# 业务功能响应异常检测算法文档

## 一、设计原理

业务功能响应可能会发生如下异常：response响应时间超时、response的返回协议大小异常。

考虑到同一个业务功能页面的响应内容大多会相似，故此处不考虑检测每一个业务功能的响应内容异常检测。

## 二、数据源

### （1）训练数据

#### 训练样本生成/采集方式

获取hdfs上的业务系统最近一个星期的https记录，以此作为训练样本。

#### 数据源地址

Hdfs：10.130.10.62

路径：hdfs://10.130.10.62:9000/usr/hadoop/nta/http/

#### 数据源格式

|  |  |
| --- | --- |
| 字段 | 备注 |
| flow\_id | id |
| srcip | 源ip |
| dstip | 目的ip |
| host | 域名 |
| protocol | 协议 |
| request\_date | 请求时间 |
| response\_date | 响应时间 |
| request\_urL | 请求url |
| request\_uri | 请求uri |

### （2）检测数据

#### 检测数据生成/采集方式

获取hdfs上的业务系统最近一小时的https记录，以此作为检测数据。

#### 数据源地址

Hdfs：10.130.10.62

路径：hdfs://10.130.10.62:9000/usr/hadoop/nta/http/

#### 数据源格式

|  |  |
| --- | --- |
| 字段 | 备注 |
| flow\_id | id |
| srcip | 源ip |
| dstip | 目的ip |
| host | 域名 |
| protocol | 协议 |
| request\_date | 请求时间 |
| response\_date | 响应时间 |
| request\_urL | 请求url |
| request\_uri | 请求uri |

## 三、特征处理

1）业务功能定义

业务功能指业务系统中带有提交参数的子功能或URL。对于常见的URL，一般语法格式为(带方括号[]的为可选项)：

protocol :// hostname[:port] / path / [;parameters][?query]#fragment

我们将其写成以下格式：

protocol://host:path?key1=value1&key2=value2...

**定义方法一：**

由于http/https有get和post两种请求方式。对于get型请求，我们直接获取URL来处理；对于post型请求，请求参数分为两种：url上的参数和表单上的参数。我们将url上的参数和表单的参数拼接一起，作为总体参数，即构造出了带有完整参数的URL。

这里我们定义的业务功能有以下三点的要求：

①相同host

②相同path

③在三个不同IP都访问了这个host+path的前提下，存在共同使用的key（可选）

对于满足以上三点的url，构造“host+path+共同使用的key”以此作为业务功能。具体例子如下：

IP1-URL1:https://113.108.195.66:28080/WorkFlowArea/WorkFlowRun/Index?action=add&format=normal&display=1&username=232445467

IP2-URL2:https://113.108.195.66:28080/WorkFlowArea/WorkFlowRun/Index?action=add&format=normal&username=231145467

IP3-URL3:https://113.108.195.66:28080/WorkFlowArea/WorkFlowRun/Index?action=add&username=23221467

IP4-URL4:https://113.108.195.66:28080/WorkFlowArea/WorkFlowRun/Index?action=del&username=23221207

那么最终定下来的业务功能是：

113.108.195.66:28080/WorkFlowArea/WorkFlowRun/Index?action=add

**定义方法二：**

客户配置。

2）请求参数的定义

①GET型

对于GET型请求，url上的参数就是其请求参数。

请求网址例子：

https://113.108.195.66:28080/WorkFlowArea/WorkFlowRun/Index?flowid=7885774a-b16e-4073-9532-b3295ce8169f&stepid=624480e4-da39-460a-86cf-a134f33f68ab&instanceid=8b78638d-2f8f-431e-a936-b5015ff9fffd&taskid=6ef99934-925c-40fc-9105-37404412d8b8&groupid=51955654-e37b-4eef-b612-f09e647c059b&display=1

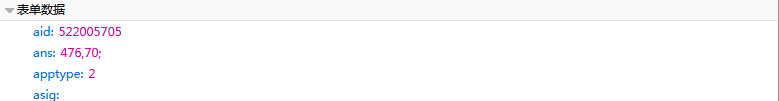
②POST型

对于POST型请求，需要结合url中的参数和表单数据，以此作为完整请求参数。

请求网址例子：

<https://ssl.captcha.qq.com/cap_union_new_verify?random=1521622298522>

表单数据：



## 四、算法模型

### （1）算法名称

无

### （2）算法详情

无

## 五、训练过程详情步骤

### （1）训练周期

一周训练一次

### （2）训练步骤

注意：此处使用到《业务功能参数异常检测算法文档》中的基线结果数据中的业务功能uri字段。

①响应时间基线

Step1：识别出业务功能后，筛选出对应的http协议记录。该记录已经关联了query对应的response信息；

Step2：选出request\_date和response\_date，计算两者之差，作为该业务功能的响应时间，得到一个响应时间的list；

Step3：计算响应时间list的mean值、var值，mean+3var作为响应时间基线。

②响应报文大小基线

Step1：识别出业务功能后，筛选出对应的http协议记录。该记录已经关联了query对应的response信息；

Step2：选出content\_length，此作为响应http报文的大小。每一个业务功能将得到一个报文大小的list；

Step3：计算响应报文大小list的mean值、var值，mean+3var作为响应报文大小基线。

③将基线结果存储到postgrel

## 六、检测过程详情步骤

### （1）检测周期

一小时检测一次

### （2）检测步骤

注意：需要训练好基线后，才能运行检测部分的代码。

① 响应时间检测

Step1：当新的业务功能数据产生，过滤出对应的response记录；

Step2：选出request\_date和response\_date，计算两者之差，作为该业务功能的响应时间；

Step3：判断新的响应时间是否在基线时间的mean+3va内，若不在，则进行响应时间异常告警。

② 响应报文大小检测

Step1：当新的业务功能数据产生，过滤出对应的response记录；

Step2：选出content\_length，此作为响应http报文的大小；

Step3：判断新的响应报文大小是否在基线的mean+3va内，若不在，则进行响应报文大小异常告警。

③ 保存结果

## 七、输出结果

### （1）模型保存位置

基线结果保存到Postgrel，表结构为

|  |  |
| --- | --- |
| id | Row id |
| app\_utility | 业务功能（没用到此字段） |
| app\_uri | 业务功能中的uri |
| app\_param | 业务功能中的参数（没用到此字段） |
| param\_name | 除业务功能外的参数名称（没用到此字段） |
| param\_length | 参数长度范围（没用到此字段） |
| param\_type | 参数类型（没用到此字段） |
| response\_baseline | 响应基线值（3σ） |
| response\_type | 响应基线类型（size表示响应报文大小基线，time表示响应时间基线） |
| time | 保存时间 |
| result\_type | 结果类型（0表示参数检测结果，1表示响应结果） |

### （2）检测结果

#### 保存方式

保存到POSTGRESQL或HDFS

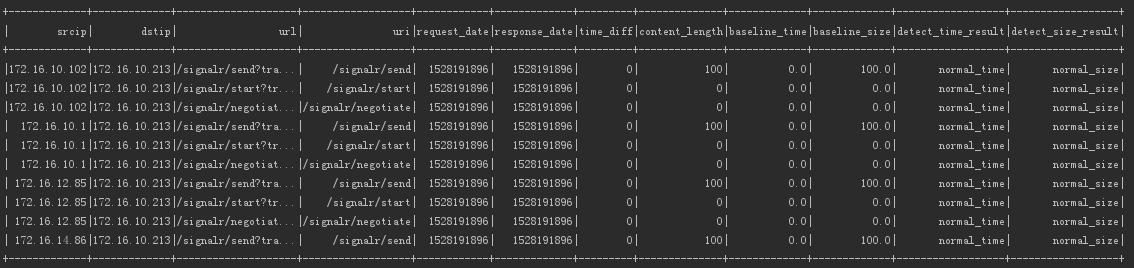
#### 保存位置

POSTGRESQL：10.130.10.22

#### 结果格式

|  |  |
| --- | --- |
| 字段 | 备注 |
| flow\_id | id |
| srcip | 源ip |
| dstip | 目的ip |
| host | 域名 |
| protocol | 协议 |
| request\_date | 请求时间 |
| response\_date | 响应时间 |
| request\_url | 请求url |
| app\_utlity | 业务功能 |
| app\_uri | 请求uri |
| app\_para | 参数字符（没用到此字段） |
| extra\_app\_para\_baseline | 参数的基线  （key1=value1&len\_min-len\_max&char#key2=value2&len\_min-len\_max&char）（没用到此字段） |
| parameter\_detect\_result | 参数检测结果（异常长度、异常类型）（没用到此字段） |
| baseline\_time | 响应时间基线 |
| baseline\_size | 响应报文大小基线 |
| detect\_time\_result | 响应时间检测结果 |
| detect\_size\_result | 响应报文大小检测结果 |
| result\_type | 结果类型（0为参数检测结果，1为响应检测结果） |

#### 结果示例



## 模型评估

无