需求规格说明书

**酒店收集分析**

**小组：**

**姓名： 黄锦波 学号： 2017324236**

**姓名： 郭奇东 学号： 2017324217**

**姓名： 谢兴悦 学号： 2017324204**

**姓名： 沈小玉 学号： 2017324218**

**姓名： 谢金英 学号： 2017324225**

**版本：V1.0**

**发布时间：2020-09-11**

**岭南师范学院 信息工程学院**

**深圳信盈达科技有限公司 人才实训学院**

**文件变更记录**

\*A - 增加 M - 修订 D - 删除

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **版本** | **作者** | **日期** | **修订描述** | **变更记录** |
|  | 黎清湖 | 2019/03/05 |  | 撰写初始版本 |
|  | 陈君林 | 2019/03/06 | M | 修订需求说明书格式 |
| V3.0 | 陈君林 | 2019/03/27 | M | 修订项目背景，功能需求概述，功能列表说明 |
| V3.1 | 黄文轩 | 2019/04/06 | M、A | 修订项目背景，增加项目总结 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**目录** *(居中, 小二号粗宋体)*

*(目录要由Word系统自动生成)*

[1 引言 5](#_Toc20910)

[1.1 编写目的 5](#_Toc20798)

[1.2 读者对象 5](#_Toc1317)

[1.3 编写原则 5](#_Toc18715)

[1.4 术语与缩写解释 5](#_Toc14312)

[2 项目概述 6](#_Toc28259)

[2.1 项目背景 6](#_Toc27009)

[2.2 项目目标 7](#_Toc26343)

[2.3 项目范围 7](#_Toc803)

[2.4 客户与最终用户介绍 7](#_Toc3860)

[3 功能需求概述 7](#_Toc12164)

[4 功能需求详细说明 8](#_Toc3800)

[4.1 功能列表说明 8](#_Toc18086)

[4.2 业务分析—综合查询 9](#_Toc5837)

[4.2.1 功能概述 9](#_Toc5326)

[4.2.2 功能点描述 9](#_Toc4061)

[4.3 业务分析—全量筛选 9](#_Toc15920)

[4.3.1 功能概述 9](#_Toc6310)

[4.3.2 操作流程 10](#_Toc19750)

[4.3.3 功能点描述 10](#_Toc18287)

[4.4 业务分析—系统筛选 11](#_Toc9959)

[4.4.1 功能概述 11](#_Toc7287)

[4.4.2 操作流程 11](#_Toc17169)

[4.4.3 功能点描述 12](#_Toc12242)

[4.5 业务分析—热点分析 12](#_Toc9361)

[4.5.1 功能概述 12](#_Toc16191)

[4.5.2 功能点描述 12](#_Toc18083)

[4.6 业务分析—交叉分析 13](#_Toc9112)

[4.6.1 功能概述 13](#_Toc1798)

[4.6.2 功能点描述 13](#_Toc12999)

[4.7 业务分析—语音特征分析 14](#_Toc7830)

[4.7.1 功能概述 14](#_Toc31336)

[4.7.2 操作流程 14](#_Toc997)

[4.7.3 功能点描述 14](#_Toc24706)

[4.8 业务分析—聚类规则 15](#_Toc12185)

[4.8.1 功能概述 15](#_Toc25341)

[4.8.2 操作流程 15](#_Toc14471)

[4.8.3 功能点描述 16](#_Toc13240)

[5 性能需求说明 17](#_Toc11754)

[系统运行指标 17](#_Toc7793)

[高可用性要求 17](#_Toc2097)

[6 其它需求说明 18](#_Toc20195)

[其它需求 18](#_Toc3367)

[新系统开发、运行条件与限制 18](#_Toc188)

[7总结 19](#_Toc10894)

(目录行距20磅固定值，宋体，五号)

# 引言

## 编写目的

编写本文档是为了帮助项目小组明确本项目的详细需求，以细化项目计划，评估项目成本，实现系统设计和开发，编写测试用例，减少需求变更，保证项目开发成功。

## 读者对象

本文档适合以下人员阅读：

* 本项目科技与业务的项目经理；
* 本项目相关业务需求人员和测试人员；
* 本项目相关科技设计和开发人员；
* 本项目相关数据中心支持人员；
* 本项目相关行领导和业务部门领导；
* 科技部相关领导；
* 科技研发管理部、风险合规部以及相关内外审计；
* 经授权的本项目相关合作公司人员；其他相关人员。

## 编写原则

本文档将作为测试和需求变更的依据。对于本文档指明应达到的需求系统未实现的，属于缺陷；对于本文档未细化的需求，若开发完成后需要调整，属于需求变更，需发起项目变更管理流程（小于一人天的调整，诸如界面字体大小颜色样式等，无需发起项目变更流程）。整体开发过程中，需求变动不得超过10%。

## 术语与缩写解释

【简要说明本文档中涉及的专门术语、容易引起歧义的概念、关键词缩写及其他需要解释的内容。】

|  |  |
| --- | --- |
| **缩写、术语** | **解 释** |
| 语音转译 | 将录音文件转译成录音文本 |
| 热词 | 出现频率较高的关键词 |
| 抽样 | 抽取录音样本 |
| 关键词 |  |

# 项目概述

## 项目背景

客服热线是电信运营商客户服务的主要平台，其服务量通常占运营商客户服务的85%以上，在运营商进入全业务竞争阶段后，服务质量的竞争也日益关键。需要以改善客户感知为着力点，积极采取各种举措，构建为客户导航的服务体系，实现服务质量的持续优化。目前客户热线主要包括传统按键模式和人工服务，传统按键模式因业务的增加导致用户的放弃率逐渐上升，已经越来越无法满足全业务发展的需要。同时随着客户基数的增多，导致人工服务的压力迅速增长，由此带来的人工成本及管理成本居高不下。

为提升用户感知、降低人工成本，本文重点对电信运营商构建智能语音客服平台进行探讨，主要从以下几个方面进行研究分析：传统服务模式与智能语音交互模式对比分析、语识别技术、智能语音客服系统架构、评价标准、未来展望等。

近年来，语音识别技术的应用日益成熟，已经从原本简单的语音转文本(Speech to Text)转化为各类更高级、对业务价值更大的应用，而其中非常重要的一项应用就是语音分析。

目前呼叫中心的录音主要是为了满足合规的需要，并在客户与企业发生纠纷时提供佐证而规避风险；另一层面，呼叫中心专门的质量管理团队会通过对录音的随机采样抽听，对座席的服务表现进行评价，是衡量座席每月的绩效表现的重要因素。但是，由于人工调听需要占用大量的时间，业界普遍的录音抽检的概率均只在 1%左右，而剩余的 99%的录音都没有发掘利用。

随着客服中心客户服务中心的不断发展，电话呼入、呼出服务等各类服务渠道是客服中心和消费者等用户接触的最常规的途径。多渠道的拓展以及业务量的增长，客服中心队伍也日渐扩大。日常的人工质检已无法满足业务增长的需要，通过引进智能语音分析系统，围绕关键词自动将录音转为结构化文本索引，对关键词进行检索、筛选和归类。通过系统挖掘、分析有价值信息，快捷、简单、方便、准确地定位焦点问题，快速反应客户需求。

质检人员人工抽查测听方式进行录音质检，目前存在以下问题：

1. 录音利用率低：仅有不到1%的录音被抽检
2. 质检方式效率不足：每条录音质检耗时4分钟，发现问题比例小于5%
3. 专项质检能力弱：对于专业业务质检和客户需求分析无法承载
4. 录音中90%以上的有价值的客户反馈缺乏有效检查和分析手段

智能语音技术（包括语音识别、语音合成和语音分析等）及应用的日益成熟，为呼叫中心服务运营水平的进一步提升拓宽了通道，使得开展智能语音质检和分析成为可能。

文本分析是指用自然语言处理、文本挖掘以及计算机语言学等方法来识别和提取原素材中的主观信息.简单地说是对带有情感色彩的主观性文本进行分析、处理、归纳和推理的过程。这是基于Python的SnowNLP库的文本分析。SnowNLP(NLP, Natural Language Processing)是python中用来处理文本内容的，主要包括文本分类, 语音识别, 中文自动分词, 机器翻译, 文本校对、信息抽取等方面的技术。当然也可以使用功能更强大的jieba分词库，对中文有着强大的分词能力，它的优点在于不仅支持三种分词模式（精确模式、全模式、搜索引擎模式），而且还支持自定义词典。此外，信息抽取 (IE, Information Extraction) 可从非结构化或半结构化文本中快速识别并提取关键信息，在这里指的是关键词的提取。

## 项目目标

全国集中的智能语音客服，给用户带来更加新颖便捷的体验、降低人工服务成本，将逐步成为运营商客户服务的主门户主渠道。智能语音客服核心功能是提供自然语言对话式服务，在此基础上进一步开展语音数据挖掘和业务拓展。a）在线路资源、语料库方面：可以为中小企业提供智能语音交互的线路资源依赖于沉淀的丰富语料库（尤其全国各类方言、语素），配置相应行业的术语，即可对行业客户提供语互服务。b）在语音能力、业务能力方面：可以对内外部应用提供语音识别能力、语音合成能力、语音转文本能力；对内部应用除提供语音能力外，也可提供配合查询、办理、咨询等业务能力。比如，运营商的手机客户端使用智能语音客服的语音能力和业务能力，向用户提供语音交互式自助服务。c）在语音大数据挖掘方面：通过对全网用户的语音交互日志进行分析，挖掘用户的潜在需求、对业务发展的建，如遇大规模网络或业务投诉及时告警等。d）语音识别能力方面：语音识别率是一个不断提升的积累过程，借助于人工智能自学习不断优化语音语料库和识别引擎，将识别率逐步提升至100%。

为积极利用智能语音技术实现更加便捷、自然的自动语音服务和更加高效、精细的运营管理体系，实现在服务行业“合规性”质检工作“看听结合”，扩大业务分析覆盖、提升分析效率、降低质检成本，解决特定行业客户的精细化和覆盖全面性问题；同时利用语音搜索进行相关行业客户关注的热点问题。以此来通过本项目的建设实现上述目标。

## 项目范围

智能语音分析，关注基于语音的业务分析。

智能语音分析：针对现在广电录音日渐增多，提高客服的服务质量，为企业提供全量录音的质量监督及对热点业务的深入分析。

为了满足以上需求实现以下功能：综合查询、全量筛选、系统筛选、交叉分析、语音特征分析、热点分析、聚类规则。

## 客户与最终用户介绍

本项目所建设的系统智能语音分析系统，主要用户为电信公司语音业务员。

# 功能需求概述

**产品角色列表**

【提示：阐述本产品的各种角色及其职责。各种角色的具体行为将在功能性需求中描述。】

|  |  |  |
| --- | --- | --- |
| **角色名称** | **人员组成** | **需求描述** |
| 质检主管 | 质检管理组各组长 |  |
| 质检员 | 各质检小组成员 |  |
| 班组长 | 管理各质检小组领导成员 |  |
| 坐席 | 语音业务员 |  |

# 功能需求详细说明

## 功能列表说明

|  |  |  |
| --- | --- | --- |
| 语音分析系统功能清单 | | |
| 一级菜单 | 二级菜单 | 功能说明 |
| **分析** |  |  |
|  | 综合查询 | 快捷查询目标录音 |
|  | 全量筛选 | 根据配置条件以及录音时间范围筛选录音 |
|  | 系统筛选 | 根据配置条件以及采集周期和运行周期来配置任务 |
|  | 热点分析 | 分析多条件维度下录音中的高频词汇 |
|  | 交叉分析 | 展示不用维度下的录音或通话时长的分布图 |
|  | 语音特征统计 | 用于统计各部门、个人之间语速，静默，重叠次数以及回合占比的数量占比情况 |
|  | 聚类规则 | 根据录音样本和关键字产出不同的主题 |

## 业务分析—综合查询

### 功能概述

快捷查询目标录音，同一页面输入筛选条件，点击查询即可查看结果，显示结果可以选择按详情展示和按数据显示。一般可用于在别的地方已得知某一查询条件，如录音编号，手机号码或者少数关键字的情况下。

运用场景：当你从其他相关系统接收到了一份工单信息，只填写了客户来电号码和来电时间，这时候就可以运用综合查询直接按录音时间和来电号码查找对应录音即可。

### 功能点描述

* **查询**

质检主管按照特定的条件和抽样方式进行录音抽样，比如录音属性、客户属性、员工属性及语音特征等。抽样结果可以选择直接分配给质检人员或加入样本池。

* **保存策略**

当筛选条件选择完成后，点击“策略保存”，为筛选策略命名，下一次可以直接从“查看已保存策略”中选择，然后点击“查询”，得到对应的录音清单。

* **导出**

按条件筛选出录音清单，可以选择导出录音清单的信息，导出excel表，包含录音编号、录音时间、通话时长等详细信息。

* **详细信息**

输入查询条件显示录音清单后，质检主管可以点击录音查看录音的详细信息，详细信息页左侧显示了语音的基本信息和语音特征，右侧显示客户信息和事后录入单。

## 业务分析—全量筛选

### 功能概述

全量筛选可理解为“即时任务”，可根据配置条件以及录音时间范围筛选录音。全量筛选按照预设的全量筛选策略对选定范围内的全部录音进行全量筛选，找出有问题的疑似录音清单，对有问题的录音进行标识。

### 操作流程



### 功能点描述

* **新建倒是筛选**

进入全量筛选，可以选择新建全量筛选项目，配置全量筛选的条件及筛选模板，保存后新建项目。

* **新建全量筛选模板**

新建全量筛选项目之前，可以先配置全量筛选模板，通过设定关键词、静默、语速、重叠以及情绪值完成模板。

设置全量筛选模板的路径：业务配置——全量筛选模板。

* **执行任务**

在全量筛选条件设置完成后点击“开始任务”，系统开始按条件进行筛选录音

* **执行结果**

全量筛选“执行状态”为“已完成”后，点击“执行结果”在下方“项目执行情况”区域，显示已检出录音详情。

* **加入样本池**

筛选出来的录音清单，可以将全选或部分放入样本池暂存，后续进行统一分配质检等。

* **模板操作**

对已经执行的任务，可以进行“复制”“修改”“删除”操作，“复制”是新建一条与当前相似筛选条件的任务时的便捷操作，“修改”是对当前任务的筛选条件进行修改，“删除”是删除当前筛选任务。

## 业务分析—系统筛选

### 功能概述

系统筛选可理解为“循环任务”，可根据配置条件以及采集周期和运行周期来配置任务，任务执行开始之后会根据任务配置的运行周期循环执行，找出有问题的疑似录音清单，对有问题的录音进行标识。

### 操作流程



### 功能点描述

* **新建筛选任务**

进入系统筛选，可以选择新建系统筛选任务，配置系统筛选的条件及筛选模板，保存后完成新建筛选策略。下次可以直接调用，在“查看已保存策略”选择，然后点击“启用”后执行。

运行周期：定义该筛选策略多少天运行一次。

采集周期：该策略的录音采集的时间，是在今天之前的多少天。

* **新建筛选模板**

新建系统筛选任务之前，可以先配置系统筛选模板，通过设定关键词、静默、语速、重叠以及情绪值完成模板。

设置全量筛选模板的路径：业务配置——系统筛选模板。

* **查询**

质检主管可根据时间范围对系统筛选项目进行查询。

* **筛选任务详情**

查询结果展示筛选任务详情，包括项目执行基本情况（检出数量、状态、检出率）和具体的录音清单，录音清单可以选择按详情展示或者按数据展示，点击录音查看录音的详细信息。

## 业务分析—热点分析

### 功能概述

主要用于分析多条件维度下录音中的高频词汇。

运用场景：希望查询针对A产品的投诉录音中出现的高频词，在此场景下主维度可选择A产品，副维度选择投诉，即可展示出此场景下出现的高频词汇以及他们的环比和在此时间范围内的趋势变化。

### 功能点描述

* **查询**

质检员主管可以根据主维度和副维度对录音进行周期性查询，查询结果用词频统计、趋势图和词语信息直观显示。

* **词频统计**

质检主管根据主维度和副维度对录音进行周期性查询，查询后筛选出词库里面相关的关键字，统计出关键字的词频和环比进行显示。

* **添加关键词**

质检主管可以在“业务配置-词库管理”中添加、修改以及删除业务关键字和停词水词。在词频统计中，可以对关键字添加关键字描述。

* **添加水词**

质检主管可以在“业务配置-词库管理”中添加、修改以及删除停词水词。在词频统计中，可以对关键字添加水词，添加水词之后该词不会再出现在词频统计中。

* **趋势图**

系统根据主维度和副维度对选定时间周期的录音进行分析，筛选出高频率的词语，用折线统计图直观显示，横坐标是时间，纵坐标是词频。质检主管可以选择性对词汇不进行显示。

* **词语信息**

词语信息对词语词频情况进行显示，鼠标置于词语上面能看到该词语的词频。

## 业务分析—交叉分析

### 功能概述

主要用于展示不用维度下的录音或通话时长的分布图。

运用场景：在外呼营销中，希望查找产品在全国各省中的销售情况，此场景下主维度便可以选择区域，副维度选择营销成功，即可展示出全国各省中产品营销成功的录音分布图。

功能点描述

* **查询**

质检员主管可以根据主维度和副维度对录音的录音量或者是录音平均时长进行周期性查询，查询结果用视图直观显示。

* **主维度**

质检主管根据主维度的选择对查询数据以直方图和折线统计图的形式进行显示。显示出来的结果可以选择“保存图片”进行保存。

直方图横坐标是主维度，纵坐标是所选择的统计类型（录音量/平均时长）。

折线统计图横坐标是时间，纵坐标是统计类型（录音量/平均时长）。主维度元素可以选择不进行。

* **副维度**

质检主管根据副维度的选择对查询数据以直方图形式进行显示。显示出来的结果可以选择“保存图片”进行保存。

直方图横坐标是主维度，纵坐标是所选择的统计类型（录音量/平均时长）。

* **筛选配置**

筛选配置可以更改折线统计图的显示方式。横坐标可以选择按天、按小时显示，纵坐标可以选择按录音数量、按总通话时长、按平均通话时长三种方式进行显示。

## 业务分析—语音特征分析

### 功能概述

对静默、语速、叠音通话占比，影响服务质量的客服语音特征进行统计，输出问题录音及问题人员、问题组织的清单并导出。主要用于统计各部门、个人之间语速，静默，重叠次数以及回合占比的数量占比情况。

### 操作流程



### 功能点描述

* **静默分析**

点击静默分析标签页，定义静默类型及时间，按部门或者班组维度展现出统计数据结果和问题录音清单，可以选择全部或者部分导出录音到excel表格。

* **语速分析**

点击语速分析标签页，定义平均语速及时间，按部门或者班组维度展现出统计数据结果和问题录音清单，可以选择全部或者部分导出录音到excel表格。

* **语音重叠分析**

点击语音重叠分析标签页，定义平均语音重叠词数及时间，按部门或者班组维度展现出统计数据结果和问题录音清单，可以选择全部或者部分导出录音到excel表格。

* **通话占比**

点击通话占比标签页，定义排序方式及时间，按部门或者班组维度展现出统计数据结果和问题录音清单，可以选择全部或者部分导出录音到excel表格。

## 业务分析—聚类规则

### 功能概述

以录音和关键字为基础，结合算法，挖掘产出录音中关联性较高的主题以及典型问法。

运用场景：根据录音样本和关键字产出不同的主题，例如录音样本为营销成功类，执行聚类规则后可产出营销成功场景下相关的主题和典型问法。

### 操作流程



### 功能点描述

* **新建聚类规则**

质检主管选择新建聚类规则，给聚类规则命名，设定显示主题词数（5-9），限定语句数，选择分析对象（全部/客户/客服）。

* **训练样本**

质检主管可以对聚类规则进行样本训练，样本训练分为加入训练样本和移除训练样本两部分。训练样本必须大于50条才可以进行聚类。

* **呼入录音**

进入呼入录音，质检主管可以从录音属性、员工属性、客户属性以及语音特征四个方面进行录音查询，查询结果可以按详情显示或者按数据显示。录音清单可以选择性加入训练样本。

* **呼入已选录音**

进入呼入已选录音，质检主管可以根据客户姓名、坐席姓名以及坐席工号进行录音查询，查询结果可以按详情显示或者按数据显示。录音清单可以选择性进行移除训练样本，同时也可以选择导出录音（excel）。

* **执行样本**

点击“开始”，系统执行训练样本。训练样本的数量必须在50条以上，启用状态的词表个数需在[100,2000]范围内。

* **执行结果**
* **主题聚类**

主题聚类根据聚类规则对训练样本进行聚类，显示结果格式为“主题：词1-词2-词3-词4-词5-词6-词7（词频）”。

主题聚类的结果后面的数字，表示主题词在多少个句子中出现（不是多少条录音）。

* **典型问法**

典型问法根据聚类规则对训练样本进行聚类，显示结果格式为“词1-词2-词3-词4-词5-词6-词7（语句数量）”。

* **词表维护**

再进入“词表维护”启用/停用关键词。词表中的关键词受词库管理中的“业务关键词”、“停词水词”的影响。加入到“停词水词”的词，不会出现在词表中。

* **查看日志**

日志记录用户对聚类的操作记录。

# 性能需求说明

## 系统运行指标

每物理核CPU每小时转译3小时的录音数据量。

提供热词统计功能，可以自动发现录音中存在客户和客服的热词。

提供自定义模型建立的工具，可根据不同的业务要求，通过关键词或组合的配置，实现分析模型的建立。

支持语音转化成文本能力，系统上线后，经过多轮优化迭代，对于业务录音汉语普通话语音转写效果全文与关键字正确率分别达到80%和80%以上。

## 高可用性要求

指在规定条件下和给定时间内[计算机系统](http://www.hudong.com/wiki/è®¡ç®æºç³)正确运行（①程序不为故障所破坏或停止;②结果不包括由故障所引起的错误；③执行时间不超过一定的限度；④程序运行在允许的领域内。）的概率。

系统能够满足：平均无故障间隔时间MTBF大于90天。

# 其它需求说明

## 其它需求

涉及算法：排序算法，各语音文本按日期排序，精确时间查询定点分析

## 新系统开发、运行条件与限制

暂无。

# 7总结

本系统融合了云计算、语音处理、商业智能和互联网技术，以实现语音数据的高效转写、分析统计和数据挖掘等广泛应用为目的，面向语音数据密集型的话务行业，提供数据处理所需资源、能力和应用服务的云平台，能够提升企业的客服质量。

目前机器处理的文本情感分析还只是仅仅从数学统计角度融合一些自然语言的语法规则, 完成极性分类、语法分析的任务, 而情感是一个相对复杂的研究客体, 交给机器与算法处理只是用数学量化的基本模型完成简单任务，对大量数据隐含的情感特征和语言表达的统计和排查只是初步解决了分析中的小问题，如何让机器对人类情感进行精确的把握和分析，甚至是通过分析让机器模仿、创造出与人类的情感交流方式是文本分析更深入的问题。