**详细设计说明书**

**Detailed Design Specification**

**编号**： **TMP-DBD**

**版本 1.0**

|  |  |  |  |
| --- | --- | --- | --- |
| 作者： |  | 日期： |  |
| 审批： |  | 日期： |  |

**变更记录**

|  |  |  |  |
| --- | --- | --- | --- |
| 日期 | 版本 | 变更说明 | 作者 |
|  | 1.0 | 创建 |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

目录

[1 引言 1](#_Toc112819340)

*[1.1 编写目的](#_Toc2074070726)* [1](#_Toc2074070726)

*[1.2 背景](#_Toc952133778)* [1](#_Toc952133778)

*[1.3 术语定义](#_Toc1611753049)* [1](#_Toc1611753049)

*[1.4 参考资料](#_Toc374771285)* [1](#_Toc374771285)

[2 详细设计说明 3](#_Toc211450344)

*[2.1 时序异常检测Web前台系统](#_Toc1907979470)* [3](#_Toc1907979470)

[2.2 用户登录功能设计（编号：AD1001） 3](#_Toc1185135286)

*[2.3 时序异常检测后台系统](#_Toc657925877)* [4](#_Toc657925877)

[3 系统出错时备选设计 5](#_Toc366916336)

[3.1 问题 5](#_Toc1337308615)

# 引言

## *编写目的*

本说明书使用类图，伪代码等手段，详细说明各个模块的设计构思，是开发人员的重要参考，在开发阶段，应严格按照本说明书进行开发。

## *背景*

|  |  |  |  |
| --- | --- | --- | --- |
| 项目名称 | 时序异常检测系统 (Time series anomaly detection System) | | |
| 项目简称 | TSADS | 项目编号 | IGS09001 |
| 任务提出者 |  | | |
| 开发者 | 池浪，史蒙蒙 | | |
| 用户 | 系统运维人员 | | |
| 同其他系统的关系 |  | | |

## *术语定义*

时间序列：时间序列（或称动态数列）是指将同一统计指标的数值按其发生的时间先后顺序排列而成的数列。时间序列分析的主要目的是根据已有的历史数据对未来进行预测。经济数据中大多数以时间序列的形式给出。根据观察时间的不同，时间序列中的时间可以是年份、季度、月份或其他任何时间形式。

异常检测：异常检测是指将用户正常的习惯行为特征存储在数据库中，然后将用户当前的行为特征与特征数据库中的特征进行比较，如果两者的偏差足够大，则说明发生了异常。异常检测方法首先定义一组系统处于“正常”情况时的数据，如CPU利用率、内存利用率、文件校验和等然后进行分析确定是否出现异常。

## *参考资料*

[1]Lynda Khiali,Mamoudou Ndiath,Samuel Alleaume,Dino Ienco,Kenji Ose,Maguelonne Teisseire. Detection of spatio-temporal evolutions on multi-annual satellite image time series: A clustering based approach[J]. International Journal of Applied Earth Observations and Geoinformation,2019,74.

[2]Mengqi Zhang,Xin Jiang,Zehua Fang,Yue Zeng,Ke Xu. High-order Hidden Markov Model for trend prediction in financial time series[J]. Physica A: Statistical Mechanics and its Applications,2019,517.

[3]Stuart K. Grange,David C. Carslaw. Using meteorological normalisation to detect interventions in air quality time series[J]. Science of the Total Environment,2019,653.

[4]Wei Lu,Wei Zhou,Dan Shan,Liyong Zhang,Jianhua Yang,Xiaodong Liu. The linguistic modeling of interval-valued time series: A perspective of granular computing[J]. Information Sciences,2019,478.

[5]Katarzyna Kaczmarek-Majer,Olgierd Hryniewicz. Application of linguistic summarization methods in time series forecasting[J]. Information Sciences,2019,478.

[6]Valdério Anselmo Reisen,Adriano Marcio Sgrancio,Céline Lévy-Leduc,Pascal Bondon,Edson Zambon Monte,Higor Henrique Aranda Cotta,Flávio Augusto Ziegelmann. Robust factor modelling for high-dimensional time series: An application to air pollution data[J]. Applied Mathematics and Computation,2019,346.

# 详细设计说明

## *时序异常检测Web前台系统*

### 用户登录功能设计（编号：AD1001）

#### 程序描述

设置用户登录模块，分为三类用户：Administer、User和Others。Administer拥有最高权限，可以管理系统接入的数据库，通过增删数据库，选择系统需要检测哪些数据库，同时Administer可以查看系统运行状态，系统相关日志记录会暴露给Administer，Administer可以根据相关日志进行分析和管理系统运行状态，包括暂停运行、调节运行参数等。User具有查看系统实时异常检测状态的权限以及向系统中提交样本及进行分段异常检测。Others只具有查看实时异常检测权限。

#### 功能概述

主要包括三个功能点。

1. 实时监控接入时序异常检测系统的数据库。通过Administer操作向系统中接入时序的数据库，由系统监控数据库的数据变动情况，实时读取数据库中的增量数据，通过后台相关时序异常检测算法对增量数据进行分析，最后将增量数据及分析结果发送给前台显示出来。
2. 通过Administer和User上传时序样本，对单样本文件进行时序分析并产生分析报告。主要侧重于对存疑数据进行单独分析，并产生算法测试结果。
3. 对选定的历史时间段进行时序异常检测，依赖于用户对系统中某一阶段产生的数据进行异常检测，产生异常结果报告。

#### 输入项

主要包括两类输入项：

1. 系统接入的数据源实时产生的在线数据。
2. 用户对系统进行操作时产生的操作指令数据。

#### 输出项

主要包括两类输出：

1. 时序异常检测实时数据产生的走线图及异常点标注。
2. 对单样本和选中时间段进行单独分析时产生的异常分析结果和日志文件。

#### 处理逻辑

数据库将实时采集的数据存入，后台将数据库中采集的数据取出，进行数据格式化后由后台任务调度模块将数据传递给时序异常检测算法模块进行检测，算法模块将检测结果反馈给任务调度模块，然后由任务调度模块将分析数据及异常数据打包分发给前端进行显示。

#### 接口

主要包括三个接口：

1. 后台和数据库的DAO层接口，
2. 后台和算法模块之间的处理逻辑接口，
3. 后台和前端进行交互的HTTP接口

#### 存储

使用ElasticSearch和Influxdb两种数据库。

一、ElasticSearch主要用于运行日志的存储和查询。

二、Influxdb主要用于对时序数据的存储和查询。

## *时序异常检测后台系统*

# 系统出错时备选设计

### 问题

1. 后台系统有多个线程跑算法时，可能会影响运行效率，导致运算变慢。
2. 后台要保持和数据库维持连接不中断，以保证能够实时检测，如果发生中断，需要具备断点续连功能。

#### 备用技术

1. 使用全局模式，后台使用一个全局变量，由一个manager程序将实时数据交付算法运行，然后由manager将检测结果分发给各类前端进行显示。这样就隔离了前端对后台算法模块的影响，将多个前端的操作归由后台一次处理，然后分发处理结果。
2. 增加一个数据库中断异常处理模块，发生中断异常时，保存运行上下文，并写入运行日志和恢复文件中，并产生告警。当排除异常后，由恢复文件进行恢复操作，完成断点续连。

#### 可能的影响

1. 使用全局模式时，消除了用户的操作参与感，由后台进行统一管理，设计不够人性化。
2. 无法做到自愈，有可能无法及时产生恢复文件，导致产生数据移位等错误情况。