角色分类引擎源代码概要设计说明

# 概述

## 1.1 简介

角色分类程序的主要功能是将聚类的结果分词提取特征之后，通过SVM模型预测出输入文本的角色类别信息，并将预测结果写入到输出文件。

## 1.2 读者对象

本文档是对角色分类引擎的内部代码流程和相关技术点的整体说明，因此读者最好满足如下几点条件后再读此文档

1）有一定的JAVA开发基础

2）对语义理解和SVM算法有一定了解

## 1.3 术语定义

SVM，support vector machines：支持向量机

## 1.4 参考资料

Paper：

基于概率语义分布的短文本分类 作者：马成龙

svm：https://zhuanlan.zhihu.com/p/77750026

# 2. 总体设计

角色分类引擎模块流图如图所示：

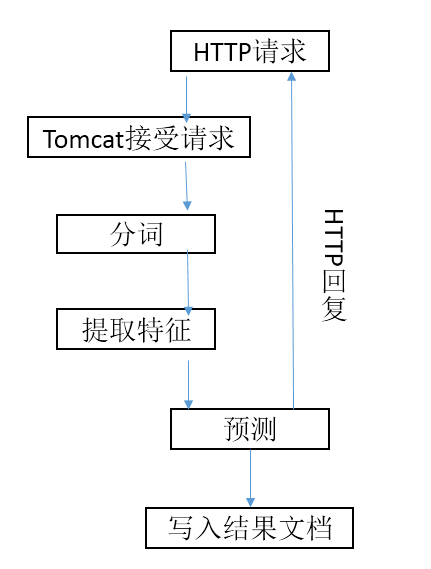


图2.1 角色分类引擎流程图

如上图所示，引擎采用http协议发送请求，通过角色分类模块的处理，最终以http协议返回最终结果。

# 3. 可靠性设计

1）函数返回值规范，对应程序里的异常（如：文件打开失败等），通过函数返回值的形式逐层向上反馈，避免直接退出引擎；

2）支持多线程异步调用，加快处理速度。

# 4. 扩展性设计

## 4.1 支持跨平台

目前引擎支持intel、ARM平台上的编译运行。

# 5. 安全性设计

引擎对各类异常进行判断，有一定的容错机制。

# 6. 易用性设计

引擎接口设计简洁，方便开发者快速熟悉引擎接口调用流程，快速完成接口集成；

引擎同时提供了接口测试程序，可供开发者参考。