**中国移动四川公司 语义识别模块**

**功能需求说明书**

|  |  |  |  |
| --- | --- | --- | --- |
| 编 写 |  | 编写 时间 | 2019-01-17 |
| 审 核 |  | 审核 时间 |  |
| 文档版本 | V0.2 | | |

**文档修订记录**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **版本号** | **描述** | **著者** | **审阅者** | **日期** |
| V0.1 | 初稿，根据本周与移动网优中心李丹老师沟通会议 | 钟闻威 |  | 2019-01-17 |
| V0.2 | 根据1月与李丹老师需求确认 | 钟闻威 |  | 2019-02-11 |
|  |  |  |  |  |
|  |  |  |  |  |

目录

[第1章 引言 4](#_Toc1316837)

[1.1 编写目的 4](#_Toc1316838)

[1.2 读者对象 4](#_Toc1316839)

[1.3 缩略词 4](#_Toc1316840)

[1.4 参考资料 4](#_Toc1316841)

[第2章 项目概述 4](#_Toc1316842)

[2.1 建设目标 4](#_Toc1316843)

[2.2 产品描述 5](#_Toc1316844)

[2.3 产品功能 5](#_Toc1316845)

[2.4 系统运行环境 6](#_Toc1316846)

[2.4.1 硬件环境 6](#_Toc1316847)

[2.4.2 软件环境 6](#_Toc1316848)

[2.5 业务系统总体结构 6](#_Toc1316849)

[2.6 与外部系统关系 6](#_Toc1316850)

[第3章 项目约束 6](#_Toc1316851)

[3.1 技术约束 6](#_Toc1316852)

[3.2 业务约束 6](#_Toc1316853)

[3.3 进度约束 6](#_Toc1316854)

[3.4 质量约束 7](#_Toc1316855)

[3.5 性能约束 7](#_Toc1316856)

[3.6 其他约束 7](#_Toc1316857)

[第4章 业务需求 7](#_Toc1316858)

[4.1 数据监控（数据质量保障系统） 7](#_Toc1316859)

[4.1.1 功能设计 7](#_Toc1316860)

[4.1.2 流程图 8](#_Toc1316861)

[4.2 投诉问题分类（投诉智能分类助手） 9](#_Toc1316862)

[4.2.1 功能设计 9](#_Toc1316863)

[4.2.2 数据呈现 11](#_Toc1316864)

[4.2.3 用户界面 12](#_Toc1316865)

[4.2.4 示例（更新） 12](#_Toc1316866)

[4.3 低满意度用户识别（用户情绪识别助手） 13](#_Toc1316867)

[4.3.1 功能设计 13](#_Toc1316868)

[4.3.2 数据呈现 14](#_Toc1316869)

[4.3.3 用户界面 14](#_Toc1316870)

[4.3.4 示例（更新） 14](#_Toc1316871)

[4.4 用户行为特征分析 16](#_Toc1316872)

[4.4.1 功能设计 16](#_Toc1316873)

[4.4.2 数据呈现 16](#_Toc1316874)

[4.4.3 用户界面 17](#_Toc1316875)

[4.5 用户满意度调查（没数据，待定） 17](#_Toc1316876)

[4.5.1 功能设计 17](#_Toc1316877)

[4.6 管理驾驶舱 17](#_Toc1316878)

[4.6.1 功能简介 17](#_Toc1316879)

[4.6.2 数据项 18](#_Toc1316880)

[4.6.3 用户界面（示例） 18](#_Toc1316881)

[第5章 附录 19](#_Toc1316882)

[5.1 附录1：功能列表 19](#_Toc1316883)

# 引言

## 编写目的

为了保证软件开发的质量、需求的完整与[可追溯性](https://www.baidu.com/s?wd=%E5%8F%AF%E8%BF%BD%E6%BA%AF%E6%80%A7&tn=SE_PcZhidaonwhc_ngpagmjz&rsv_dl=gh_pc_zhidao)，编写此文档。通过此文档，以保证业务需求提出者与需求分析人员、开发人员、测试人员及其也相关利益人对需求达成共识。

## 读者对象

系统开发人员、测试人员以及其他与项目相关的人员。

## 缩略词

|  |  |  |
| --- | --- | --- |
| 缩写 | 英文描述 | 中文描述 |
|  |  |  |

## 参考资料

《投标标书-技术分册-证通电子v20181203V1》

《语义识别需求分析V1.02》

# 项目概述

## 建设目标

本项目主要建设目标如下：

(1) 建立一个语义识别语料库与知识图谱构建模块，作为以后全面应用智能化的基础，包括语义识别、语音识别、图像文字识别等功能架构，建立丰富的移动网优业务为核心的语音语料库、知识图谱。需要实现数字化大脑，支持跨平台应用，支持自然语言处理与多轮对话理解；

(2) 实现一段文字的语义识别，如对话场景，或一篇长文章等，可以从中抽取重要的信息如时间、地点、投诉类型，识别文字情绪等；

(3) 能够把一段对话音频转换成文字。

(4) 提供专题分析功能，针对特定专题，支持跨平台的数据提取、整理和报表、PPT报告输出；

(5) 多平台数据串联，在不改动现有系统接口的情况下，实现跨系统的数据交互，从而节省现有系统改动频率，增强系统稳定性；

(6) 基础数据、网络、平台的监控。调用现有监控手段，7\*24小时不间断监控网络情况，及时通过多种渠道进行报警，确保网络稳定。

## 产品描述

语义识别模块根据用户与10086客服对话的全量录音以及翻译的文本，结合投诉工单数据，运用自然语言处理和数据挖掘技术对录音进行分析，并给录音文本打上多维度的标签；并基于情绪识别和用户行为特征识别出低满意度客户,根据服务态度和服务质量等维度来分析客户不满意的原因。

## 产品功能

语义识别模块包括数据监控、投诉问题智能挖掘分类、低满意度用户识别、用户行为特征分析、用户满意度调查、管理驾驶舱等功能子模块。

## 系统运行环境

### 硬件环境

### 软件环境

## 业务系统总体结构

## 与外部系统关系

# 项目约束

## 技术约束

## 业务约束

## 进度约束

本项目要求在合同签订2个月内到货完成安装；合同签订6个月初验，投诉工单智能分类准确率达到 %，低满意度用户情绪识别准确率达到 %；试运行3个月以上终验。

## 质量约束

## 性能约束

## 其他约束

# 业务需求

## 数据监控（数据质量保障系统）

### 功能设计

该子模块对用户与客服对话录音文件、翻译文本等原始数据进行实时监控，定时汇总报表，并对异常情况进行预警。

#### 数据增量监控

实时监控（每天）新增用户与客服对话录音文件、翻译文本等的数量。

#### 数据时延监控

实时监控用户与客服对话录音文件、翻译文本等原始数据上传的时延，以及从录音文件上传到翻译文本上传之间的时间差。

#### 数据匹配监控

实时监控用户与客服对话录音文件、翻译文本、投诉工单等数据的匹配。计算公式如下：

录音与工单匹配率（有工单且有录音文件） = （匹配上录音的工单数 / 全量工单数） \* 100%

文本与工单匹配率（有工单且有转录文本） = （匹配上文本的工单数 / 全量工单数） \* 100%

录音与文本匹配率（有录音文件且有转录文本） =（匹配上文本的录音数 / 全量录音）\* 100%

工单录音一对多发生率（一个工单号对应有多个录音文件） =（发生工单录音一对多的工单数 / 全量工单数） \* 100%

#### 监控日报表和预警

时延监控报表分类别按时延长度分别统计录音文件和翻译文本总数，例如30分钟，1小时，2小时，3小时等。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 30分钟 | 1小时 | 2小时 | 3小时 | 总计 |
| 现场 |  |  |  |  |  |
| 诉求 |  |  |  |  |  |
| 申告 |  |  |  |  |  |

表4-1 时延监控报表模板

时延标准暂定录音和翻译文本上传时延分别小于1小时，超出时延标准的异常情况需进行统计。

异常监控报表按不匹配类型，如有录音无文本、有录音无工单、工单录音一对多等，统计异常数量。

以上异常统计汇总生成报表后，经过审核后推送。（前期发送到个人微信）

### 流程图

图4-1 数据监控业务流程图

## 投诉问题分类（投诉智能分类助手）

该子模块利用自然语言处理、数据挖掘、人工智能等技术对投诉录音文本进行语义分析，结合移动通信领域的业务特点，按照预定义的文本标签进行标注，对投诉工单进行自动化的智能分析与处理，实现快速的热点问题排序、故障分析、系统热点问题跟踪分析等，准确指明后台支撑维护资源的流向，保障有限的支撑资源准确地投入到主要矛盾的解决过程中。

### 功能设计

#### 投诉问题分类标注

下表定义了预设的文本标签，标\*的标签必须标注。支持对已有标签通过与或非逻辑运算得到的组合标签，用户也可以自定义新增标签。

|  |  |  |
| --- | --- | --- |
| 标签名称 | 类别 | 备注 |
| 分类标签\* | 基础通信、家宽、资费、业务、终端/SIM卡、其他 |  |
| 设备类型标签\* | 上网卡/上网本、手机、MIFI、CPE |  |
| 网络制式类标签\* | 2G、3G（TD）、4G（LTE）、VoLTE、5G |  |
| 基础通信类标签\* | **语音通话，互联网上网，VoLTE** |  |
| 场景1类 | 室内、室外 |  |
| 场景2类 | 乡镇、高铁、地铁、高校、美食、美景、高速、医院、交通枢纽、商业区、机场、大型居民区、酒店 |  |
| 场景位置  （不易定位区域） | **公共标签：**电梯、停车场  **乡镇：**镇区、郊区？  **高铁：**候车厅、售票厅、站台、列车  **地铁：**售票厅、站台、列车  **高校：**教学楼、实验室、图书馆、办公楼、运动场  **美食：**收银台、大厅、包房  **美景：**售票处、场馆、风景、古迹？  **高速：**收费站、服务区、路面、车辆  **医院：**挂号收费区、候诊区、诊室、检查室、治疗室、药房、病房  **交通枢纽：**候车船区、站台、码头  **商业区：**商铺、餐厅  **机场：**值机区、候机厅、登机口、到达区  **大型居民区：**高楼层、中低楼层、阳台、卧室、客厅  **酒店：**大堂、餐厅、客房 | 实际应用中需要根据用户描述抽取具体位置关键词增加标签，通过学习和训练完善 |
| 现象类标签 | **语音通话类**：无法接通、用户掉话、串线、单通、通话时断时续、干扰（回音、噪音等）  **互联网上网类：**上网时出现断线、网速慢或网页无法打开、完全不能上网  **覆盖类：**完全没有信号、信号弱、信号时强时弱、有信号但无法使用 |  |
| 用户行为标签 | 1. 短时间内同一个投诉工单、多次来电。**[敏感用户：耐受度低]** 2. 重复投诉。 3. 多次投诉，多个问题。**[投诉倾向用户：喜欢投诉]** |  |
| 不限量套餐标签 | 1. 不限量套餐限速后完全不能上网 2. 不限量套餐限速后网速慢 3. 不限量套餐限速跨月后网速未恢复 4. 不限量套餐用户流量使用习惯 |  |
| 竞对分析类标签 | 1. **预离网用户** 2. **竞对差覆盖盲点区域投诉** 3. 竞对差覆盖弱区域投诉 4. 竞对差服务类投诉 |  |
| 网络服务类标签 | 1. 完全没有响应（投诉没有用、问题依然存在） 2. 响应慢 3. 问题持续时间（15天、几个月……以上） |  |
| 自定义标签 | 1. 与或非逻辑组合标签 2. 新增标签 |  |

注：标\*的为必须标注的标签

表4-2 预设文本标签定义

#### 录音文件与工单匹配

由于全量录音的文件数大于全量工单数，需要将录音与工单进行匹配。匹配结果通过接口可供提取查询，可以识别一个工单对应多个录音文件的情况。例如：

|  |  |  |  |
| --- | --- | --- | --- |
| 录音号 | 时间 | 投诉电话 | 工单号 |
| 0813011 | 2018-12-27 08:13:01 | 15928107937 | 2018122700128 |
| 0822436 | 2018-12-27 08:22:43 | 18384350921 | 2018122700231 |

表4-3 录音与工单匹配

#### 工单标准地址匹配

当前客服录入工单地址不完整时，投诉平台利用离线地址库和地址规整功能把不规范地址转为规范地址，但会出现多个地址实际上为同一地点的情况，例如中国移动西区枢纽和百草路519号实际上是同一地点。需要增加同一地点不同地址的匹配，统一成门牌地址，匹配结果通过接口可供提取查询。

如下表，统一门牌地址为成都武侯区天仁路388号凯德天府。

|  |  |  |  |
| --- | --- | --- | --- |
| 工单记录的故障地址 | 标准地址 | 门牌地址 | 经度/纬度 |
| 【凯德天府】天仁路388号 | 成都武侯区凯德天府 | 成都武侯区天仁路388号凯德天府 | 104.066751290236/ 30.605485313908702 |

表4-4 工单标准地址匹配

### 数据呈现

#### 外部系统数据接口

对外部系统支持跨平台的数据提取、整理，以及报表、报告输出。

#### 数据可视化

1. 自定义关联查询：对预设标签、细分标签、与或非逻辑组合标签、自定义标签等进行模糊查询；
2. 查询结果自动透视分析：对按照查询标签查询出的数据，根据特征标签进行透视结果呈现。例如：根据“不限量套餐”标签查询，对查询结果，根据工单标注的其它标签进行透视，得到不限量套餐投诉现象分布情况、用户类型分布情况、场景分布情况等；
3. 查询结果按地理区域分布呈现；
4. 查询结果按时间趋势分布呈现；
5. 查询结果按用户画像分类呈现（需确认投诉平台可用数据）；
6. 后台记录操作日志和用户习惯，分析用户使用哪些标签、使用频率、解决哪些问题（问题需记录）。

### 用户界面

### 示例（更新）

|  |
| --- |
| 您好 这里 是 网络 专席 有 什么 需要 帮助 的 吗？ 好的 我 帮您 马上 看 一下 稍等 一下， 请问 昨天 你的 号码 有没有 欠费 呢 先生。 呃 对 是的 但是 我 帮您 看 一下 你的 流量 这些 我们 都是 给您 打开 的， 把 流量 打开 之后 它 有没有 出现 信号 一个 标志 呢 在 您 手机 上面。 好的 我 也 帮您 检查 一下 这些 功能 的话 确实 都是 打开 的， 我 帮您 查一下 最近 这一 个 新的 一个 使用 记录 看 一下 您 用的 是 什么 网络 稍等 一下？ 好的 我 通过 系统 查一下 您 最近 这一 个 新的 话， 这个 呢 是 四 g 网络， 我 帮您 看 一下 你在 今天 的， 哎 昨天 都用了 一部分 用的 确实 比较 少。 稍后 呢 有 一个 专门 的 一个 修复 数据 的 方法 我 给您 发 过来 一下， 如果 是 网络 临时 掉 线 的 原因 你可以 先 调整 一下， 关机 开机 试 一下， 看这 样 能不能 改成 使用 可以 把 网络 先 调整 一下 好吧。 诶 对 关机 开机 一下， 先 按照 短信 把 您 卡 的 时间 忧患 一下。 把 卡 的 数据 换 一下 我 把 兑换 的 一个 指令， 发 到 您的 手机 上 了。 稍后 呢 您 在 使用 流量 可以 先把 流量 关闭 几分钟 然后 重新 打开， 手机 是 一个 信号 这样 操作 一下 好吧。 诶 那 确实 给您 带来不便 了 非常 抱歉， 请问 还有 别的 需要 帮助 的 吗 谢谢？ 诶 不客气 稍后 呢 请 对我 的 个人 服务 做 个 评价 好吧 谢谢 好的， 祝您 生活 愉快再见。 |
| 我 这个 手机， 就 昨天 晚上 开始 就是， 那个 数据 线 没有， 没有 就是 信号 啊， 微信 收不到 可能 也 发 不出 去 也 收不到。 昨天 没有 嘛 今天 昨天 我看 到 还有 八 块 多 钱 没 欠费 我 这也 可以 头 豆 是 五十 块 的 吧。 但是 我 手机 它 就是 没有 信号 吗？ 嗯 发 微信 发 不出 去 收 也 收不到 今天 早上， 嗯 我 几点 的时候 偶尔 收到 一下 就， 很 很大 的 时间 之内， 一下 有没有 了， 有 啥子 收到 几个， 你们 的 微信 发来 别人 发 微信 的 那个 提示 音 下来 就没 有 就没 有了， 昨天 晚上 开始 就 反正 有 这个 情况。 好的。 那 这样 先生。 意思 关机 开机 呀。 嗯 卡 的 什么 啊？ 卡 的 用 款 数据。 哎。 好的 好的。 |

**标签：网络 手机 数据上网 室内 大型居民区 4G（LTE）网速慢或网页无法打开 无限量套餐不能上网**

客服建议:按照短信指令进行优化（？）

## 低满意度用户识别（用户情绪识别助手）

该子模块通过语音情感分析、情感分词、上下文语境、用户行为特征等方式对用户情绪进行识别和分析。

### 功能设计

#### 低满意度用户情绪识别

识别的方式主要包括：

1. 分析语音情感，即分析语音语调、语速、声音能量谱等识别；
2. 提取情感分词，建立情感分词库，通过训练学习完善；
3. 根据上下文语境，即根据对话上下文语境识别明示或暗示的信息，提取负向的、明确不满意的分词。例如“告你们”、“投诉”、“去通管局/工信部/法院告”、“换电信/换联通”、“销号”等；
4. 根据用户行为特征，即用户行为特征标签（敏感用户）辅助识别。如短时间内多次来电反馈同一个问题，或2小时以内来电要求由诉求激活为申告类工单，都为敏感用户。

通过以上方式识别出

1. 情绪浓度；
   1. 用数值表示，例如0-100；
   2. 表示情绪（低满意度）的强度；
2. 情绪时长；
   1. 考虑用问题持续的时间来反映情绪（低满意度）持续的时间长度，例如问题1个星期、1个月、半年等都没有解决；
   2. 考虑其它方式来反映情绪（低满意度）持续的时间；
3. 情绪效价；
   1. 情绪效价包括正向、负向、中性情绪，界面只呈现负向（低满意度）数据；
4. 情绪（低满意度）特征;
   1. 潜在投诉升级用户：识别出的有情绪（低满意度）升级倾向的用户；
   2. 离网倾向：投诉中出现“换电信/换联通”、“销号”等；
   3. 重复投诉：90天内同一用户关于同一类型问题的投诉；
   4. 敏感用户：短时间内发生同一个投诉工单对应多次来电（多个录音文件或转录文本），即工单录音一对多的情况；2小时以内来电要求由诉求激活为申告类工单；
5. 情绪（低满意度）原因；
   1. 包括没解决问题、投诉没用、无反馈、客服态度差等原因；

### 数据呈现

#### 外部系统数据接口

对外部系统支持跨平台的数据提取、整理，以及报表、报告输出。

#### 数据可视化

1. 查询低满意度用户，按地理区域分布呈现，在地图上标注规划、建设、维护、优化点（根据已知问题库），进行多维度数据的栅格化地理呈现（数据较少）或密度分层渲染（高密度连通区域，DBSCAN）；
2. 查询低满意度用户，按时间趋势分布呈现；
3. 查询低满意度用户按投诉问题智能挖掘分类标签分类呈现;
4. 查询低满意度用户按用户行为特征分类呈现；
5. 查询低满意度用户按用户画像分类呈现（需确认投诉平台可用数据）；

### 用户界面

### 示例（更新）

|  |
| --- |
| 您好 很高兴 为您 服务。 呃 就是 用， 那个 数据 流量 上网 网络 很 慢 对 吧。 呃 对不起 用户 添麻烦了 我想 麻烦 问一下 出现 那个 问题 有 多长时间 呢 是 今天 开始 的 吗 先生？ 那 这边 的话 我看 了 一下， 上面 显示 您的 那个 网络 功能 吗 状态 都 正常 那 这边 我想 冒昧 的 问一下？ 现有 十 来 天 左右 时间 呢， 嗯 现在 你们 身边 的 朋友 这些 他们 使用 网络 正 不正 常 有 看到 过 这个 情况 吗？ 呃 呃 就是 那个， 周围 的 朋友 那些 上网 网络 都很 慢 对 吧。 那你 能 帮我 提供 一下 是在 什么 地方 嘛 我 查一下 是不是 这边 的 网络 出问题 了呢？ 是 广元 市 望 上线。 是在 旺 苍 县 县城 吗 还是 在 什么 地方？ 呃 那个 具体地址 能 帮我 提供 一个 吗 我 好 查一下 当地 的 网络 嘛 可不可以？ 呃， 白马 寺 社区 是 吧。 呃， 那个 是的 话 是 寺庙 的 是 吗？ 噢 好 稍微 等一下， 呃 这个 对不起 让我 找 一个地方 是在 那个， 呃 呃 件 原 路 与 白马 路 这个 交叉 那边 是不是 这样。 呃 就 电源 路 和 白马 路 嘛 是不是 这个 也。 是 那个 旺 苍 县 的 那个 白马 是 社区 吗？ 稍微 等一下 我看 一下 啊 就是 那个 姓 王 东 路 这边 嘛 对 吧。 嗯嗯 好的 先生 稍微 等一下， 呃 现在 话 那个 感谢 耐心等待 我想 最后 问一下， 您 上面 的话， 就是 那个 显示 吗？ 有 显示 四 g 嘛 只是 网络 很 慢 对不对。 呃 您 看这 样 先生 问题 的话 了解 清楚 了， 但是 我 通过 系统 上 也 排查 了 一下， 呃 上面 显示 功能 吗 状态 都 正常 没有 接到 故障 通知 了然 后 我 把 那个 情况？ 立刻 上报 到 技术部门 检测 一下 看 一下 是不是 您 这边 的 那个 集团 数据 的 问题 但是 有 一点 需要 麻烦 先生。 稍后 的话 您 注意 观察 在 十五 分钟， 十五 分钟 过后 把手 机关 机 开机 一次， 如果 说 这边 上报 处理 过后， 最多 两个 小时， 嗯 两个 小时 过后 这段 时间 能够 恢复 正常 用 那 您 就不 用 管 它 了， 但如果 还是 没有 恢复 的话 就 打 一下， 幺零零八六 幺 九 这个 专线 联系 一下 恢复 正常 您 就不 管他了 好 嘛。 嗯 那 这边 我 现在 立刻 上报 如果 晚上 先 修复 过后 晚上 您 开 完了 休息 的话 也没 关系 可以 在 明天 早上 看 一下 效果 嘛 好吧。 诶 好 那就 麻烦 先生 对不起 了， 您 看 一下 还有 没有 其他 能够 帮助 到 您的 呀 先生。 昨天 我看 一下 有没有， 那 您 这边 的话 有没有 拨打 我们 那个 网络 专线 幺零零八六 幺 九 了。 呃 有， 他 给你 打电话 的话 是 怎么 给您 说 的 呢 先生。 噢 那个 对不起 那 现在 的话 我 就 把这 个 情况 吗？ 进行 上报 到时候 的话 您 注意 观察 一下 看 一下 情况 嘛 好吧。 嗯 那 我 现在 就 立刻 把这 个 情况 进行 上报 到时候 您 注意 观察 看 一下 那个 网络 的 一个 效果。 呃 那 这边 的话 现在 还有什么 可以 帮助 到 您的 吗 先生？ 啊 不客气 应该 的 先生， 那 祝您 生活 愉快 了 再见。 |

|  |
| --- |
| 就是 为啥 为啥 子 这个 网络 变 慢 了 网速。 嗯嗯。 有 好几 天 了 多少 钱 呢， 啊。 网速 都比 较 慢 了。 对 这边 吗？ 往 专线。 噢。 正常。 开 嘛 四十 七。 对。 对。 键 联络。 没有， 的 是 吧， 就 新 入网 办 的。 对， 噢。 对。 对， 它 也 显示 一个 一个 圈 转 嘛 说是， 就是 那个 积分 能 转。 噢。 噢。 昨天 也是 这样 子 昨天 我 就 弄 了 一个 什么 十五 分钟 关机 我 就 弄 过了？ 他 给我 打电话 了 嘛 暂时 没什么 反应 然后 就没 联系 我们？ 他 就是说 什么什么？ 没 没说 什么 挂机 了？ 听不到 他 就 挂 了 之后 联系。 啊。 好的 谢谢 啊。 |

情绪识别结果：

**原因：网速 情绪强度：一般不满意 情绪倾向：弱**

## 用户行为特征分析

### 功能设计

#### 用户行为特征分类

用户行为特征可分为高敏感用户（同一问题多次投诉）、重复投诉、投诉倾向客户（喜欢投诉）、需情感修复用户（没有明确不满意点）等，对重点关注的用户行为特征进行历史投诉数据关联分析，可以提高用户满意度。

#### 同一问题多次投诉数据分析

通过分析一个工单对应多个录音文件的数据，可以识别高敏感用户（低耐受），辅助低满意度用户识别，必要时由客服主动联系用户。

工单录音一对多的原因包括：

1. 用户重复打电话反馈同一个问题，提示用户忍受度较低；
2. 确认非网络问题或问题已解决，工单已闭环；
3. 重复投诉类型；

#### 重复投诉数据分析

对识别出来的同一用户多次投诉，对问题类型、故障地址、投诉场景位置进行分析；识别情绪指数，可以是数值，或者是冷淡、平和、激动、狂躁等情绪描述词语；根据用户描述识别用户投诉特征（偶尔投诉、经常投诉等）。

### 数据呈现

#### 外部系统数据接口

对外部系统支持跨平台的数据提取、整理，以及报表、报告输出。

#### 数据可视化

### 用户界面

## 用户满意度调查（没数据，待定）

### 功能设计

满意度的反馈目前只有分数，没有具体信息，需要对满意度调查的录音进行分析，找到具体的原因，以及用户对问卷的态度、对网络的态度。满意度调查每个季度15000次左右，需要将15000个录音转成文本。

#### 满意度调查标签定义

满意度调查的目的是找到用户不满意的原因。目前按照用户对问卷的态度、对网络的态度等维度进行分析。

|  |  |  |
| --- | --- | --- |
| 标签名称 | 类别 | 说明 |
| 对问卷的态度 | 敷衍、中性、认真 | 认真对待问卷的用户的态度更具有真实性 |
| 对网络的态度 | 不满意、还行、很满意 | 按照三个层次对网络的态度进行区分 |
| 不满意的原因 | 网速、通话 |  |

表4-5 用户满意度标签定义

## 管理驾驶舱

### 功能简介

#### 网络数据

提取网络相关的数据，进行同比和环比分析，包括日、周、月、季度、年的对比。

#### 投诉问题分类

提取投诉分类数据，进行日、周、月、季度、年的同比和环比分析。

#### 低满意度用户识别

提取用户情绪数据，进行日、周、月、季度、年的同比和环比分析。

#### 用户行为特征

提取用户行为特征数据，进行分析。

#### 用户满意度

提取用户满意度数据，进行日、周、月、季度、年的同比和环比分析。

### 数据项

### 用户界面（示例）



# 附录

## 附录1：功能列表

|  |  |  |
| --- | --- | --- |
| **一级功能** | **二级功能** | **对应功能需求描述** |
| 人机交互窗口 | 人机文字交互窗口设计 | 需要提供一个可视化的、友好的人机文字交互窗口 |
| 人与自动化机器人平台交互 | 能够支撑下述功能的人与自动化机器人、智能语义分析平台等其他系统、人工智能进行交互。方便用户通过文字指令的方式调用后台机器人完成指定的工作内容，包含但不限于以下工作内容的指令识别：网络性能指标查看、投诉相关指标查看、各类日报/周报/月报查看、各类专题分析报告查看、三方服务人员考核表查看等等 |
| 智能语义分析平台交互 |
| 交互指令识别 |
| 聊天记录存储管理 | 1. 能够提供聊天记录的查看。能够储存至少3年的记录。 |
| 聊天记录查看 |
| 单人多轮对话管理 | 2. 能够支持单人多轮对话。 |
| 问题澄清管理 | 3. 支持问题澄清，通过问题澄清，了解用户意图，识别新的任务指令。 |
| 交互响应机制设计 | 4. 能够快速响应请求。回复时间应在秒级。 |
| 多线程对话交流管理 | 5. 能够支持多线程的对话交流。能够支持同时5个对话以上。 |
| 移动端交换管理 | 6. 支持移动端的交互，支持外网交互，支持角色权限控制。 |
| 安装信息配置 | 7. 安装便捷或无需安装。 |
| 版本管理 |
| 无线网络类投诉问题智能分类 | 信息提取 | 基于客服前端实时语音识别转录的文本信息，提取投诉现象描述、故障地址信息等，基于机器学习辅助10086前端客服完成投诉问题的智能分类。 |
| 投诉现象描述类分词划分 | 1、 投诉现象描述类分词划分以及对应的词向量构建； |
| 词向量构建 |
| 故障地址分词划分 | 2、 故障地址分词划分以及对应的词向量构建； |
| 词向量构建 |
| 用户语音转化成的文本进行分类 | 3、 基于机器学习算法，对用户语音转化成的文本进行分类，在文本转写完全正确的情况下，针对单个问题的识别准确率达到90%，系统最终根据投诉处理的流程自动完成17无线网优类投诉问题分类。 |
| 单个问题的识别 |
| 无线网优类投诉问题分类 |
| 投诉问题的落单 | 4、 在现有服务助手系统基础上，辅助客服完成投诉问题的落单，能够将原有串行流程升级为并行流程。 |
| 串行流程升级为并行流程 |
| 投诉客户满意度评估与低满意度原因识别 | 客服前端实时语音识别转录的文本信息提取 | 基于客服前端实时语音识别转录的文本信息，提取投诉问题描述、故障地址信息、客户情绪分词、投诉用户满意度回访信息，构建用户满意度评估模型，识别低满意度用户，并定位低满意原因。 |
| 关键信息提取 | 1、 提取投诉原因分词、情绪分词、满意度分词、用户满意度回访数据等，基于机器学习，构建投诉用户满意度评估模型，识别低满意度用户，支持调用时延秒级，判断时延分钟级。 |
| 构建投诉用户满意度评估模型 |
| 低满意度用户识别 |
| 调用时延秒级设计 |
| 判断时延分钟级设计 |
| 来电文本信息提取 | 2、 根据来电文本信息，能够识别低满意用户的低满意原因，例如：投诉响应速度慢、没有信号、某套餐产品设置不合理等。支持调用时延秒级，判断时延分钟级。 |
| 低满意用户的低满意原因识别 |
| 低满意度用户识别准确率计算 | 3、 要求低满意度用户识别准确率达到70%。说明：识别准确率由本项目低满意度用户与集团满意度调查中的低满意用户匹配率来计算。 |
| 全网投诉用户满意度综合分析统计 | 4、 提供统计页面，分时间纬度、区域纬度、原因类别等纬度呈现全网投诉用户满意度综合分析情况。包含但不限于以下统计分析：低满意度用户趋势跟踪分析、低满意度用户区域分布情况分析、低满意原因统计情况分析等。 |
| 单用户数据查询接口 | 5、 提供结果数据共享接口，可支撑其他业务平台调用。要求至少支持单用户数据查询接口、批量用户查询接口等。 |
| 批量用户查询接口 |
| 无线投诉线条专业性语料库/知识图谱的构建 | 长文本处理 | 1. 能够处理长本文。 |
| 多种信息提取 | 2. 支持多种信息提取。 |
| 信息提取格式配置 | 3. 灵活可配置信息提取格式。 |
| 特有的语料库构建 | 4. 建立特有的语料库、问题知识库与知识图谱，能够本地化部署。 |
| 问题知识库构建 |
| 知识图谱构建 |
| 与对话识别功能交互规则设计 | 5. 能够与对话识别功能交互。 |
| 与对话识别功能交互API |
| 语料库与知识图谱构建 | 各类投诉工单数据采集 | 采集网优部门各类投诉工单、规章制度、系统使用手册、系统功能规格说明书、四川省的地址信息等 |
| 规章制度数据采集 |
| 系统使用手册数据采集 |
| 系统功能规格说明书数据采集 |
| 地址信息数据采集 |
| 数据注入 | 对采集的原始数据做预处理后注入到系统中 |
| 预处理 |
| 词法分析 | 基于大数据和用户行为的分词、词性标注、命名实体识别，定位基本语言元素，消除歧义，支撑自然语言的准确理解。以分词和词性标注为基础，分析语音命令中的关键名词、动词、数量、时间等，准确理解命令的含义，提高用户体验；通过挖掘实体之间、实体与关键词之间的关联，构建实体信息（如人物、机构）数据库；通过专名识别定位多轮对话中的核心实体，自动判断后续对话中对该实体的进一步信息需求。 |
| 数据挖掘 |
| 句法分析 | 利用句子中词与词之间的依存关系来表示词语的句法结构信息(如主谓、动宾、定中等结构关系)，并用树状结构来表示整句的的结构(如主谓宾、定状补等) ，自动分析文本中的依存句法结构信息，实现对自然语言的精准理解。对大规模非结构化文本数据进行句法结构分析，从中抽取实体、概念、语义关系等信息，帮助构建领域知识或知识图谱。 |
| 词向量表示 | 依托全网海量数据和深度神经网络技术，通过词语的向量化来实现文本的可计算，帮助快速完成语义挖掘、相似度计算等应用。词向量计算是通过训练的方法，将语言词表中的词映射成一个长度固定的向量。词表中所有的词向量构成一个向量空间，每一个词都是这个词向量空间中的一个点，利用这种方法，实现文本的可计算。 |
| 词义相似度 | A、依托全网海量数据和深度神经网络技术，通过词语向量化来计算两个词之间的相似度。用于计算两个给定词语的语义相似度，基于自然语言中的分布假设，即越是经常共同出现的词之间的相似度越高。 |
| 短文本相似度 | 依托全网海量数据和深度神经网络技术，为您提供高精度的短文本相似度服务，帮助快速实现推荐、检索、排序等应用。短文本相似度计算服务能够提供不同短文本之间相似度的计算，输出的相似度是一个介于0到1之间的实数值，越大则相似度越高。相似度数值建议在一组数据中进行整体比对选用，输出数值越大，则代表语义相似程度相对越高。 |
| 情感倾向分析 | 自动对包含主观信息的文本进行情感倾向性判断，为口碑分析、话题监控、舆情分析等应用提供基础技术支持。 |
| 意图识别 | 通过分析用户请求的依存句法结构信息，抽取其中的语义主干及相关语义成分，帮助智能实现对用户意图的精准理解。 |
| AI模型选择 | 对经过基础服务层处理后的数据提取特征值、选择合适的AI模型并进行模型训练，对训练后的结果进行验证，并反馈给基础服务层进行修正，从而提高分析的准确度 |
| 模型训练 |
| 结果验证 |
| 基础服务层修正 |
| 通用分词库 | 这类分词库是语义分析的基础，对于小规模的语义分析，一般35万个分词基本可以完成，本系统要求至少提供50万个通用分词 |
| 行业分词 | 这类分词主要是通过四川移动的工单、规章制度、系统使用手册、系统功能规格说明书等形成的，预计分词量在40万个左右 |
| 地址库 | 地址库要求覆盖四川全省的地址信息 |
| 情感分词库 | 情感分词量在10万个左右 |
| 行业知识图谱 | 通过分词处理后，以及对四川移动的工单、规章制度、系统使用手册、系统功能规格说明书分析，知识图谱在10万条，最大20维。 |
| 任务调度模块 | 任务管理增加管理 | 该语义识别平台需要支持至少10个线程的并发调用，能够动态调整服务器占用率，能够支持快速扩容。新增任务须在秒级识别出来，并且在上一个任务完成后秒级开始处理。扩容时间不超过24小时。 |
| 任务管理删除管理 |
| 任务管理修改管理 |
| 任务管理查询管理 |
| 服务器占用率统计 |
| 任务调度机制 | b）任务调度机制。 |
| 扩容策略 | c）扩容策略。 |
| 用户模块 | 用户信息录入 | 该平台设计给网优中心使用，但是用户可能来自于各个不同部门，所以需要对该平台进行相应的用户权限设置。需提供多角色权限设置，提供用户管理功能，灵活的权限控制功能。配置完成后应在分钟级生效。 |
| 角色权限设计 | a）角色权限设计。 |
| 用户权限配置策略 | b）用户权限配置策略。 |
| 日志管理模块 | 运行日志记录 | 该平台需要时刻记录运行的情况，支持多种日志输出模式，如全日志输出、错误日志输出等，方便调试与追踪。能够记录人工智能运作的情况。配置完成后应在分钟级生效。 |
| 日出输出 |
| 日出呈现与查询 |
| 错误代码设计 | a）错误代码设计。 |
| 日志配置设计 | b）日志配置设计。 |
| 语音识别（附加项） | 语音识别应用 | 1. 建立语音识别应用,能够调用多种语言库 |
| 语音识别库(标准普通话) | 2. 建立语音识别库(标准普通话)，训练机器人识别 |
| 语音识别库(四川方言) | 3. 建立语音识别库(四川方言)，训练机器人识别 |
| 补充标点符号 | 4. 能够自动补充标点符号，使得文字内容可读性高。 |
| 账号权限模块配置 | 账号信息自主查询 |  |
| 账号新增管理 |  |
| 账号删除管理 |  |
| 账号权限模块可视化配置 |  |