

智能问答算法原理及实践之路

主讲人：

腾讯数据平台部高级算法研究员 陈松坚



CONTENTS

01

智能问答算法原理

02

实践：小知客服机器人/电话机器人

03

总结和挑战



腾讯小知

总览

query

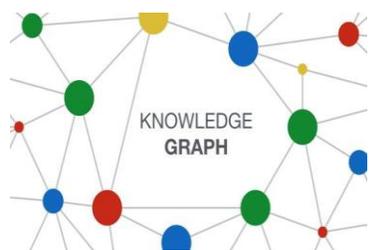
中控逻辑

response

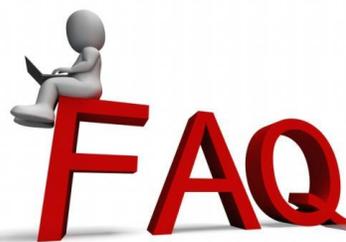
任务机器人



知识图谱机器人



FAQ机器人



闲聊机器人



阅读理解机器人



腾讯小知

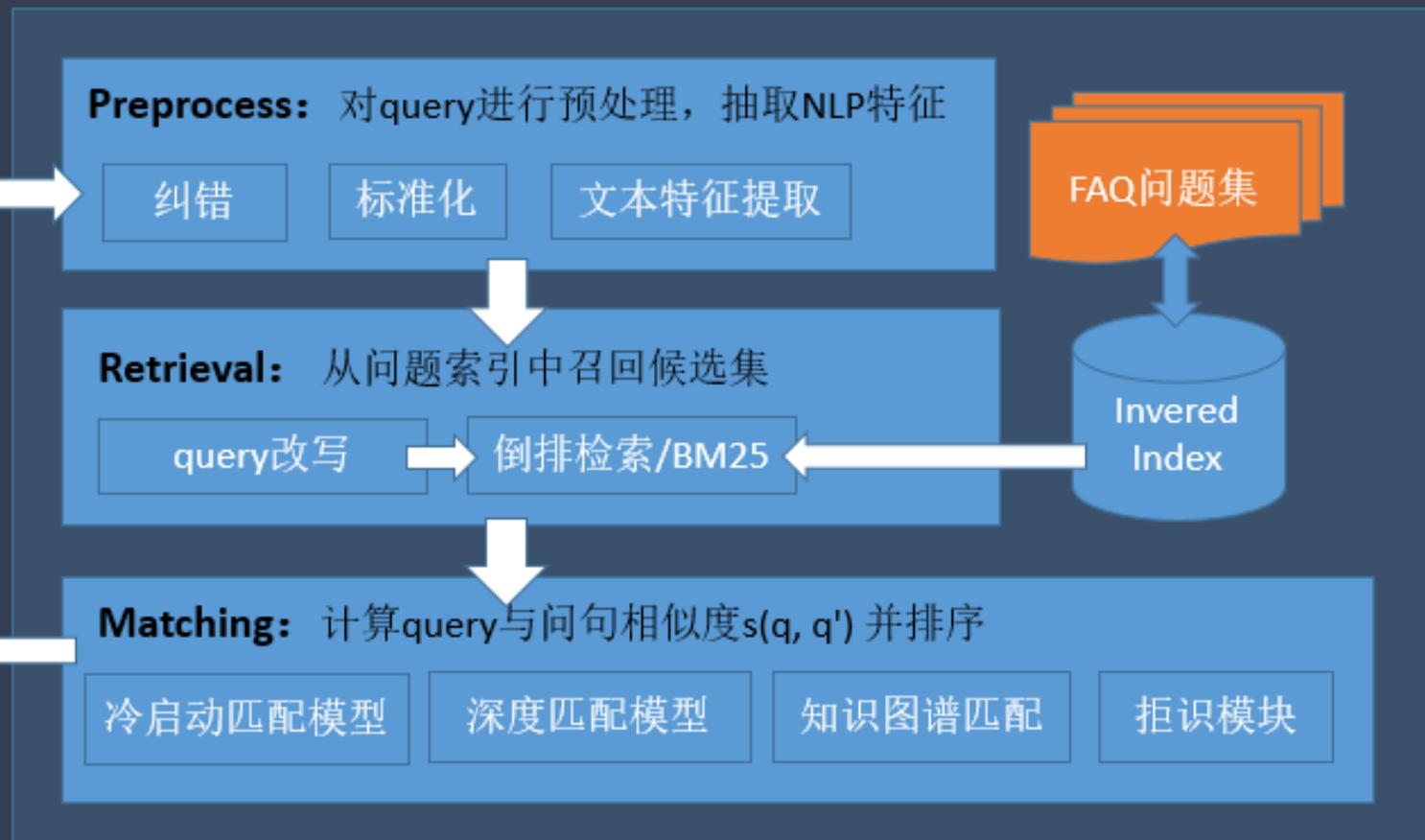
FAQ机器人



社保余额
怎么查询？



可以在微信上关
注社保公众号，
在底部菜单...

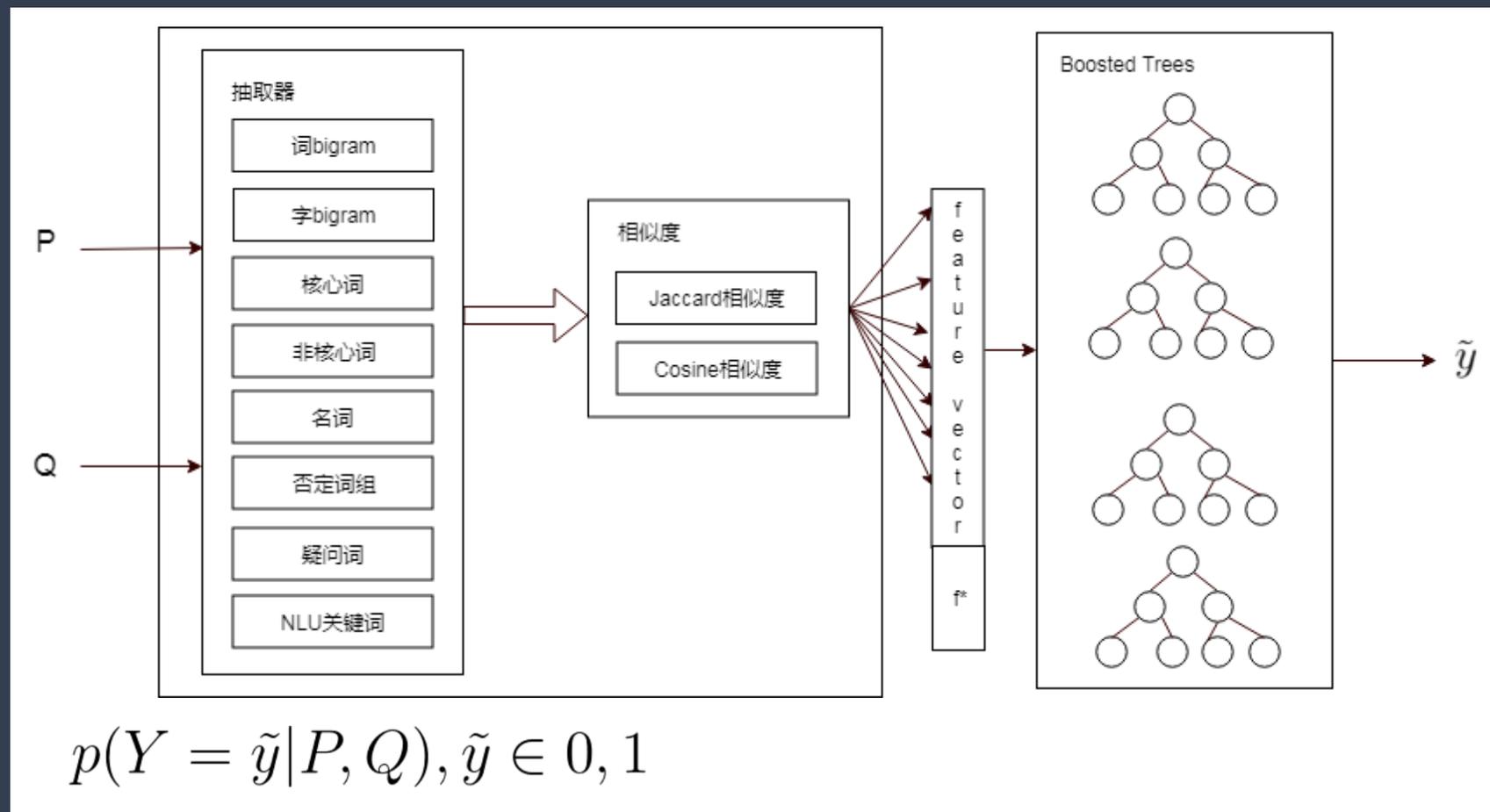


字面匹配

- 动机：解决冷启动问题

方案：

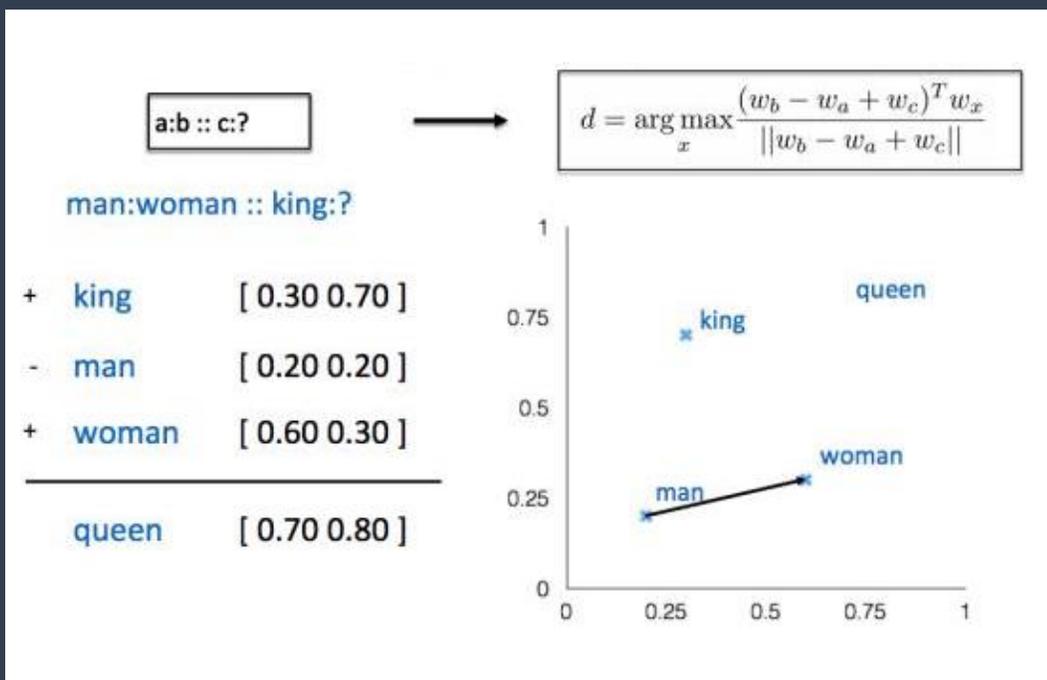
- 文本相似度特征
- xgboost非线性建模



字/句向量匹配

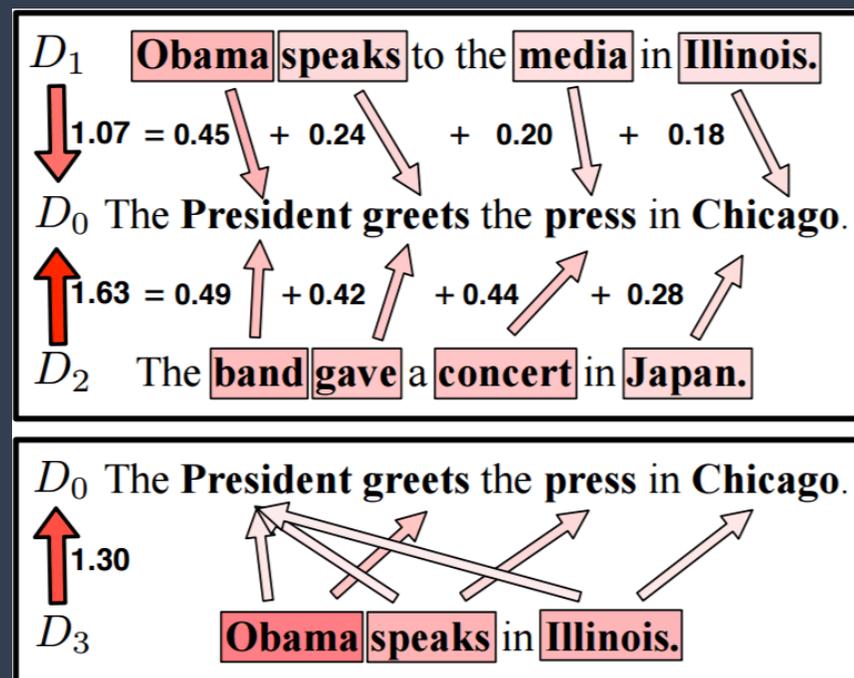
词向量:

- word2vec
- glove



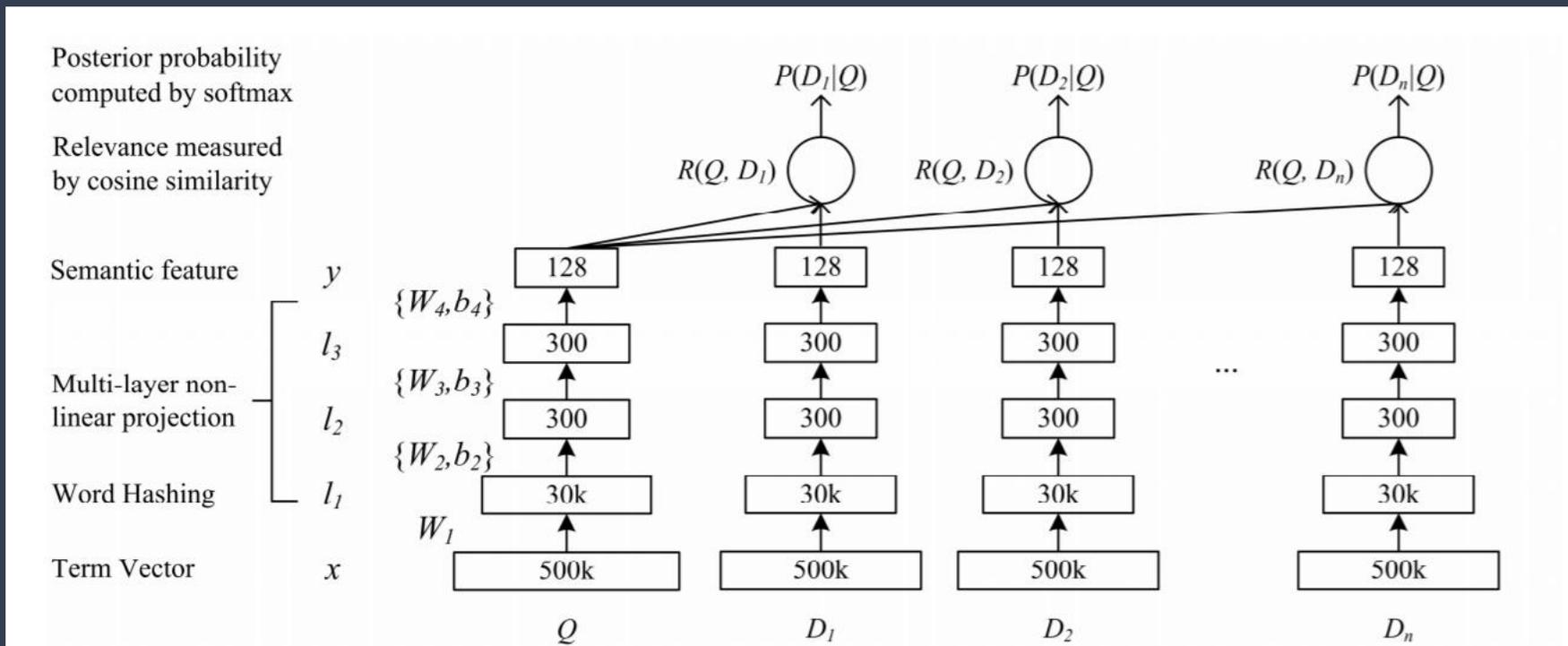
句向量:

- WMD(Word Mover's Distance)
- SIF (Smooth inverse Frequency)



深度匹配模型

- 高维稀疏term特征映射到低维语义特征
- 深层网络结构

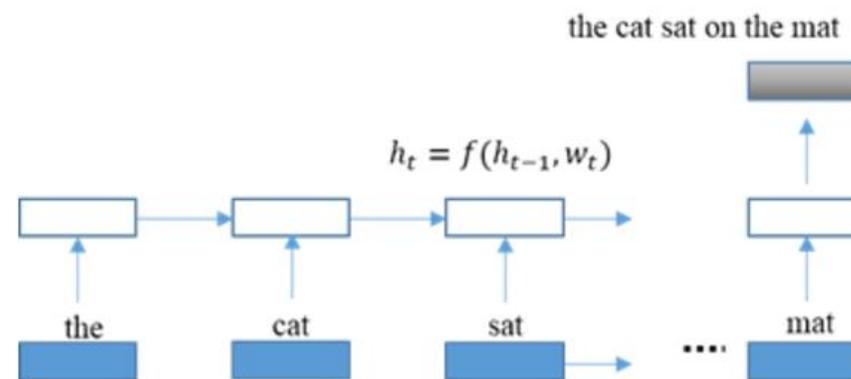
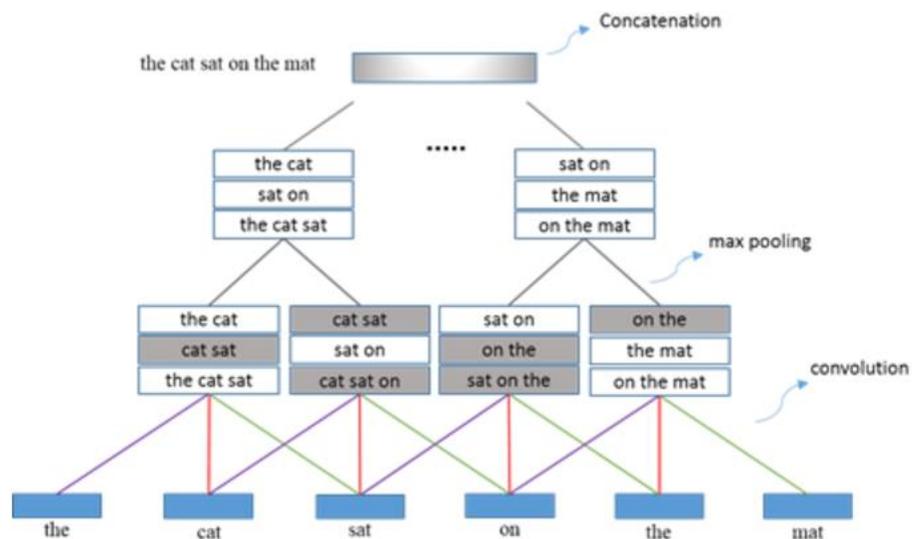


(Reference: P.S Huang, Learning deep structured semantic models for web search using clickthrough data, 2013)

深度匹配模型

- CNN，卷积神经网络建模

- RNN，循环神经网络建模



深度匹配模型

- Representation-based, 表示式建模
- Interaction-based, 交互式建模

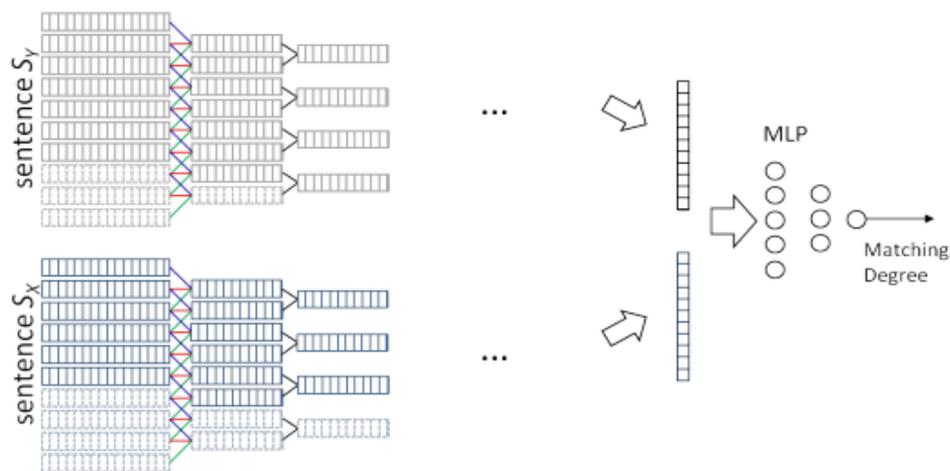


Figure 3: Architecture-I for matching two sentences.

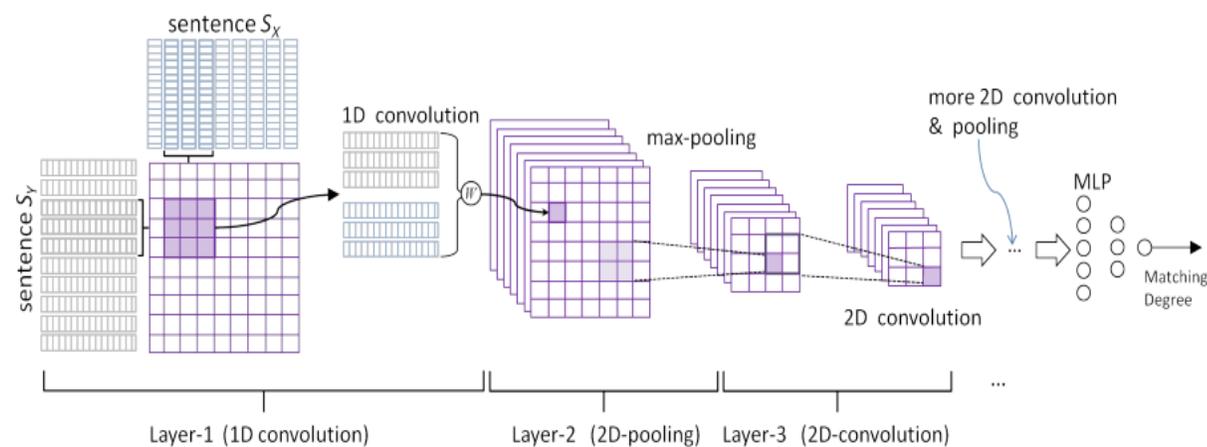
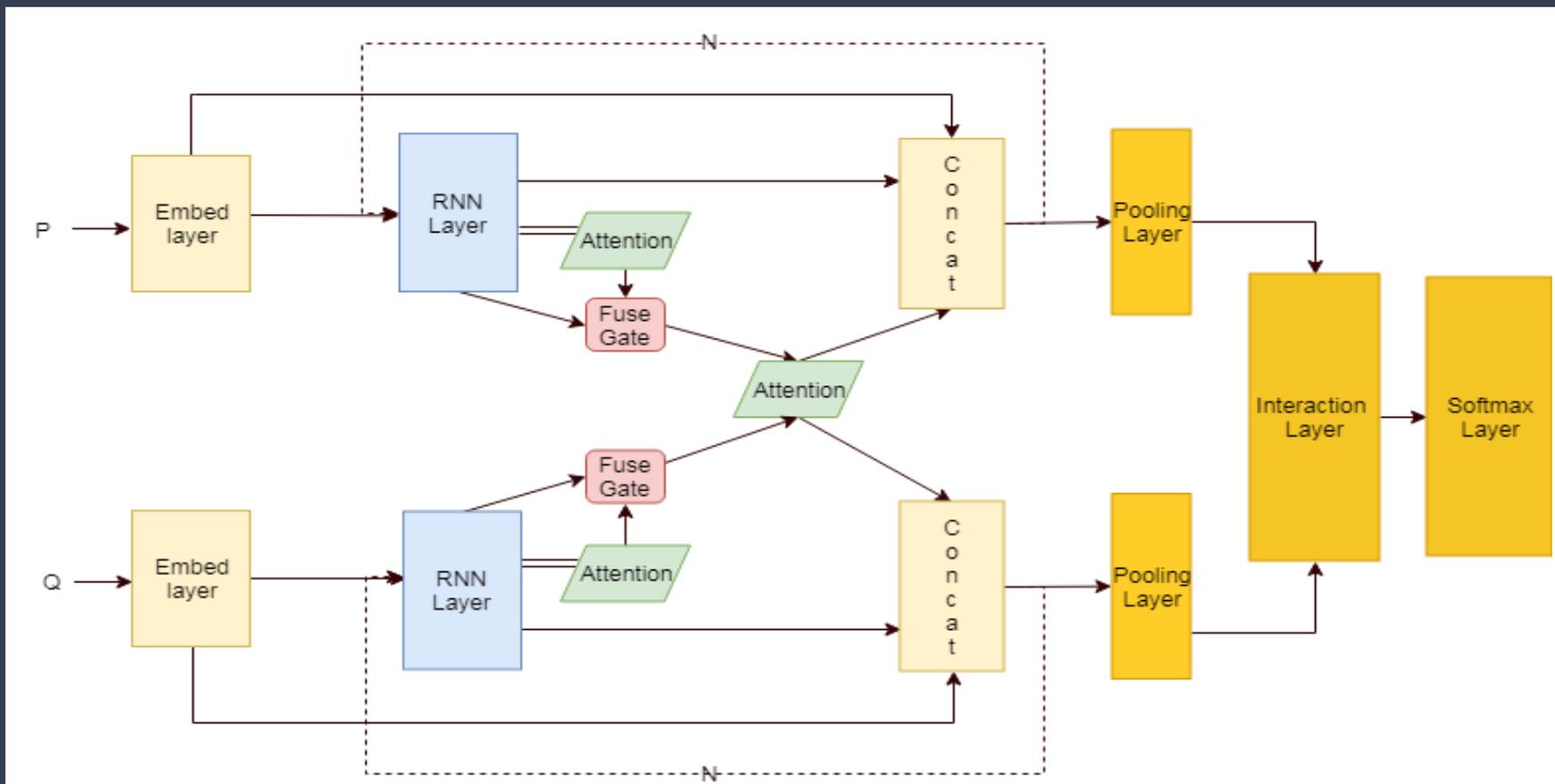


Figure 4: Architecture-II (ARC-II) of convolutional matching model

(Reference: Baotian Hu,, Convolutional Neural Network Architectures for Matching Natural Language Sentences, 2015)

深度匹配模型

- 稠密连接网络 (Densely-connected)
- 混合注意力机制 (Attention + Self Attention)



迁移学习

- 多任务联合学习：

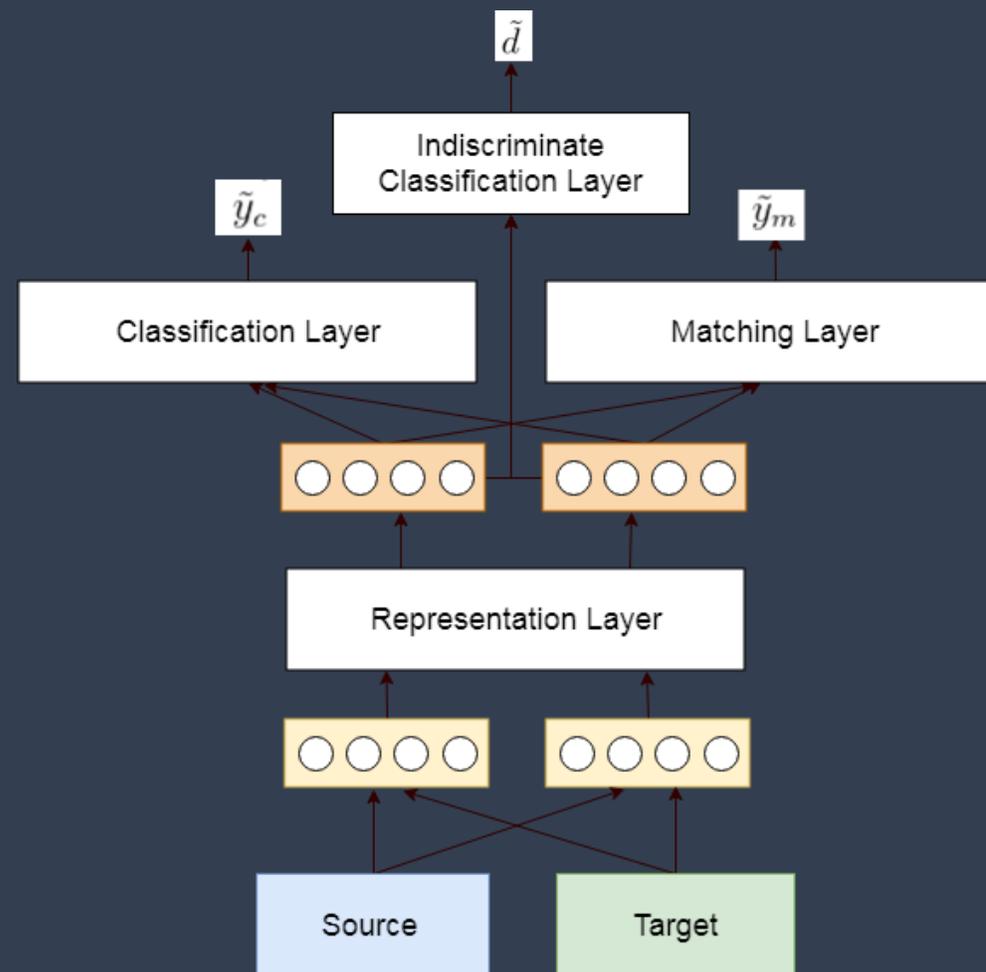
- 问句分类: 同时进行问句匹配和分类
- 问句生成: 同时进行匹配和seq2seq训练

- 多语料迁移学习

- fine-tune：上层参数微调
- adversarial loss：样本来源的混淆分类

- 预训练模型

- ELMO, BERT



多轮对话架构

- 对话管理：query -> NLU -> DST -> DPL -> NLG -> response



我想查社保余额,
社保号1234

NLU

负责意图识别和槽位实体抽取
intent={查询##社保##余额}
entity={社保号:1234}

DST (Dialogue State Tracking)

负责维护会话状态, $DST(q, s) \rightarrow s'$, 输出新状态 s'

DPL (Dialogue Policy Learning)

负责选择下一步行动, $DPL(s', KB) \rightarrow a$, 输出新行动 a 。

NLG

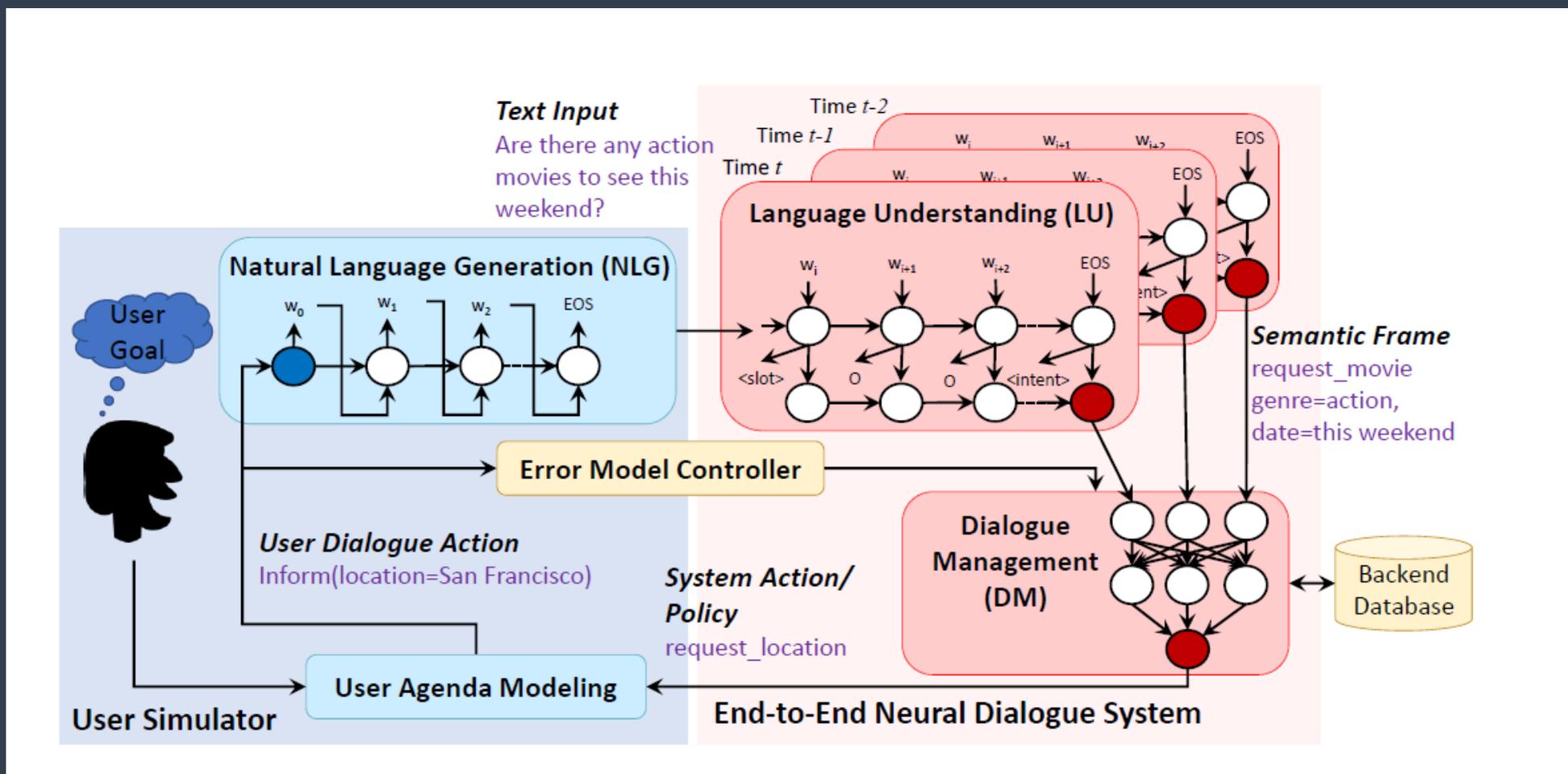
负责根据action输出回复文本

好的, 请您提供
姓名和预留手机
号后4位。

Dialogue Management

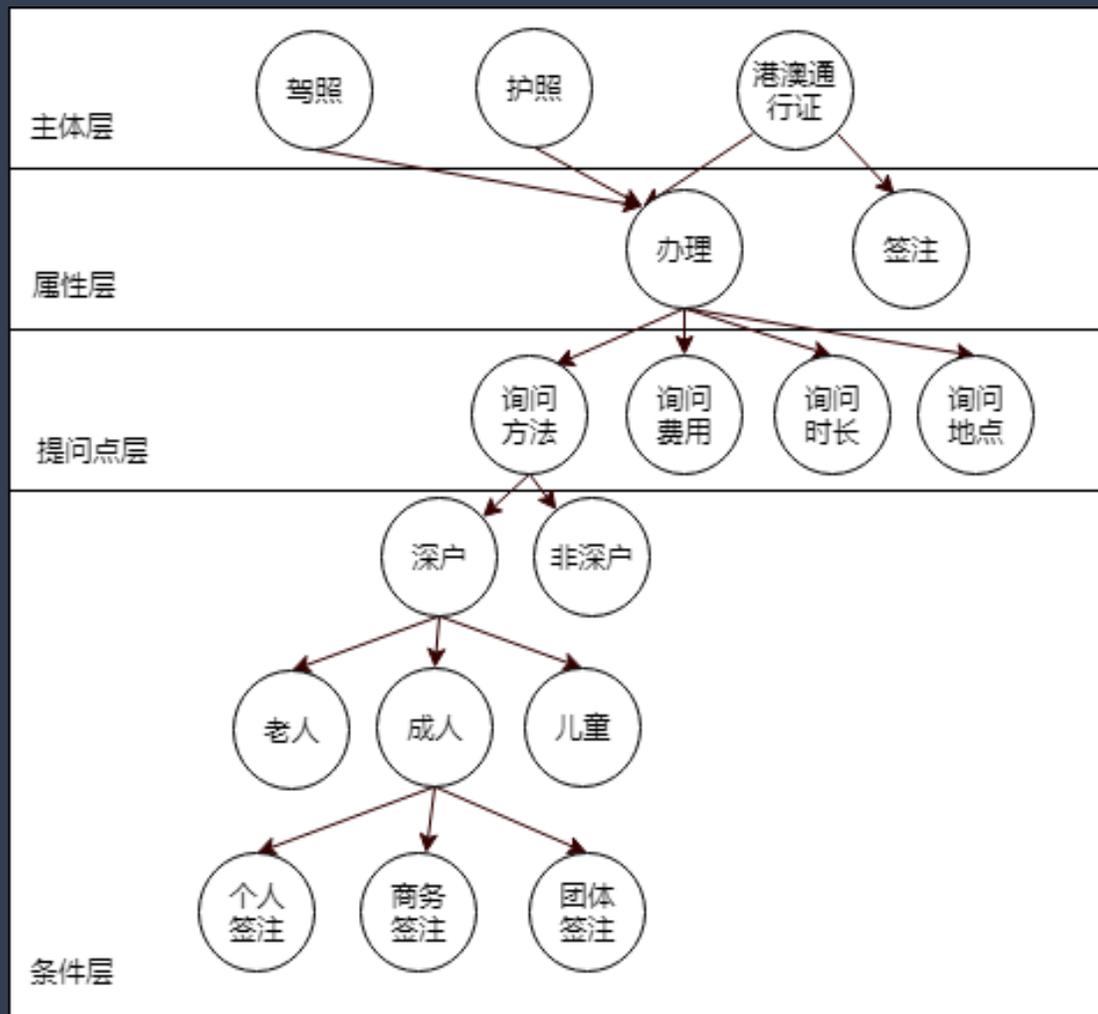
任务机器人

- End-to-end神经网络模型



- (Reference: Xuijun Li, End-to-end task-completion neural dialogue system, 2017)

知识图谱机器人




智能客服

请问您办理的是：
[港澳通行证](#)
[护照](#)
[驾照](#)

请问您是：[深户](#) [非深户](#)

请问您是：[个人签注](#) [商务签注](#)
[团体签注](#)

深户办理港澳通行证及个人签注：申请人请携带《.....申请表》。

港澳签注收费标准：往来港澳通行证每本80元.....每项次80元

我想办理

常见问题

如何取消出入境业务的预约号...

个人港澳签注如何再次办理

办理护照、通行证及签注应该...

非深户个人港澳通行证及签注...

深户港澳通行证及签注的首次...

深圳居住证签注如何办理

非本市户籍居民身份证办理如...

本市户籍居民身份证的办理补...

如何在深圳办理居住证

如何预约办理身份证

港澳通行证

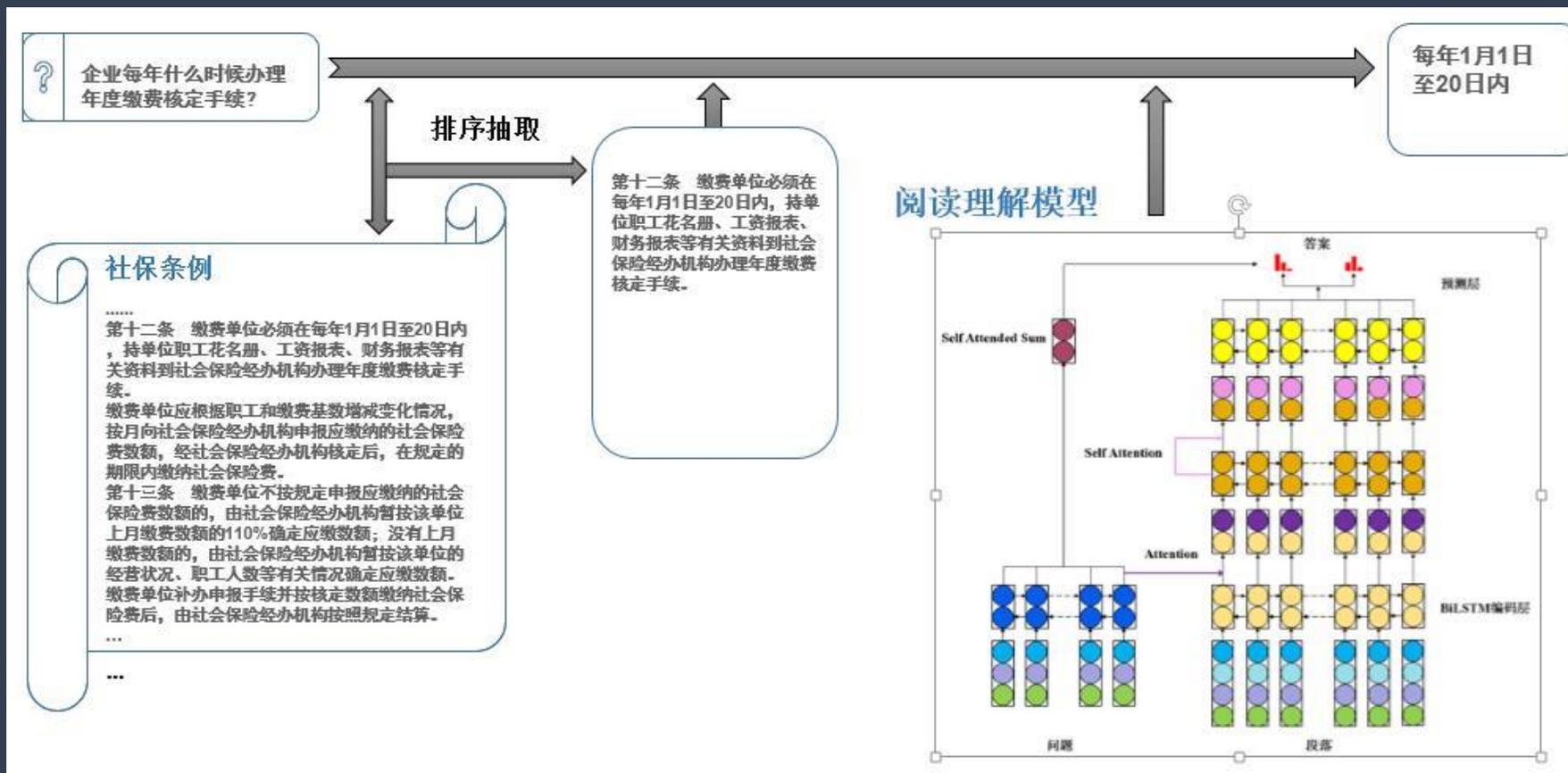
深户

个人签注

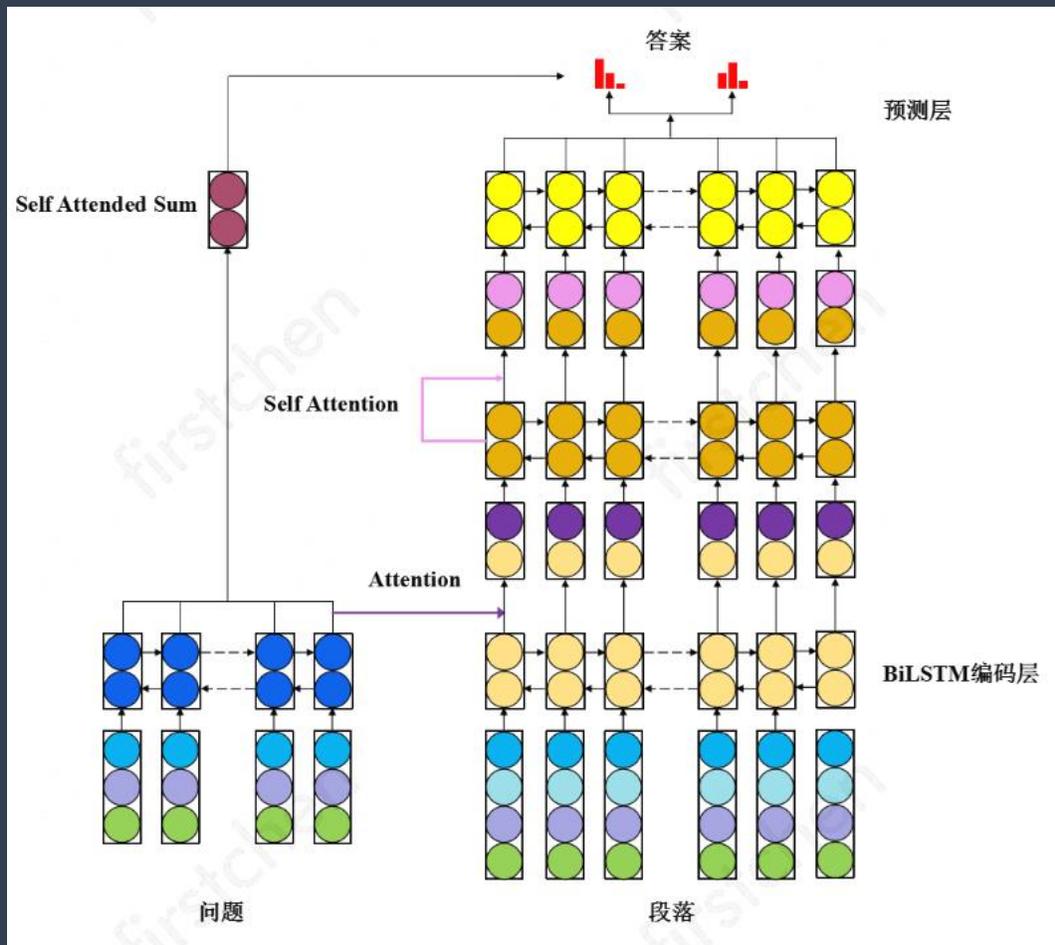
多少钱

阅读理解机器人

- 应用示意



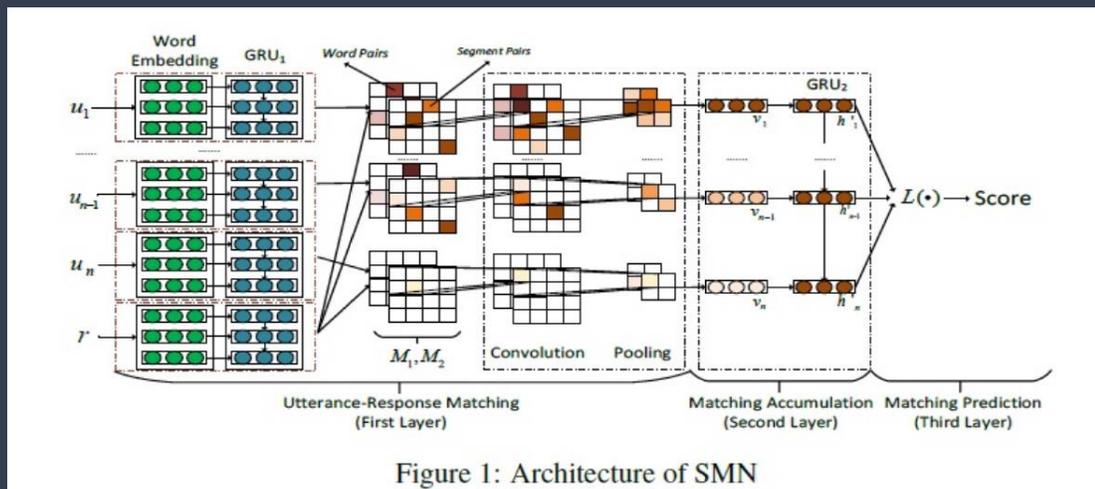
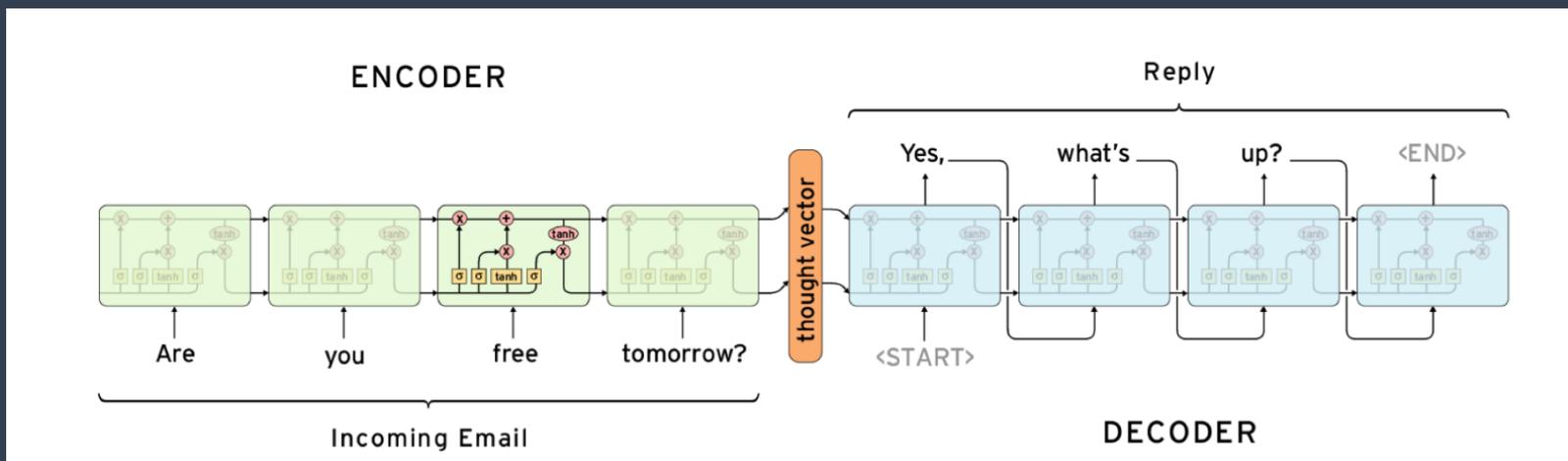
阅读理解机器人



Rank	Model	EM	F1
1 Dec 22, 2017	AttentionReader+ (ensemble) Tencent DPDAC NLP	81.790	88.163
1 Dec 18, 2017	r-net (ensemble) Microsoft Research Asia http://aka.ms/rnet	82.136	88.126
2 Nov 18, 2017	BiDAF + Self Attention + ELMo (ensemble) Allen Institute for Artificial Intelligence	81.003	87.432
3 Dec 06, 2017	SAN (ensemble model) Microsoft Business AI Solutions Team https://arxiv.org/pdf/1712.03556.pdf	79.608	86.496
4 Oct 18, 2017	Interactive AoA Reader+ (ensemble) Joint Laboratory of HIT and iFLYTEK	79.083	86.450
5 Oct 25, 2017	FusionNet (ensemble) Microsoft Business AI Solutions Team https://arxiv.org/abs/1711.07341	78.978	86.016
6 Oct 22, 2017	DCN+ (ensemble) Salesforce Research	78.852	85.996

闲聊机器人

- 检索重排
- Seq2Seq
- 上下文建模



(Reference: Yu Wu, Sequential Matching Network: A New Architecture for Multi-turn Response Selection in Retrieval-based Chatbots, 2016)

CONTENTS

01

智能问答算法原理

02

实践：小知客服机器人/ 电话机器人

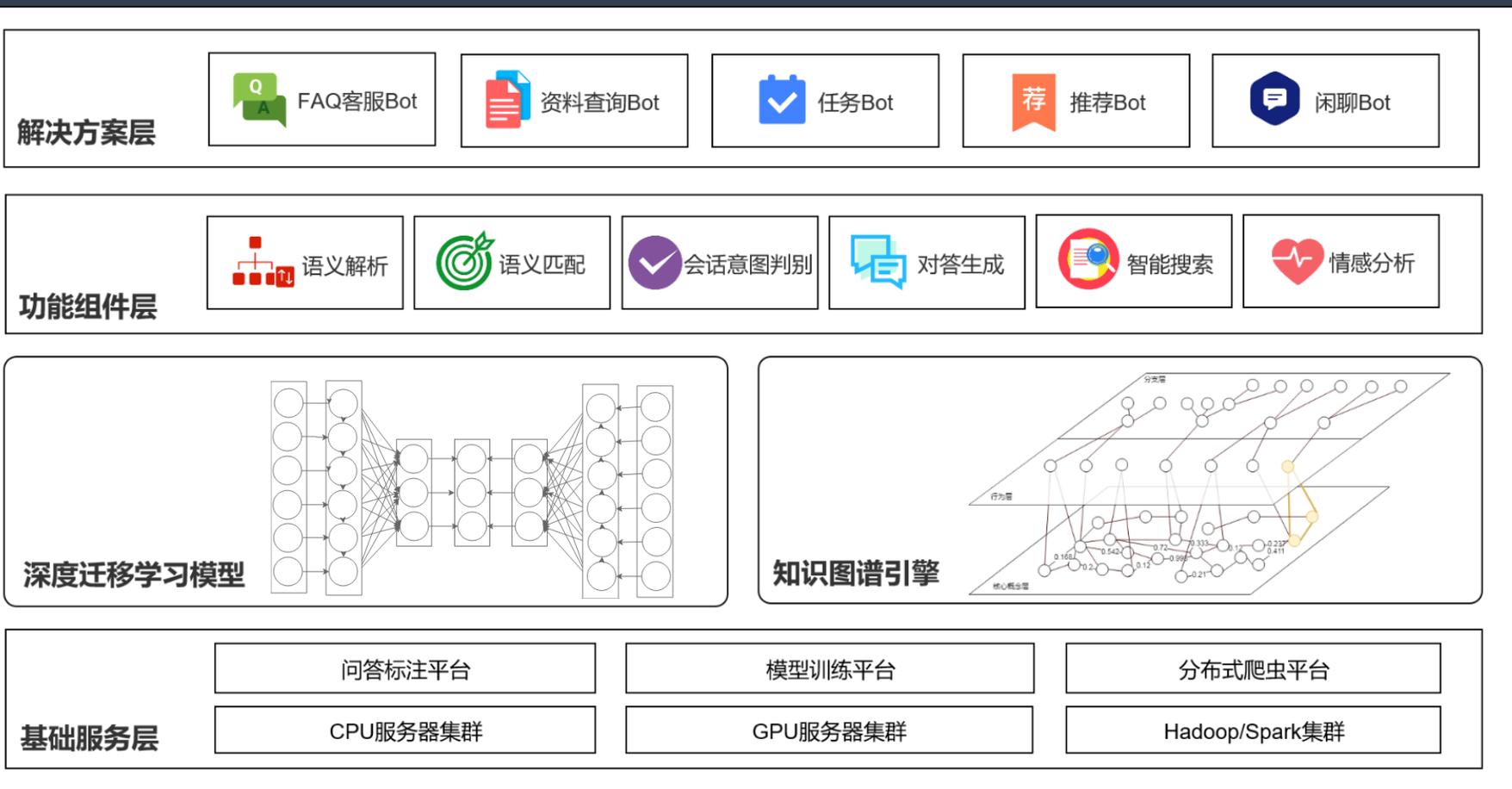
03

总结和挑战



腾讯小知

小知智能客服架构



小知客服机器人

- 知识库管理


腾讯小知


graceliu@tencent.com

知识库管理

- 知识库
- 待处理任务
- 智能学习
- 问答测试
- 客服工作台
- 留言回复
- 快捷回复语料
- 会话记录
- 系统管理
- 机器人管理
- 机器人设置
- 人工客服设置

知识库

专业问答
闲聊对话
同义词

问答分类

添加问题
批量导入
导出到Excel

输入关键字，回车检索

编辑人

知识来源

编辑时间

开始日期 ~ 结束日期

状态

	标准问题	答案	编辑人	编辑时间	状态	有效期	操作
<input type="checkbox"/>	相似问法: 4		360913530@qq.com	2018-11-19 17:02:32	启用 (永久有效)	永久有效	编辑 删除 停用
<input type="checkbox"/>	相似问法: 8		467137609@qq.com	2018-11-18 13:32:00	启用 (永久有效)	永久有效	编辑 删除 停用
<input type="checkbox"/>	相似问法: 4		陈丽敏	2018-11-16 15:52:53	启用 (永久有效)	永久有效	编辑 删除 停用
<input type="checkbox"/>			467137609@qq.com	2018-11-16 15:51:28	启用 (永久有效)	永久有效	编辑 删除 停用


在线客服

FAQ机器人

智能客服机器人

基于知识库自动准确解答访客提问



人工客服

在机器人无法回复或客户主动寻求人工协助时，客服系统将接入人工服务



智能客服

你好

您好，请问您需要咨询什么业务？

有奖举报专线电话

有奖举报电话：22222110

谢谢啦

不客气，很高兴能解决您的问题。
如果您有其他问题欢迎随时咨询人工客服

常见问题

- 如何取消出入境业务的预约号...
- 个人港澳签注如何再次办理
- 办理护照、通行证及签注应该...
- 非深户个人港澳通行证及签注...
- 深户港澳通行证及签注的首次...
- 深圳居住证签注如何办理
- 非本市户籍居民身份证办理如...
- 本市户籍居民身份证的办理补...
- 如何在深圳办理居住证
- 如何预约办理身份证

请在此输入...

按 Enter 快捷发送消息

发送



支持多轮对话

多轮追问

补充问题信息，保证回答的准确性

上下文记忆

联系上下文，准确识别用户意图

智能关联问题

推荐相关问题，提供完备信息

智能客服

办理港澳通行证

请问下，您要咨询的是：
[非深户](#)
[深户](#)
[以上都不是](#)

深户

深户办理港澳通行证及个人签注：
申请人请携带《.....申请表》。

多少钱

港澳签注收费标准：往来港澳通行证每本80元.....每项次80元

您是否还想咨询一下问题：
[各出入境大厅办公时间是什么时候](#)
[如何办理换发、补发](#)

常见问题
如何取消出入境业务的预约号...
个人港澳签注如何再次办理

请在此输入...

按 Enter 快捷发送消息

发送

人工客服工作台



会话列表

会话消息区

人工客服后台

会话功能区

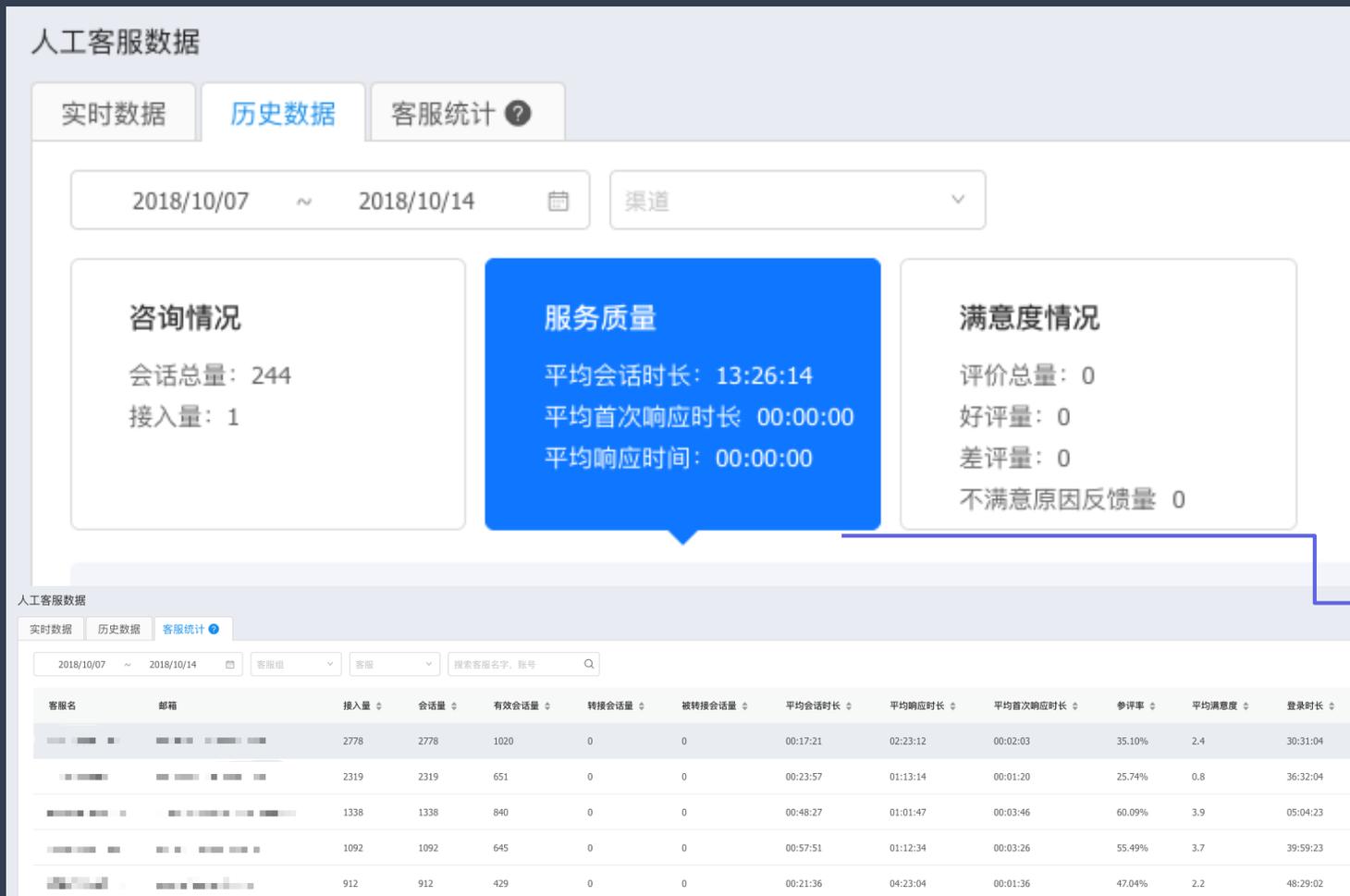
人工客服入口



App, 微信, Web三大渠道的
即时信息统一整合

运营数据统计

· 客服工作数据面板



分析问题，辅助决策

记录客服工作情况
问题统计和数据分析



准确应答

92%

一次性准确率

95%

综合准确率

95%

节省服务人力

智能AI引擎

深度+迁移学习模型

NLP算法引擎

知识图谱引擎

海量数据支持

丰富的语料积累

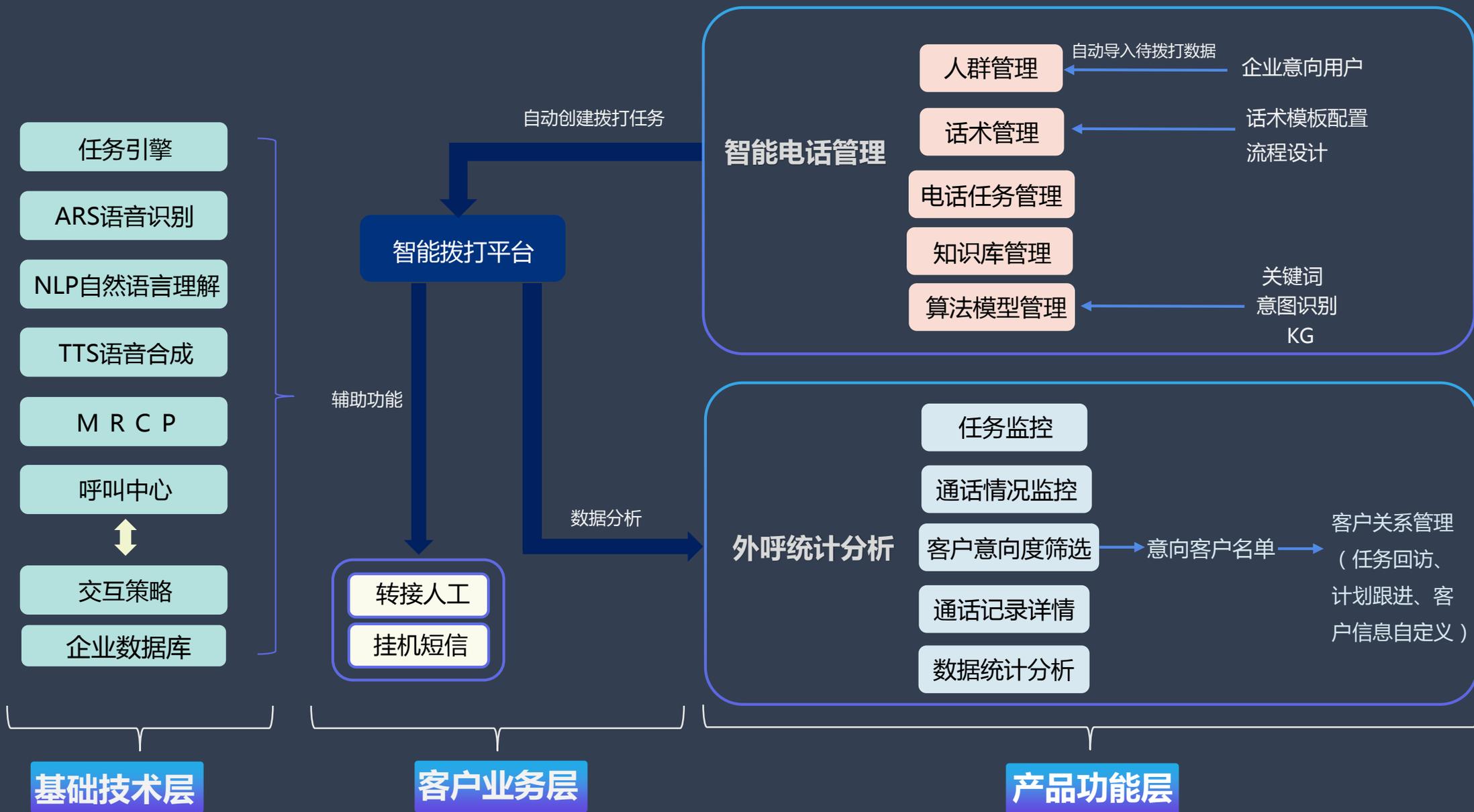
专业的知识图谱

周全的相似问题



智能电话机器人与智能客服、人工在线客服、工单系统完美融合

电话机器人系统架构

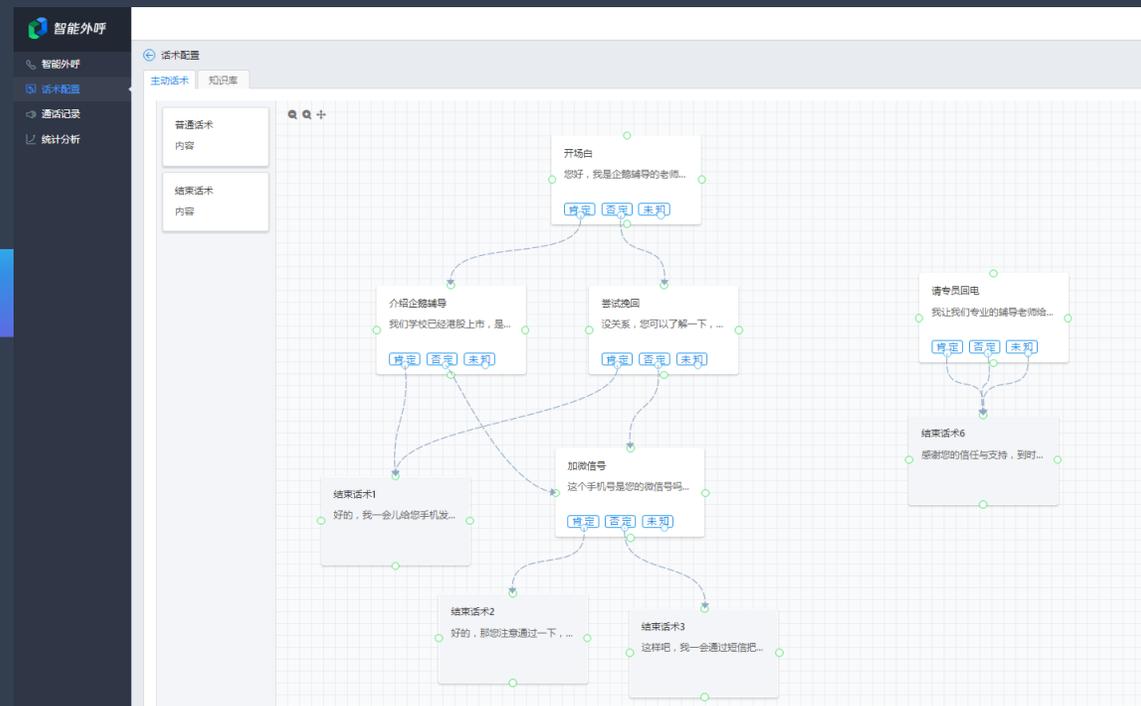


电话机器人功能亮点

- 客户业务层——智能应答

话术模板

拖拽式流程设计更方便



补充个性化问题信息，保证回答的准确性，丰富的话术模板可应对众多电话场景。

标题	问题	更新时间	操作
无法应答处理	特殊配置		默认
连续3次无法回答的后续处理	特殊配置		默认
用户不听话	特殊配置		默认
学校信息	你们是什么	2018-09-12 20:12:01	查看 编辑 删除
课程收费	上课是怎么收费的	2018-09-12 20:13:01	查看 编辑 删除
师资	老师是谁的	2018-09-12 20:14:25	查看 编辑 删除
教材	你们用的是什么教材啊?	2018-09-12 20:14:58	查看 编辑 删除
上课形式	什么类型	2018-09-12 20:17:10	查看 编辑 删除
上课时间和回放	什么时候上课	2018-09-12 20:18:03	查看 编辑 删除
上课效果	你怎么知道肯定就能补上去呢	2018-09-12 20:19:12	查看 编辑 删除
自行选择	你们课程都有哪些啊	2018-09-12 20:20:40	查看 编辑 删除
拒绝沟通	我是老人	2018-09-12 20:23:03	查看 编辑 删除
闲聊	你是机器人吗	2018-09-12 20:24:27	查看 编辑 删除
催促人工跟进	你怎么知道是我的电话的	2018-09-12 20:25:49	查看 编辑 删除
孩子问题	孩子自制力比较差	2018-09-12 20:26:58	查看 编辑 删除
压力太大	孩子没有时间以后再说吧	2018-09-12 20:27:47	查看 编辑 删除
孩子成绩好	孩子成绩很好	2018-09-12 20:28:20	查看 编辑 删除
评价不好	以前报过线上辅导平台效果不好	2018-09-12 20:30:22	查看 编辑 删除
价格太高	太贵了	2018-09-12 20:31:13	查看 编辑 删除
在忙	在忙	2018-09-12 20:34:43	查看 编辑 删除

电话机器人功能亮点

客户业务层——智能应答

意向设置

精准判断用户意向等级，实现优质线索筛选，谨防主观臆断。

客户意向评级 (意向答复复制)

- A类(强烈意向): 走到流程结束且最后一次答复为肯定。并且整体肯定占比在90%以上。
- B类(有意向):
 - 走到流程结束: 肯定占比在70%以上或否定占比在10%以下。
 - 未到流程结束: 肯定占比80%以上, 答复次数大于3次或否定占比20%以下, 答复次数大于3次。
- C类(需筛选):
 - 走到流程结束: 肯定占比在30%以上, 答复次数大于3次或者否定比100%以下, 答复次数大于3次。
 - 未到流程结束: 肯定占比在50%以上, 答复次数大于2次或者否定比100%以下, 答复次数大于3次。
- D类(需要重新发起): 1、忙状态; 2、占线; 3、无人接听、4、呼叫失败
- E类(无效号码): 1、停机; 2、空号

客户意向评级 (话题分制)

在话术制作中对每项话题设置意向分值，交互对话后根据每项话题客户的意向进行分值加成评级。

	A	B	C	D	E	F	G
	话题分类	话题	分值	机器人语句1	机器人语句2	机器人语句3	Action
2	主流程	开场白		您好,我是企鵝辅导的老师,现活动期间我们有初高中辅导班免费试听,老师都是清北一线名师,给您简单介绍一下。	您好,我是企鵝辅导的老师,耽误您一分钟,我想和您聊聊孩子学习方面的事情,您看您是否方便?		肯定->[产品介绍] 否定->[挽回] 在忙->[在忙拒绝] 拒绝沟通->[拒绝沟通拒绝]
3	主流程	产品介绍		我们学校已经港股上市,是非常可靠的线上辅导平台。我们四大高中录取率35%,一本上线率61.5%,中考状元也在我们班上。您看您是否愿意带孩子来试试?	我们学校已经在港股上市,是非常可靠的线上辅导平台。我们有初高中课程,一本上线率61.5%,本届数学、物理状元都在我们班上。您可以带孩子过来试试。		肯定->[成功结束] 否定->[微信邀约] 在忙->[在忙拒绝] 拒绝沟通->[拒绝沟通拒绝]
4	主流程	挽回		没关系,您可以了解一下,我们学校已经港股上市,是非常可靠的线上辅导平台。我们四大高中录取率35%,一本上线率61.5%,中考状元也在我们班上。您可以带孩子过来试听一下我们的课。			肯定->[成功结束] 否定->[微信邀约] 在忙->[在忙拒绝] 拒绝沟通->[拒绝沟通拒绝] 疑问->[推动知识库]
5	主流程	微信邀约		这个手机号是您的微信号吗,我们加一个微信,以后关于孩子学习方面的问题,我们在微信上随时给您解答。			肯定->[微信邀约成功结束] 否定->[微信邀约失败结束] 在忙->[在忙拒绝] 拒绝沟通->[拒绝沟通拒绝]
6	主流程	电话邀约		我让我们的专业辅导老师给您回电话,让她给您详细说说,您看可以吗?			任何回答->[电话邀约结束]
7	结束语	拒绝沟通拒绝		那真是非常抱歉打扰您了,感谢您接听我们的电话,祝您生活愉快,再见!			
8	结束语	在忙拒绝		没关系,那您先忙,我会给您发封关于我们企鵝辅导介绍的短信,您可以了解一下,那就不打扰您了,祝您生活愉快,再见!			
9	结束语	成功结束		好的,我一会给您手机发封短信,您在链接中领取一下试听课程就可以了,感谢您的信任与支持,祝您生活愉快,再见!			
10	结束语	微信邀约成功结束		好的,那您注意通过一下,感谢您的信任与支持,这边就不打扰您了,祝您生活愉快,再见!			
11	结束语	微信邀约失败结束		这样吧,我会通过短信把我们课程具体相关资料发给您,您可以了解一下,就不打扰您了,祝您生活愉快,再见!			
12	结束语	电话邀约结束		感谢您的信任与支持,到时候再联系,祝您生活愉快,再见!			
13	态度	肯定		【没问题、有、对、OK、不错、yes、轻me、好、行、要、可以、你说、你讲、你介绍、你继续、请讲、怎么去、怎么买、了解下、嗯、哦、喂、喂、喂、明白、知道、想了解、有兴趣、不需要、没需求、不用、不要、没有、不、没兴趣、没必要、不想弄、不想用、不想聊、不想了解、不想知道、不感兴趣、不用介绍、好麻烦、算了、没意思、没什么用、不情愿、不考、在忙、在开车、在开会、不方便说话、不方便聊天、没时间、赶飞机、没空、没时间、在上课、不方便、晚点给我电话、晚点联系我、我现在不舒服、有空再说、以后再说、等我忙完再【已经报了、我的孩子还小、没有孩子、孩子没上学、作了、孩子很大了、别烦我、我要挂了、骗子、不想聊、很烦、烦不烦、打过了、打过多少次、打了好几次、骚扰电话、骚扰【等等、等一下、稍后说			
14	态度	否定					
15	态度	在忙					
16	态度	拒绝沟通					
17	态度	打断					

在主流程中,客户态度的判断基于客户应答话语里的关键词。多维度多方面设置意向判断词,精准识别用户态度。

CONTENTS

01

智能问答算法原理

02

实践：小知客服机器人/电话机器人

03

总结和挑战



腾讯小知

• 总结：

- 单轮FAQ机器人：深度语义匹配/迁移学习/知识图谱
- 多轮任务机器人：对话管理/槽位提取/端到端学习/强化学习
- 阅读理解机器人/开放域闲聊机器人

• 挑战：

- 如何结合规则和模型？
- 如何进一步利用外部世界知识（如知识图谱）？
- 知识图谱如何做推理？
- 如何结合前三者？



网易云课堂

MEET THE BEST
OF YOURSELF

遇见最好的自己



断水流大师
向您推荐

LeetCode 百题斩(1~3)

解惑者学院

¥5.28 ~~¥9.90~~

BAT面试必问，大佬带你刷题！每天刷3题，有效提升编程能力，秒变offer收割机！算法是计算...



长按扫码 进入课程

机器学习 解惑者

全网最新最全机器学习、深度学习、NLP、CV资料免费下载

唐宇迪、吴恩达、优达、七月、小象视频课程应有尽有

更有Google、Amazon、BAT
大佬答疑解惑，带你轻松学习AI

赶快扫码加入吧

