

团体标准

T/CES XXX-XXXX

移动端智能交互多轮对话流程的设计规范

Design specification for intelligent interactive multi round dialogue process
of mobile terminal

(征求意见稿)

XXXX-XX-XX 发布

XXXX-XX-XX 实施

中国电工技术学会 发布

目 次

1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 移动端智能交互多轮对话处理流程.....	4
4.1 语音识别模块.....	5
4.2 自然语言理解模块.....	5
4.3 对话管理模块.....	5
4.4 知识管理模块.....	5
4.5 自然语言生成模块.....	5
4.6 语音合成模块.....	5
5 智能交互多轮对话流程规范.....	6
5.1 话术设计规范.....	6
5.2 对话管理规范.....	7
5.3 场景设计规范.....	9
5.4 必要的兜底回答和引导.....	9
5.5 及时反馈.....	10
5.6 超时退出.....	10

前 言

为规范移动端智能交互多轮对话流程规范,为智能交互对话流程工作开展提供指导规范,制定本文件。

本文件按照 GB/T1.1—2020《标准化工作导则 第 1 部分:标准化文件的结构和起草规则》给出的规则起草。

本文件由四川中电启明星信息技术有限公司提出。

本文件由中国电工技术学会标准工作委员会能源智慧化工作组归口。

本文件起草单位:国网信息通信产业集团有限公司、四川中电启明星信息技术有限公司、国网重庆市电力公司。

本文件主要起草人:李强、宋卫平、吕跃春、叶林峰、周孔均、赵峰、倪平波、李炳森、李欢欢、邓大建、赵峰、田鹏、徐小云、杨帆、周平。

本文件为首次发布。

1 范围

本标准规定了移动端智能交互多轮对话的处理流程和设计规范。

本标准适用于设计、开发移动端智能交互多轮对话组件的相关人员。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 5271.34-2006 信息技术 词汇 第 34 部分：人工智能 神经网络

GB/T 5271.28—2001 信息技术 词汇 第 28 部分 人工智能 基本概念与专家系统

GB/T 5271.31-2006 人工智能机器学习

GB/T 21023 中文语音识别系统通用技术规范

GB/T 36464（所有部分） 信息技术 智能语音交互系统

3 术语和定义

下列术语和定义适用于本文件。

3.1

词法分析 lexical Analysis

找出词汇的各个词素，从中获得语言学的信息，主要任务是词性标注和词义标注。

3.2

句法分析 Parsing

发现句子内部结构的方法，显式地发现句子中可能存在的各种谓词-论元的依存关系。

3.3

语义分析 Semantic Parsing

句子或文本中识别出意义块（meaning chunk），确定语言所表达的真正含义或概念，并尝试将其转换为某种数据结构的过程（将自然文本映射成计算机可处理的结构化表示），包括深层语义分析（deep semantic parsing）与浅层语义分析（shallow semantic parsing），后者常用方法是语义角色标注（semantic role labeling）。

3.4

分词 Word Segmentation

指对字符序列进行分块处理的过程，其输出结果由分开的有意义的词元组成，是文本分析的基础性步骤。

3.5

意图 Intention

是从承诺实现的愿望中选取的当前最需要完成或者最适合完成的一个，是当前主体（agent）将要正在实现的目标，它是属于思维状态的意向方向。当前意图对主体（agent）得到当前动作具有指导性的作用。

3.6

槽值 Slot

指智能交互系统的记忆。它们以 key-value 的形式存储，可用于存储用户提供的信息以及收集到的关于外部世界的信息。大多数时候，槽值会影响对话的进展。不同的行为有不同的槽值类型。

3.7

命名实体 Named Entity

1) 狭义讲，命名实体是唯一一个体的专有名称。狭义命名实体具备三个充要属性。属性 a. 符号性：命名实体是一个区别性称谓，具有代号性质；命名实体并不排除蕴含义，但蕴含义不是命名实体的必然属性。属性 b. 个体性：命名实体所指称的事物不可向下分类，因为类是对群体区分的结果（但任何事物都是一个过程，都是一个和合态，所以，既可以按阶段划分，也可以分解成部分）。属性 c. 唯一性：命名实体所指称的事物是世上独一无二的，通常不以数量词修饰；一旦饰以数量词，命名实体便转义了。

2) 广义讲，除狭义命名实体之外，命名实体还包括对一类事物的非概念性称谓。对一类事物的命名可划分为概念性命名和非概念性命名两种方式。概念性命名是通过自然语义和构词法来规定事物内涵的，如：月门、拱桥；非概念性命名是通过不表明事物内涵的饰词（一般充当命名实体饰名）来指称一类事物的，如：“丁达尔现象，一二次设备”。广义命名实体一般只具备符号性，不具备个体性和唯一性。

3.8

话术 Words

用户与智能设备对话的形式。

3.9

多轮对话 Multi-turn Dialogue

在一定时间和交互次数内，智能交互系统每次发送内容时，考虑对话的上下文，从上下文情景中分析最合理的回答，准确提供回复。

3.10

动作 Action

在识别到意图后，系统根据用户的意图进行的下一步操作。

3.11

自然语言理解 Natural Language Understanding, NLU

自然语言理解是所有支持机器理解文本内容的方法模型或任务的总称。其在文本信息处理系统中扮演者非常重要的角色，是推荐、问答、搜索等系统的必备模块。

3.12

自然语言生成 Natural Language Generation, NLG

能够根据一些关键信息及其在及其内部的表达形式，经过一个规划过程，来自动生成一段高质量的自然语言文本。

3.13

对话管理 Dialog Manager, DM

根据用户当前的输入，以及对话上下文，决定系统下一步最佳响应。对于任务型 DM，其职责是通过一致性的对话交互，完成用户的对话的对话目标。

3.14

智能交互 Intelligent Interaction

智能交互一般指智能语音交互。智能语音交互是基于语音输入的新一代交互模式，通过说话就可以得到反馈结果。

3.15

语音识别 Automatic Speech Recognition, ASR

将人的语音转换为机器可以识别并理解的文本或命令的过程。

3.16

模型 Model

系统或其他事物及过程的抽象表示。

3.16

模版 Template

是指在设计方案时的固定格式，是将一个事物的结构规律予以固定化、标准化的成果，它体现的是结构形式的标准化

3.17

语音合成 Text To Speech, TTS

将文字转换为声音输出的过程。

4 移动端智能交互多轮对话处理流程

智能交互多轮对话技术作为人工智能领域地核心技术，即将成为新的和谐人机交互方式。作为新一代交互系统，智能交互也面临很多难题。准确理解的用户的需求，并在用户能忍受地时间内完成用户的需求，使用户满意是智能交互的目标。根据用户输入的表达，包括语义表达和语义解析，做出满足用户期望的语义表达，满足用户需求。多轮对话中为了清晰明确的理解用户的意图和需求，将对话建模过程中缺少的信息形成一个填槽值的过程，槽值(slot)就是多轮对话对话当中将初步用户意图转化为明确用户指令所需要补全的信息。一个槽值与任务处理中所需要获取的一种信息相对应。槽值不应当是有顺序的，缺什么槽值就向用户询问什么信息。这些需要开发人员和设计人员自己设计。下图为移动端智能交互多轮对话处理流程：

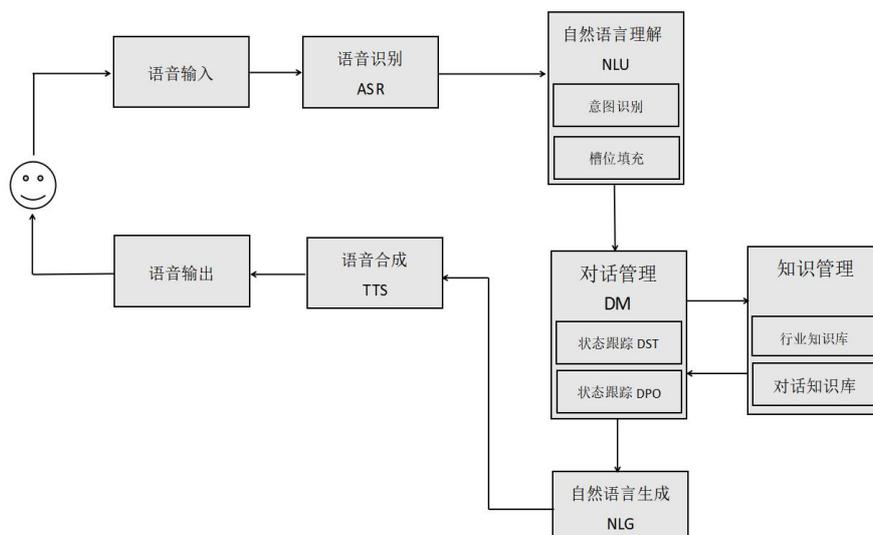


图 4.1 移动端智能交互多轮对话系统

下面对移动端智能交互多轮对话系统模块功能进行详细阐述：

4.1 语音识别模块

该模块的主要任务是将人类的语音中的词汇内容转换为计算机可读的输入，例如二进制编码或字符序列。语音识别技术就是让机器通过识别和理解把语音转变为相应的文本或命令的技术。该技术主要包括：特征提取技术、模式匹配技术、模型训练技术。

4.2 自然语言理解模块

自然语言理解的目的是将用户的输入映射到预先根据不同场景定义的语义槽中，通常包括意图识别和槽位填充两个任务。自然语言理解应尽可能完整、清晰和准确地将用户输入转化为计算机能够理解的形式，NLU 模块的准确性对对话系统的质量有很大影响。

4.3 对话管理模块

对话管理（Dialog Management, DM）控制这人机对话的过程，DM 根据对话历史信息，决定此刻对用户的反应。最常见的应用还是任务驱动的多轮对话，用户带着明确的目的如订用车、订票等，用户需求比较复杂，有很多限制条件，可能需要分多轮进行陈述，一方面，用户在对话过程中可以不断修改或完善自己的需求，另一方面，当用户的陈述的需求不够具体或明确的时候，机器也可以通过询问、澄清或确认来帮助用户找到满意的结果。

对话管理模块主要任务是对话状态维护（Dialog State Tracking, DST），对话状态记录了当前填充的槽植（Slot），即将执行的动作（Action）以及即将填充的槽植（Slot）。用数学形式表达为， $t+1$ 时刻的对话状态 $S(t+1)$ ，依赖于之前时刻 t 的状态 S_t ，和之前时刻的 t 的系统行为 A_t ，以及当前时刻 $t+1$ 对应的用户行为 $O(t+1)$ 。可以写成 $S(t+1) \leftarrow S_t + A_t + O(t+1)$ 。

4.4 知识管理模块

知识管理作为一个接口，为智能交互多轮对话提供一个知识查询功能，其主要有行业知识库和对话知识库，其中行业知识库中存储大量本行业的相关知识，供用户查询；对话知识库作为一个存储用户需求的模块，可将其作为语料库，为模型的泛化，改善系统的交互能力提供支撑。

4.5 自然语言生成模块

自然语言生成模块的主要任务是将 DM 模块输出的抽象表达转化为句法合法、语义准确的自然语言句子。一个好的应答语句应该具有上下文的连贯性、回复内容的精准性、可读性和多样性。

4.6 语音合成模块

语音合成是将智能交互系统产生的文字信息转变为人类听得懂的、流利的口语进行输出的技术。

5 智能交互多轮对话流程规范

5.1 话术设计规范

在与智能设备交互过程中，用户通常使用比较固定的句式，当在话术设计时，从中可以发现一些规则，避免走入不必要的对话路径，从而导致用户挫败感或者对交互不满意。

话术设计要求包括：

- (1) 准确：表述要避免笼统模糊，没有歧义。

表 1 话术设计准确要求样例

	Not Recommended	Recommended
系统	"你可以这样说： 1.【日程查询】'查询所有日程' 2.【日程查询】'请输入关键词，如： 周例会'"	"你可以这样说： 1.【日程查询】'查询所有日程' 2.【日程查询】'请输入关键词'"
用户	请输入关键词	周例会
系统	进入下一题请说“继续”	说“继续”进入下一题
用户	继续进入下一题	继续

- (2) 简洁：表述要简单明了，避免信息量过大，一次提供选项不超过 3 个。

表 2 话术设计简洁要求样例

	Not Recommended	Recommended
系统	欢迎使用 AI 助手，我为您提供现场违章行为检测功能，可以检查现场违章行为，并将违章行为上传到违章数据集；拍摄变电站缺陷功能，检查变电站缺陷，并上传到缺陷数据集；语音笔记功能，记录开会语音等；拍摄输电通道隐患功能，并上传到输电隐患数据集。	欢迎使用 AI 助手，我为您提供了现场违章行为检测功能，拍摄变电站缺陷功能，检查变电站缺陷，语音笔记功能；拍摄输电通道隐患功能。
用户	上传到缺陷数据集	拍摄输电通道隐患

- (3) 意图确认：在不确定用户意图时进行”意图确认“。

表 3 意图确认设计样例

	Not Recommended	Recommended
用户	现场违章行为	现场违章行为
系统	好的,正在为您进行现场违章行为拍摄	请问您是需要进行现场违章行为拍摄还是询问现场违章行为又能哪些?
用户		现场违章行为拍摄
		好的,正在为您进行现场违章行为拍摄

(4) 需求未满足时, 不轻易终结对话。

表 4 需求未满足设计样例

	Not Recommended	Recommended
系统	正在为您完成表单资料收集和提交	正在为您完成完成表单资料收集和提交
用户	好的	好的
系统	已完成表单资料收集,工作事项标题为: 计量资产领料单。	已完成表单资料收集,工作事项标题为: 计量资产领料单。是否需要修改标题? 如需修改, 请对我说: XXX 领料单
用户 (我想修改表单名字)	需要

(4) 自然: 措辞口语化, 避免使用专业的、晦涩的语言; 句式自然, 使用对话标记; 统一内容提供多样性表达。

(5) 友好: 主动认错, 不责怪用户

(6) 人格化: 适时幽默

5.2 对话管理规范

即使你完美的设计了你的智能交互系统, 用户也会不可避免的对你的智能交互系统说一些意料之外的话语。

(1) 多轮对话的容错性

a. 系统需要有相应的机制来优雅处理意图识别失败、用户意外输入这类问题。如:

当前系统提供了查询天气，查询日程，添加日程，查询联系人等相关业务支持，但不提供找工作的能力，此刻可以针对这类问题进行统一友好回复。

表 5 容错性设计样例

	Not Recommended	Recommended
系统	"你可以这样说： 1.【日程查询】'查询所有日程' 2.【日程查询】'请输入关键词，如：周例会'"	"你可以这样说： 1.【日程查询】'查询所有日程' 2.【日程查询】'请输入关键词，如：周例会'"
用户	我想在 Rasa 申请一份工作	我想在 Rasa 申请一份工作
系统（不知道接下来怎么办）	我知道您在找工作，但恐怕我还不能掌握这项系统

b. 系统需要对预测置信度较低的动作进行回滚，并对用户进行统一回复。

如：在设计智能交互系统时，设置可配置的动作（Action）置信度阈值，当用户发送意料之外的消息，这种行为可能会将他们引向未知的对话路径，从而导致预测动作置信度低于设置阈值，此时需回退对话，并给用户一个友好反馈，例如：“对不起，我不明白您的意思”。

注：在统一回复中，大多数设计话术来自于真实对话场景里。

（2）消除歧义

对话是双方共同的行为，双方必须不断地建立共同地基础，也就是双方都认可地事物地集合。共同基础可以通过听话人依靠或者确认说话人地话断来实现。确认行为由弱到强地 5 种方法有：持续关注（continued attention）、相关邻接贡献（relevant nextcontribution）、确认（acknowledgement）、表明（demonstrate）、展示（display）为了让智能交互系统有机会了解用户想要什么，您通常会希望它通过询问澄清问题来试图消除用户信息的歧义。

Two-Stage 回退是用来处理低 NLU 置信度在多个阶段使用一下顺序：

step 1: 用户消息被低置信度分类

要求用户确认意图

step 2: 用户确认或者拒绝意图

如果用户确认了，对话继续，不采取进一步的 fallback 步骤

如果用户拒绝了，要求用户重新措辞消息

step 3: 用户会重新表达他们的意图

如果消息高置信度识别为意图，对话继续，就好像用户一开始就有这个意图

如果重新措辞的用户消息仍然具有低置信度，则要求用户确认意图

step 4: 用户确认或否认改写后的意图

如果用户确认了，对话继续，就好像用户从一开始就有这个意图

如果用户拒绝了，则会触发一个最终的备用操作（例如，将任务移交给人类）。默认的最终回退操作是调用 `action_default_fallback`，此操作导致智能交互系统发出默认响应（`utter_default`），并重置对话状态。

5.3 场景设计规范

（1）一个给予用户的”愉悦路径“，即可以用最简单的方式完成任务的路径，不会过于复杂。

（2）其他完成任务的路径，能够使用户完成和”愉悦路径“同样的任务。这样可能会是多样化的，因为有的用户会选择一次只说一部分信息，二其他用户可能会一次把信息全说完。

（3）需要进行调整的对话场景，例如系统无法支持或不能理解用户的请求。

（4）用户中途退出，以及用户完成任务后对话结束的场景。需要考虑到如何对话的结束足够明确。

（5）对用户的问候语，以及引出对话操作的方式。

5.4 必要的兜底回答和引导

在用户需要帮助的时候提供必要的帮助说明和引导，其作用是让用户知道当前对话进入兜底状态，对用户进行安抚并让用户知道接下来怎么办。有以下几点原则：

（1）应答要清晰明确

（2）意图不明时要优先确认意图，确认无果在退出

表 6 兜底回答和引导设计样例

	Not Recommended	Recommended
系统	“Hi，您好，我是小 i，您的随身智能助手。我能为您提供日程管理、知识库查询、申请用车、发起流程、位置导航、	“Hi，您好，我是小 i，您的随身智能助手。我能为您提供日程管理、知识库查询、申请用车、发起流程、位置导航、扫描实物二维码查询设备资料服务，快来试试吧！ 您可以这样对我说：1.【日程管理】“明天

	扫描实物二维码查询设备资料服务，快来试试吧！”	上午九点在 19 楼 1 会议室参加周例会”“日程查询” 2.【知识库查询】“查询下什么是供应商大会” 3.【流程申请】“我要用车”“我要请假” 4.【位置导航】“我要去市北供电公司” 5. 通过对话框的“+”号，拍一张设备实物二维码发给我。 “
用户	位置导航	查询下什么是供应商大会
系统	对不起不明白您的意思	好的，正在为您查询，请稍等

5.5 及时反馈

智能交互系统除了具备功能实用、运行稳定、响应流畅、界面美观之外，能提升用户体验的另一个至关重要的因素就是反馈。用户在与系统进行交互的过程中，无论多轮对话成功与否，系统要给予及时友好的反馈。

表 7 及时反馈成功设计样例

	Not Recommended	Recommended
用户	我要申请用车	我要申请用车
系统	没有回复	好的，小 i 正在为您申请用车

表 8 及时反馈失败设计样例

	Not Recommended	Recommended
用户	我要申请用车	我要申请用车
系统	没有回复	您好，小 i 目前还不能提供申请用车功能

5.6 超时退出

在设计智能交互系统时，需要考虑多轮对话历史信息，历史信息保留在存储介质上，会根据用户 id 进行存储，一般选用快速响应存储介质，如：Redis，MongoDB,ES 等存储空间，系统每次接受到用户的 message，会根据用户 id 召回历史信息（如槽值、意图），如果用户多时没有输入，则需退出当前对话场景，否则会加大存储介质压力。一般定义的超时退出时间在 3 分钟，用户没有任何沟通会自动退出多轮，各智能交互系统可以根据自己情况酌情调整。

