

从“连接”到“交互”

阿里巴巴-iDST 资深专家
孙健/千诀

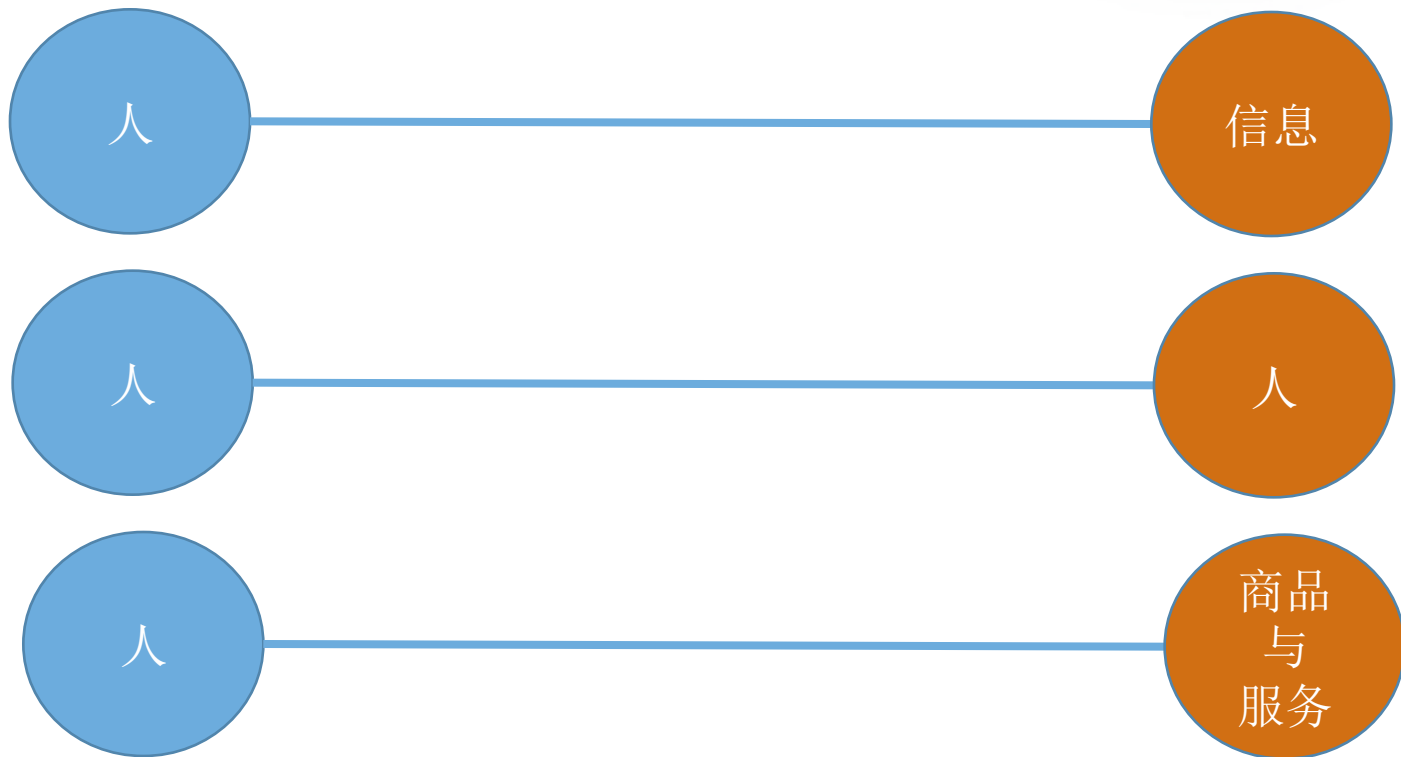


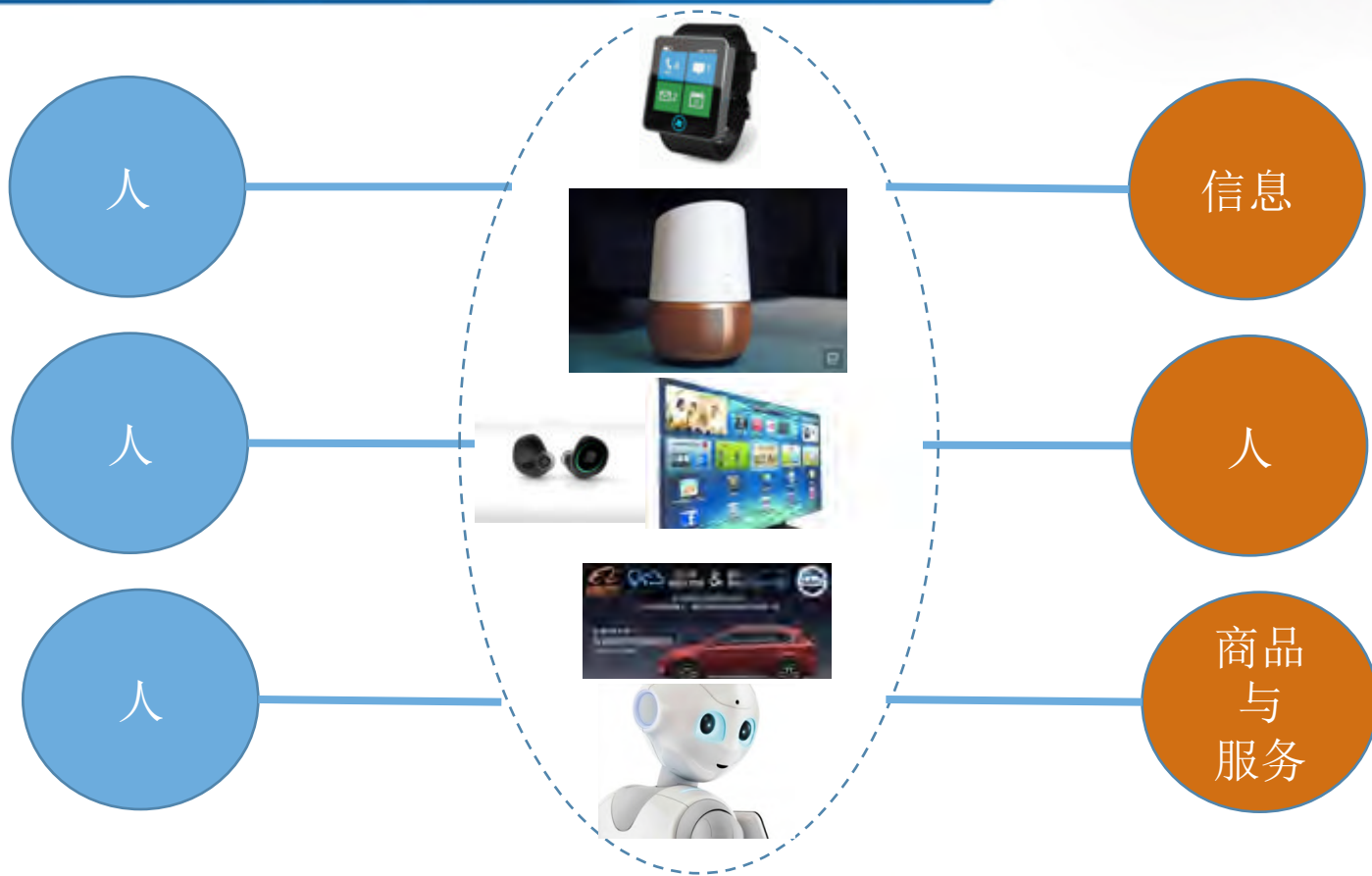
从“连接”到“交互”的趋势

阿里在智能对话交互方向的进展

开放和赋能合作伙伴

智能对话交互生态的范式思考

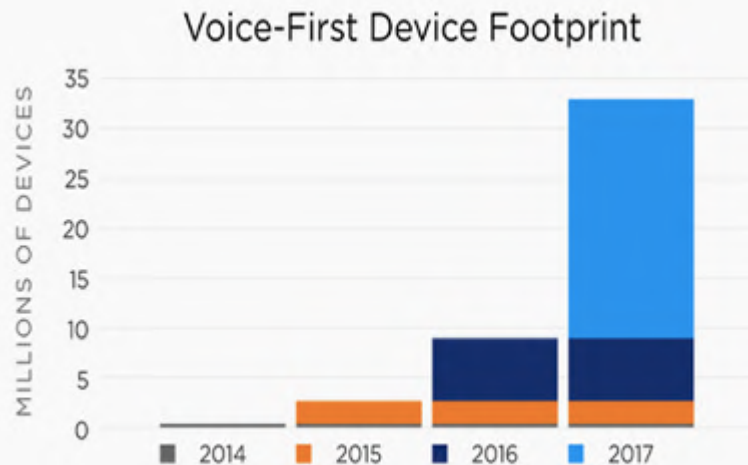




- 可穿戴设备
 - 智能眼镜、智能耳机、智能手表...
- 智能家居
 - 智能电视、智能音箱、智能冰箱、智能服务机器人...
- 智能出行
 - 智能汽车、智能后视镜、智能记录仪...

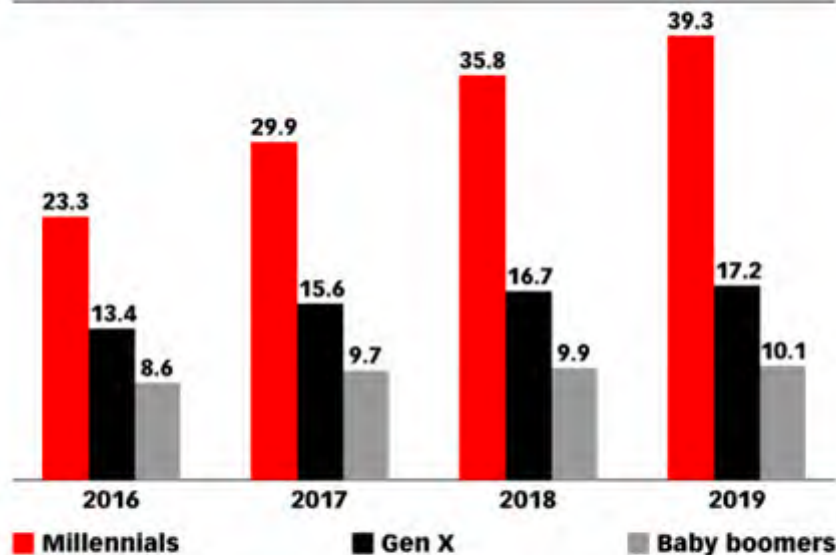
年份 美国语音设备出货量

2015	170万台
2016	650万台
2017	2450万台



VoiceLabs Analysis combined with research from CIRP, KPCB and InfoScout

US Voice-Enabled Digital Assistant Users, by Generation, 2016-2019
millions



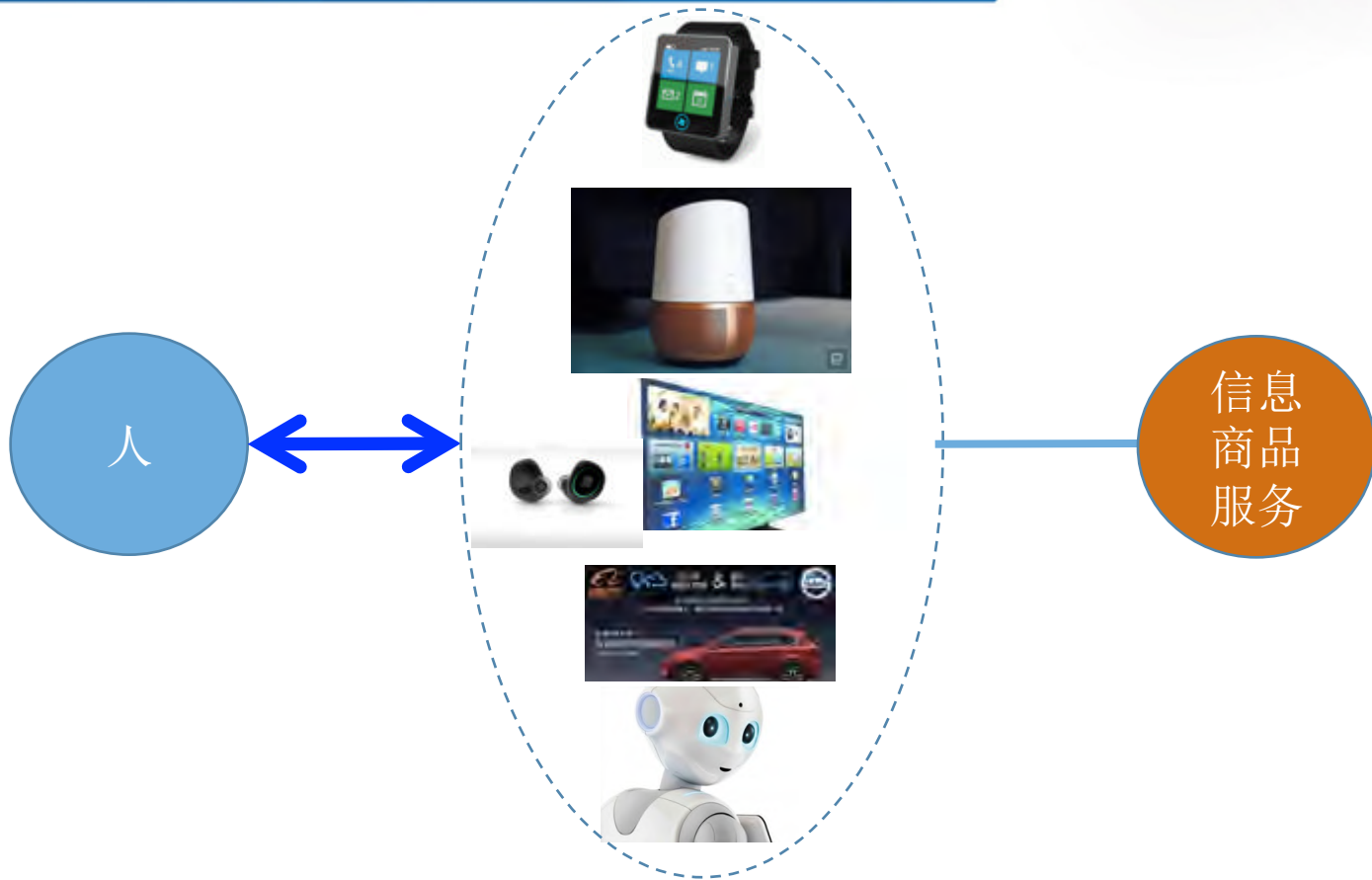
Note: individuals who use voice-enabled digital assistants at least once a month on any device; millennials are individuals born between 1981-2000, Gen X are individuals born between 1965-1980 and baby boomers are individuals born between 1945-1964

Source: eMarketer, April 2017

年龄	2017年
千禧一代 (1981-2000)	29.9M
X世代 (1965-1980)	15.6M
婴儿潮 (1945-1964)	9.7M

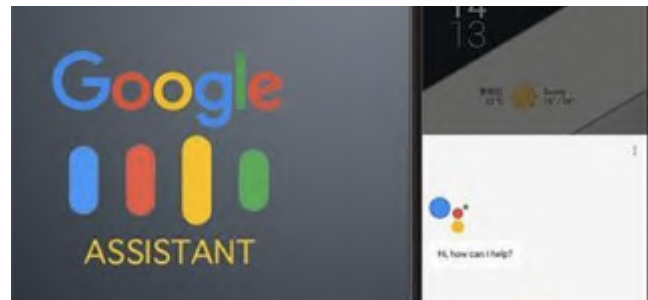
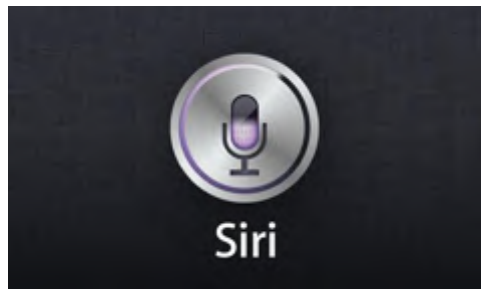
从“连接”到“交互”的趋势

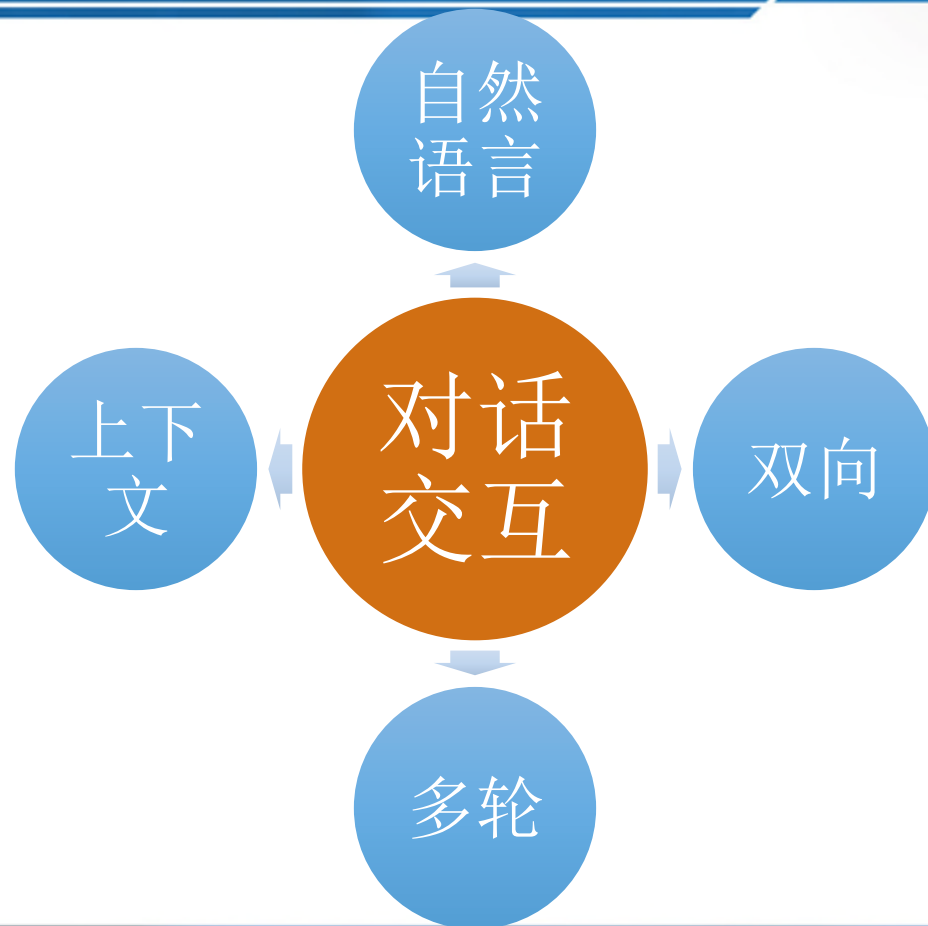
CCTC 2017 中国云计算技术大会
Cloud Computing Technology Conference 2017



业界在对话交互方向的布局

CCTC 2017 中国云计算技术大会
Cloud Computing Technology Conference 2017





从连接到对话交互背后的本质改变

CCTC 2017 中国云计算技术大会
Cloud Computing Technology Conference 2017

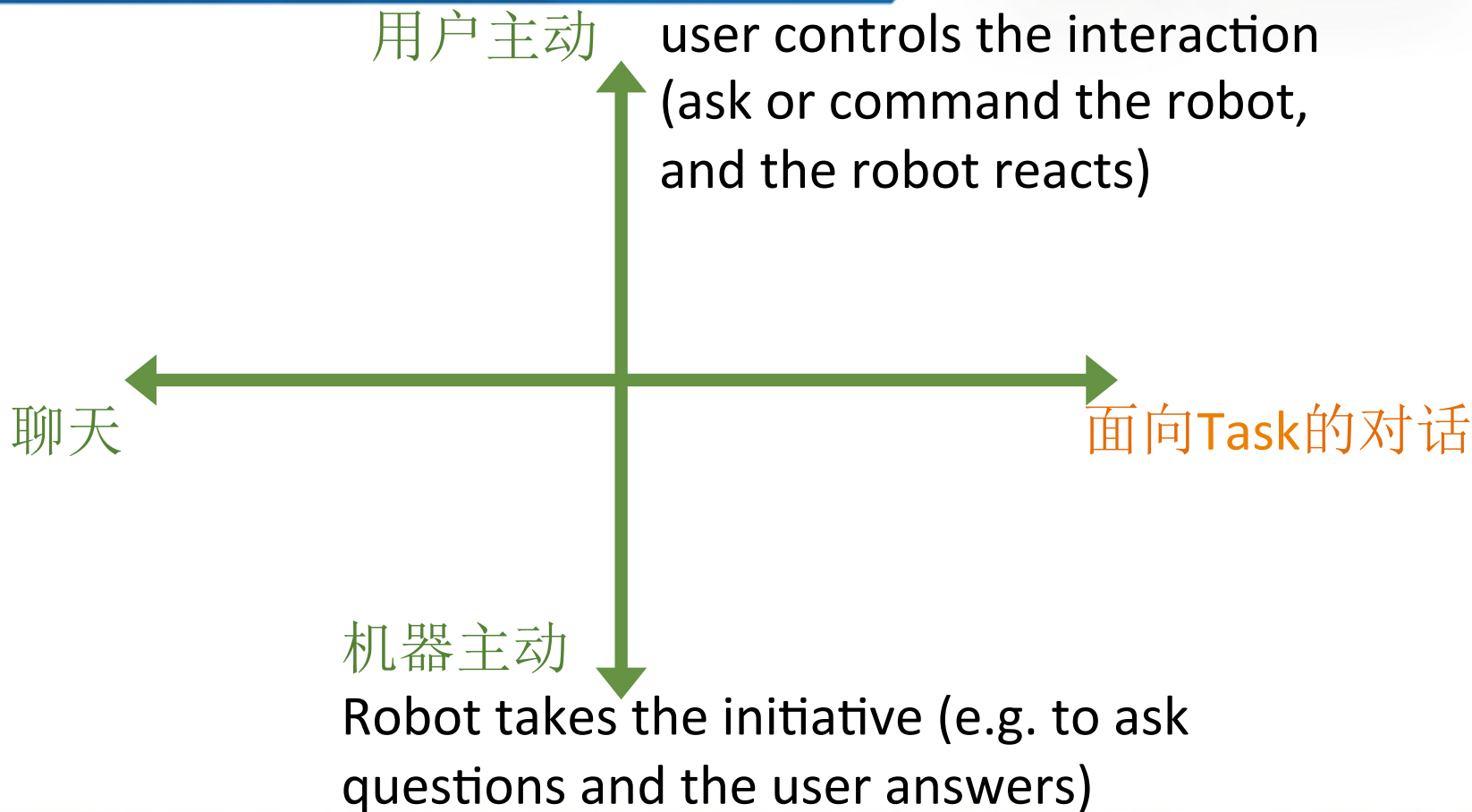


受限->确定性

开放->不确定性

连接

对话交互



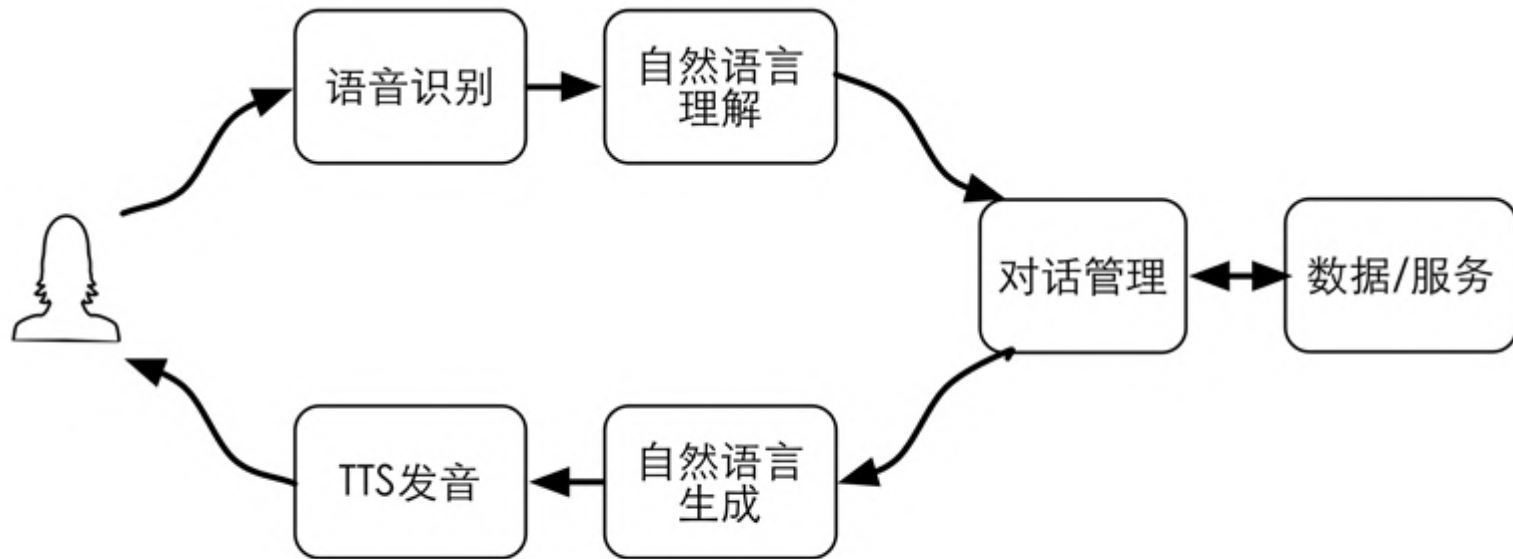


从“连接”到“交互”的趋势

阿里在智能对话交互方向的进展

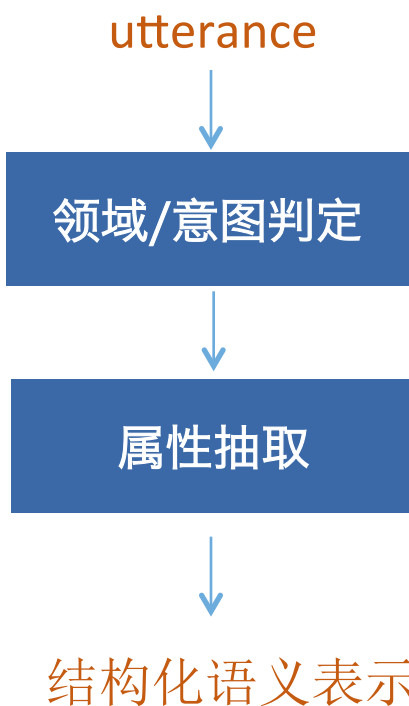
开放和赋能合作伙伴

智能对话交互生态的范式思考

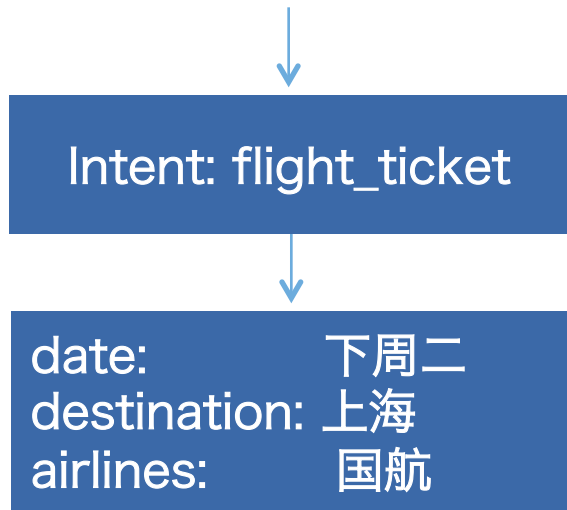


对话系统架构简图

- 问题：将人的语言形式化为结构化、完整的语义表示



买一张下周二去上海的飞机票，国航的



- 多样性

- 我要听大王叫我来巡山
- 给我播大王叫我来巡山
- 我想听歌大王叫我来巡山
- 放首大王叫我来巡山
- 给唱一首大王叫我来巡山
- 放音乐大王叫我来巡山
-

- 歧义性

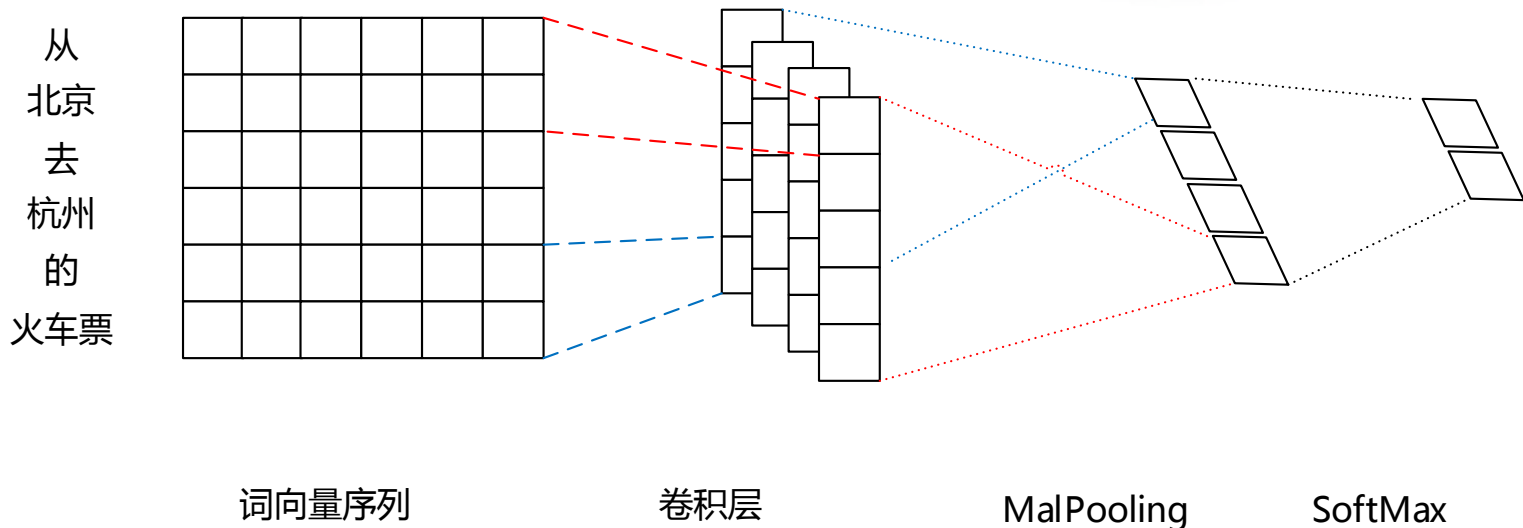
- 我要去拉萨

- 鲁棒性

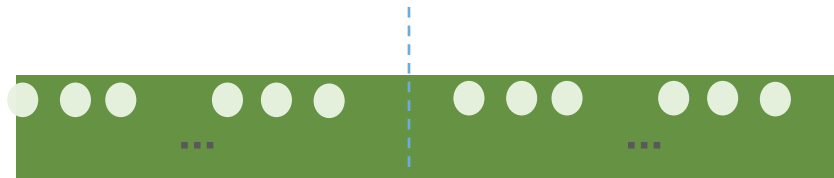
- ASR错误：大王叫我来新山
- 多字：大王叫让我来巡山
- 少字：大王叫我巡山
- 别称：熊大熊二（指熊出没）
- 不连贯：我要看那个恩花千骨
- 噪音：全家只有大王叫我去巡山咯
- ...

- 上下文

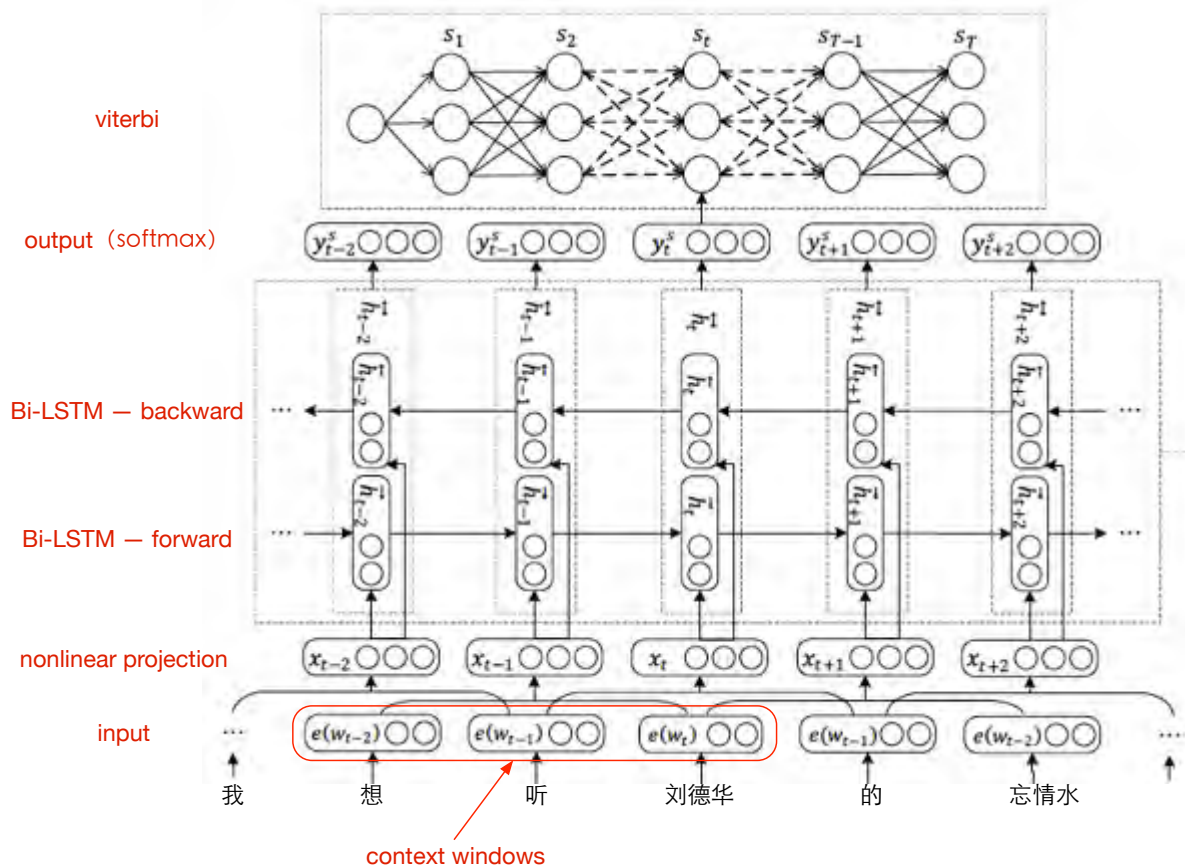
- 模型



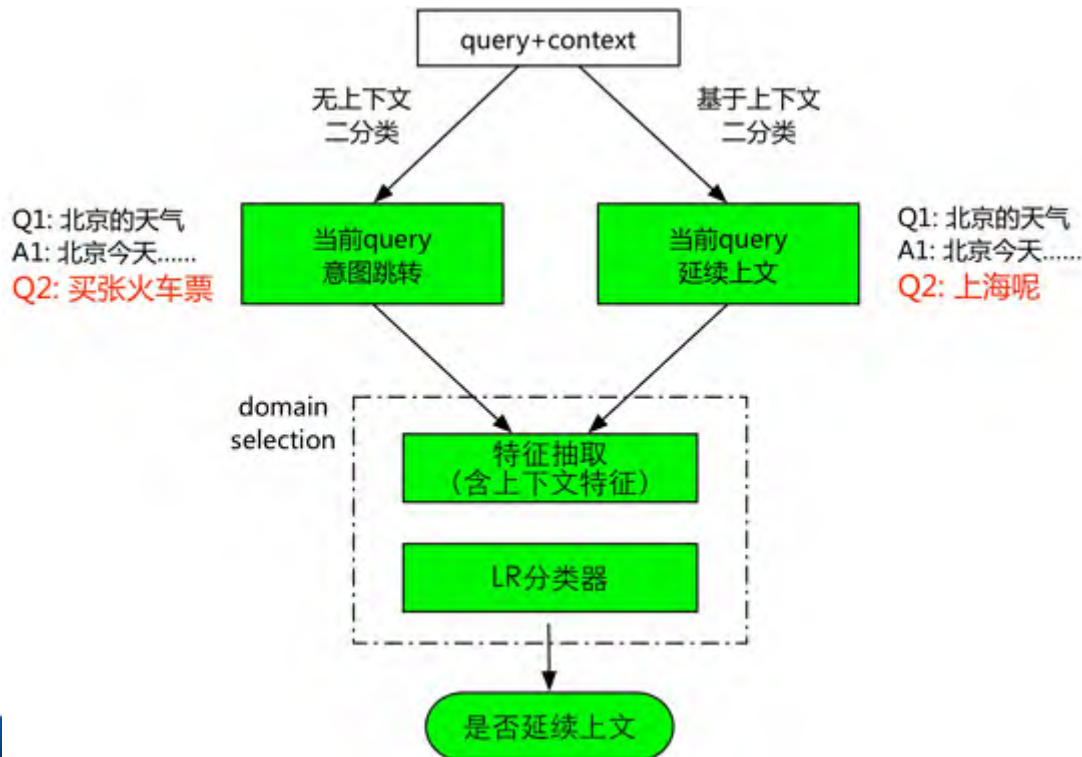
- 把词的语义知识融入到分类中
 - 大王叫我来巡山、爱探险的朵拉



- 怎样让模型更鲁棒？
 - 口语：我想去额如家快捷，崇文门附近的那个必胜客
 - 叠字：我要听要听刘德华的歌，北京北京天气怎么样
 - 随机：那英的内新歌，我要听岳云鹏卡



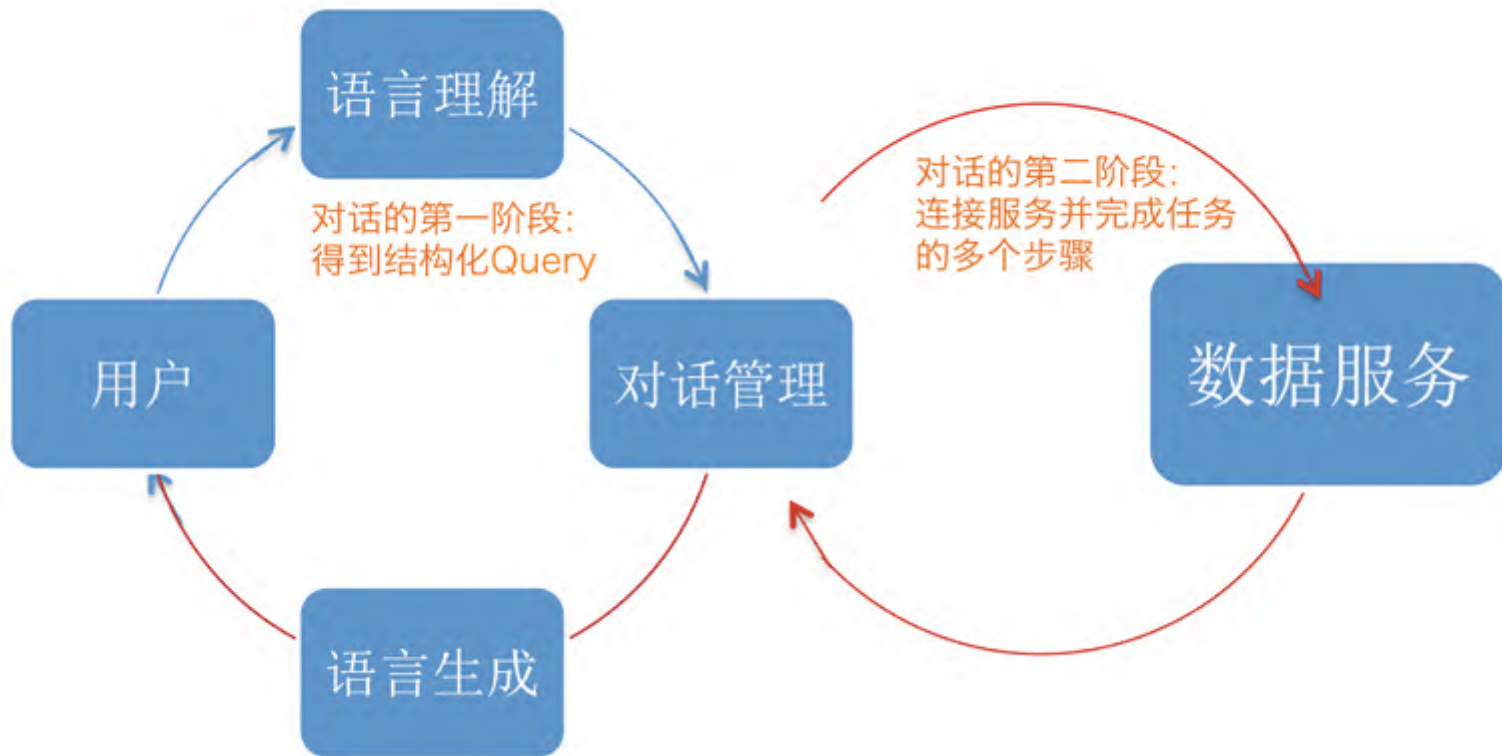
- 抽象为：在上下文场景下，判断当前句子是延续上文domain还是新起domain？即**二分分类器**



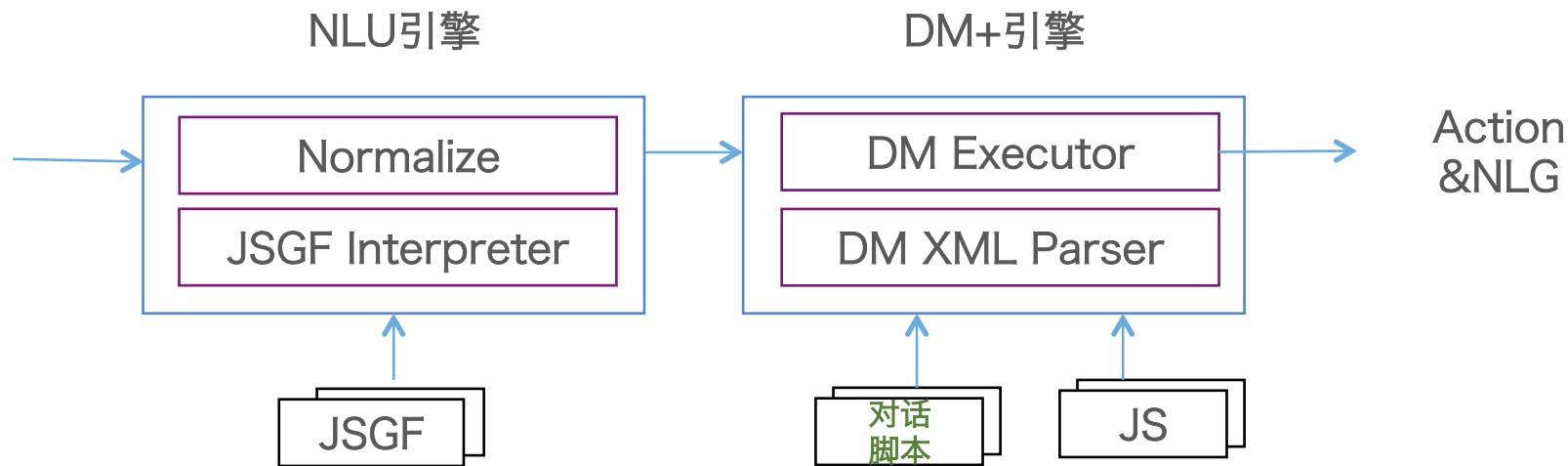


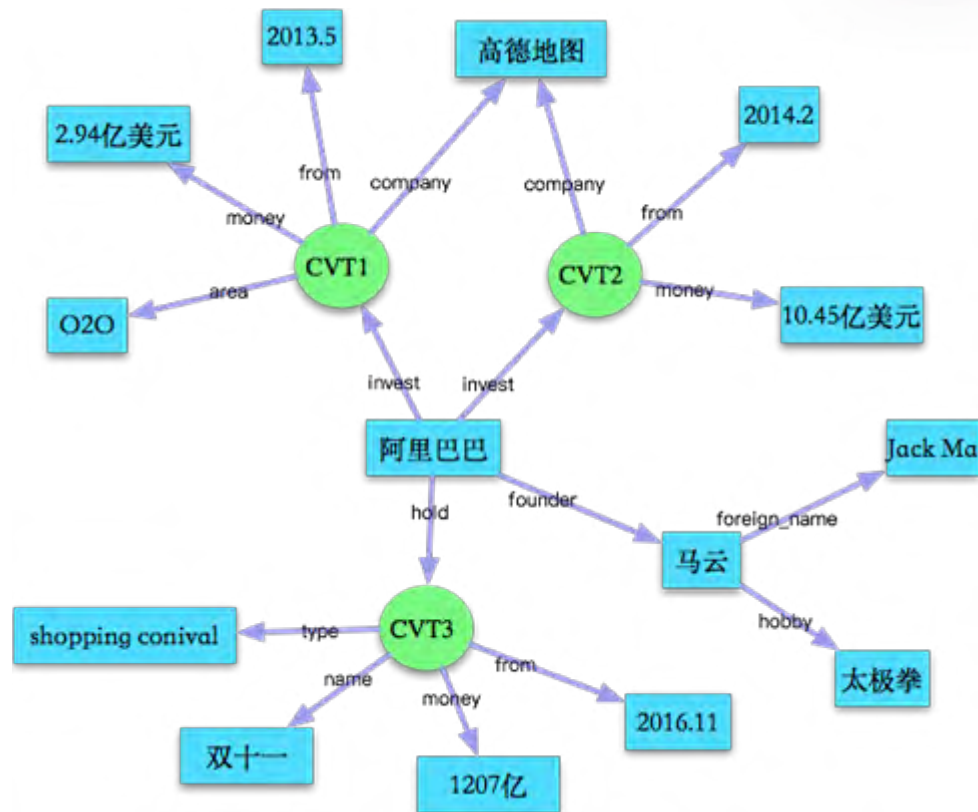
- 用于指导语言生成
 - E.g. ask_slot(time), “请问你什么时间出发？”
- 用于调用数据服务
 - E.g. goto_service (train)
- 用于客户端的执行
 - E.g. open_app, set_reminder

用户	机器人
我要去杭州，帮我订一张火车票	<code>ask_slot(date_time)</code>
明天上午	<code>goto_service(train)</code>
第二个	<code>select_item(2)</code>
购买	<code>goto_pay</code>

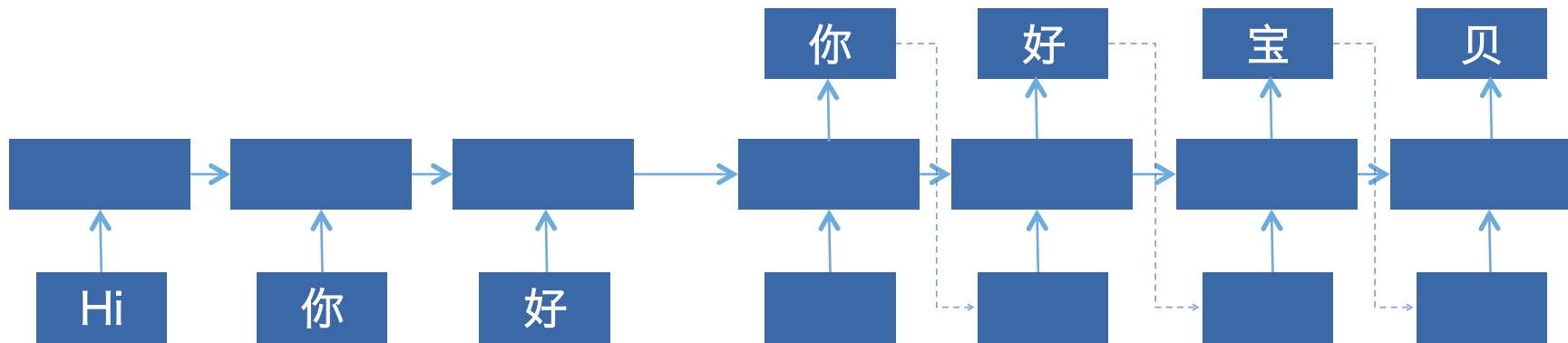


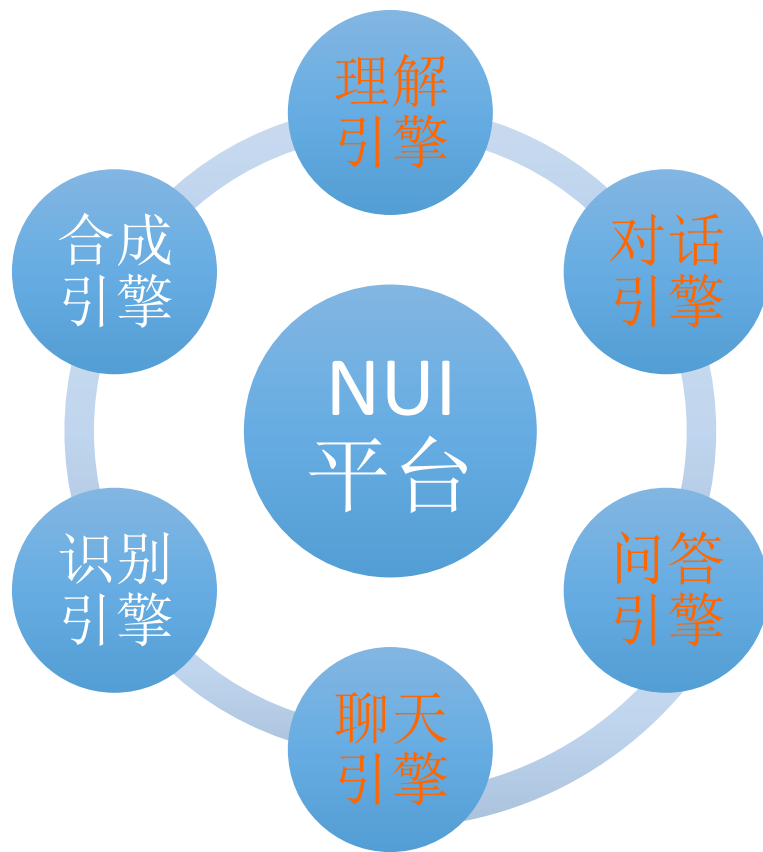
- 面向task flow的对话描述语言
- 对话引擎的中断和返回机制
- 面向开发者的Open Dialogue解决方案





- 基于 $\langle k, v \rangle$ 对的聊天引擎
- 基于seq2seq模型的生成式聊天引擎







出行

- 地图
- 导航
- 实时路况
- 天气

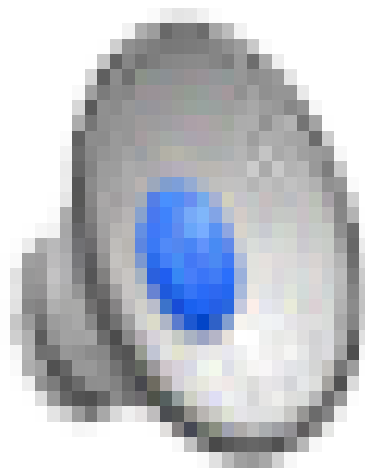
娱乐

- 音乐
- 收音机
- 有声读物

控制

- 打电话
- 设置
- 车辆控制







从“连接”到“交互”的趋势

阿里在智能对话交互方向的进展

开放和赋能合作伙伴

智能对话交互生态的范式思考

语音交互生态圈（合作伙伴）

智能手机

智能电视

互联网汽车

智能音箱机器人

其他智能设备

行业解决方案和产品（合作伙伴）

服务智能化解决方案

网上法庭解决方案

视频云语音解决方案

.....

开发者平台（SDK\API）

客户端解决方案

NUI 客户端引擎和SDK

语音硬件解决方案

NUI services

NUI服务平台

语音识别

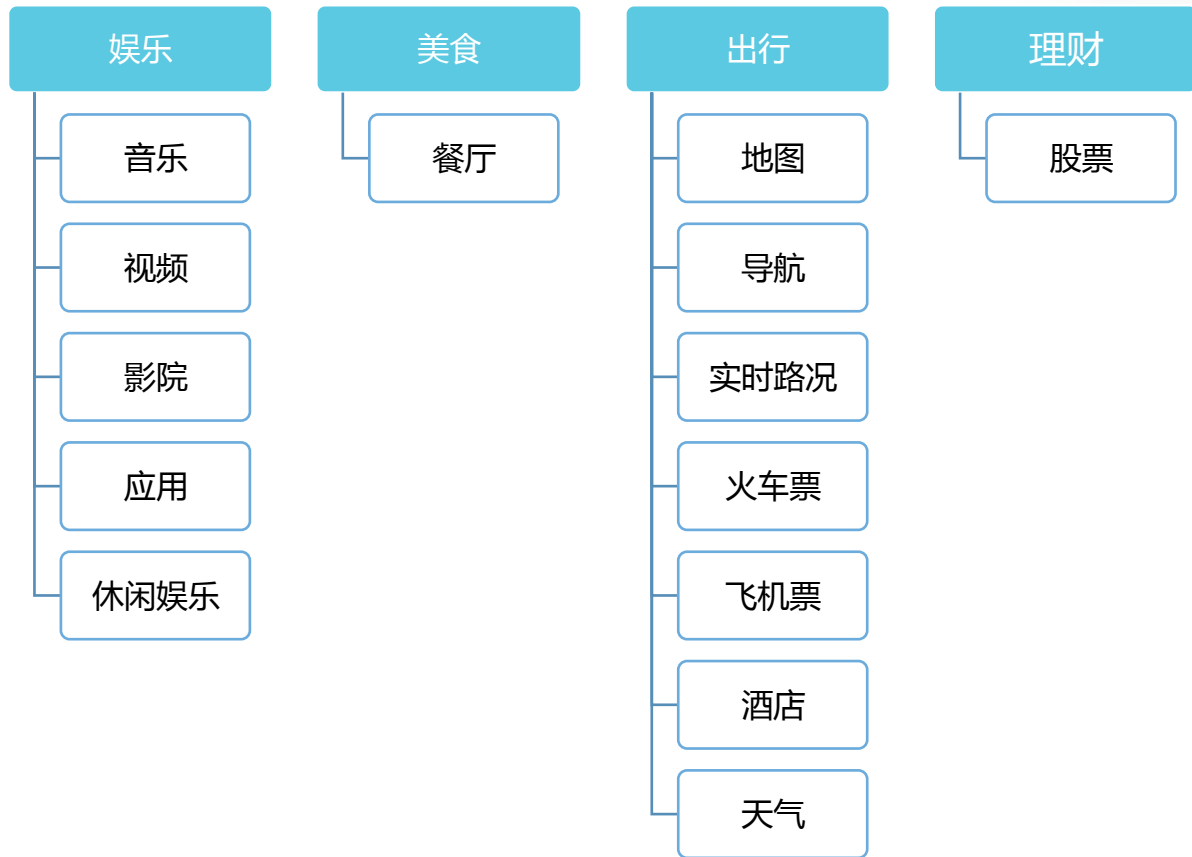
语音合成

语言理解和对话

知识问答

个性化、场景化定制工具

智能语音核心技术研发



理解引擎的定制

- 词典
- grammar

对话引擎的定制

- XML

聊天和问答引擎的定制

识别引擎的定制

- 对话交互cover的领域少
- 对话交互的体验还不够好

体验相对较好



范式1：自研

范式2：平台

扩展不够快

交互与业务耦合



体验不够好

语言理解是瓶颈
缺乏有效的评测

范式1：自研

范式3：自研+平台

范式2：平台

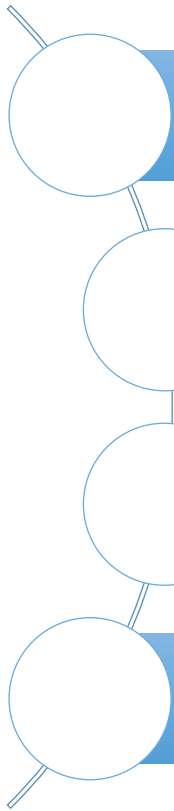
用户：坚持用户体验为先

产品：产品和交互设计想办法降低不确定性

语言：语言理解的鲁棒性和领域扩展性

交互：打造让机器持续学习能力

数据：数据闭环；接入尽可能多的数据服务



从“连接”到“交互”的趋势

阿里在智能对话交互方向的进展

开放和赋能合作伙伴

智能对话交互生态的范式思考

多谢