OTA场景中的两类对话机器人及其实现方案

2020-05-22

分享人:王创峰

大纲

- OTA场景的两类机器人需求:售前与售后
- 两类场景的性质分析
- 信息引导类建模
- 解决问题类建模
- 总结

两类场景的区别和性质

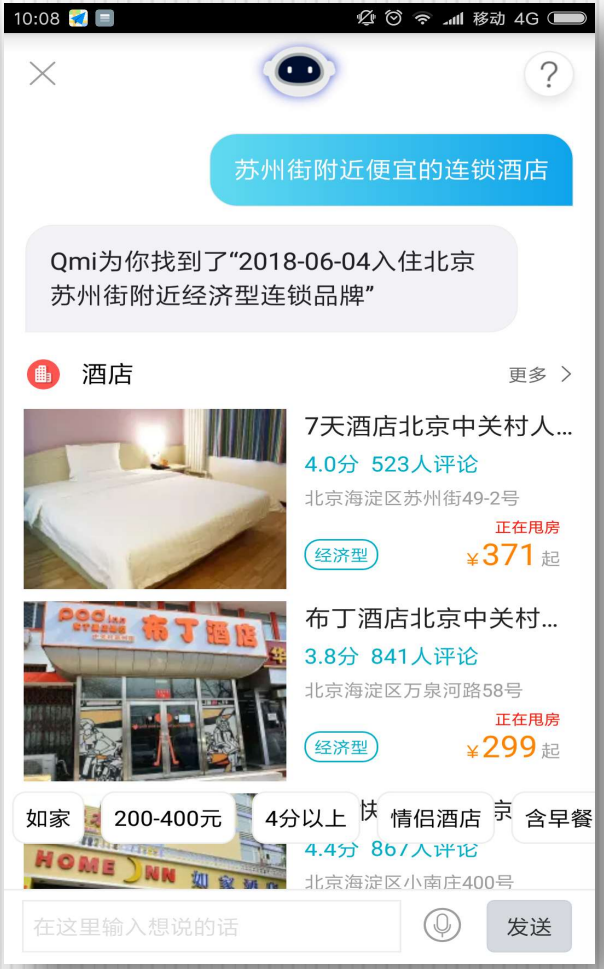
	售前	售后
类别	信息引导类	解决问题类
诉求	替代翻页,搜索	替代客服
case	我说话,你给我我要的信息	我碰到问题了,你帮我解决,或者告诉我怎么解决.
性质	新的用户界面	解放人力
思考	2C: 界面是命脉,交出去要小心 2B: 参考2C的担忧,是问题也是机会	2C: 人机可替代,协同工作 2B: 壁垒不足,甲方分分钟自己搞

售前,引导类

- 输入: 我要啥啥啥
- 输出: 查询API和参数
- 解决方案: 填槽-状态机
- trick:
  - 纠错
  - 词典
  - 场景间槽位继承
  - 根据业务侧重某些意图
  - 闲聊后置
  - 限制能力
  - …

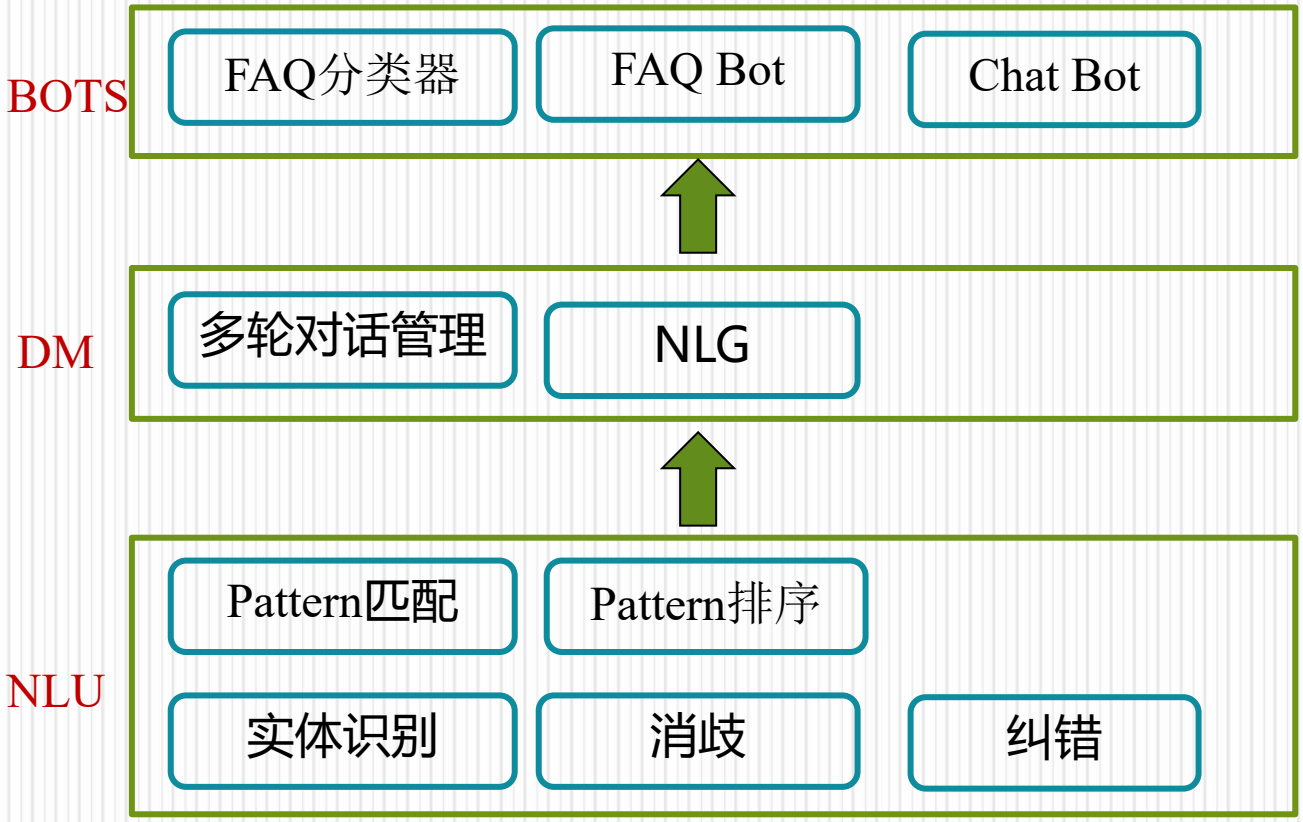
QMI遗容


中国联通 4G 16:55 30%
搜索旅行地/酒店/景点
520 学生专
拆红包 领现金
酒店
特价酒店
海外酒店
客栈·民宿
高端酒店
您有300元红包
机票
低价机票
汽车票·船票
火车票
专车·租车
学生机票3折起
旅游
景点·门票
旅游团购
攻略·美食林
周边游
上海200元券
借钱·信用卡
保险·索赔
购物·外汇
签证·WiFi
出境游
接送机
一日游
智慧旅游
定制游
家庭游
自由行
特价游
更多
首页
订单
发现
服务
我的


10:08 移动 4G
苏州街附近便宜的连锁酒店
Qmi为你找到了“2018-06-04入住北京苏州街附近经济型连锁品牌”
酒店
更多
7天酒店北京中关村人...
4.0分 523人评论
北京海淀区苏州街49-2号
正在甩房
经济型
¥371 起
布丁酒店北京中关村...
3.8分 841人评论
北京海淀区万泉河路58号
正在甩房
经济型
¥299 起
如家
200-400元
4分以上
情侣酒店
含早餐
在这里输入想说的话
发送


中国联通 4G 10:17 79%
下周一到上海机票
请问您从哪个城市出发?
北京
Qmi为你找到了“2018-06-11从北京出发到上海机票”
2018-06-11从北京出发到上海机票
圆明园
外滩附近的酒店
北京周边游
请输入你想说的话

QMI架构



FAQ、闲聊处理

怎么选座？
你是谁？

上下文管理及槽位填充

机票: {出发地：北京
,目的地：上海}

实体及意图识别

意图：机票、酒店、火车
要体：城市、地标、酒店名

难题

- 危险的闲聊
  - 无法区分被判定为闲聊的问题是否真的涉及业务
  - 先过业务逻辑
  - 闲聊越简单越好
- 长句子
  - 包含多个成分
- 语音杂音录入

售后,解决问题类

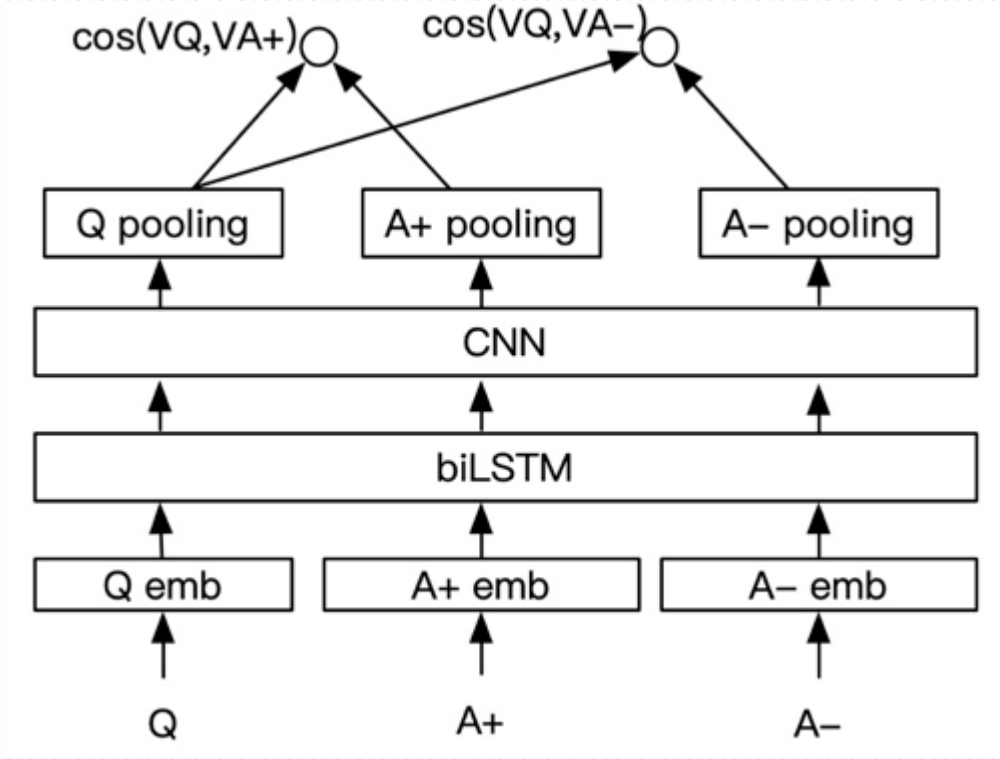
- 特点: 有了标准解决方案的才适合这个场景
- 输入: 我的啥啥怎么回事
- 输出: 标准问题
- 解决方案: 语义相似性匹配
  - DSSM
  - BERT + triplet
- trick:
  - 区分场景,压缩备选空间

售后客服机器人(第四代)


1:34
客服
您还没有订单~
在线客服
智能专属客服，轻松解决订单问题
常见问题
机票
酒店
火车票
门票
度假
用车
账号
其他
首页
订单
攻略
客服
我的

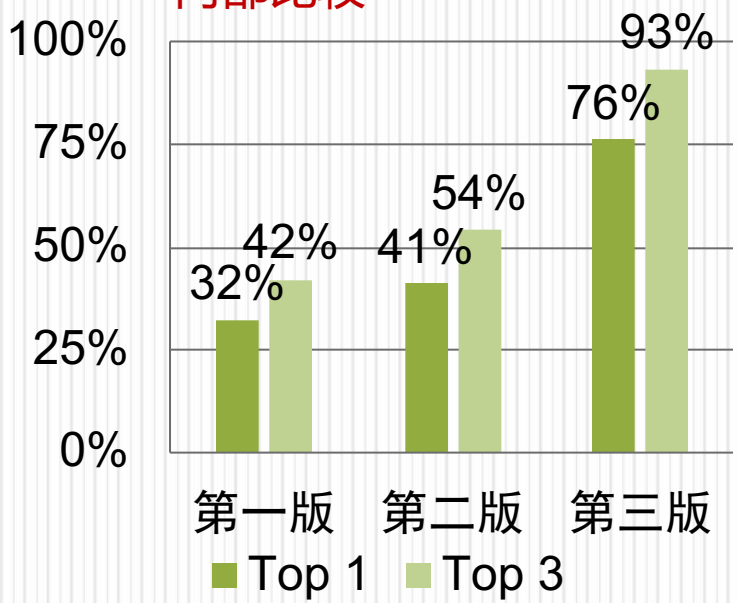

中国联通 4G
上午11:34
74%
在线客服
可以开发票吗
您是遇到了以下问题吗？请直接点击查看
补开发票
发票抬头能开什么
发票由谁提供
Qmi没有解决您的问题？找人工客服
补开发票
您好，您的订单已退款。若订单退款时收取手续费，可联系酒店补开手续费发票，酒店电话：
酒店信息咨询
订单操作
发票报销
请输入聊天内容

售后机器人模型结构(第三代)

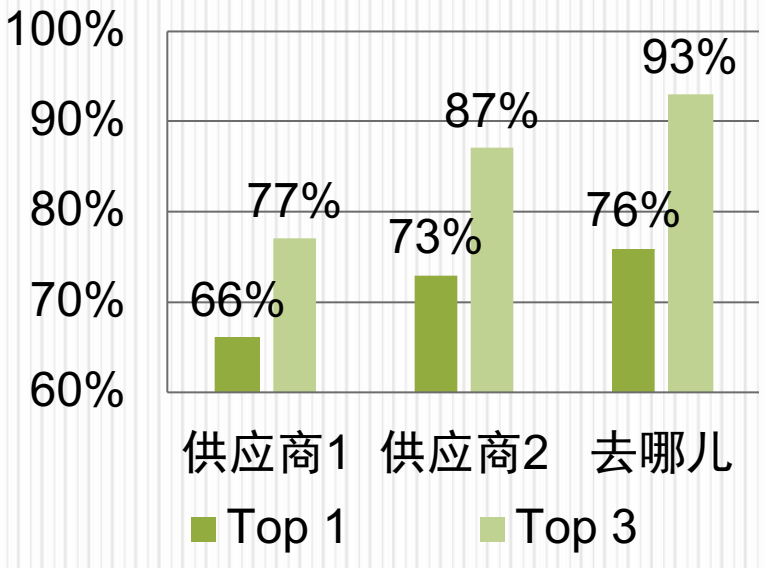


FAQ Bot：售后机器人准确率

内部比较



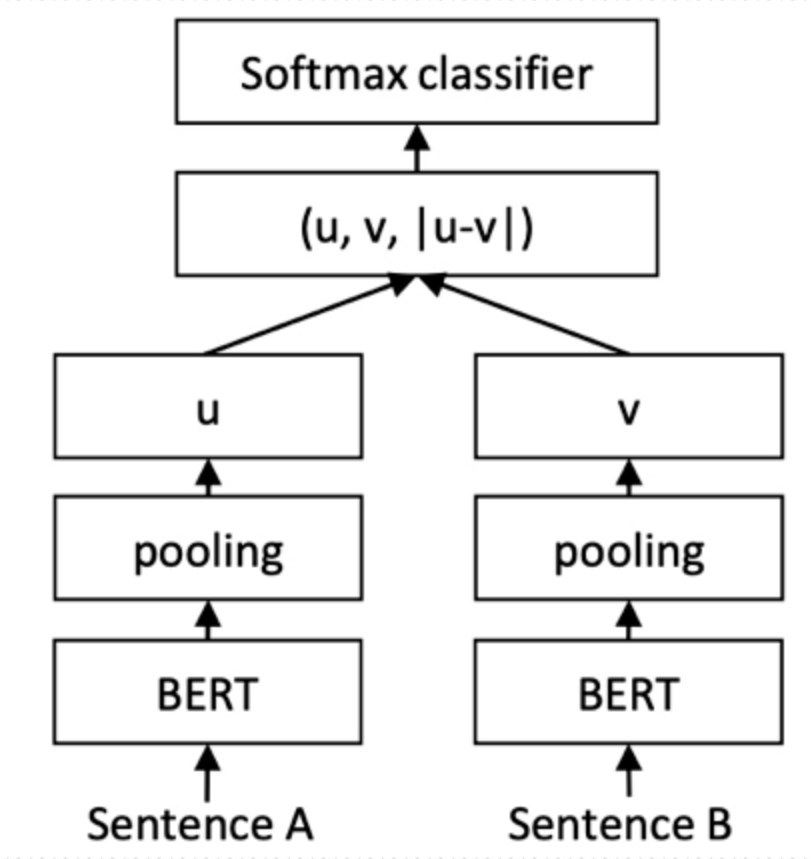
第三方比较



不同版本准确率： 第三版比第一版Top 1准确率高出 44%

不同供应商准确率：去哪儿自研机器人Top 1准确率比供应商1高10% 比供应商2高3%

售后机器人模型结构(第四代)



准确率比较

	P@1	P@3	P@5
base	0.6046	0.8207	0.8692
DSSM	0.5632	0.7913	0.8641
ESIM	-	0.3231	0.5095
BERT	0.3619	0.6147	0.7435
base + BERT	**0.7626**	**0.9332**	**0.9723**

推理性能比较

	batch@1	batch@10	all
BERT(8 核 cpu)	0.220 s	1.303 s	
4-Layer+BERT(8 核 cpu)	0.1271 s	0.3528 s	
ALBERT(4 层)	0.030551	0.203293	
ALBERT(12 层)	0.256034	1.600066	
cuBERT(12 层)	0.207686	0.915885	
(base)BILSTM+CNN(cpu)			0.012
CNN + BERT (gpu)	0.125090	0.166560	
CNN + BERT (cpu)	0.219316	0.854623	
CNN+ALBERT(4,max_seq=64) cpu	0.030551	0.203293	
CNN+ ALBERT (12,max_seq=64) cpu	0.256034	1.600066	
CNN +onnx(cpu)		1.654162	
CNN +cuBERT	0.207686	0.915885	
STS-BERT(gpu)			0.016921
STS-BERT (cpu)			0.049080

总结

- 聊天机器人的两类场景
  - 新用户界面: 替代搜索和翻页
    起到信息引导的作用
  - 解放人力: 替代客服
    把标准动作自动化
- 信息引导类机器人
  - 复杂繁琐,但有各种trick提升效果
  - 任重道远
- 标准问题的自动问答机器人
  - 能够从各种NLP语义表示的进步中获益
  - 复杂模型的性能是个问题

致谢与部分参考资料

- 致谢
  - 前Leader: 李兆海
  - 前小组一起调参的伙伴们

- 参考:
  - Qunar技术嘉年华(2018)Qdata专场《Qmi智能机器人》
  - 《深度学习在酒店售后智能问答场景实践》https://mp.weixin.qq.com/s/PPUh13SVvk3R0Tn0DgxWlQ