聚类引擎源代码概要设计说明

# 概述

## 1.1 简介

### 1.1.1引擎功能介绍

离线客服聚类引擎主要应用于智能客服领域，处理单声道语音，通过聚类算法，标记出同一个说话人的语音分段，通过语义理解，判断说话人所属类别，便于上层业务层进行质检分析。

适用场景及性能：

1）适用于单声道、两个人通话的客服语音

2）目前正常实网客服场景分割性能达到85%-90%左右

### 1.1.2引擎输入介绍

（1）资源文件，包括：

1）配置文件configure.cfg

2）d-vector模型及相关特征提取配置文件Decode.cfg、feat.cfg

（2）语音数据及识别结果\_sent.txt文件

编码格式要求：8k\_16bit\_pcm。

引擎的输入为语音及对应的识别结果\_sent.txt文件。

### 1.1.3引擎输出介绍

语音目录下会输出\_1.txt、\_2.txt及\_ABC.txt，分别为第一个人对应的分段语音、第二个人对应的分段语音以及最后的综合的聚类结果。

## 1.2 读者对象

本文档是对聚类引擎的代码流程和相关技术点的整体说明，因此读者最好满足以下几个条件：

1）有一定的C和C++语言基础；

2）了解相关聚类算法。

## 1.3 术语定义

AHC，Agglomerative Hierarchical Cluster，凝聚层次聚类

K-means，K均值聚类

Spectral Cluster，谱聚类

## 1.4 参考资料

Paper：Deep Speaker Feature Learning for Text-independent Speaker Verification

# 2. 总体设计

聚类引擎的基本框架图如下：



图2.1 聚类引擎基本架构图

语音数据首先经过语音识别后，被切分为分段语音，传统方法采用LSP特征，计算GLR距离，采用层次聚类和K-means方法进行聚类，得到相应的聚类簇；随着深度学习的发展，开始采用神经网络特征向量d-vector来表征说话人特征，并引入谱聚类算法对聚类结果进行修正，提升聚类正确性。

目前引擎采用神经网络d-vector特征，主要模块包括：

特征提取：

1） 提取60维fbank特征

2） 将60维fbank特征输入到神经网络，神经网络采用TDNN网络结构，提取隐层特征作为说话人特征向量，即d-vector

距离公式：

1） 神经网络d-vector 采用cosine离散余弦公式，计算分段之间的相似度

聚类模块：

1） 首先采用层次聚类算法确定聚类中心

2） 利用K-means算法对聚类中心进行修正

3） 在采用神经网络d-vector特征时，我们还引入了谱聚类算法，可一定程度上提高聚类算法的准确性。

# 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、内存分配、释放异常等），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。

## 4.2 支持SSE、neon指令加速

SSE指令加速：X86架构CPU特有的，通过宏定义“USEIPPFUNCTION”开启。

NEON: ARM架构CPU特有， 通过宏定义“USE\_NEON”开启。

# 5. 安全性设计

引擎对各类异常进行判断，如语音格式、语音长短等，保证程序正常运行。

# 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。