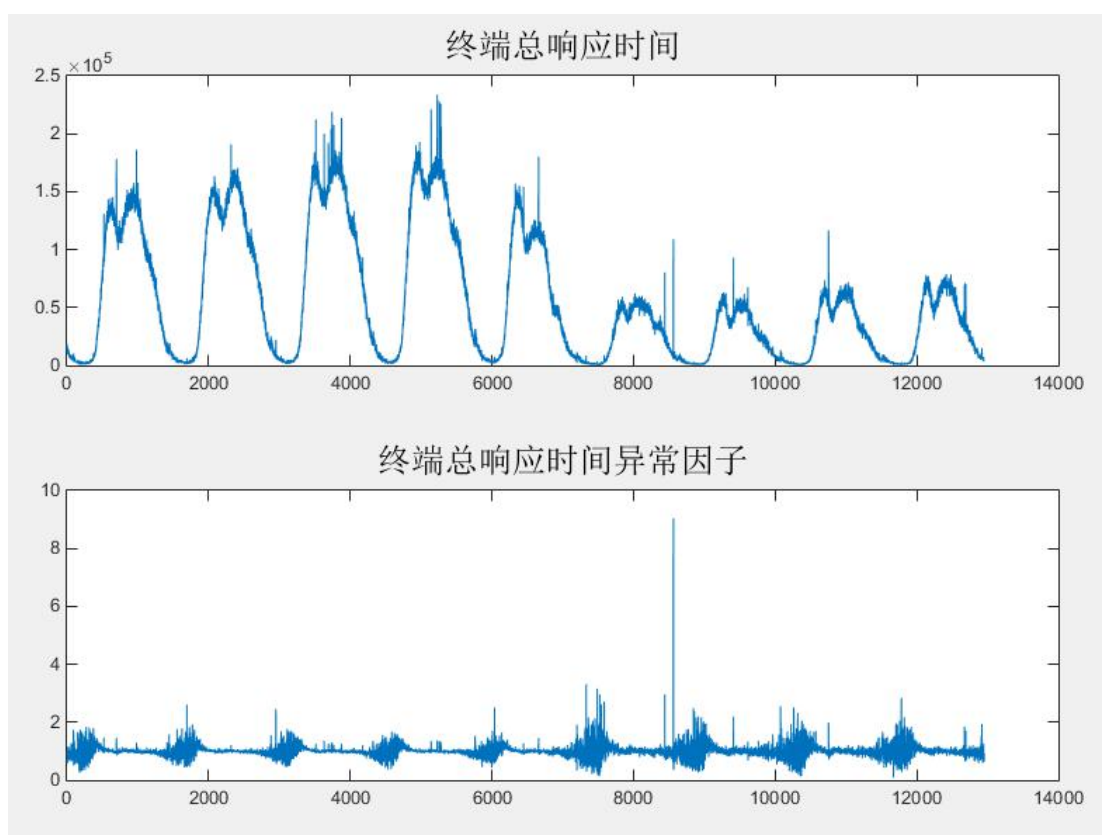


图 18

小波变换分析：

小波分析模型即对数据同时在时域和 变换域展开，通过定量分析，得到每一时刻的异常状态。

小波变换

$$F(w) = \int_{-\infty}^{\infty} f(t) * e^{-iwt} dt \quad \Rightarrow \quad WT(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) * \psi\left(\frac{t-\tau}{a}\right) dt$$

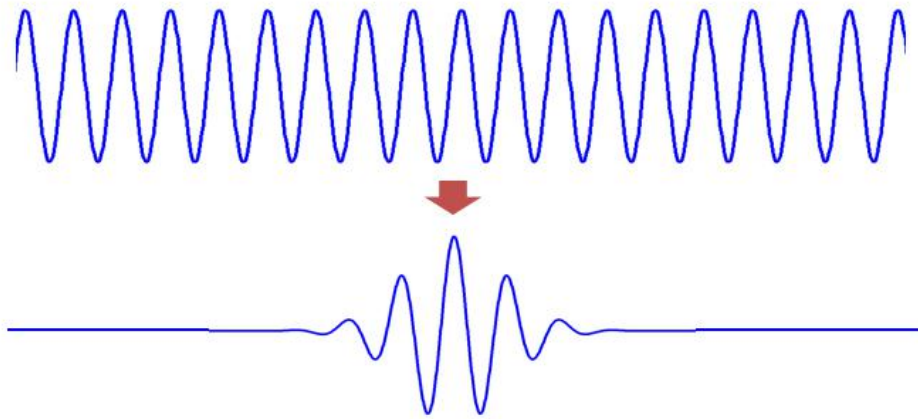


图 19

选取 Daubechies(dbN)小波作为基波,对响应时间进行小波变换得到时域-变换域,如图:

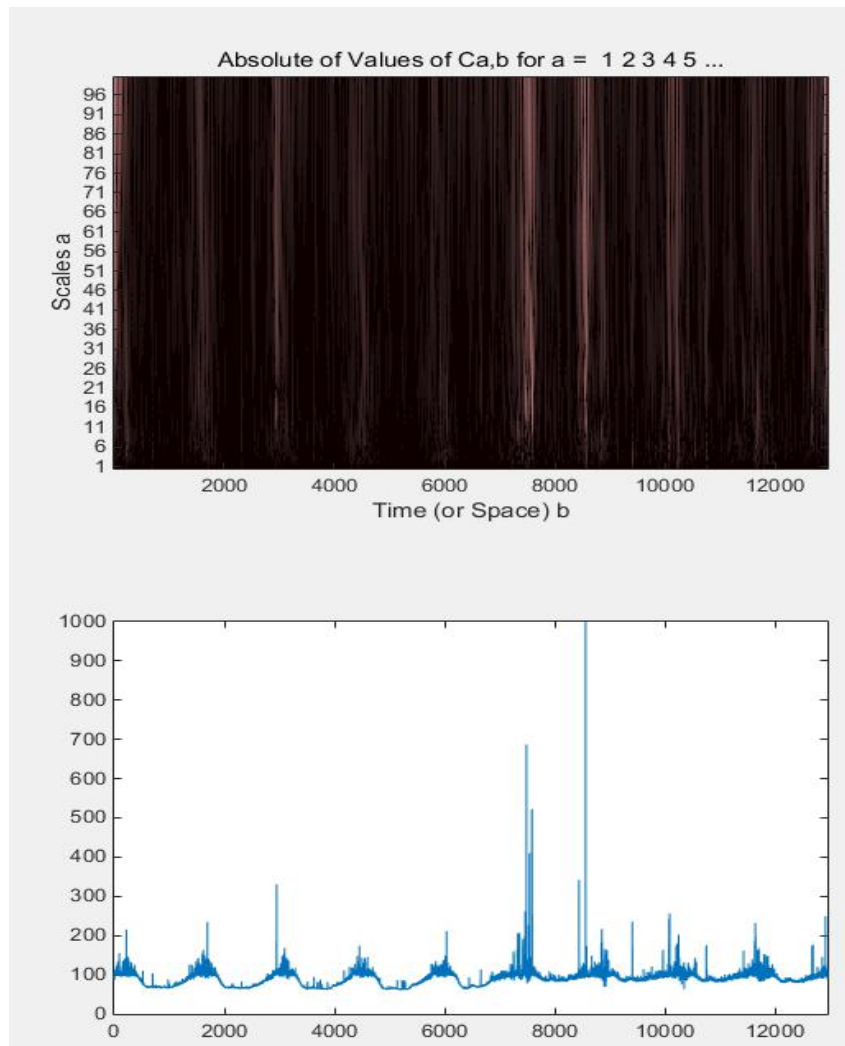


图 20

可以看出在数据异常时，时域-变换域图能够反映出，但效果不如数字滤波器得到的异常因子直观。

4. 3. 3 模型检验:

通过数字滤波法对一月份的响应时间提取响应时间异常因子:

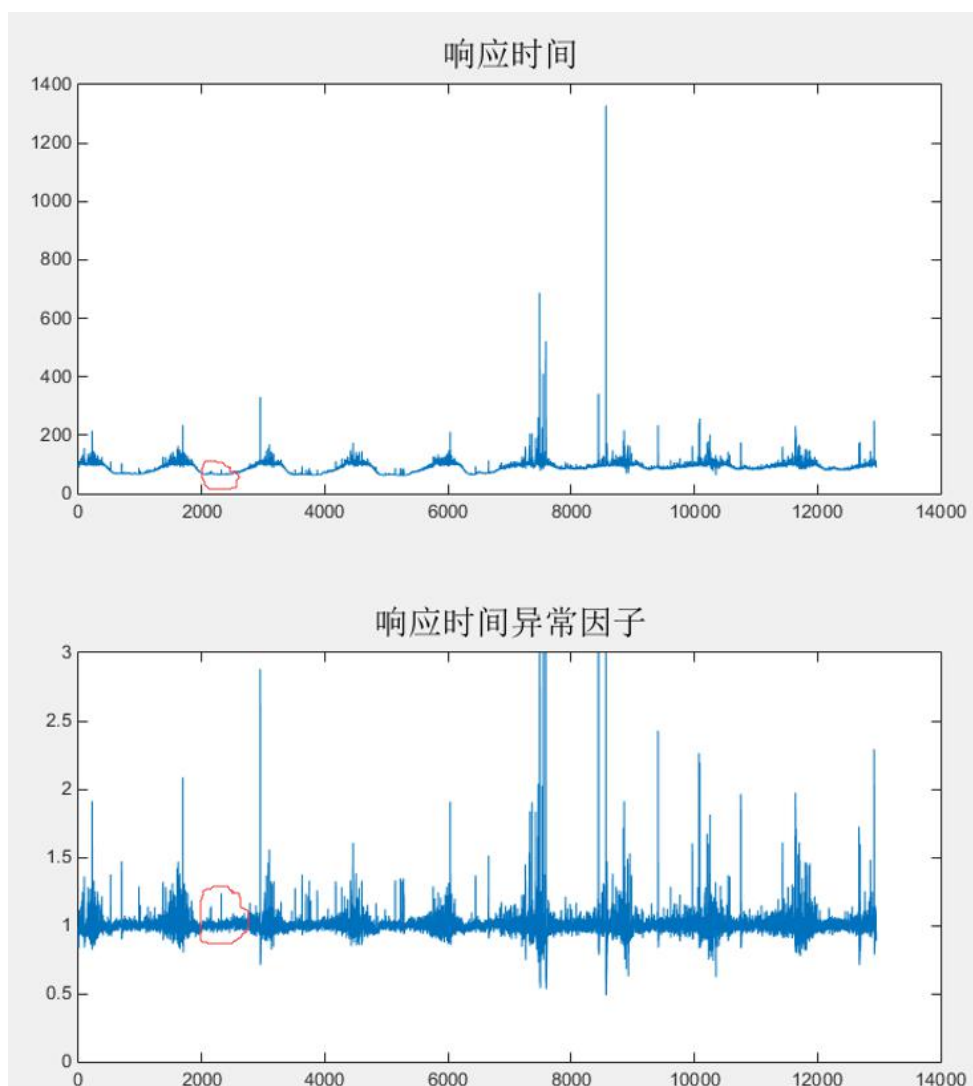


图 21.1 月份响应时间和响应时间异常因子

红圈内响应时间有着微弱的异常变化，用传统阈值判断法很难体现出来异常，然而异常因子则可以显著的表达出来该电的异常，而本身就很显著的异常在异常因子上的表达和微弱的异常的表达相差并不大，可见异常因子对微弱异常有着很高的敏感度。

4.3.4 模型评价：

数字滤波分析法可以准确的反应动态数据中的异常数据，且对微弱的异常变化有很高的灵敏度，便于结合动态阈值进行报警系统的建立。但对极端值的表达则显不足，

小波变换法虽然可以直接的反应每一时刻的异常状态，但由于异常状态表达不够直观，且计算量略大，故效果不如数字滤波法。在后文报警系统的建立中，主要依赖数字滤波法得出的异常因子。

4.4 模型四：ATM 交易系统故障的波及影响模型：

4.4.1 问题分析

ATM 交易系统故障不仅仅给当下的本次交易带来交易时间延迟等的影响，也在其他方面给整个交易人群的交易状态带来了附加的波及影响。由题中所给数据可以发现，当 ATM 交易系统故障的严重性趋于一定高度时，交易状态无法在交易系统恢复相对正常的情况下立即作出同步性应激，反而是会有一定程度的滞后，并且，当 ATM 交易系统故障的严重性超出一定的范围，使交易的成功率下降到一定的程度，并保持低交易成功率一段时间后，交易状态会出现“滞空”期，即在一段时间内无人交易。研究 ATM 交易系统故障的波及影响模型，是为了更好地预判 ATM 交易系统故障后的附加影响，以便能够尽早对系统的异常监测方案进行合理的建议与影响评估。

4.4.2 模型的建立

(1) 为了进行后续的动态排队模型的讨论，本文先对已有数据，从数学分析和软件验证顾客到达 ATM 机时间满足泊松分布：

设 N_t 表示在时间区间 $[0, t)$ 内顾客到达 ATM 机的数目 ($t > 0$)

$$P_n(t_1, t_2) = P\{N(t_2) - N(t_1) = n\} \quad (t_2 > t_1, n \geq 0)$$

显然 $P_n(t_1, t_2)$ 符合下列三个条件：

- ① 在不重叠的时间区间内到达数是相互独立的，我们称这性质为无后效性。
- ② 对充分小的 Δt ，在时间 $[t, t + \Delta t)$ 内有顾客到达的概率与 t 无关，而约与区间长 Δt 成正比，即

$$P_1(t, t + \Delta t) = \lambda \Delta t + o(\Delta t)$$

其中 $o(\Delta t)$ 表示当 $\Delta t \rightarrow 0$ 时，是关于 Δt 的高阶无穷小。 $\lambda > 0$ 是常数，它表示单位时间有顾客到达的概率，称为概率强度。

- ③ 对于充分小的 Δt ，在时间区间 $[t, t + \Delta t)$ 内有 2 个或 2 个以上顾客到达的概率极小，以至于可以忽略，即

$$\sum_{n=2}^{\infty} P_n(t, t + \Delta t) = o(\Delta t)$$

在上述条件下，我们研究顾客到达的数目 n 的概率分布。

由条件②我们总可以取时间由0算起，并简记 $P_n(0, t) = P_n(t)$

由条件②和③容易推出在 $[t, t + \Delta t)$ 区间内没有供应商发生问题的概率为

$$P_0(t, t + \Delta t) = 1 - \lambda \Delta t + o(\Delta t)$$

在求 $P_n(t)$ 时，用通常建立未知函数的微分方程的方法，先求未知函数 $P_n(t)$ 由时刻 t 到 $t + \Delta t$ 的改变量，从而建立 t 时刻的概率分布与 $t + \Delta t$ 时刻概率分布的关系方程。

对于区间 $[0, t + \Delta t)$ ，可分成两个互不重叠的区间 $[0, t)$ 和 $[t, t + \Delta t)$ 。现在到达顾客总数是 n ，分别出现在这两个区间上，不外下列三种情况：

$$\begin{cases} A: [0, t) \text{ 内个数为 } n; \\ B: [0, t) \text{ 内个数为 } n-1; \\ C: [0, t) \text{ 内个数为 } i; (i = 0, 1, \dots, n-2) \end{cases}$$

三种情况互不相容，故 $P_n(t + \Delta t)$ 应是上述三种情况的概率之和：

$$P_n(t + \Delta t) = P_n(t)(1 - \lambda \Delta t) + P_{n-1}(t)\lambda \Delta t + o(\Delta t)$$

$$\frac{P_n(t + \Delta t) - P_n(t)}{\Delta t} = -\lambda P_n(t) + \lambda P_{n-1}(t) + \frac{o(\Delta t)}{\Delta t}$$

令 $\Delta t \rightarrow 0$ ，得下列方程，并注意到初始条件，则有

$$\begin{cases} \frac{dP_n(t)}{dt} = -\lambda P_n(t) + \lambda P_{n-1}(t), n \geq 1 \\ P_n(0) = 0; \end{cases}$$

当 $n=0$ 时，没有B、C两种情况，所以得

$$\begin{cases} \frac{dP_0(t)}{dt} = -\lambda P_0(t) \\ P_0(0) = 1; \end{cases}$$

解上述两式，得到

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, t > 0, n = 0, 1, 2, \dots$$

$P_n(t)$ 表示长为 t 的时间区间内到达顾客的个数为 n 个的概率，即 N_t 服从泊松分布。

下面本文结合题中所给的数据，用软件拟合验证的方法，来验证顾客到达 ATM 机时间满足泊松分布，得出结果如下：

对泊松分布的验证过程：

表 4 单样本 Kolmogorov-Smirnov 检验

N		55060
Poisson 参数 a、b	均值	1065.17
	绝对值	0.417
	正	0.417
	负	-0.251
Kolmogorov-Smirnov Z		0.663
渐近显著性（双侧）P		0.772

表检验结果表明，顾客到达 ATM 机的时间分布的渐进显著性 $P=0.772$ 大于 0.05，于是验证了顾客到达 ATM 机的时间符合泊松分布，也验证了顾客在 ATM 机排队等待交易满足排队模型。

（2）建立 M/M/1 型的多因素的动态排队模型：

ATM 机的交易模式是一对一服务，顾客到达服从泊松分布，每个顾客接受服务的时间服从负指数分布，而且满足了先到先服务的原则。客观问题背景下的 ATM 机数量是固定的，排布位点也是固定的，为了方便研究后续的问题，不妨假设每一台 ATM 机的交易状态大致相同，于是对多台 ATM 机的交易状态研究就可以近似成对单台 ATM 机交易状态研究。这样一来，就可以将其划归成 M/M/1 型的多因素的动态排队模型。根据该 M/M/1 型的多因素动态排队模型，可知当队首的顾客的交易过程发生故障，导致的时间上的延误，也会间接地附加给后面排队的顾客，建立的多因素的动态排队模型如下：

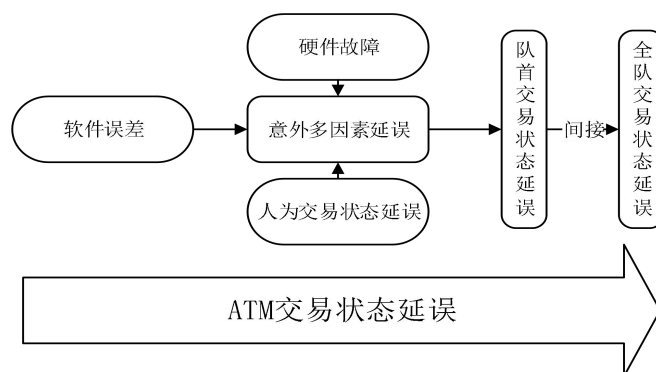


图 22

(3) 当 ATM 机器发生故障时，正在使用的第一位顾客的交易时间被延误，从而间接地导致其后面所有顾客的交易时间被延误，而每个顾客被延误的程度，影响不尽相同，于是我们将其看成一个波及模型，假设其影响的削弱程度都相同（即波及因子 D 相同），并结合故障时的影响因素，再由排队论推导出的经验公式，可以通过这些从而构造出波及效应的模型。其中特别需注意的是，当 ATM 机故障短时间内连续重复发生时，会导致交易的滞空状态，也就是波及效应的过度状态。结合本题的数据，本文主要从成功率以及响应时间两个角度对动态排队模型下的波及效应作出研究，建立的波及模型图如下：

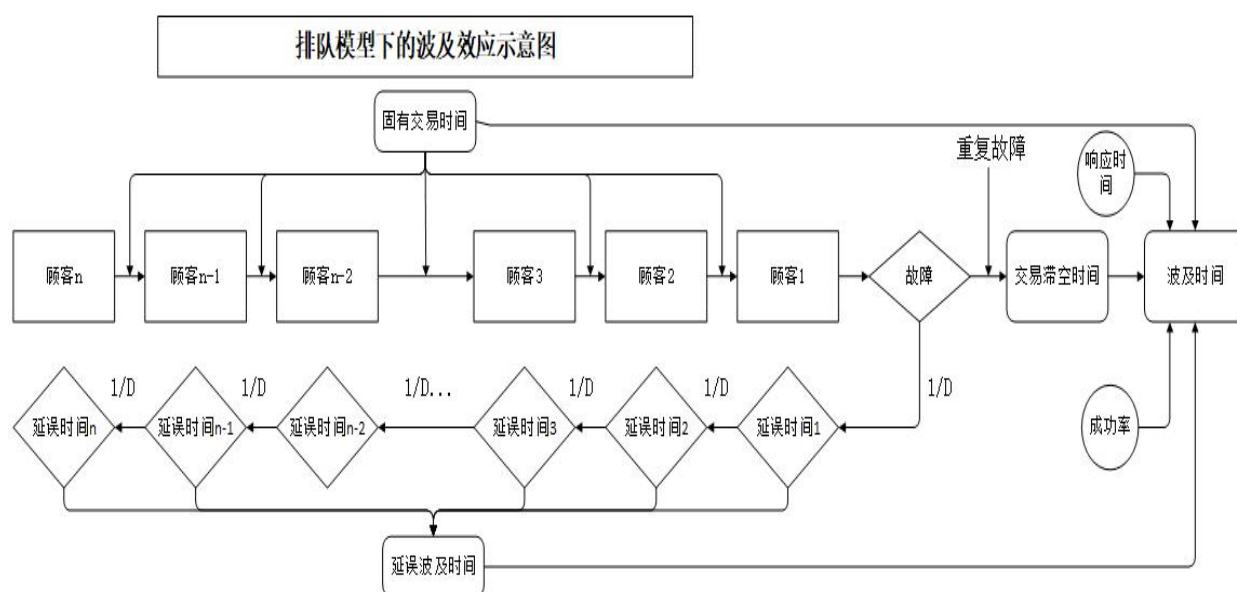


图 23

4.4.3 模型的求解

根据以上的动态排队模型，可得如果一个序列服从泊松分布，那么它的序列间隔服从负指数分布，即如果顾客到达 ATM 机的时间呈现出泊松分布，可以推导出因 ATM 发生故障而导致的波及时间服从均值为 $1/\lambda$ ，方差为 $1/\lambda^2$ 的指数分布，根据经验公式，可以得到以下表达式：

$$t = k_1 e^{y_1 x_1} + k_2 e^{y_2 x_2} - k_3 e^{y_1 x_1 + y_2 x_2}$$

$$\tau = \left(1 + \frac{1}{D} + \frac{1}{D^2} + \dots + \frac{1}{D^n}\right)t \quad t \leq T_{th}$$

$$\tau = \left(1 + \frac{1}{D} + \frac{1}{D^2} + \dots + \frac{1}{D^n}\right)t + C_s \quad t > T_{th}$$

式中：

k_1 ---- 交易状态异常时交易量对交易状态影响权重

k_2 ---- 交易状态异常时成功率对交易状态影响权重

k_3 ---- 交易状态异常时交易量与成功率相关系数

y_1 ---- 比例系数

y_2 ---- 比例系数

C_s ---- 深度故障导致的交易滞空时间

n ---- 排队人数

D ---- 波及影响因子

t ---- 总延误时间

τ ---- 总波及时间

T_{th} ---- 临界延误时间

通过对 ATM 交易故障情况的参数（交易量，成功率，响应时间）的相关性分析与主成分分析，可以得到以下图表：

表 5

Person (相关性)	Q (交易量)	S (成功率)
Q (交易量)	1	0.521
S (成功率)	0.521	1
显著性 (双测) = 0.027		

表 6

成分矩阵	成分
Q (交易量)	0.734
S (成功率)	0.955
R (响应时间)	-0.958

通过主成分分析与相关性分析，可近似认为在交易故障的状态下：

$$k_1 = 0.734 \quad k_2 = 0.955 \quad k_3 = 0.521$$

通过软件拟合求参，可进一步得到式中的 $y_1 = 0.048$ $y_2 = -0.002$ ；一般地，为了研究简便性，常将波及影响因子 $D = 2$ ；则有：

$$t = 0.734e^{0.048x_1} + 0.955e^{-0.002x_2} + 0.521e^{0.048x_1 - 0.002x_2}$$

$$\begin{aligned} \tau &= \left(1 + \frac{1}{2} + \frac{1}{2^2} + \dots + \frac{1}{2^n}\right)t \quad t < T_{th} \\ \text{故} \quad \tau &= \left(1 + \frac{1}{2} + \frac{1}{2^2} + \dots + \frac{1}{2^n}\right)t + C_s \quad t \geq T_{th} \end{aligned}$$

4.4.4 模型的补充与推广

满意度函数对 ATM 机交易故障导致的波及问题的深度研究：

纵观整个问题，因 ATM 机交易故障，而导致的排队波及问题，宏观层次上是动态排队论的波及效应的体现，而更本质上，是波及时间对顾客的满意度的影响，结合题意可知，当延误时间超过一定范围时，顾客的满意度也下降到了一定程度，由此导致顾客离开，造成了宏观上体现出来的“滞空”现象，为了研究这一问题，本文先引入心理曲线的概念及其定义：

心理曲线：

$$u(t) = 1 - e^{\left(\frac{t}{\lambda}\right)^2}$$

其中：

$u(t)$ ---- 心理曲线值

t ---- 等待时间

λ ---- 与等待时间相关的实力因子

结合心理曲线的定义以及波及效应的滞空作用，本文引出了抱怨度函数，对后续问题展开研究：

$$f(t) = \begin{cases} e^{at^2} - 1 & t < T_{th} \\ 1 - e^{-bt^2} & t \geq T_{th} \end{cases}$$

其中：

$f(t)$ -----满意度逆函数值

t -----等待时间

a, b -----调和系数

T_{th} -----临界延误时间

该函数当 $t \rightarrow 0$ 时， $f(t) \rightarrow 0$ ，符合等待时间越小，抱怨度越小的客观事实；当 $t \rightarrow \infty$ 时， $f(t) \rightarrow 1$ ，符合平均等待时间越大，抱怨度越大的事实；且分界点值 T_{th} 的设定，与波及效应函数 τ 对应，更能体现模型对排队满意度论下波及效应模型的深度刻画。并且，通过该函数在 $t = T_{th}$ 点处的连续性和一阶导数的连续性，可以定出常数 a 与常数 b 的值。由于本题所给数据有限，无法确定 T_{th} 的值，倘若如果通过增设更多数据进行问题分析，就可以进一步将此问题量化。

4.4.5 模型的结论

本模型重点分析了 ATM 机发生故障的情况下，排队顾客的波及影响以及深入层次的满意度上的改变。在泊松验证的基础上，本文通过动态排队论模型的建立与波及效应模型的联立，结合具体数据，分析并得出了题中交易故障状态下的波及时间影响，最终归结于满意度改变，层层深入地系统后续的异常监测方案提出建议与完善。

4.4.6 模型的评价

将模型计算结果和理论分析，与广州市的几家银行公布的 ATM 数据较比，基本吻合。不足之处在于，该题中所给数据不够细化，因此模型无法进一步完善。

4.5 模型五：报警系统的建立

模型（1）：基于 K-Means 聚类分析的初步异常检测

4.5.1.1 问题分析

聚类分析指的是把各种抽象的数据对象集根据彼此的相似性划分为不同类别的过程。相似性大的数据归为一类，相似性小的数据归为不同类。K-means 算法是硬聚类算法，优化数据点到聚类中心的某种距离，一般采用欧氏距离。K-means 算法通过测算对应某一初始聚类中心的欧式距离，对聚类中心进行调整，从而得到最优距离的聚类。SPSS 由于其强大的数据处理分析功能，是进行聚类分析的首选。本文采用 SPSS 进行聚类分析，分析异常的交易状态。

4.5.1.2 模型的建立与求解

由于交易量的聚类分析在本题中没有实际意义，而前后时刻交易量的差更能体现故障的发生与否，于是本文对数据集中前后时刻交易量差值的绝对值、响应时间和成功率分别进行了 K-means 聚类分析，聚集数为 2 个，得到其初步的异常值。本文以二月为例进行了分析：

1. 对于响应时间进行聚类分析，得到的结果如下：

表 7

最终聚集中心

聚集	1	2
响应时间	90	6117

每一个聚集中的观察值数目

聚集 1	40312
聚集 2	8
有效	40320
遗漏	0

从结果可以看出，二月的数据中通过对响应时间进行聚类分析得到的异常值有八

个，这八个异常值对应响应时间最大的八个异常时间段。

1、对于成功率进行聚类分析，得到的结果如下：

表 8
最终聚集中心

聚集	1	2
成功率	0.8789	0.9629
每一个聚集中的观察值数目		
聚集 1	1911	
聚集 2	38409	
有效	40320	
遗漏	0	

从结果可以看出，二月的数据中通过对成功率进行聚类分析得到的异常值有1911个，这1911个异常值对应响应时间最大的1911个异常时间段。

4.5.1.3 模型的评价

通过基于 K-Means 聚类分析，可以由单个参数多个角度进行异常值的检测与分析，并且能初步地做出一些判断和分析。然而显然通过两次聚类分析得到的结果产生的偏差较大，这是由于聚类分析对各个参量之间的不同难以做出综合的考虑。还有由于聚类分析更偏重于数据分类方面的意义，与实际的交易故障的关系并不是很明确。另外在实际生活中，倘若每次都用聚类分析进行实时的故障分析与检测，不仅耗时多，而且也导致了不必要的资源浪费。更有聚类分析只从参数单方面的情况考虑，没有考虑参数之间的联系与综合影响，因此在实际应用中也并不够灵活。

模型（2）动态阈值报警系统的建立

4.5.2.1 优化模型分析

通过数据分析可知，当交易量较大时，响应时间和成功率都较为稳定，这是由于基数较大时，响应时间即使出现异常数据，也被大量正常数据“稀释”下来，

显得不那么异常。而对于成功率，根据伯努利大数定律也可知交易量大时，“成功率”依照概率收敛于实际成功概率，从而稳定的维持在较高水准。所以在设立报警阈值时，不能设立一个固定值，而是应该设立一个动态的值。

基于交易量的变化：

业务量陡降标志着分行侧网络传输节点故障，前端交易无法上送请求。通过将实际交易量与数据分析拟合出的交易量分布做对比便可得知不交易量下降程度，并可以以此设立报警模型。

基于失败量的响应时间异常因子动态阈值设法：

由于之前建立的数字滤波模型得到的异常因子可以显著的体现响应时间的异常，所以应给异常因子设立动态阈值。为了解得一个合理的动态阈值，本文做出如下假设：

ATM 机较为稳定，报警次数应与交易失败次数近似呈正比。

（其中失败次数 $f = (1-s)q$ ）

为了满足这一先验条件，需要使报警次数和失败次数的相关系数越大越好。而动态阈值函数的应该设立成一个具有很强一般性的多参函数，从而能够很好的满足所设立的先验条件。本文选用多项式函数来作为动态阈值函数，理由如下：

大部分函数可以级数展开成多项式函数，故多项式函数有较强的一般性。

多项式函数求导方便，很容易运用算法实现先验条件。

多项式函数计算简单，有效的减轻报警系统的运算量。

基于参考数据的动态阈值设法：

将参考数据做为动态阈值的设置依据，建立报警系统。适合异常数目少，且显著的数据。比如终端响应时间。

4.5.2.2 模型的建立与求解

基于交易量的变化：

定义交易量前后相邻时刻差值为交易量落差，计做 d 。

$$d(i) = q(i+1) - q(i)$$

d 的大小标志着交易量的下降程度。

绘制 d 随时间变化曲线图：

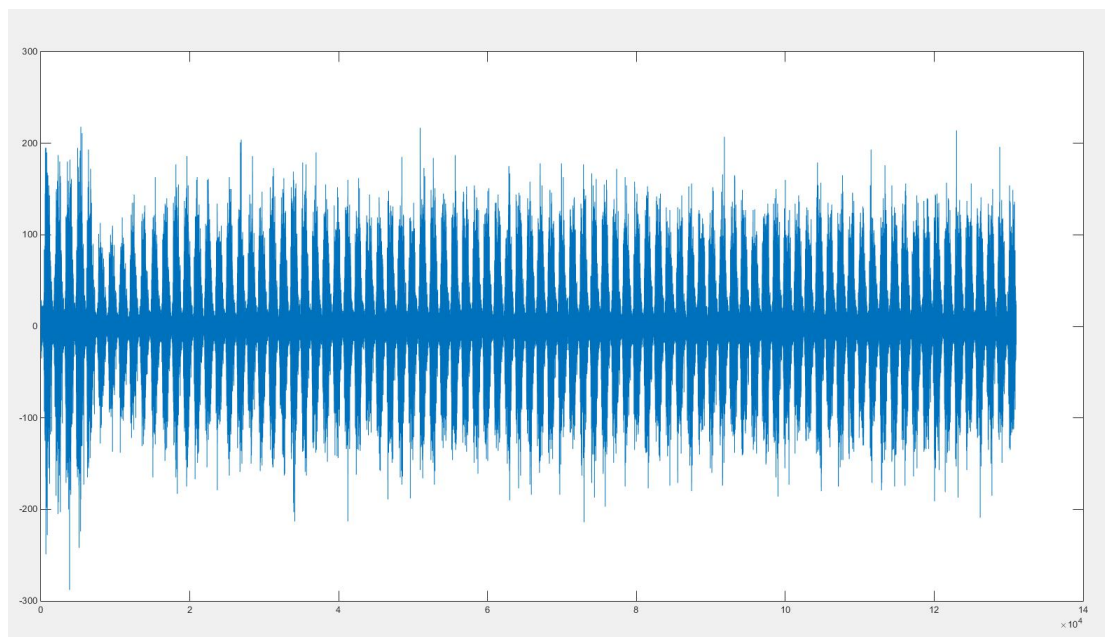


图 24

通过观察发现 d 的分布较为均匀且有周期规律，并无明显异常值。
通过统计分析，得到 d 的最大值为 218, 发生在 1 月 26 号晚高峰时。

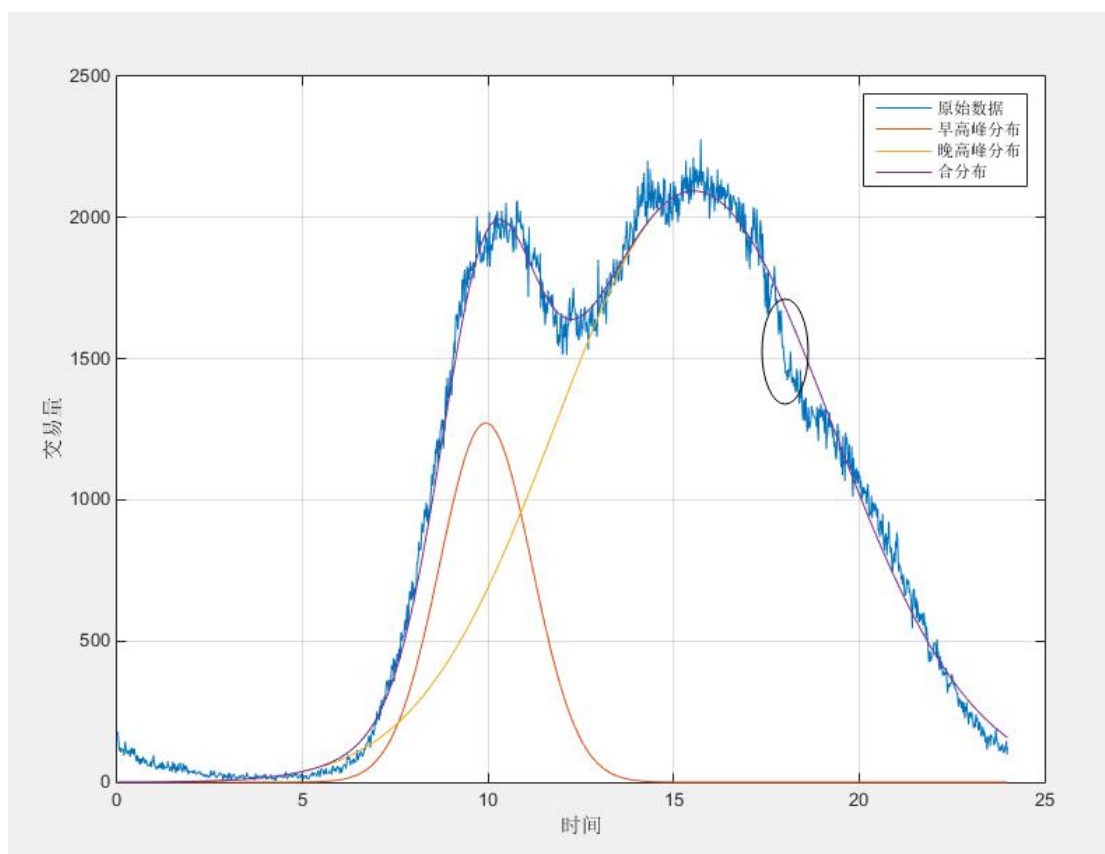


图 25 1 月 26 号交易量

图中黑圈部分即是下降程度最大的时刻，然而这个下降程度和基数比起来并不显著，与合分布图下降率也较为相近，而且每一天的这一时刻附近都有类似的下降情况，（可参见图 11 1 月 25 号交易量变化图），可见这是一个正常现象。

综上所述，交易量的陡降不是 ATM 机交易故障的主要矛盾。

基于失败量的响应时间异常因子动态阈值设法：

响应时间的异常是据中心后端处理系统应用进程异常，导致交易失败或响应缓慢的信号，而响应时间异常因子能够显著表达响应时间的异常，所以对响应时间异常因子建立阈值有很大的意义。

利用梯度上升算法使报警次数和失败次数的相关系数最大

设立动态阈值函数为 4 次多项式函数：

$$th_{\alpha_r}(i) = a_0 + a_1 f(i) + a_2 f^2(i) + a_3 f^3(i) + a_4 f^4(i)$$

其中：

$a_0 a_1 a_2 a_3 a_4$ -----待定系数

f -----失败次数

th_{α_r} -----响应时间敏感因子动态阈值

则报警系统则为逻辑式：

$$alert(i) = (\alpha_r(i) > th_{\alpha_r}(i))$$

而由于“>”逻辑式难以去做求导处理，不利于用算法实现提升报警次数和失败次数的相关系数。所以在此我们理由常用的 S 型函数 sigmoid 函数来代替“>”逻辑式，及

$$alert(i) = \frac{1}{1 + e^{-10(\alpha_r(i) - th_{\alpha_r}(i))}}$$

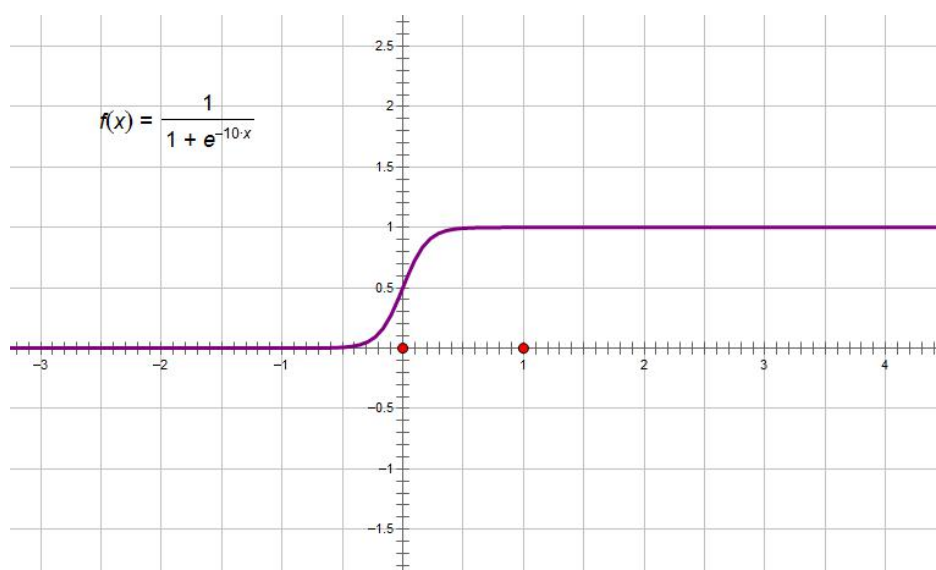


图 26 sigmoid 函数

接下来就可以得到报警次数和失败次数的相关系数：

$$co = \frac{\sum_i (\text{alert}(i) - \overline{\text{alert}}) (f(i) - \bar{f})}{\sqrt{\sum_i (\text{alert}(i) - \overline{\text{alert}})^2 \sum_i (f(i) - \bar{f})^2}}$$

其中：

co-----相关系数

为使 co 相关系数最大，及使

$$c = \frac{[\sum_i (\text{alert}(i) - \overline{\text{alert}}) (f(i) - \bar{f})]^2}{\sum_i (\text{alert}(i) - \overline{\text{alert}})^2} \text{最大即可}$$

$$\text{令 } \theta = [a_0 \ a_1 \ a_2 \ a_3 \ a_4]$$

$$\Rightarrow \overrightarrow{\text{grad}} c(\theta) = 0$$

c 对待定系数求偏导有：

$$\frac{\partial c}{\partial a} = \frac{2 \sum_i (f(i) - \bar{f}) x(i) \sum_i (f(i) - \bar{f}) \frac{\partial x(i)}{\partial a} \sum_i x(i)^2 - 2 \sum_i x(i) \frac{\partial x(i)}{\partial a} [\sum_i (f(i) - \bar{f}) x(i)]^2}{(\sum_i x(i)^2)^2}$$

其中：

$$x(i) = \text{alert}(i) - \overline{\text{alert}}$$

具体对每个系数求导：

$$\begin{aligned} \frac{\partial x(i)}{\partial a_0} &= - \frac{10e^{10(\alpha(i) - th_{a_r}(i))}}{e^{20(\alpha(i) - th_{a_r}(i))} + 2e^{10(\alpha(i) - th_{a_r}(i))} + 1} + \left(\frac{10e^{10(\alpha(i) - th_{a_r}(i))}}{e^{20(\alpha(i) - th_{a_r}(i))} + 2e^{10(\alpha(i) - th_{a_r}(i))} + 1} \right) \\ \frac{\partial x(i)}{\partial a_1} &= - \frac{10e^{10(\alpha(i) - th_{a_r}(i))} f(i)}{e^{20(\alpha(i) - th_{a_r}(i))} + 2e^{10(\alpha(i) - th_{a_r}(i))} + 1} + \left(\frac{10e^{10(\alpha(i) - th_{a_r}(i))} f(i)}{e^{20(\alpha(i) - th_{a_r}(i))} + 2e^{10(\alpha(i) - th_{a_r}(i))} + 1} \right) \\ \frac{\partial x(i)}{\partial a_2} &= - \frac{10e^{10(\alpha(i) - th_{a_r}(i))} f(i)^2}{e^{20(\alpha(i) - th_{a_r}(i))} + 2e^{10(\alpha(i) - th_{a_r}(i))} + 1} + \left(\frac{10e^{10(\alpha(i) - th_{a_r}(i))} f(i)^2}{e^{20(\alpha(i) - th_{a_r}(i))} + 2e^{10(\alpha(i) - th_{a_r}(i))} + 1} \right) \\ \frac{\partial x(i)}{\partial a_3} &= - \frac{10e^{10(\alpha(i) - th_{a_r}(i))} f(i)^3}{e^{20(\alpha(i) - th_{a_r}(i))} + 2e^{10(\alpha(i) - th_{a_r}(i))} + 1} + \left(\frac{10e^{10(\alpha(i) - th_{a_r}(i))} f(i)^3}{e^{20(\alpha(i) - th_{a_r}(i))} + 2e^{10(\alpha(i) - th_{a_r}(i))} + 1} \right) \\ \frac{\partial x(i)}{\partial a_4} &= - \frac{10e^{10(\alpha(i) - th_{a_r}(i))} f(i)^4}{e^{20(\alpha(i) - th_{a_r}(i))} + 2e^{10(\alpha(i) - th_{a_r}(i))} + 1} + \left(\frac{10e^{10(\alpha(i) - th_{a_r}(i))} f(i)^4}{e^{20(\alpha(i) - th_{a_r}(i))} + 2e^{10(\alpha(i) - th_{a_r}(i))} + 1} \right) \end{aligned}$$

利用梯度上升原则做如下过程：

```
for i=1:n
 $\theta_{i+1} = \theta_i + \alpha \cdot \text{grad } C(\theta_i)$ 
end
```

其中 α 为步长。

由于最后解得的阈值会使报警次数和失败次数的相关系数很大，这导致报警次数会较多，所以需要将最后的阈值函数进行线性放缩，使报警次数控制在合理的范围内。

经过梯度上升算法，最终得到多项式阈值函数系数为：

$$\theta = [2 \quad -2.67 \times 10^{-4} \quad 2.22 \times 10^{-8} \quad -1.23 \times 10^{-12} \quad -5.3176 \times 10^{-15}]$$

再利用阈值进行报警即可，下图为一月报警情况：

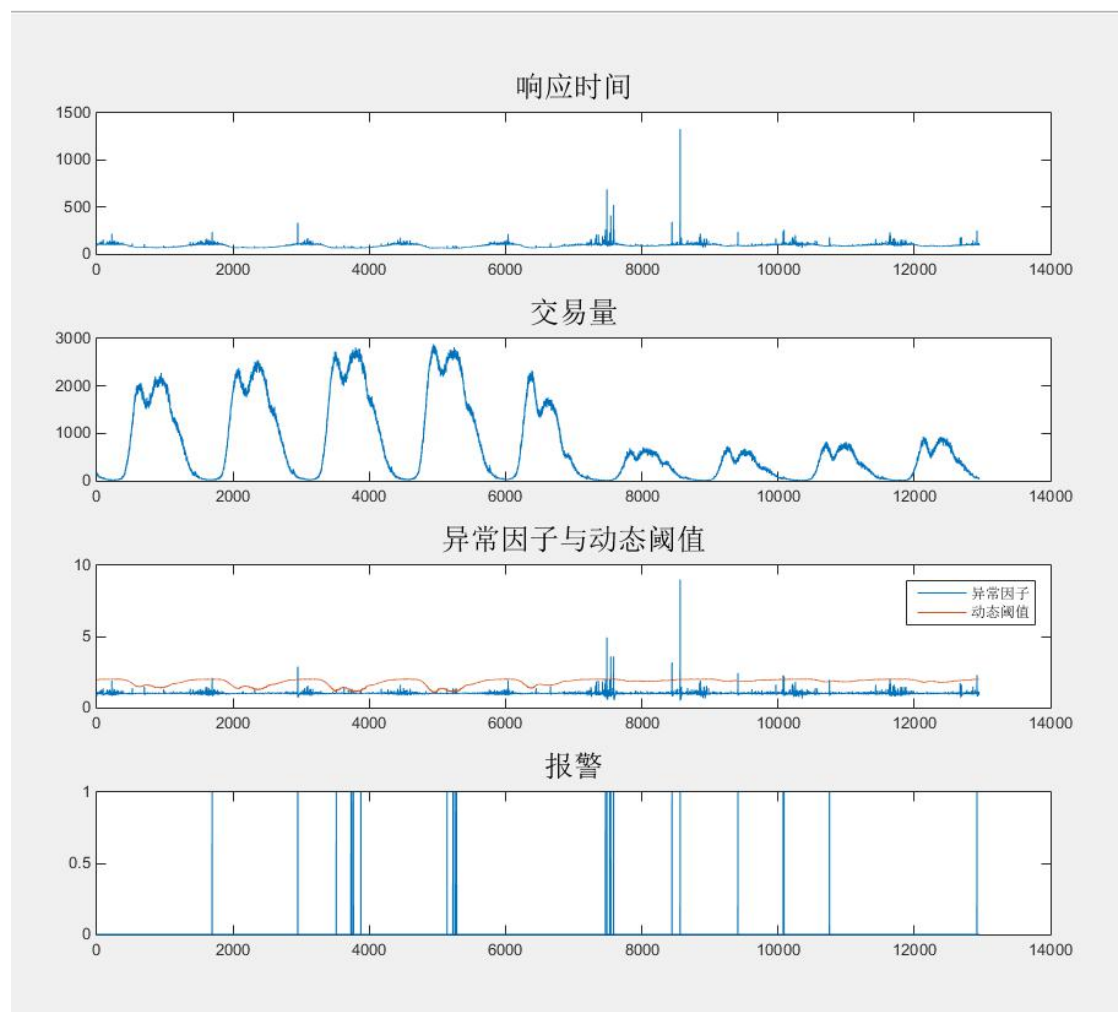


图 27 一月份报警情况

基于参考数据的动态阈值设法：

这里对异常数据相对较少，且异常显著的终端总响应时间设立阈值，

运用低通数字滤波器模型：

$$y(k) = 0.2x(k) + 0.8y(k-1)$$

对终端总响应时间进行两次滤波，计最后得到的终端响应时间参考数据为 trc 。

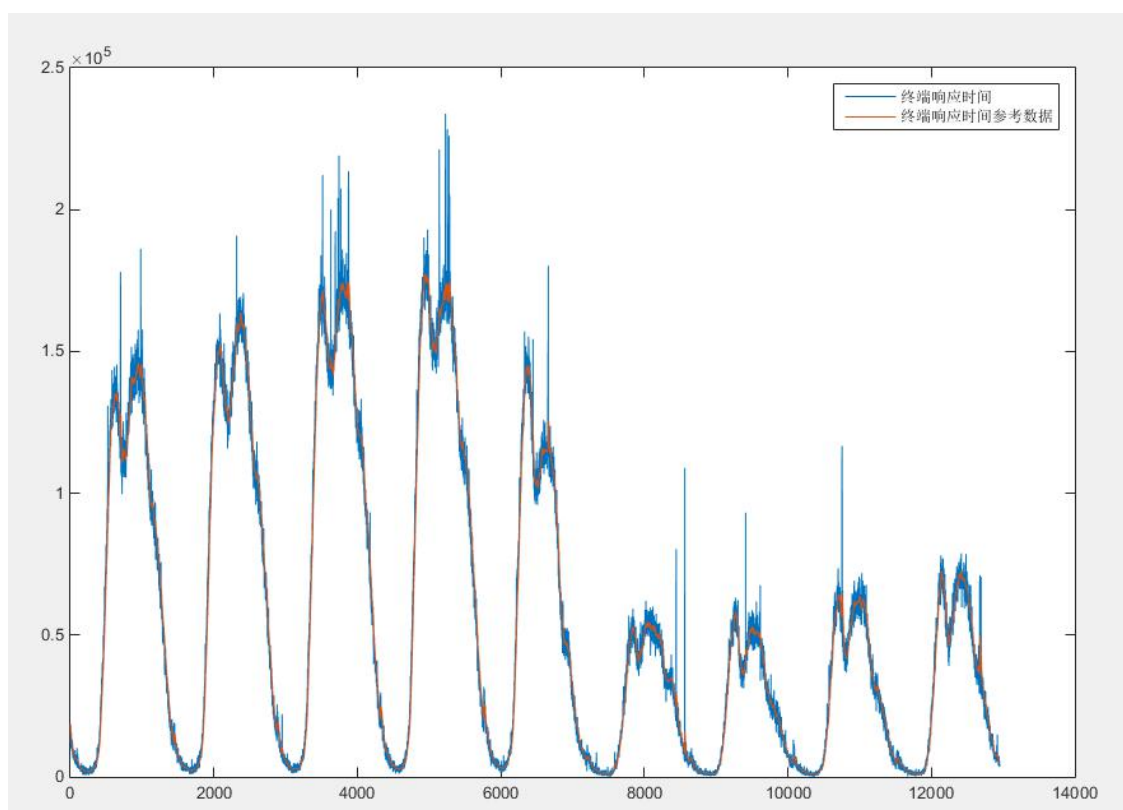


图 28 终端总响应时间及参考数据

为了保证报警次数的合理性，定义阈值为：

$$th_{tr} = trc + c$$

其中：

th_{tr} -----终端总响应时间阈值

trc -----参考数据

c -----待定常数

显然报警次数是随 c 的变化而变化的， c 越大报警次数越少， c 越小报警次数越多。从报警系统的稳定性考虑，应该尽量使阈值附近的数据越少越好，也就是说在 c 产生微小的变化后，报警情况几乎不发生改变。为了达到这一点，应该

找到报警次数关于 c 的函数变化曲线。

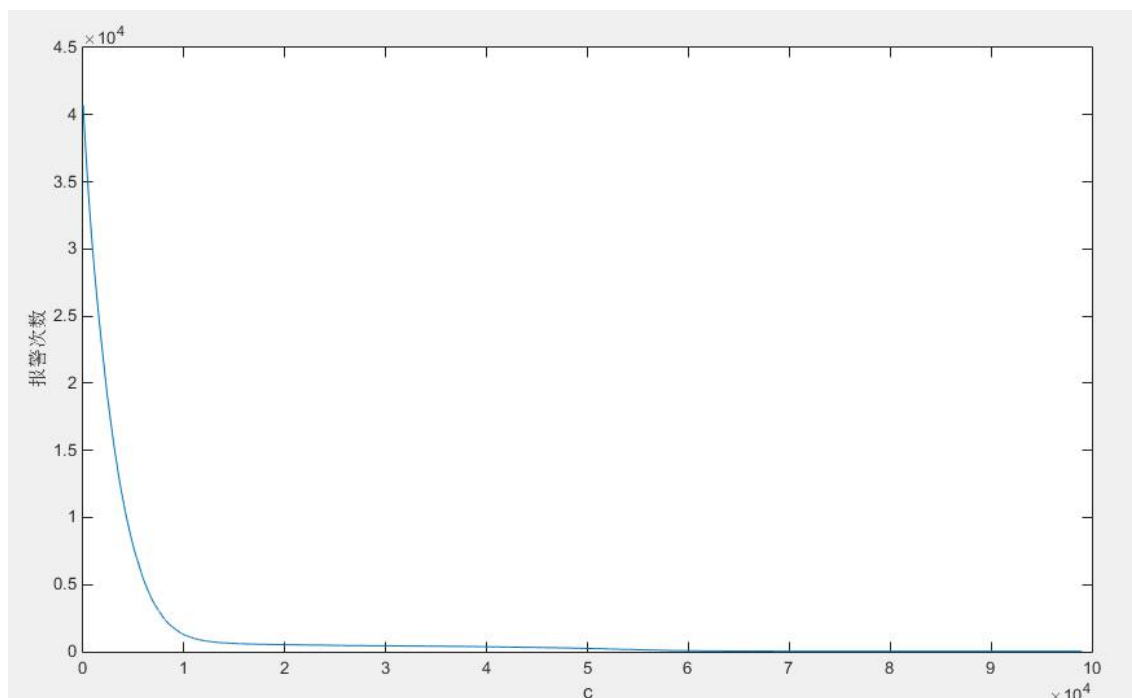


图 29 报警次数关于 c 的函数变化曲线

经过统计, $c \in [29437, 30589]$ 时, 报警次数一直维持在 448 次。这是 c 最长的满足报警次数维持不变的区间。将 c 取该区间中点 30013 可以保证几乎没有终端总响应时间接近阈值, 从而使报警系统有着很高的稳定性。综上, 设立终端总响应时间阈值为:

$$th_{tr} = trc + 30013$$

一月份报警情况如下:

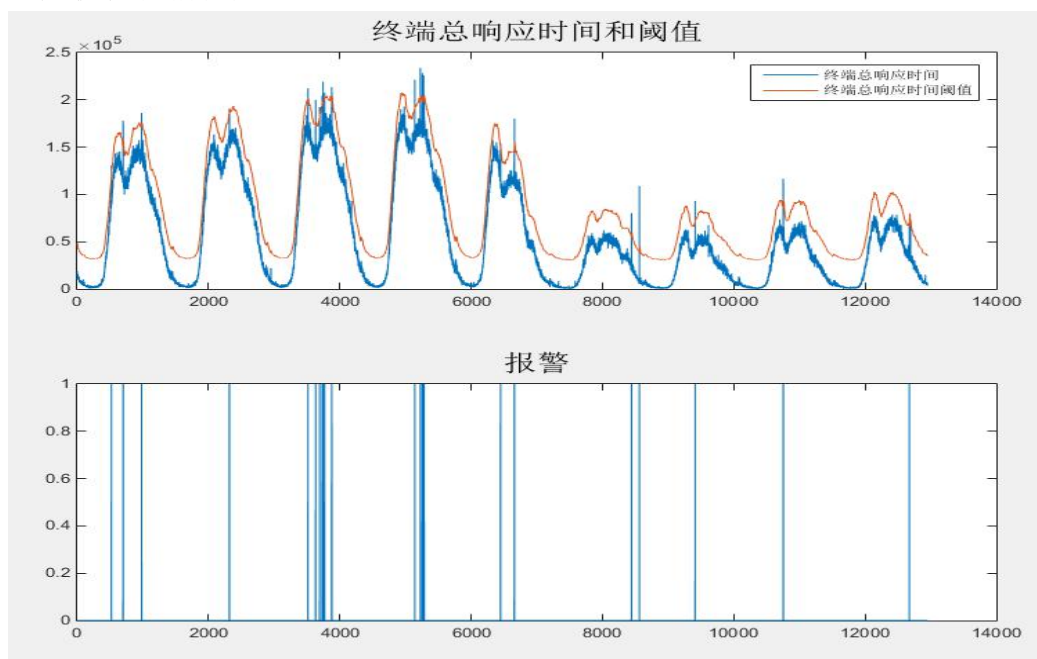


图 30 一月份报警情况

同理，对失败次数用相同的方法可以得到失败次数阈值为：

$$th_f = fr + 106$$

其中 th_f 是失败次数阈值， fr 是失败次数经过低通数字滤波器两次滤波后的值。

报警等级系统的建立：

报警等级系统可以反映报警的类别以及报警的严重性，报警的类别是由异常的种类决定的，例如失败次数过多很有可能是分行侧参数数据变更或者配置错误，数据中心后端处理失败率增加，终端总响应时间异常可能导致数据中心后端处理系统异常（如操作系统 CPU 负荷过大）引起交易处理缓慢。响应时间的异常则是数据中心后端处理系统应用进程异常，导致交易失败或响应缓慢。严重性则体现在两个方面，其一是短时间内出现大量报警，其二是交易数据有极端值的出现。从这两个方面着手，本文将报警划分为 3 个等级，第一个是普通报警，第二个是由短时间内大量的普通报警所触发的高一级别的报警，第三个是有极端值出现所触发的最高等级报警。

首先普通报警应是之前所建立的响应时间异常因子报警、终端总响应时间报警、失败次数报警三者取并。由短时间内大量的普通报警所触发的高一级别的报警，则应是根据普通报警秘籍程度所定。由于极端值大小很反常，而且异常数据分析中所得的异常因子对极端值的表达略显不足，故本文直接通过固定阈值鉴别极端值。

通过观察普通报警情况，本文人为设定当 10 分钟内发生了 3 次以上的报警，则出发高等级报警。极端值报警设定为响应时间>2000ms,失败次数>200 次，终端总响应时间>400s。

4.5.2.3 模型的检验

为了体现极端值报警，故选取有较多极端值出现的四月交易数据进行检验，检验结果如下：

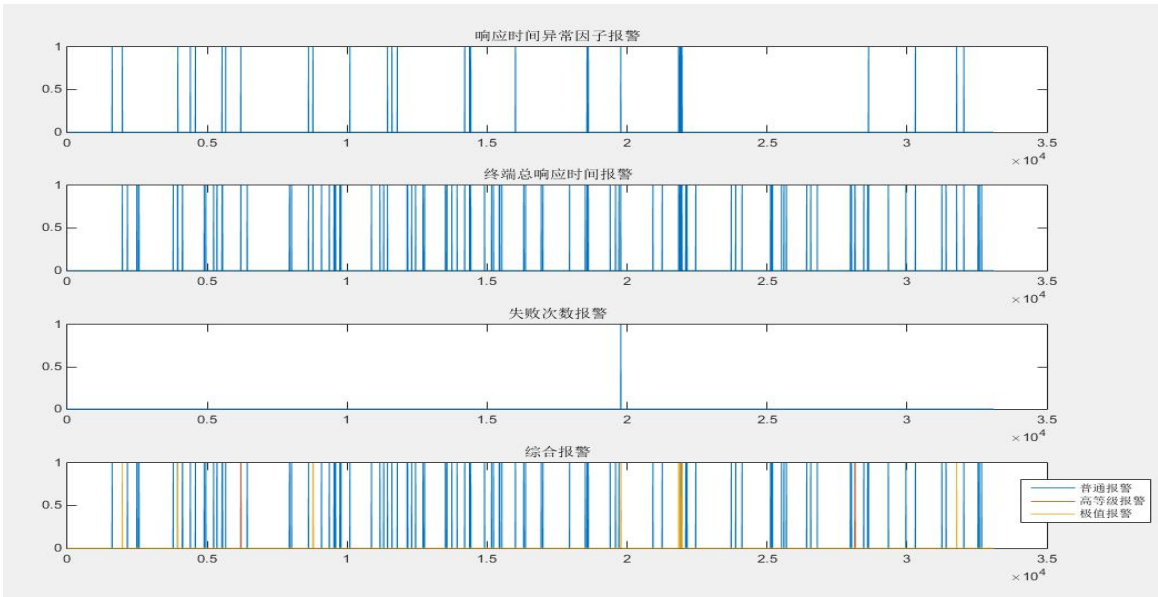


图 31 四月份报警图

高等级报警细节图：

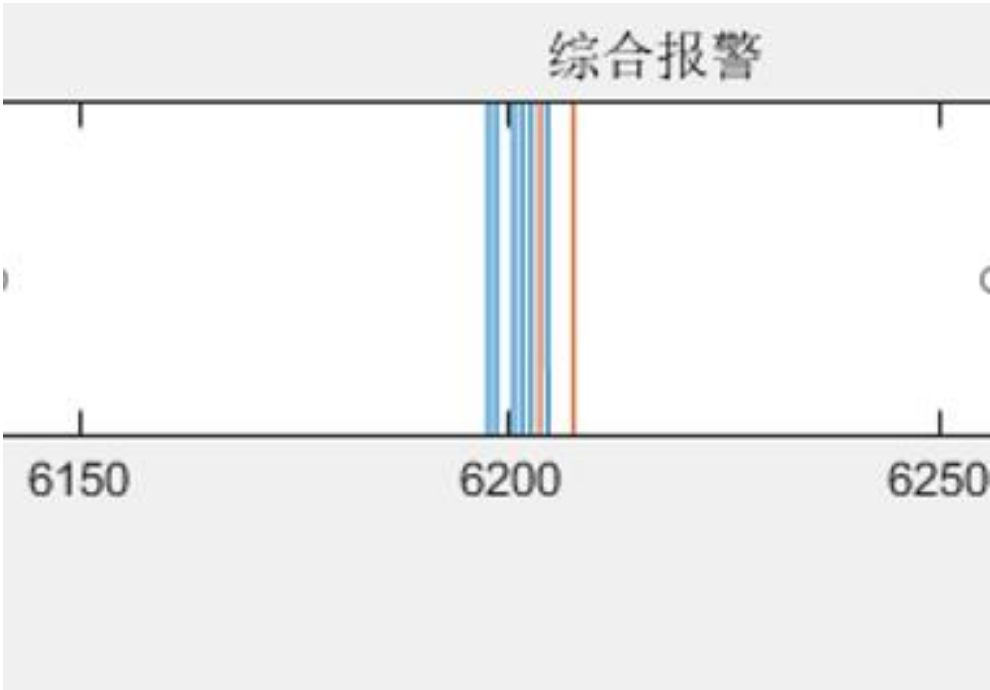


图 32

表 9 四个月的综合报警情况表：

	数据量	普通报警次数	高级报警次数	极端值报警次数	报警比例
1 月	12954	32	0	0	0.25%
2 月	40320	125	11	13	0.31%
3 月	44367	158	28	19	0.35%
4 月	33102	168	45	22	0.51%

通过之前的数据观察，4月份出现了多个严重极端值，最终报警比例达到0.51%，而一月份数据没有极端值的出现，报警比例也很低，只有0.25%，这是符合客观规律的。而且高等级报警与极端值报警有着很高的重复比例，可见极端值的出现会出发一连串异常，这一点也是符合客观规律的。

4.5.2.4 模型的评价

本文报警系统模型充分综合利用了交易量、响应时间、成功率交易数据和异常数据分析模型下的异常因子的内在关系，能够实现准确的报警，做到不误报、不漏报。除了利用交易数据的关系外，还通过ATM机的稳定性良好的客观事实，设立了报警与失败次数关系的先验条件，并从先验条件出发，构建巧妙的算法，得到合理的报警系统。在建立报警系统的同时，也考虑到了报警系统的稳定性，并用遍历的方法设立了稳定合理的阈值。为报警系统设立了报警等级，区分出不同的报警状态，有着很高的实际意义。

五、需要的进一步背景资料和数据

现在已有部分月份的交易量，响应时间，成功率的数据，已经知道了交易状态异常可能由于网络堵塞或前段故障等原因引起，但仍缺少一些资料。若能获取一下资料，对本文的模型改进会有很大帮助。

1. 成功率对应的实际故障类别判断标准。比如是仅仅操作到最后无响应或报错才算失败，还是某一操作相应了极长时间后仍然正确完成了任务，但也算失败。

2. 银行对出现异常后的处理速度。异常状态的处理对故障时段的数据也有很大影响。若可获取此部分数据，便可以进一步完善模型，使波及效应部分的分析更贴近实际情况。

3. ATM 的工作方式及信息传递的相关资料。若给出银行 ATM 交易系统的具体运行过程。通过这个可以了解 ATM 在使用过程中，发生故障的独立成分与相互影响成分，通过这些分析，便于故障时的变化以及监测报警系统的设置参量也相对可以加以完善。具体如若能获取 ATM 的前端故障记录便可对前端故障问题进行深入分析，若能够获取 ATM 信息传输网络的具体信息，则可以将由网络堵塞而引起的系统异常这一部分做更加具体有效的分析。

4. 与所给数据类型相同的，银行近几年的全年数据。掌握这些数据，一方面，提升了数据量，有助于更有说服力更准确地确定特征参数以及对特征参数的应用，另一方面，通过多年的累计同一相对时间的数据，可以对第二问的研究预测提供更前倾性的研究。

5. 给出单位时间所每台 ATM 机对应的各个顾客交易的时间，成功率以及响应时间。在对待异常的评估与监测时，平均状态，容易让监测系统收到的数据被平均，从而忽略占比较小的异常状态的发生，从而导致异常监测系统的漏报。增加这些数据，可以提高异常监测系统的灵敏度与准确性。并且，在对于本题波及效应问题上，也会有显著的帮助。

6. 网络占用率。当网络过于繁忙的时候，前端的交易请求无法顺利传递到中心后端。我们定义

$$\text{网络占用率} = \frac{\text{某时刻网络传输数据总量}}{\text{网络带宽}}$$

网络占用率越高时，前端的请求越难传达，当带宽满负荷运作时，交易请求无法上传，可判断出由于交易量过大而引起的网络通信部分故障。

7. 每笔交易对应的金额。对于 ATM 机来说，不同大小交易金额会引起不同程度的计算操作，其正常相应时间也不同。一般来说，越大额度的交易其相应时间应越长。如果可以获得每笔交易对应的金额，就可以更好的将其实际响应时间与期望的理论相应时间作比较，提高异常检测的准确率。

六、模型总结

本文从数据的分析及整理出发，根据数据的特征作出极具特点的图像。并依据图像的上数据的分布规律及其排列特征，建立初始的特征参数分析模型，从而得出一定的结论，为后续的深入分析打下铺垫。本文着重对交易量与时间两项指标的对应关系以及节假日与非节假日的参数差异进行了深入地分析。针对交易量与时间二者的对应函数关系，本文提出了梯度下降法的拟合方法，将复杂的问题转化为编程性的软件拟合问题，拟合出了贴强度高，成效好的函数模型，将交易量与时间的宏观定性模型升华为有据可循、清晰明了的函数定量模型，充实了特征参数的研究体系。针对节假日与非节假日的参数差异，本文进行了统计学上的假设检验，并借助 MATLAB 等软件，客观而具体地分析并得出了参数差异的具体结论，进一步完善了特征参数的多角度理论体系。

本文对异常数据的分析考虑到了数据本身的动态性，利用了滤波的数学模型，得到了交易数据的参考数据，并依据参考数据构造出反应交易数据异常状态的异常因子，旨在对异常数据做出较为静态的敏感表达。为报警系统的建立做出铺垫。

报警系统的建立则利用了异常分析中所得的异常因子和报警分布于失败量相关性的客观规律，运用算法，拟合出合理的动态阈值函数，可以使误报、漏报的现象尽可能降低。本文还通过分析异常的种类和严重性，设立了多级报警，能够反映出异常的类比和严重程度，对 ATM 机的维护有着很强的现实意义。

特别地，本文还从故障对顾客等待时间的波及影响角度出发，先是进行了泊松验证下的动态排队论模型的分析，对问题进行了概念化的宏观表示。然后结合波及效应的原理，给出详细的计算以及分析过程，并拟合出了最后的经验公式，对故障报警系统的建立与完善提供更好的定性分析。进一步深入分析后，本文将波及问题本质归结于抱怨度函数的研究，进一步提高了故障报警系统的预判能力与后果评估体系。

七、模型评价

(1) 本文以实际所给数据为依据，结合图像分析，函数分析等多种方法，并应用 MATLAB 等分析软件，得出由浅入深的特征参数体系，并由特征参数体系对后续的报警系统等深入问题分析铺垫了基础。

(2) 在函数拟合的过程中，本文没有采用一般形式的多项式或者矩估计法等拟合方法，而是根据题目数据分布的特点、图像排布的规律，创新性地应用了梯度下降的方法对函数进行拟合，体现了适合度高，实用性强，专用性精的特点。

(3) 在报警系统的结构层次上，本文采用了模型优化的建立模式。从初始模型的宏观估计，到优化模型的细致构想，体现了模型的层次性思考过程。

(4) 在报警系统的建模过程中，本文采用了滤波分析的方式，使得报警系统灵活性更强，效果更好。并抛弃了传统固定阈值的方式，采取动态阈值的考量，体现了对问题考虑的动态性与全面性。

(5) 本文从人文心理角度出发，将波及问题的本质归结于满意度、抱怨度等的改变，并作出了详细的分析，使建模体系更加完善。

八、参考文献

- [1] 刘丰年. 一种基于 UML-Petri 网的 ATM 自动取款机建模方法[J]. 宁波职业技术学院学报, 2016, 20 (1) : 91-94.
- [2] 俞乐克. 基于遗传编程的数据挖掘分类和聚类算法的研究与实践[D]. 上海: 华东师范大学, 2009: 1-51.
- [3] 王 焘 顾泽宇 张文博 徐继伟 魏 峻 钟 华. 一种基于自适应监测的云计算系统故障检测方法[J]. 计算机学报, 2016, 39: 1-16.
- [4] 张成虎 赵小虎. 基于小波分析的可疑金融交易时间序列研究[J]. 现代管理科学, 2009, (7) : 102-104.
- [5] 胡寿松 韩智修. 二阶向量差分方程的数字滤波设计[J]. 数据采集与处理, 1988, 3 (1) : 76-81.
- [6] 柴洪峰 李 锐 王兴建 叶家炜. 基于数据挖掘的异常交易检测方法[J]. 计算机应用与软件, 2013, 30 (1) : 165-170.
- [7] 刘卓军 李晓明. 基于时间序列建模和控制图的异常交易检测方法[J]. 数学的实践与认识, 2013, 43 (10) : 89-95.
- [8] Mehta M, Agrawal R, Rissanen J. SLIQ: A fast scalable classifier for data mining[A]. Lecture Notes in computer Sci Proc of the 5th Int Conf on Extending Database Tech [C]. Avignon, France, 1996: 18-33.

九、附录

MATLAB 代码：

数据分析：

```

plot(q4/10);
hold on;
plot(r4);
hold on;
plot(s4*100)
legend('¼»Ö×Á¿','İiÓ|Ê±¼ä','³É¹|ÂÊ')
num=zeros(1,9);
qd=1;
qd1=zeros(1,9);
thitaz1=zeros(9,6);
for date=123:123
    i=date-122;
    qd1(i)=qd;
    num(i)=sum(date1==date);
    d1=q1(qd:qd+num(i)-1);
    qd=qd+num(i);

    %plot(d1);
    t=1:num(i);
    tt=t';
    x=d1;
    b=0.2;
    l=length(x);
    a=[1 -0.8];
    n=[0:l-1];
    %ÂË²`Ç°ÐòÁÐ??

    imp=[1;zeros(l-1,1)];
    h=filter(b,a,imp);%?filter?°-ÊýÊ¹ÓÃ?¼«x?±ä³ÉÂË²`Ç°ÐòÁÐ??
    w=conv(x,h);%?×ö³Ë»ý??
    yc=w(1:l);%ÂË²`ÐòÁÐ??

    z=filter(b,a,x);%?¼»ÐÐÂË²`??
    plot(tt/1440*24,d1);
    xlabel('Ê±¼ä');
    ylabel('¼»Ö×Á¿');

```

```

hold on;
%plot(z);
grid;
x1=z(1:750);
x2=z(750:num(i));
[xm1 tm1]=max(x1);
s1=abs(x1-xm1*exp(-0.5));
[r ts1]=min(s1);
sgm1=ts1-tm1;
y1=xm1*exp(-(t-tm1).^2/2/sgm1^2);
%plot(y1);

[xm2 tm2]=max(x2);
s2=abs(x2-xm2*exp(-0.5));
[r ts2]=min(s2);
sgm2=ts2-tm2;
y2=xm2*exp(-(t-tm2-750).^2/2/sgm2^2);
%plot(y2);
%plot(y1+y2);

thita=[xm1*0.6 xm2 tm1 tm2+750 sgm1 sgm2];

for ii=1:10000
da=sum((z-thita(1)*exp(-(tt-thita(3)).^2/2/thita(5)^2)-thita(2)*exp(-(tt-thita(4)).^2/2/thita(6)^2))*2.*-exp(-(tt-thita(3)).^2/2/thita(5)^2));
db=sum((z-thita(1)*exp(-(tt-thita(3)).^2/2/thita(5)^2)-thita(2)*exp(-(tt-thita(4)).^2/2/thita(6)^2))*2.*-exp(-(tt-thita(4)).^2/2/thita(6)^2));
dm1=sum((z-thita(1)*exp(-(tt-thita(3)).^2/2/thita(5)^2)-thita(2)*exp(-(tt-thita(4)).^2/2/thita(6)^2))*2.*(-thita(1)*exp(-(tt-thita(3)).^2/2/thita(5)^2)).*((tt-thita(3)).^2/2/thita(5)^2));
dm2=sum((z-thita(1)*exp(-(tt-thita(3)).^2/2/thita(5)^2)-thita(2)*exp(-(tt-thita(4)).^2/2/thita(6)^2))*2.*(-thita(2)*exp(-(tt-thita(4)).^2/2/thita(6)^2)).*((tt-thita(4)).^2/2/thita(6)^2));
dsgm1=sum((z-thita(1)*exp(-(tt-thita(3)).^2/2/thita(5)^2)-thita(2)*exp(-(tt-thita(4)).^2/2/thita(6)^2))*2.*(-thita(1)*exp(-(tt-thita(3)).^2/2/thita(5)^2)).*((tt-thita(3)).^2/thita(5)^3));
dsgm2=sum((z-thita(1)*exp(-(tt-thita(3)).^2/2/thita(5)^2)-thita(2)*exp(-(tt-thita(4)).^2/2/thita(6)^2))*2.*(-thita(2)*exp(-(tt-thita(4)).^2/2/thita(6)^2)).*((tt-thita(4)).^2/thita(6)^3));
thita=thita-0.00001*[da db dm1 dm2 dsgm1 dsgm2];
end
thitaz1(i,:)=thita;

```

```

plot(tt/1440*24,thita(1)*exp(-(tt-thita(3)).^2/2/thita(5)^2));
hold on;
plot(tt/1440*24,thita(2)*exp(-(tt-thita(4)).^2/2/thita(6)^2));
%hold on;

plot(tt/1440*24,thita(1)*exp(-(tt-thita(3)).^2/2/thita(5)^2)+thita(2)
*exp(-(tt-thita(4)).^2/2/thita(6)^2));
hold on;
legend('Ö-ÊËËÝÝ','ÔÇ,ß·ä·Ö²¼','íí,ß·ä·Ö²¼','°İ·Ö²¼');
end
pp=zeros(1,10000);
l=length(r1);
for i=1:l
    sit=round(yinzi(i)*1000);
    if(sit<1000)
        pp(sit+1)= pp(sit+1)+1;
    end
end
plot(pp);

x=zeros(1,300);

for i=1:300
    x(i)=sum(round(r1)==i);

end

subplot(2,2,1)
plot(x);
grid on;
ylabel('ÆµËÝ');
xlabel('İiÓ|Ê±¼ä');

legend('İiÓ|Ê±¼äÆµËÝ·Ö²¼');

for i=1:300
    x(i)=sum(round(r2)==i);

end

subplot(2,2,2)
plot(x);

```

```

grid on;
ylabel('μ');
xlabel('iÓ|ä');

legend('iÓ|äμ·Ö²');
for i=1:300
    x(i)=sum(round(r3)==i);

end

```

```

subplot(2,2,3)
plot(x);
grid on;
ylabel('μ');
xlabel('iÓ|ä');
legend('iÓ|äμ·Ö²');

for i=1:300
    x(i)=sum(round(r4)==i);

end

```

```

subplot(2,2,4)
plot(x);
grid on;
ylabel('μ');
xlabel('iÓ|ä');
legend('iÓ|äμ·Ö²');

```

```

d1=q1(1:1440);
plot(d1);
t=1:1440;
tt=t';
x=d1;
b=0.2;
l=length(x);
a=[1 -0.8];
n=[0:l-1];
%Ä²·Ç°òÁ??

```

```

imp=[1;zeros(1-1,1)];
h=filter(b,a,imp);%?filter?°-ÊÿÊ¹ÓÃ?½«x?±ä³ÉÃË²`Ç°ÐòÁÐ??
w=conv(x,h);%?×Ö³Ë»Ý??
yc=w(1:1);%ÃË²`ÐòÁÐ??

z=filter(b,a,x);%?¼øÐÐÃË²`??

xlabel('n');
ylabel('x?y?yc');

hold on;
plot(z);
grid;
x1=z(1:750);
x2=z(750:1440);
[xm1 tm1]=max(x1);
s1=abs(x1-xm1*exp(-0.5));
[r ts1]=min(s1);
sgm1=tm1-ts1;
y1=xm1*exp(-(t-tm1).^2/2/sgm1^2);
%plot(y1);

[xm2 tm2]=max(x2);
s2=abs(x2-xm2*exp(-0.5));
[r ts2]=min(s2);
sgm2=tm2-ts2;
y2=xm2*exp(-(t-tm2-750).^2/2/sgm2^2);
%plot(y2);
%plot(y1+y2);

thita=[xm1 xm2 tm1 tm2+750 sgm1 sgm2];

for ii=1:10000
da=sum((z-thita(1)*exp(-(tt-thita(3)).^2/2/thita(5)^2)-thita(2)*exp(-(tt-thita(4)).^2/2/thita(6)^2))*2.*-exp(-(tt-thita(3)).^2/2/thita(5)^2));
db=sum((z-thita(1)*exp(-(tt-thita(3)).^2/2/thita(5)^2)-thita(2)*exp(-(tt-thita(4)).^2/2/thita(6)^2))*2.*-exp(-(tt-thita(4)).^2/2/thita(6)^2));
dm1=sum((z-thita(1)*exp(-(tt-thita(3)).^2/2/thita(5)^2)-thita(2)*exp(-(tt-thita(4)).^2/2/thita(6)^2))*2.*(-thita(1)*exp(-(tt-thita(3)).^2/2/thita(5)^2)).*((tt-thita(3))^2/2/thita(5)^2));
dm2=sum((z-thita(1)*exp(-(tt-thita(3)).^2/2/thita(5)^2)-thita(2)*exp(

```

```

-(tt-thita(4)).^2/2/thita(6)^2))*2.*(-thita(2)*exp(-(tt-thita(4)).^2/2/thita(6)^2)).*((tt-thita(4)).^2/2/thita(6)^2));
dsgm1=sum((z-thita(1)*exp(-(tt-thita(3)).^2/2/thita(5)^2)-thita(2)*exp(-(tt-thita(4)).^2/2/thita(6)^2))*2.*(-thita(1)*exp(-(tt-thita(3)).^2/2/thita(5)^2)).*((tt-thita(3)).^2/thita(5)^3));
dsgm2=sum((z-thita(1)*exp(-(tt-thita(3)).^2/2/thita(5)^2)-thita(2)*exp(-(tt-thita(4)).^2/2/thita(6)^2))*2.*(-thita(2)*exp(-(tt-thita(4)).^2/2/thita(6)^2)).*((tt-thita(4)).^2/thita(6)^3));
thita=thita-0.00001*[da db dm1 dm2 dsgm1 dsgm2 ]
end

plot(thita(1)*exp(-(tt-thita(3)).^2/2/thita(5)^2));
hold on;
plot(thita(2)*exp(-(tt-thita(4)).^2/2/thita(6)^2));
hold on;
plot(thita(1)*exp(-(tt-thita(3)).^2/2/thita(5)^2)+thita(2)*exp(-(tt-thita(4)).^2/2/thita(6)^2));

x=zeros(1,300);

for i=1:300
    x(i)=sum(round([r1;r2;r3;r4])==i);

end
grid on;
plot(x);
grid on;
ylabel('ÆµËý');
xlabel('ÏiÓ|Ê±äÆµËý·Ö²¼');

legend('ÏiÓ|Ê±äÆµËý·Ö²¼');

b=0.20;
a=[1 -0.8];

%ÄË²·Ç°ÐòÁÐ??
x=f;
x0=x-x(1);
l=length(x);

```



```

z=filter(b,a,x0)+x(1);%?¼øÐÐÂË²``??
z0=z-x(1);
z1=filter(b,a,z0)+x(1);%?¼øÐÐÂË²``??
xlabel('n');

```

```

subplot(2,1,1);
plot(x);
title('ÏiÓ!Ê±ä', 'fontsize',18)
hold on;

```

```

%xlabel('n');
%ylabel('x?y?yc');
%plot(z);
hold on;
%plot(z1);
grid;
subplot(2,1,2);
yinzi=x./z1;
plot(yinzi);

title('Òì³fÒò×Ó', 'fontsize',18);

```

异常分析:

```

x=r1(1:1441)-r1(1);
b=0.2;
l=length(x);
a=[1 -0.8];
n=[0:l-1];
%ÂË²``Ç°ÐòÁÐ

```

```

z=filter(b,a,x);%¼øÐÐÂË²``

```

```

z=filter(b,a,z)+r1(1); %???
%???

hold on;
plot(r1(1:1441));
hold on;
z=z(1:1441);
plot(z);
legend('ËµËÿ', 'Ë²²´îóËÿ');

N=12594;
H=1;
a=1;
P=0.04;
Q=0.04;
x=r1';
y=r1';

x_fil(1)=r1(1);
p_fil(1)=0.04;
for t=2:N
    x_pre(t)=a*x_fil(t-1);
    p_pre(t)=a*p_fil(t-1)*a'+P;
    K(t)=p_pre(t)*H'*inv(H*p_pre(t)*H'+Q);
    x_fil(t)=x_pre(t)+K(t)*(y(t)-H*x_pre(t));
    p_fil(t)=(1-K(t)*H)*p_pre(t);
end

plot(x_fil-100, 'r'); %,
hold on;
plot(y, 'g');
hold on
plot(x);

subplot(2,1,1);
c = cwt(r1,1:100, 'db4', 'plot');
subplot(2,1,2);

plot(r1);
axis([0 12954 0 1000]);
%subplot(3,1,3);
%plot(sum(c));

```

```
%axis([0 12954 0 30000]);
```

报警系统:

```
b=0.20;
```

```
a=[1 -0.8];
```

```
x=r;
```

```
x0=x-x(1);
```

```
z=filter(b,a,x0)+x(1);%?%?%?%?%?
```

```
z0=z-z(1);
```

```
xlabel('n');
```

```
subplot(4,1,1);
```

```
plot(x);
```

```
title('ÿiÓ|Ê±ä', 'fontsize', 18)
```

```
hold on;
```

```
%plot(z);
```

```
%Ê²`Ç°ÐòÁÐ??
```

```
z1=filter(b,a,z0)+z(1);%?%?%?%?%?
```

```
%xlabel('n');
```

```
%ylabel('x?y?yc');
```

```
%plot(z);
```

```
hold on;
```

```
%plot(z1);
```

```
subplot(4,1,2);
```

```
yinzi=r./z;
```

```
plot(yinzi);
```

```
title('Òì³£Òò×Ó', 'fontsize', 18);
```

```
%yz
```

```
yztt=2*[1 1/-6000 1/6000^2/2 -1/6000^3/6 -1/6000^4/24];
```

```

yz=yztt(1)+yztt(2)*q+yztt(3)*q.^2+yztt(4)*q.^3+yztt(5)*q.^4;
alert=1./(1+exp(-10*(yinzi-yz)));%sigmoid函数; yinzi>yz;±
subplot(4,1,3);
plot(alert);
xgxx=mean((alert-mean(alert)).*(q-mean(q)))/sqrt(var(alert)*var(q));
while(yztt(1)+yztt(2)*3000+yztt(3)*3000^2+yztt(4)*3000^3+yztt(5)*3000
^4>1&&length(xgxx)<50)
Y=1./(1+exp(-10*(yinzi-yz)))-mean(1./(1+exp(-10*(yinzi-yz))));
Ya=-1./(1+exp(-10*(yinzi-yz))).^2.*(20*q.*exp(-10*(yinzi-yz)))+mean(1.
/(1+exp(-10*(yinzi-yz))).^2.*(20*q.*exp(-10*(yinzi-yz))));
Yb=-1./(1+exp(-10*(yinzi-yz))).^2.*(20*q.^2.*exp(-10*(yinzi-yz)))+mea
n(1./(1+exp(-10*(yinzi-yz))).^2.*(20*q.^2.*exp(-10*(yinzi-yz))));
Yc=-1./(1+exp(-10*(yinzi-yz))).^2.*(20*q.^3.*exp(-10*(yinzi-yz)))+mea
n(1./(1+exp(-10*(yinzi-yz))).^2.*(20*q.^3.*exp(-10*(yinzi-yz))));
Yd=-1./(1+exp(-10*(yinzi-yz))).^2.*(20*q.^4.*exp(-10*(yinzi-yz)))+mea
n(1./(1+exp(-10*(yinzi-yz))).^2.*(20*q.^4.*exp(-10*(yinzi-yz))));
da=(sum(2*(q-mean(q)).*Y)*sum((q-mean(q)).*Ya)*sum(Y.^2)-sum(2*Y.*Ya)
*(sum(Y.*(q-mean(q))))^2)/(sum(Y.^2))^2/10^38;
db=(sum(2*(q-mean(q)).*Y)*sum((q-mean(q)).*Yb)*sum(Y.^2)-sum(2*Y.*Yb)
*(sum(Y.*(q-mean(q))))^2)/(sum(Y.^2))^2/10^38;
dc=(sum(2*(q-mean(q)).*Y)*sum((q-mean(q)).*Yc)*sum(Y.^2)-sum(2*Y.*Yc)
*(sum(Y.*(q-mean(q))))^2)/(sum(Y.^2))^2/10^38;
dd=(sum(2*(q-mean(q)).*Y)*sum((q-mean(q)).*Yd)*sum(Y.^2)-sum(2*Y.*Yd)
*(sum(Y.*(q-mean(q))))^2)/(sum(Y.^2))^2/10^38;
yztt=yztt+[0 da db dc dd];
yz=yztt(1)+yztt(2)*q+yztt(3)*q.^2+yztt(4)*q.^3+yztt(5)*q.^4;
alert=1./(1+exp(-10*(yinzi-yz)));%sigmoid函数; yinzi>yz;±
xgxx=[xgxx
mean((alert-mean(alert)).*(q-mean(q)))/sqrt(var(alert)*var(q))];
display(xgxx);

```

end

```

subplot(4,1,4);
yztt=yztt*0.8;
yztt(1)=2;
yz=yztt(1)+yztt(2)*q+yztt(3)*q.^2+yztt(4)*q.^3+yztt(5)*q.^4;
plot(alert);
ttt=1:300;
yyy=yztt(1)+yztt(2)*ttt+yztt(3)*ttt.^2+yztt(4)*ttt.^3+yztt(5)*ttt.^4;
t1=yinzi>yz;

```

b=0.20;

```

a=[1 -0.8];

%ÄË²`Ç°ÐòÁÐ??
x=tr;
x0=x-x(1);
l=length(x);

z=filter(b,a,x0)+x(1);%?¼øÐÐÄË²`??
z0=z-x(1);
z1=filter(b,a,z0)+x(1);%?¼øÐÐÄË²`??
xlabel('n');

subplot(2,1,1);
plot(x);

title('ÖÕŒ×Üïíó|Ê±¼ä°íãÐÖµ','fontsize',18)
hold on;
plot(z1+30013);
legend('ÖÕŒ×Üïíó|Ê±¼ä','ÖÕŒ×Üïíó|Ê±¼äãÐÖµ');

subplot(2,1,2);
yinzi=tr./z1;
t2=(x>(z1+30013));
plot(t2);
title('±´¼¯','fontsize',18);

%xxx=-3:0.001:3;
%yyy=0.2./(1-0.8*exp(-1j.*xxx));
%absyyy=abs(yyy);
%plot(xxx,absyyy);
%legend('ÄË²`Æ÷ÆµÆ×Çúİß')

b=0.20;
a=[1 -0.8];

%ÄË²`Ç°ÐòÁÐ??
x=f;
x0=x-x(1);
l=length(x);

```

```

z=filter(b,a,x0)+x(1);%?¼øÐÐÂÊ²´??
z0=z-x(1);
z1=filter(b,a,z0)+x(1);%?¼øÐÐÂÊ²´??
xlabel('n');

subplot(2,1,1);
plot(x);

title('ÖÖŒË×ÜÏìÓ!Ê±¼ä°íãÐÖµ','fontsize',18)
hold on;
plot(z1+100);
legend('ÖÖŒË×ÜÏìÓ!Ê±¼ä','ÖÖŒË×ÜÏìÓ!Ê±¼äãÐÖµ');

subplot(2,1,2);
yinzi=f./z1;
t3=x>((z1+106));
title('±´¼¯','fontsize',18);

n=zeros(1,100000);
for i=1000:100000

    n(i)=sum(x>(z1+i));

end

%plot(n);
%ylabel('±´¼¯´îÊÝ');
%xlabel('c');
jishu=zeros(1,1000)
for i=100:1000
    jishu(i)=sum(n==i)
end

subplot(4,1,1);
plot(t1(97912:131013));
title('ÏìÓ!Ê±¼äÒì³£Òð×Ó±´¼¯');
subplot(4,1,2);
plot(t2(97912:131013));

```

```
title('ÖÛË×Üïó|Ê±¼ä±´³´');
subplot(4,1,3);
plot(t3(97912:131013));
title('Ê¸°Ü´ÎÊý±´³´');
t=t1|t2|t3;
subplot(4,1,4);
plot(t(97912:131013));
hold on;
tt=zeros(1,131013);
for i=10:131013
    a=i-9;
    if(sum(t(a:i))>3)
        tt(i)=1;
    end
end
plot(tt(97912:131013));
hold on;
ttj=(r>2000)|(tr>400000)|(f>200);
plot(ttj(97912:131013));
legend('ÆÕÍ´³´³´','¸,ÛpÈ¼±´³´','¼«Öµ±´³´');
title('×Ô°Í±´³´');
```