



对话技术与应用

崔安颀

2019/12

目录

CONTENTS

1. 对话技术的主要挑战
2. 对话应用场景举例
3. 应用案例
4. 公司简介

传统方法：模版

- 没有任何系统可以避免大量的手写模版
- 精准、快速的开发环境是成功关键
- 正则文法 / 上下文无关文法

精确 vs 鲁棒

• 统计模型、深度模型

<i>Human</i>	我渴了怎么办	I am thirsty, what do I do?
<i>Machine</i>	喝水	Drink water.
	(1) 我有点渴了怎么办	I am <i>a bit</i> thirsty, what do I do?
<i>Human</i>	(2) 我 x 有点渴了怎么办	I x a bit thirsty, what do I do?
	(3) 我来有点渴了怎么办	I <i>come</i> a bit thirsty, what do I do?
	(4) 我中国有点渴了怎么办	I <i>China</i> a bit thirsty, what do I do?
<i>Machine</i>	多喝水	Drink more water.
<i>Human</i>	我北京有点渴了怎么办中国	I <i>Beijing</i> a bit thirsty, what do I do <i>China</i> ?
<i>Machine</i>	喝王老吉	Drink Wong Lo Kat (a popular Chinese herbal tea drink).
<i>Human</i>	蝴蝶我北京有点渴了怎么办中国	<i>butterfly</i> I <i>Beijing</i> a bit thirsty, what do I do <i>China</i> ?
<i>Machine</i>	喝点水	Drink a bit water.
<i>Human</i>	蝴蝶我北京有点饿了怎么办中国	<i>butterfly</i> I <i>Beijing</i> a bit hungry, what do I do <i>China</i> ?
<i>Machine</i>	吃东西	Eat something.
<i>Human</i>	蝴蝶在北京有没有怎么办中国	<i>butterfly</i> in Beijing are there any <i>what do I do China</i> ?
<i>Machine</i>	有啊, 在北京	Yes there are (butterflies), in Beijing.

Xiong, Kun, Anqi Cui, Zefeng Zhang, and Ming Li. "Neural contextual conversation learning with labeled question-answering pairs." arXiv preprint arXiv:1607.05809 (2016).

“精准” 理解



知识 = 理解?

- 人类说话两种模式 (Daniel Kahneman: Thinking, fast and slow) :
 - 第一种模式: 直觉反应
 - 第二种模式: 想想再说, 三思而后行
- 上一句 => 语义 => 下一句



上下文

- 指代消解
- 缺失
- 消歧

Detection	A: 加拿大总理是谁? (Who is the prime minister of Canada?)
	B: 贾斯汀鲁铎. (Justin Trudeau.)
Detection	A: 他[People]老爸是谁? (Who is his[People] father?)
	Resolution 贾斯汀鲁铎老婆是谁? (Who is Justin Trudeau's father?)
Detection	A: 绵中离我们近吗, 我朋友去那里[Location]上学. (How far is Mian High school? My friend is studying there[Location].)
	B: 打车过去△[Location]十分钟 (Ten minutes cab △[Location].)
Detection	A: 打车钱他[People]报销我就去△[Location] (I will go to △[Location] if he[People] reimburses me the cab fee.)
	Resolution A: 绵中离我们近吗, 我朋友去绵中复读 (How far is Mian High school? My friend is studying in Mian High school.)
Detection	B: 打车过去绵中十分钟. (Ten minutes by cab to Mian High school.)
	A: 打车钱我朋友报销我就去绵中. (I will go to Mian High school if my friend reimburses me the cab fee.)
Detection	A: 麦当劳的薯条是什么做的?为什么△[Noun phrase]凉了吃会涩呢? (What are the ingredients in McDonalds' fries? Why do △[Noun phrase] taste bitter when cooling down?)
	Resolution 为什么麦当劳的薯条凉了吃会涩呢? (Why do McDonalds' fries taste bitter when cooling down?)
Detection	A: 法国人喜欢生一点的牛排. (The French prefer steaks rare.)
	B: △[People]为什么不吃熟的△[Noun phrase] (△[People]Why not eat well-done △[Noun phrase]?)
Detection	Resolution 法国人为什么不吃熟的牛排? Why don't the French like well-done steaks?

目录

CONTENTS

1. 对话技术的主要挑战
2. 对话应用场景举例
3. 应用案例
4. 公司简介

从“对话”技术展开

客服机器人

流程自动化

质检

助手

业务应用领域

社交媒体文本分析

文档语义问答

对话辅助分析

机器问答

语义解析

交互闲聊

多轮对话

阅读理解

AI能力建设

意图分类

情感计算

语义相似度

文本结构化

核心算法技术

深度神经网络

统计机器学习

文法解析

基础机器学习平台

字词句表示

上下文表示

知识图谱

领域语料库

高维特征

低维特征

智能化客服业务场景总览

对外：消费者服务



对内：业务管理

多终端
智能客服

产品营销
推荐

社交媒体
舆情分析

人工客服
风险控制

生活助手

客服质检

智能外呼

坐席辅助

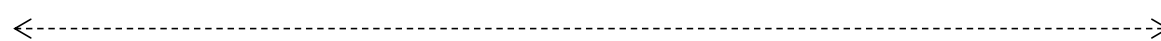
客服业务
标准化

工单自动化

用户画像

安全规程
培训考核

减少人工



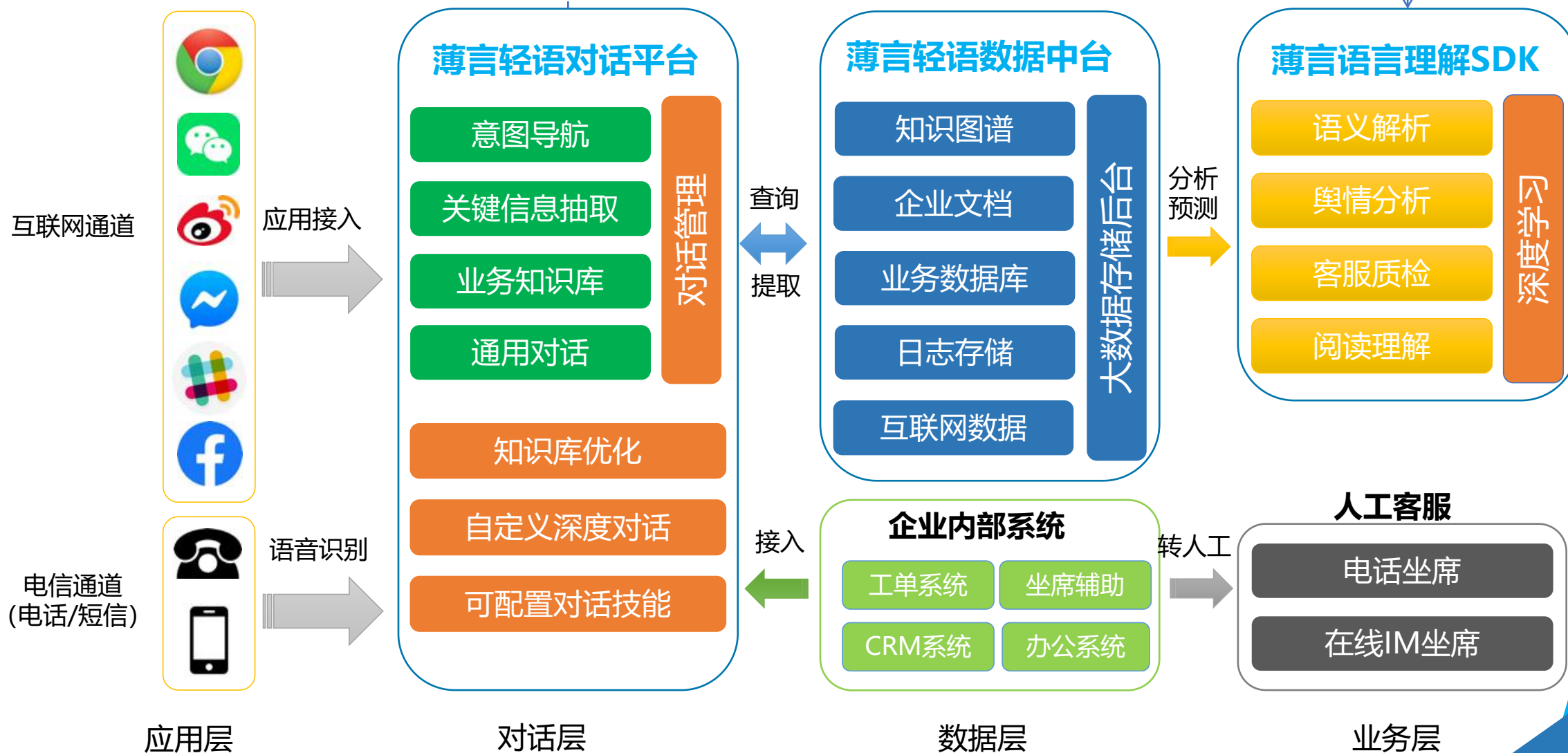
提升技能

产品定位



集成部署方案

云部署方案



目录

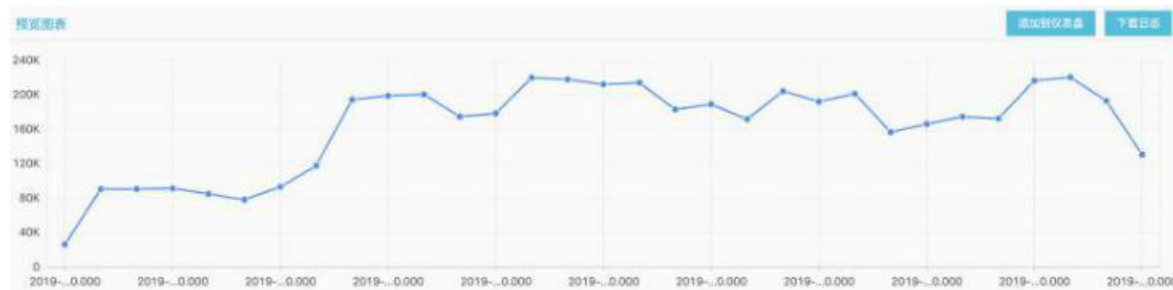
CONTENTS

1. 对话技术的主要挑战
2. 对话应用场景举例
3. 应用案例
4. 公司简介

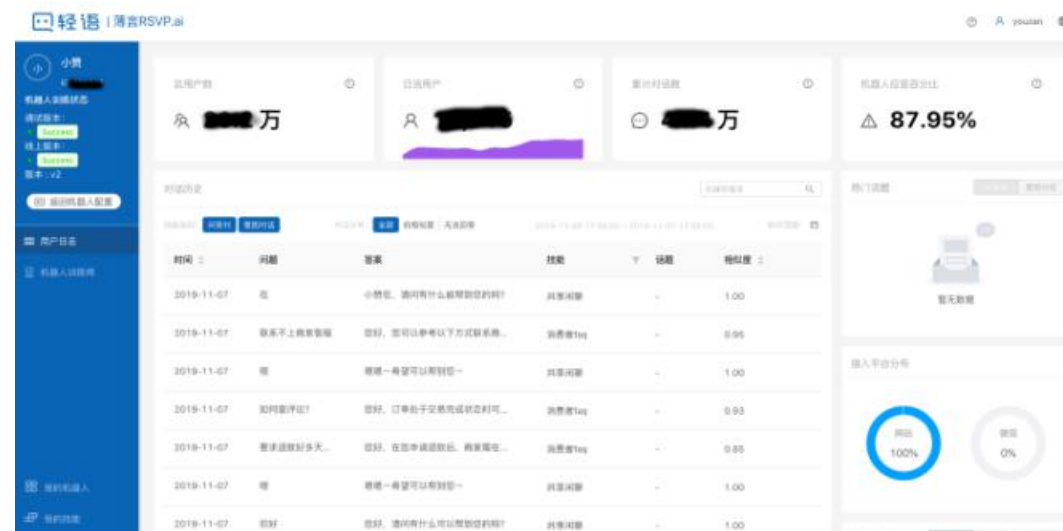
案例：电商智能客服（多终端）

有赞 中文智能客服对话和搜索机器人

- 接入上万个商户，服务于数十万消费者
- 实时处理客服对话



坐席辅助API调用



后台日志分析

案例：餐饮点单机器人（产品营销推荐）

- 餐饮点单
- 复杂条件查询
- 多元化营销方式



用户：江南一品店，一杯橙小西，中杯。

客服：好的，江南一品店取货。1份橙小西，中杯，一共17元



客服：恭喜你，点单成功，获得本期神秘暗号“一骑红尘仙女笑”。

用户：什么？

客服：凭借此暗号可在我们任意门店领取一杯新品“仙女笑”哦！

用户：这么好？

客服：嗯，季节限供新品。小贱老板说，今天就想给你送福利，看好你哦！

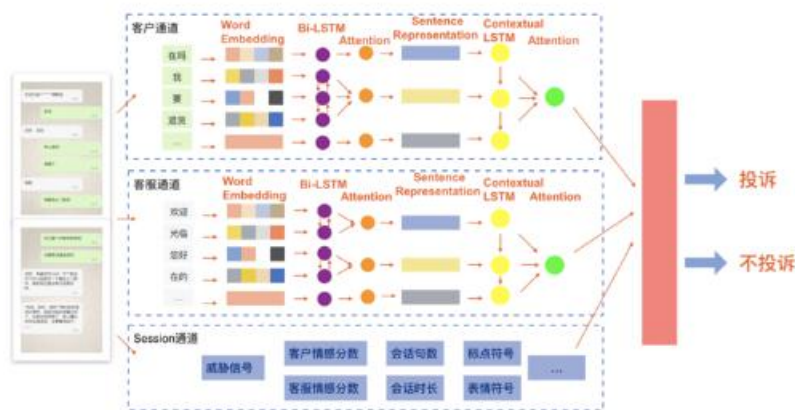
用户：真棒！

案例：客户投诉风险监控系统（客服风控）



自然语言理解服务

- 实时分析客户-客服对话（百万轮/日）
- 识别投诉威胁信号
- 预测产品质量风险、投诉用户



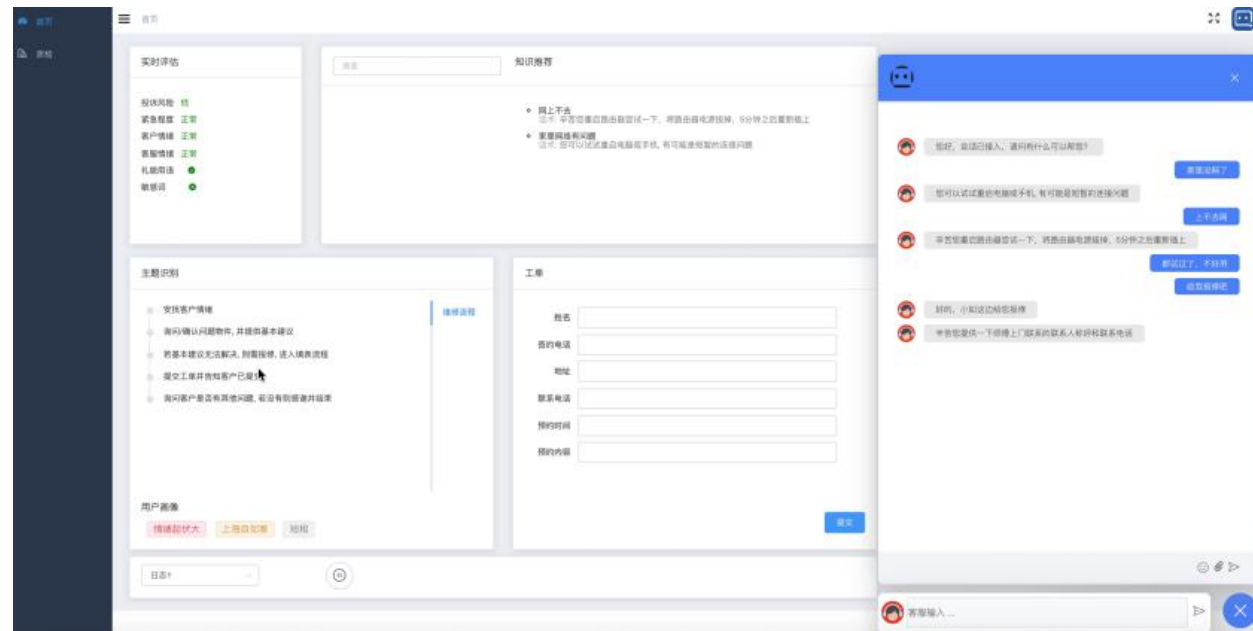
多通道异构数据对话分析



案例：物业管理坐席辅助、工单自动化

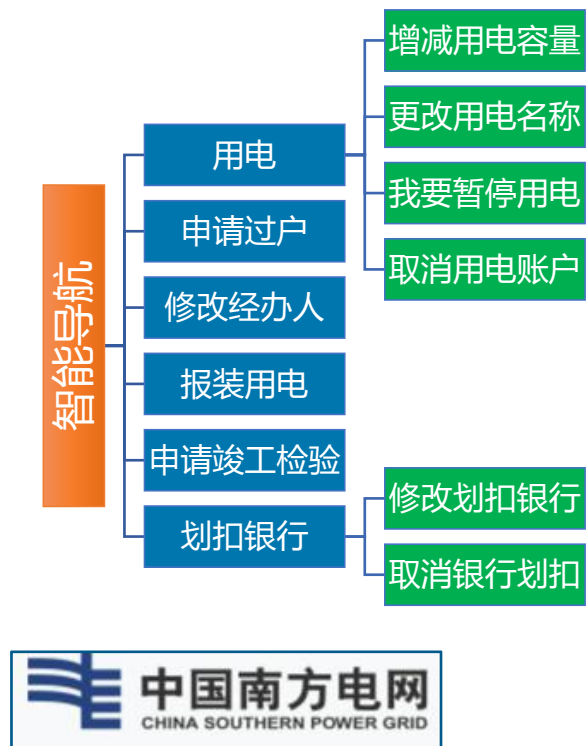


- 处理多类别物业场景（报修、投诉.....）
- 功能点：
 - 流程识别
 - 知识推荐
 - 用户画像
 - 情感评估
 - 实时质检
 - 自动工单（派单）



案例：企业客服（生活号）

• 营业厅、微信客服



我要开电费单

好的，请问是哪个月的电费单

上两个月的

请输入您的账单号或刷卡

2019xxxxxxxxx

2019年4月和5月电费单已打印，请从单据出口取走您的电费单

请问还有什么可以帮您

怎么交电费

您可以在营业厅窗口办理，或者关注微信公众号XXXX进行微信支付

我要销户

已为您取号，个人业务XXX号，前方还有2人排队，请稍候

我们小区怎么停电了

2019年5月6日上午6点线路检修发现故障，预计下午2点恢复用电，非常抱歉为您带来不便

案例：安全规程培训（模拟考核）

- 功能1：机器解答人的问题
- 功能2：机器生成题库



Q: 雨天如何操作室外高压设备

A: 应使用有防雨罩的绝缘棒，并穿绝缘靴、戴绝缘手套

7.3.6 操作的基本要求

- 7.3.6.1 停电操作应按照“断路器—负荷侧隔离开关—电源侧隔离开关”的顺序依次进行，送电合闸操作按相反的顺序进行。不应带负荷拉合隔离开关。
- 7.3.6.2 应按操作任务的顺序逐项操作。
- 7.3.6.3 雷电天气时，不宜进行电气操作，不应就地电气操作。
- 7.3.6.4 操作机械传动的断路器或隔离开关时，应戴绝缘手套。没有机械传动的断路器、隔离开关和跌落式熔断器，应使用绝缘棒进行操作。
- 7.3.6.5 更换配电变压器跌落式熔断器熔丝，应先将低压刀闸或高压隔离开关或跌落式熔断器拉开，验明无电后，方可进行。
- 7.3.6.6 雨天操作室外高压设备时，应使用有防雨罩的绝缘棒，并穿绝缘靴、戴绝缘手套。
- 7.3.6.7 验明线路无电压后，应验明线路无电压，验明时间应在停电操作后，验明时间应在停电操作后。
- 7.3.6.8 高压开关柜手车拉至“检修”位置后，应确认隔离挡板已封闭。
- 7.3.6.9 操作后应检查各相的实际位置，无法观察实际位置时，可通过间接方式确认该设备已操作到位。
- 7.3.6.10 发生人身触电时，应立即断开有关设备的电源。

单选题

问题

低压电压用于配电的交流系统中____及其以下的电压等级。

A. 1000V B. 2000V C. 220V

无接地引下线的杆塔，应在接地线时，可采用临时接地体。临时接地体的截面不应小于____。

A. 190mm² B. 200mm² C. 210mm²

在架空绝缘导线附近作业时，工作人员与绝缘架空导线之间的最小距离不应小于____。

A. 1.2m B. 0.8m C. 0.4m

____及以上电压等级作业还应戴面罩。

A. 750kV B. 1000kV C. 1250kV

线路检修后，应保持导线及人体对下方带电导线的安全距离比规定的安全距离数值增大____。

A. 0.5m B. 1m C. 1.5m

在一经合闸即可送电到工作地点的断路器、隔离开关及跌落式熔断器的操作处，均应悬挂____的

A. 合闸时注意检查 B. 禁止合闸，线路有人工作 C. 止步，高压危险

一条线路。同一电气连接点的几条线路同杆架设且同时停送电的几条线路上工作，可填写同一

工作票。

A. 第一种 B. 第二种 C. 带电作业

填写电力线路____工作票时，不必履行工作许可手续

A. 第一种 B. 第二种 C. 带电作业

一个工作负责人不应同时执行____以上工作票

A. 一张 B. 两张 C. 三张

多选题

问题

作业现场使用的带电作业工具应放置在____上

A. 防潮的帆布 B. 工作设备 C. 绝缘物

恶劣天气条件下巡线和事故巡视时，应根据实际情况配备必

A. 防护用品 B. 自救器具 C. 药品

对同杆塔架设的多层、同一横担多回线路验电时，应____。

A. 先验低压，后验高压 B. 先验上层，后验下层 C. 先

验工作票签发人或工作负责人，应根据现场的____等具体情

A. 安全条件 B. 施工范围 C. 工作需要

判断题

问题

承包工程中，工作票可实行双方签发形式。

事故紧急处理，应做好安全措施，必须填写操作票。

工作负责人、专责监护人应始终在工作现场，对工作班成员进行监护。线路停电工作时，工作负责人、专责监护人应始终在工作现场，对工作班成员进行监护。

在工作过程中遇有恶劣天气或其他威胁到工作人员安全的情况时，工作负责人或专责监护人可下令停止

工作。

工作间断时，工作地点的全部接地线可保留不变。

无接地引下线的杆塔，应在接地线时，可采用临时接地体。

配合停电的线路不能只在工作地点附近装设一处接地线。

测量杆塔、配电变压器和避雷器的接地电阻时，不能在线路和设备带电的情况下进行。

接地线拆除后，应恢复至工作前的状态。

停电设备的各端应有明显的断开点，或应有能反映设备运行状态的电气和机械等指示，可在只经断

源的设备上工作。

可以用缠绕的方法进行接地或短路。

杆塔上有人工作时，可以调整或拆除拉线。

同杆塔架设的多回线路交叉档内，下层线路带电时，上层线路应进行放、撤等（电）线的工作。

架空绝缘导线可以视为绝缘设备。

做好绝缘措施后可以穿越未停电接地的绝缘导线进行工作。

在低电压带电导线采取绝缘措施时，工作人员不能穿鞋。

上杆前，应先分清相线、零线，选好工作位置。人体可以同时接触两根导线。

在带电作业过程中如遇设备突然停电，应视设备的带电情况，工作负责人应及时与线路运行单位或调度

运行维护单位或值班调度员联系，不得擅自离开工作地点。

等电位工作人员应穿戴全套屏蔽服，各部分连接良好，应通过屏蔽服、接空

合电容器的电流接地点。

在瓷横担上可以挂作业，在转动横担的瓷横担上挂作业时，应先将瓷横担固定。

带电、接空载线路，工作人员应戴护目镜，并采取消弧措施，应带绝缘手套、接引线。

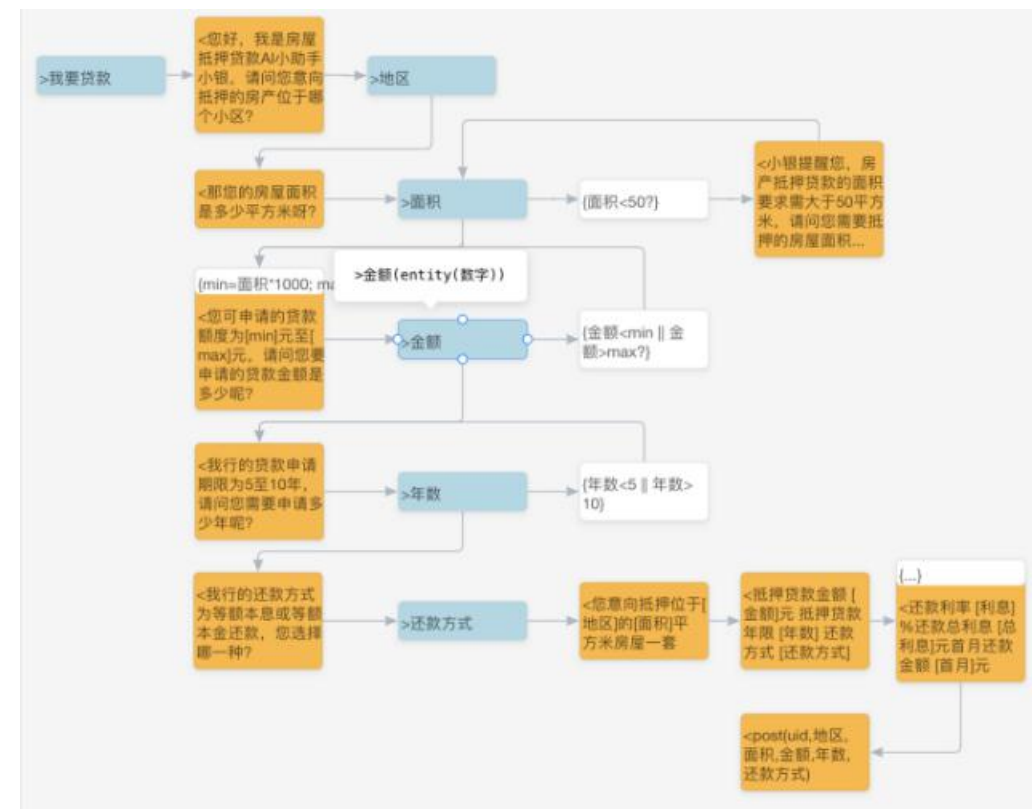
未采取绝缘措施前，应在跌落式熔断器下头新装、调换熔断器或更换、接接电熔断器。

问答题

变电站（电力系统）电力系统的一部分，它集中在一个指定的地方，主要包括哪些设备。

案例：多属性产品咨询

- 银行、保险咨询
- 多条件、多轮次
- 推荐对应产品



贷款业务筛选流程（示例）

现状：缺乏上下文

案例：企业助手

- 社保助手
- 用户之声



宁波市人力资源和社会保障局
WELCOME TO RSJ.NINGBO.GOV.CN



欢迎来到社保中心，我是你的社保助手小鱼，请问有什么需要咨询？

异地就医怎么报销

好的，请问您是否办理过异地就医？

没有，以前没办过

好的，下面为您介绍未办理过异地就医应该如何报销的具体流程。

若您是急诊药费，则可以报销，所需材料有...

若非急诊，则社保认为就医具备提前筹划的条件，应先办理异地就医才能报销。

回复“1” 查看办理异地就医后如何报销；

回复“2” 查看如何办理异地就医。

1

为您介绍如何办理异地就医。

第一步，请先确定您想就诊的异地医院；
第二步，...

请问还有什么可以帮您？

没有了谢谢

案例：生活场景多轮对话

- 航空订票
- 多条件、多轮次
- 根据后台数据动态调整



订一张去杭州的飞机票

好的，请问您打算哪天出行？

后天星期天上午的

是从您的所在地北京(Beijing)出发吗？

是的

好的，正在为您查询：6月9日星期天，北京飞往杭州的航班，上午12:00前起飞：

10:35 13:00 ¥780
首都T2 萧山T3 经济舱3.3折
东航 MU5132 | 空客320(中)

10:50 13:15 ¥2060
首都T3 萧山T1 两舱特惠
国航 CA1702 | 空客321(中) 公务舱2.9折

11:45 14:00 ¥1350
首都T3 萧山T1 经济舱5.6折
国航 CA1704 | 空客321(中)

可以通过航空公司和起飞时间选择

国航10点50这一班吧

为您预定：6月9日星期天10:50北京首都国际机场飞往杭州萧山国际机场的中国国际航空CA1702班机，请确认。

没问题

好的，正在出票中

目录

CONTENTS

1. 对话技术的主要挑战
2. 对话应用场景举例
3. 应用案例
4. 公司简介

公司介绍

薄言：全球领先的中文自然语言理解和对话技术提供商

- 成员来自于康奈尔、哈佛、MIT、清华、北大、中科大、滑铁卢等大学
- 拥有谷歌、百度、Twitter等一线互联网公司研发经验
- 产品服务于智能硬件、电商、银行、保险公司和大型国企



核心团队 (1)



李明
创始人、总裁

(接受加拿大总督颁奖)

- 滑铁卢大学终身教授
- 加拿大皇家科学院院士
- ACM和IEEE Fellow
- 教育部长江讲座教授
- 国家千人计划专家
- 清华大学客座教授
- 2010年获加拿大最高科学奖 (Killam Prize)
(历来获此奖的唯一中国大陆学者)

核心团队 (2)



Jimmy Lin 博士
首席科学家

滑铁卢大学终身教授，麻省理工学院博士，曾任职马里兰大学、Twitter数据科学家，是机器学习和自然语言处理领域的世界知名学者。



熊琨
联合创始人 | CEO

滑铁卢大学硕士、清华大学学士（计算机系），曾任职Telenav（硅谷公司，NASDAQ: TNAV）移动广告平台研发部，是应用机器学习和自然语言处理的专家。



王靖波
联合创始人 | COO

北京大学硕士、学士（数学系），曾任职创维（数字电视）、联想、摩托罗拉等公司，拥有丰富的技术团队管理和智能产品运营经验。



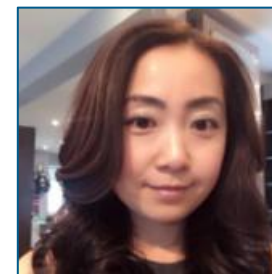
张泽峰 博士
联合创始人 | CTO

中科院计算所博士，中国科技大学少年班“算法天才少年”，计算机理论和算法的专家，曾代表中国国家队参加机器人世界杯（RoboCup）。



崔安颀 博士
联合创始人 | AI负责人

清华大学博士、学士，滑铁卢大学博士后，曾在卡耐基梅隆大学和新加坡国立大学访问，研究方向为情感计算、统计数据挖掘、自然语言处理等。发表专著、论文数十篇。



汤静
联合创始人 | CFO

滑铁卢大学学士（金融专业），曾任职于加拿大汇丰银行，创立科技和投资公司，通晓北美金融、法律环境。

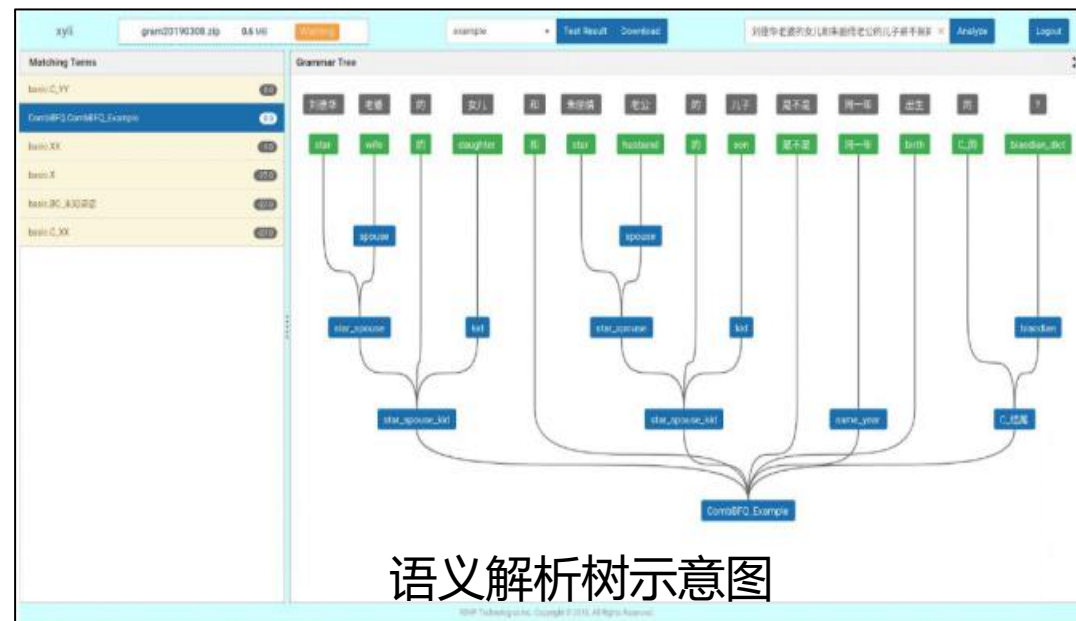


瞿宜情
合伙人 | 商务总监

清华大学学士（计算机系），曾任空中网、Roxbeam研发工程师。拥有超过5年的公司管理和大客户销售经验。

技术优势：最快的语义解析引擎

- 基于上下文无关文法，**表达能力强**
- **标注快、精度高、耗时少**，适用于实用系统
 - 业界常见方法：深度或统计模型、正则文法、词网格等
- 已覆盖一百多个中英文场景，为大企业内部系统做支撑



技术优势：海量的知识图谱构建

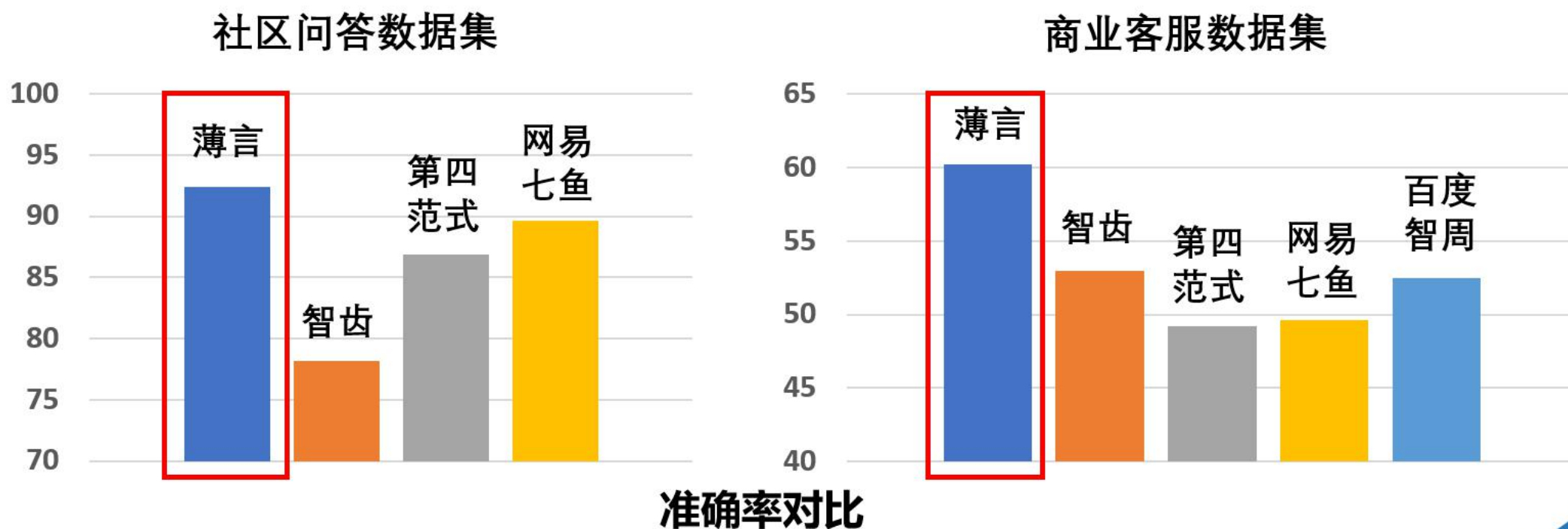
- 结合语义解析引擎构建知识图谱
- 数量：中英文千万级别
- 覆盖垂直领域：影视、文学、生活、文学、儿童教育、社保、能源、保险等数十个，**应用于相关行业客户**



领域知识库（部分）

技术优势：最准确的文本语义相似度

- 通用模型：适用于**全行业**平台解决方案
 - 即使不做领域优化，在专有数据集上也优于其他业内系统
- 已应用于华为、有赞等企业，处理数据量来自**百万人次**的顾客

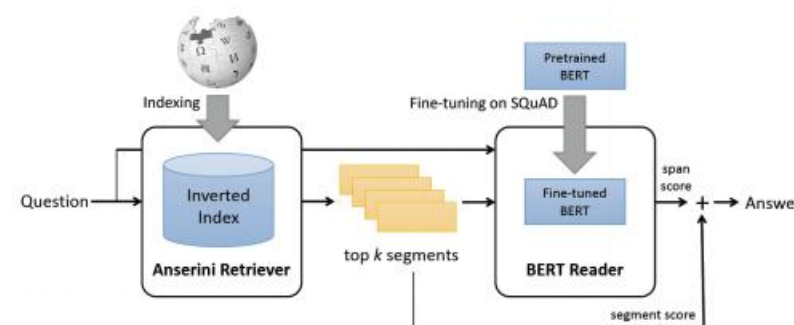


技术优势：通用的开放域机器问答

- 让机器学会读文章（做阅读理解）



人工整理
自动构建



- 2019年初曾获微软MS MARCO比赛第一名



论文颁奖

行业评价

- 行业内NLP研究实验室在各个方向有自己的技术壁垒，因为语义和场景有关，需要长时间的积累和打磨，不会马上出现一家独大。李明教授深耕算法领域多年，尤其是复杂度和信息距离理论处在世界第一梯队。

——朱小燕（清华大学计算机系教授，博士生导师，智能技术与系统国家重点实验室主任）

- NLP行业内竞品较多，华为看重薄言的独特优势——将CFG技术路线作为探索方向之一，具备领先行业的上下文文法编辑器；李明教授作为行业资深科学家，所带领的团队也十分优秀；华为有意愿在未来，与薄言共同落地一些成熟产品。

——李岩（华为企业发展部高级投资经理）

- 薄言在较短时间给出优异的解决方案，技术实力和行业积累远超京东自有100多人的NLP团队，在京东智慧客服课题中，薄言所提供的解决方案比京东自身方案的效果高出40%。京东已把薄言列为长期NLP供应商，对薄言的科研底蕴和职业素养十分青睐。

——杜永青（京东大数据与智能供应链事业部二级部门负责人）

行业评价

- 李明教授是非常知名的国际上最早一批从事AI工作的学者，行业现在语义通用标注方法为人工数据标注，效率比较低，成本高，同样的一句话，变化语序就需要对应不同个范式。薄言拥有语料生成器，标注效率高，成本低。百度，科大讯飞，思必驰，云之声，捷通华声，三角兽等，在进行语素测试时，使用语言模板的情况下，各个企业差距不大。但是不使用模板的情况下，薄言的表现最为突出。

——杨永成（小米音频产品线负责人，小米电视副总裁）

- 李明教授在学术领域有较高声誉，360主动联系到李明教授后，与薄言进行了合作。360比较了薄言、三角兽、微软的产品，在问答领域，薄言在数据量较少的情况下和其他企业的测试结果不相上下，可以看出技术实力不俗。NLP不存在通用技术，在细分垂直领域，小公司与大公司之间出现更多是合作关系。

——李志勇（前360智能音箱负责人）

期待合作

www.rsvp.ai
qingyu.rsvp.ai



RSVP.ai