文章编号：1005-8451（2015）11-0035-03

通过日志比对快速定位业务问题的方法

苏　燕
（哈尔滨铁路局　信息技术所，哈尔滨　150008）

摘　要：本文提出一种通过日志比对快速定位业务问题的方法，通过构建高效的日志处理流程，过滤噪音数据，对关键字进行人工识别，实时地对日志信息进行解析，准确定位业务问题，并根据业务问题提供解决方法，提升了业务故障处理效率，降低了人为维护带来的不便。

关键词：日志；数据库；语义识别；故障处理
中图分类号：U293.22：TP39　　**文献标识码：**A

Rapid positioning for business problems by log comparison

SU Yan
(Institute of Information Technology, Harbin Railway Administration, Harbin　150008, China)

Abstract: This paper proposed a method of rapid positioning for business problems by log comparison. It was built a high efficiency treatment process, filtered the noise data, recognized the keywords artificially, analyzed the log information in real time, posited the business problems accurately, provided solutions according to the business problems, improved the efficiency of business failure process greatly, reduced the inconvenience for artificial maintenance.

Key words: log; database; semantic identification; fault treatment

随着中国铁路客票发售和预订系统（简称：客票系统）的迅速发展，网络规模不断扩大，主机、网络设备、应用软件数量不断增多，系统复杂性已经到了运维监控人员难于完全掌控的程度。由于系统复杂、模块增加、更新频繁，业务故障问题时有发生，造成业务中断，用户感知差。若仅依赖现有的网络、设备、应用、业务监控难于真正快速发现业务故障问题。因此，当前的监控已经不能适应企业用户目前和未来业务发展的要求，而这些业务问题故障经常隐藏在某些日志信息中，急需建立一套通过日志快速发现故障问题的管理系统，来迅速解决业务问题。

1　现有开发模式分析

为了发现系统或业务故障与风险，传统方式主要有两种：（1）通过监控系统，实时通过代理（AGENT）或无代理协议，采集系统和业务数据，以一种报文或消息（TRAP）方式主动发送故障或问题，最后进行告警处理；（2）收集系统或业务日志信息，通过关键词对日志信息进行审计，发现问题，进行告警处理。针对业务问题的监控与故障定位，当前的方案一般有以下几种：

（1）在硬件设备部署无代理或AGENT，通过采集操作系统、业务应用等数据信息进行系统监控，发生故障时，进行告警处理；

（2）通过模拟用户操作的方式，周期性模拟用户处理相关业务，监测业务系统是否正常；

（3）实时监控网络信息，还原用户业务操作，监控是否发现问题。

目前的技术方案中，有很多快速发现业务故障与问题的方法，如果是通过监控系统和业务，发现问题经常迟于问题出现，引起客户投诉后才开始部署相关监控，制定相关故障问题监控规则。为应对市场的瞬息万变，业务系统需要经常的更新，而目前的业务故障定位往往处于被动局面，缺少一套有效的方法。

一方面，由于铁路售票业务系统的迅速发展，系统和网络规模迅速扩大，主机、网络设备、应用软件数量不断增多，业务资源访问、操作量不断增加，

收稿日期：2015-04-10
作者简介：苏　燕，工程师。

造成系统复杂，当发生业务或系统故障时，很难定位问题的根本原因，影响售票业务，旅客投诉量大；

另一方面，传统的监控只是针对历史发生过故障进行监控，难于定位问题的根本原因，特别是更新或新部署的系统和业务，更难于发现故障和问题，只有被动的等待问题发生，再让系统、业务等所有的人员全部核查，费时费力，且效果一般。

2 技术实现方案

针对现有技术存在的问题，通过专门日志管控平台，收集系统平台、数据库、中间件与应用程序的日志，把收集的日志进行分类，即数据库类型、中间件类型、操作系统，并进行规范化处理；针对日志按类型进行关键词分析，如果发现异常关键字，即进入风险处理流程；完成标准信息比对后，再与标准封装模板进行模糊比对，找出差异文本，针对差异项，过滤正常差异处理内容。如时间差异，对当前处理数值项范围进行分析，超出范围即进入风险处理流程;最后进行异常语义识别，如果识别成功，则认为有相关风险，进入风险处理流程；下文着重介绍日志处理平台的架构和处理机制。

日志管控平台主要分数据采集层、数据处理层和数据分析层，通过数据处理引擎将需要接入的各类数据引入，对日志进行分类，按每一类应用，建立日志分析比对模板与分析流程，以及日志信息模糊比对风险识别的数学算法、语义分析的方法，快速通过日志信息发现业务问题,整体方案如图 1 所示。

步骤 1：通过专门日志收集平台，收集系统平台、数据库、中间件与应用程序的日志。由于日志信息较大，一般按文件指针、时间戳、或以文件比对方式进行增量的收集。

步骤 2：把步骤 1 中收集的日志进行分类处理，如按数据库类型、中间件类型、操作系统、应用系统与环节类型，进行规范化处理，去掉时间、空行等相关信息，形成规范的日志分析文本。

步骤 3：在完成步骤 2 日志文本规范化后，首先针对日志按系统情况进行异常关键词分析，也称异常关键词排查，如果发现与配置中有相同的异常关键词，与传统处理一样，即进入风险处理流程。

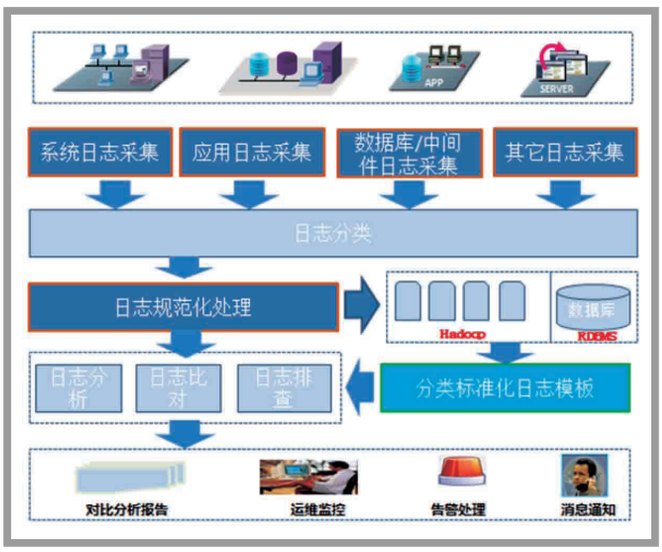

图1 日志监控处理平台

步骤 4：在完成上述排查后，如果没有发现系统或业务问题，再与标准封装模板进行比对，找出差异文本。针对差异项，去除正常差异处理内容，如时间差异，在基本的业务处理数值项范围进行分析，超出范围即进入风险处理流程。

步骤 5：在完成步骤 4 的差异比对与分析后，进行异常语义识别，识别的流程如图 2 所示，如果识别成功，则认为有相关风险，进入风险处理流程。

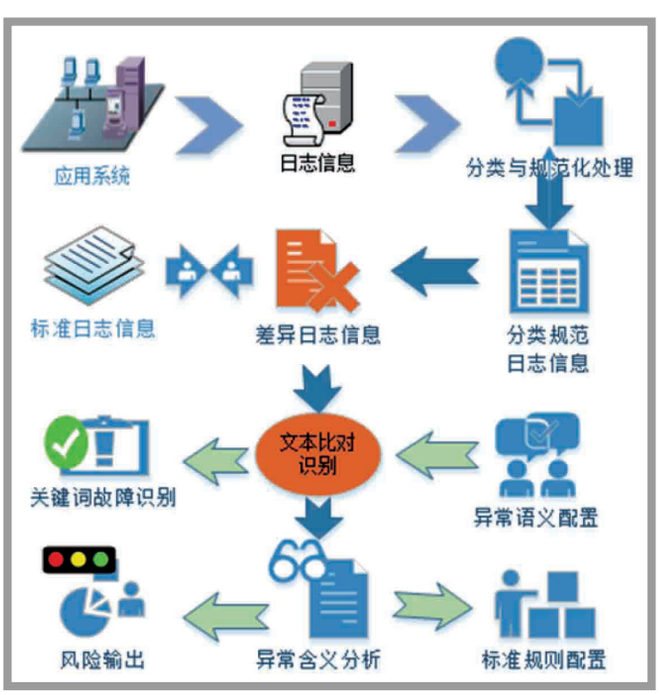

图2 异常语义识别流程

如图 2 所示，在异常文本识别时，首先需要读取用户的异常语义配置信息。异常语义配置如表 1 所

示，包括风险编号、风险故障、风险故障等级、风险 WORD、识别权重、关键标识、识别方式、风险率下限值、风险率等信息。

按风险编号进行循环，在循环内，按风险 WORD 进行逐一处理，处理的方法为“识别方式”，一般值为“包括”、“排除”等；处理成功即为 1，否则为 0；再检查当前风险 Word 的识别权重、关键标识，综合以上信息，确定某一日志的风险概率计算公式，如式（1）：

$$P=k\bullet(A_1\bullet\omega_1+A_2\bullet\omega_2+\cdots+A_n\bullet\omega_2)\bullet 1/m \quad (1)$$

说明：

（1）m 表示 A_i 取非零值的个数；

（2）k 的定义如下：若日志中存在关键风险 WORD，k 值为 1，否则 k 值为 0；

（3）A_i 的定义如下：若日志中存在关键词与其匹配，A_i 取值 1，否则取值 0。

系统整体处理流程如图 3 所示。

进行日志分类，即数据库类型、中间件类型、操作系统，并进行规范化处理；针对日志按类型进行关键词分析，如果发现异常关键字，即进入风险处理流程；完成标准信息比对后，再与标准封装模板进行模糊比对，找出差异文本，针对差异项，去除正常差异处理内容，如时间差异，接着对当前处理数值项范围进行分析，超出范围即进入风险处理流程；最后进行异常语义识别，如果识别成功，则认为有相关风险，进入风险处理流程。

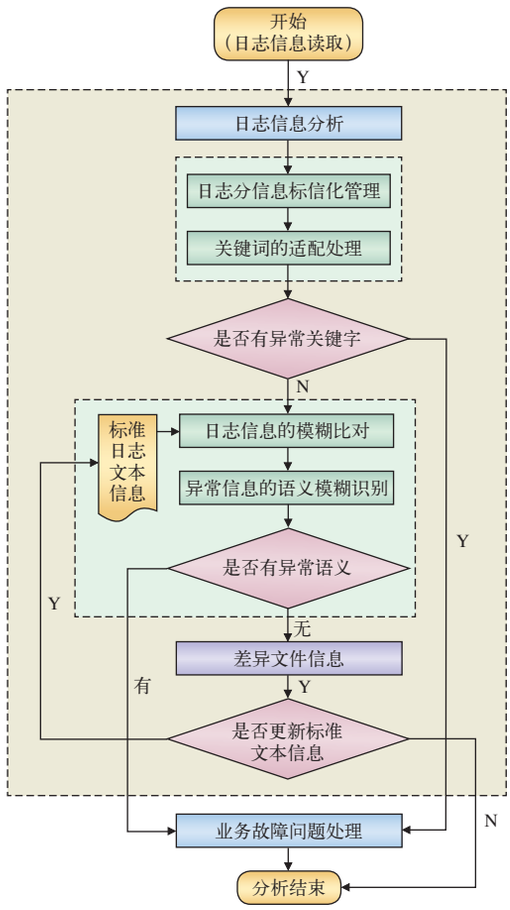

图3　系统处理流程图

3　结束语

本文通过构建一种高效的日志处理流程，准确、实时的对日志信息进行解析。由于对日志信息进行规范化与标准化处理，与传统的日志异常关键词分析方法相比，处理效率提升了 27%。此外，通过采用异常语义配置与分析算法，业务故障诊断准确率从原来的 19% 提高到 78% 以上，提升了 4 倍左右，具有一定的实用价值。

表1　异常语义配置信息表

风险编号	风险故障	风险等级	风险WORD (A_i)	识别权重(ω_i)	关键标识 (K)	识别方式	风险率下限 (P)	风险率(P)
A001	CTMS进程异常	1	rating_3010_gsm	1	Y	包括	80%	
A002	CTMS进程异常	1	slop over	0.6	N	包括	80%	
A003	CTMS进程异常	1	break down	0.5	N	包括	80%	
C001	Sybase数据库回滚段不可用	1	SYB-01545	1	Y	包括	80%	
C002	Sybase数据更新异常	3	bad SQL grammar	0.6	N	包括	80%	

参考文献：

[1] 孙林超，陈　群．专业的工作日志管理软件使用与比较分析报告 [EB/OL].http://wenku.baidu.com, 2014-08-04.

[2] 影儿悠悠．如何使错误日志更加方便地排查问题 [EB/OL]. http://blog.jobbole.com/84752/ , 2015-03-09.

[3] guangyue_qin. 开源日志系统比较：scribe、chukwa、kafka、flume[EB/OL].http://wenku.baidu.com, 2014-11-04.

责任编辑　王　浩