

携程AI和推荐系统的云化实践

AI+数据，开启公众旅行智慧

什么样的技术服务，难以云化？

和业务高度耦合的复杂技术体系



携程AI实践



AI的云化



集团个性化云



旅游技术的未来



携程AI实践

客服机器人、小诗机



Input:

景点写诗

城市写诗

天气、季节

心情

照片写诗

人名写诗

格律:

宝塔诗

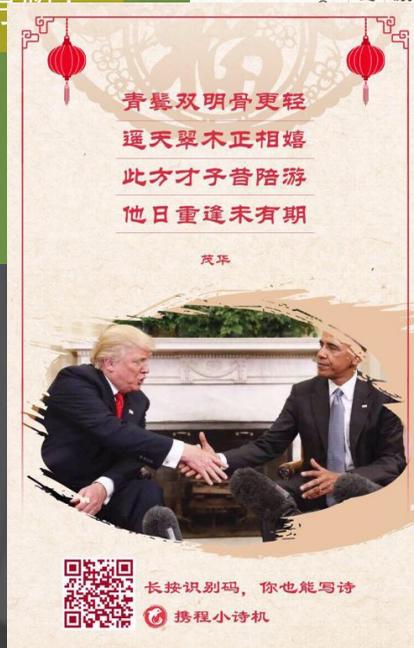
律诗

绝句

景物识别
&成诗：
红楼；
梅花；



性别及表情
识别：
严肃；
青年男性。



Input:

景点写诗

城市写诗

天气、季节

心情

照片写诗

人名写诗

格律:

宝塔诗

律诗

绝句

照片地点
知识库：
西湖；
西湖的知识库



时间写诗：
秋；
傍晚；
天气知识库；



郭德纲
老郭诙谐通道术，
融和春意动剑眉。
大话天仙妙绝伦，
喜嗔语默皆相宜。

人物画像：
职业；
作品；
身材；风格



基于情境的感情：
春节前最后一天；

人机盲测 第3名(11场)

与上海地区诗人比，略胜

	冠军	亚军	季军
赵玉玮	5	0	1
邹竞夫	2	6	2
AI LEE	2	3	2
卿云子	2	2	2
沈以昕	0	0	2
张一双	0	0	0

	思念	孤独	夏之愉悦	秋之悲伤	北京的雾	杭州的雨	黄山的雪	青海湖的晴	沙-正宝塔	沙-倒宝塔	景点 愉悦
赵玉玮	7.45	7.155	6.72	6.47		6.52	6.73			6.68	6.72
邹竞夫	8.37	7.575	6.705	6.49	8.22	6.41	6.685	6.32	6.425	6.24	6.375
AI LEE	7.885	8.14	6.195	6.535	7.61	6.225	6.07	6.4	6.15	5.91	6.4
卿云子		8.395	6.315		8.655		6.6			6.365	6.515
沈以昕	7.725	6.725	5.85	6.32	7.565	6.395			5.61		6.31
张一双	7.115			6.155			6.325	5.87	5.175		6.115

目前 远超竞品

宝塔寺、情感、季节、天气、人物、绝句、律诗、城市、美食

西湖
凉风乱入杭城路，
秋水孤山似旧年。
遍地桂花停过客，
侵帘绿影满湖边。

西湖
绿树余杭花草畔，
通都月色谁自怜。
喜听潇雨鸣枫叶，
安得凉风起桂筵。
久靠林荫堤岸伫，
长依湿地碧波边。
西湖桥下潺湲水，
凭有孤山却不前。

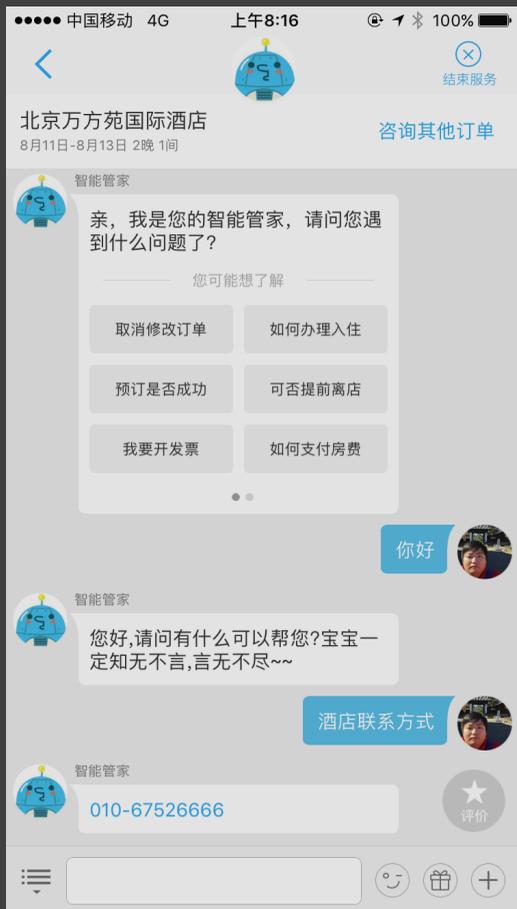
西湖.夏
重云深处引溪涧，
湖水孤山路远长。
遥爱断桥残夏早，
风清杨柳映通方。
莲花倚槛雨生雾，
荷叶绕船橹走旁。
林缝落光波潋潋，
寺边湿地色苍茫。

西湖.雪
城巷夜空却白昼，
西湖山色故依然。
敢忘湿地凌冬约，
不负君梅岁竹缘。
万里重云寒水边，
满城娉树雪中天。
萧疏枫叶扁舟岸，
灵隐冷风谁顾怜。

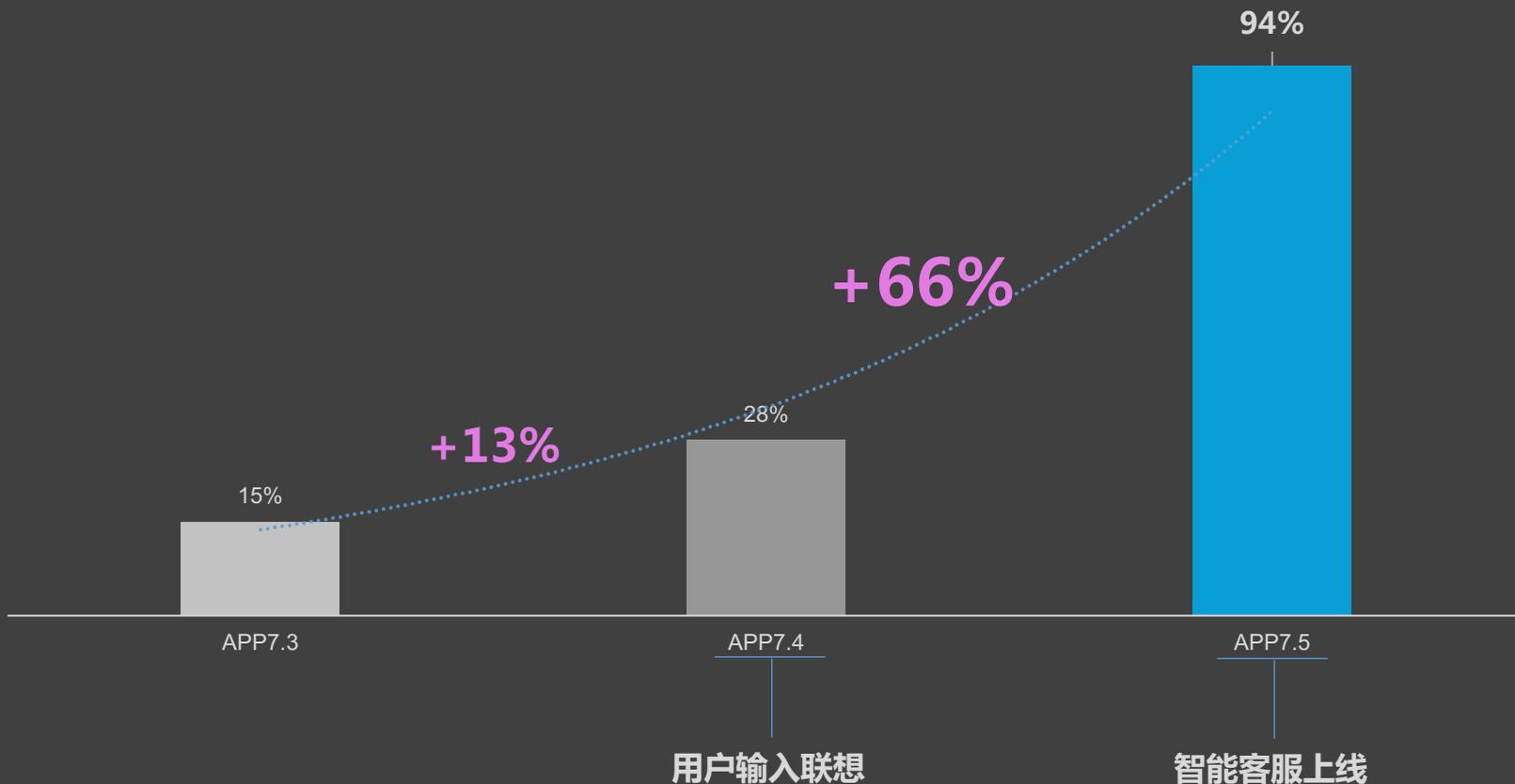
澜
惊湍
水流漫
鲸波海观
江涌万卷蟠
涛濑掀翻浪寒
激荡澎湃渺云端

香惹燕雀凉醉颊
玉屑梅蕊翠靥
偷香团扇妾
芳情细贴
花重叠
春叶
蝶

智能客服（整体解决率达到80%左右）



酒店客服可自助场景解决率

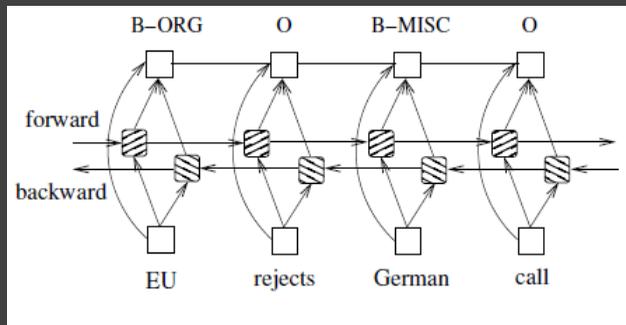


应用架构



命名实体识别

- 应用场景
 - ✓ 售前售后景点、目的地等识别
 - ✓ 旅游名词识别
 - ✓ KB-QA中识别主题信息



- 常用模型

Model	优势
RNN	获取长远的上下文信息，拟合非线性能力
CRF	考虑整个句子的局部特征的线性加权组合，目标是 $p(y x;w)$ 使得序列的联合概率最大
HMM	基于马尔科夫性质
LSTM+CRF ^[1]	利用两个模型的优势

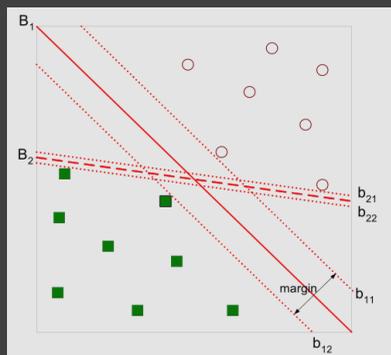
```
我想下个月5号去杭州看西湖
ParseResult:
{
  u'City': {
    u'start': [],
    u'unknow': [],
    u'dest': [{u'type': u'city', u'id': 17, u'name': u'杭州'}]
  },
  u'IsDiscount': -1,
  u'error': u'',
  u'Poi': [{u'id': 91557, u'name': u'西湖'}],
  u'Date': {
    u'StartDate': u'2017-10-5',
    u'EndDate': u'',
    u'Days': u'',
    u'UnknowDate': u''
  },
  u'TripType': [],
  u'Level': -1,
  u'TopicLabel': []
}
```

原句: 我要开发票: 衢州恒业汽车部件有限公司, 税务识别号9133080275590610X5。邮寄地址: 浙江省衢州市柯城区三江东路2号卢森堡A座浙江青风律师事务所戴剑波收, 手机13305703919
 解析: 我要开发票: [institution], 税务识别号9133080275590610X5。邮寄地址: [address][person]收, 手机[mobile]

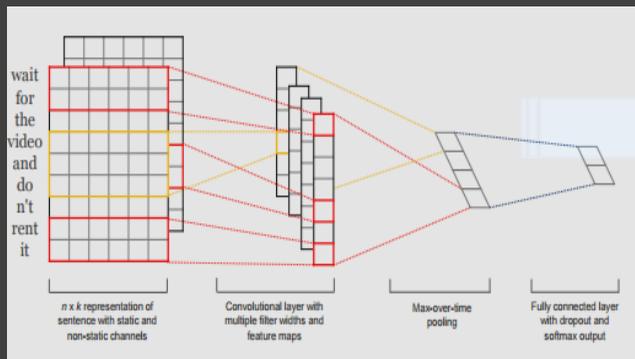
[1]. Z Huang, W Xu, K Yu. Bidirectional LSTM-CRF Models for Sequence Tagging - arXiv preprint arXiv:1508.01991, 2015

用户意图识别

- 基于用户问题，使用意图分类模型进行用户意图的判定，从而将用户问题定位到某个子领域。
- 辅助进行进一步的用户语义解析
- 基于意图信息，进行用户对话管理和状态分析

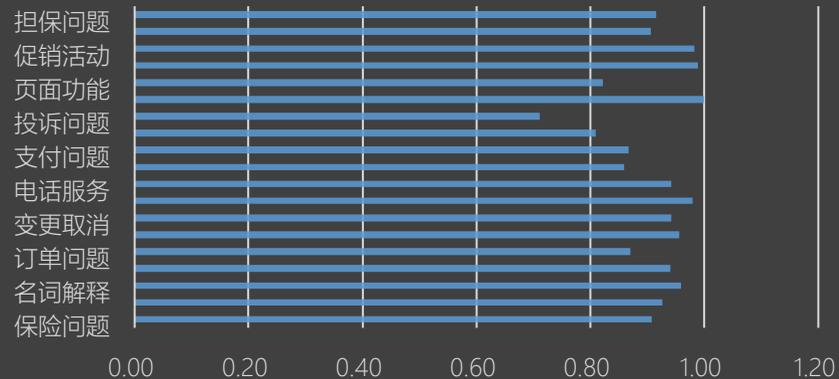


SVM分类模型

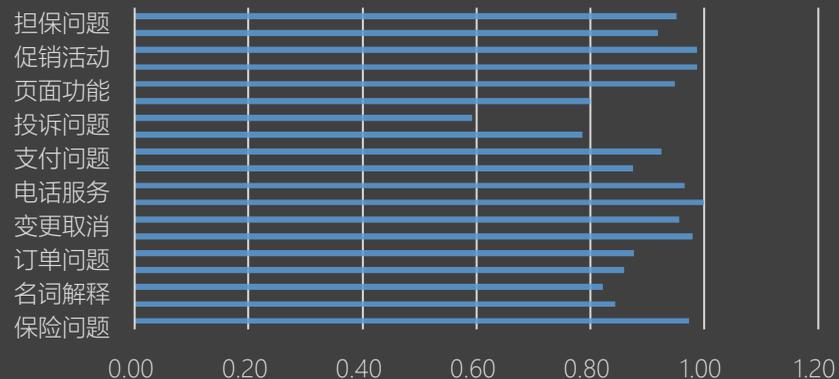


CNN分类模型

Precision



Recall

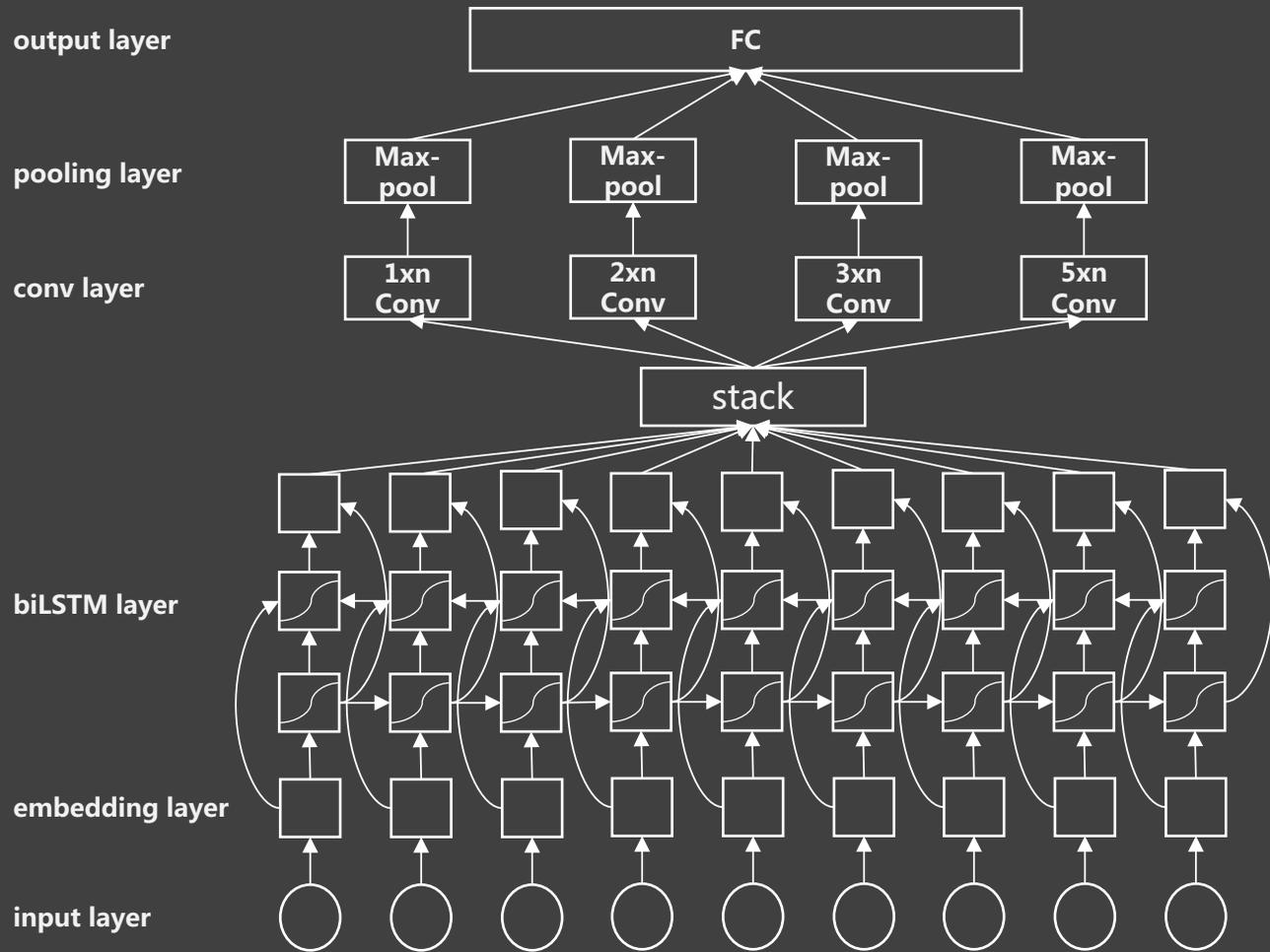
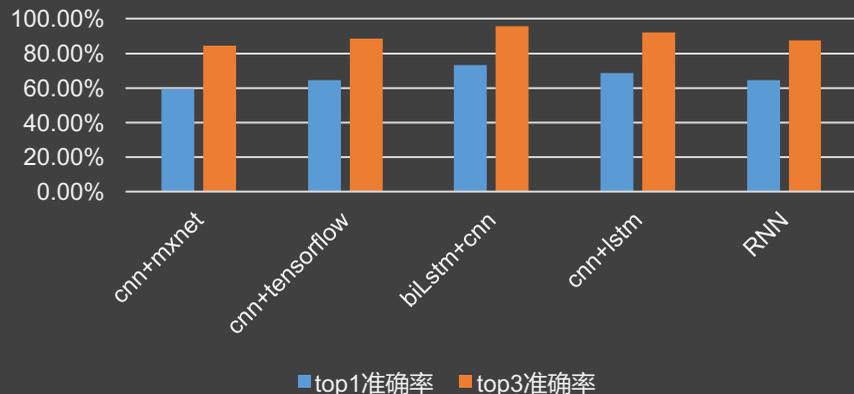


标准Q匹配

优势

- biLSTM可以从forward和backward两个方向获取上下文信息
- CNN模型对平移、缩放等具有不变形，可以较好检测出特征区域，鲁棒性好。
- 最终，通过多组实验比较，发现biLSTM+CNN离线效果最佳。

算法对比

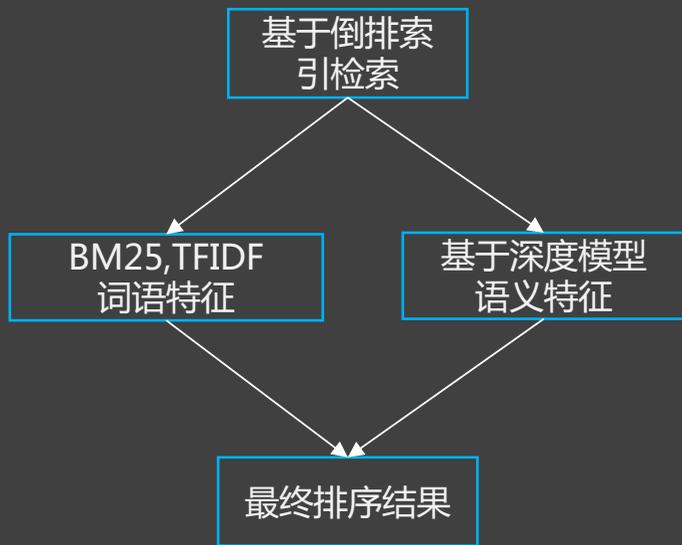


基于biLSTM-CNN的QA匹配

[1] Applying Deep Learning to Answer Selection: A Study and An Open Task Minwei Feng, Bing Xiang, Michael R. Glass, Lidan Wang, Bowen Zhou ASRU 2015
 [2] Lai S, Xu L, Liu K, et al. Recurrent Convolutional Neural Networks for Text Classification[C]//AAAI. 2015, 333: 2267-2273.
 [3] Wen Y, Zhang W, Luo R, et al. Learning text representation using recurrent convolutional neural network with highway layers[J]. arXiv preprint arXiv:1606.06905, 2016.

QA-检索

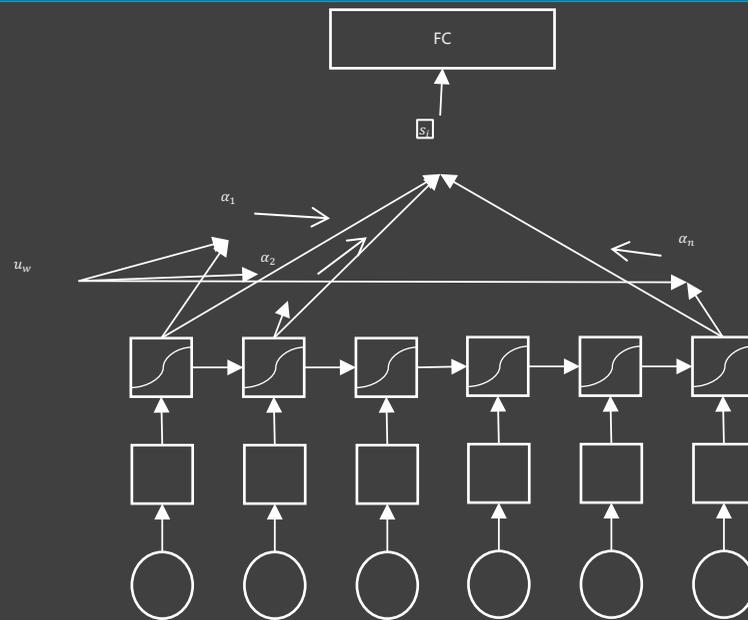
- 类似个性化推荐，分两步：检索+排序
- 适用QA量大的场景，例如闲聊，问答检索等
- 基于attention机制可以提升QA关联度



$$u_{it} = \tanh(W_w h_{it} + b_w)$$

$$\alpha_{it} = \frac{\exp(u_{it}^T u_w)}{\sum_c \exp(u_{it}^T u_w)}$$

$$s_i = \sum \alpha_{it} h_{it}$$



基于检索的QA

>>北京的景点有什么
北京旅游景点之一：北海公园：位于故宫西北角，始建于十世纪初，是中国现存最完整的古代园林之一，公园面积70余公顷，其中水面有39公顷。
北京旅游景点之二：大钟寺。原名觉生寺，座落在北京城的北三环西路，建于1733年。因寺内藏有明代永乐年间铸的大钟，俗称大钟寺，永乐大钟高近7米，外径3.3米，重达46.5吨；钟身铸满佛教经文，约23万字。钟声清朗，方圆百里可闻。
北京旅游景点之三：天坛。明、清两代帝王祭天祈谷的地方，建于1420年。坛墙南方北圆，象征“天圆地方”。
北京旅游景点之四：十三陵。位于八达岭长城东南不远的天寿山下，这里埋葬着明代的13个皇帝。从1409年始建长陵起，到1644年修建思陵止，历时200余年。1956年，中国考古工作者发掘出万历皇帝朱翊钧的陵墓——定陵，发现了地下宫殿，除棺椁外，还出土了3000余件珍宝。北京旅游景点之五：雍和宫。在北京城区东北角，建于1694年，原为清第三位皇帝雍正皇帝的府邸，1744年改建为喇嘛庙。庙中藏有18米高的白檀木大佛，由一整棵檀木雕成。

>>上海到杭州怎么走方便
动车

>>我要去内蒙古，要注意什么
内蒙古旅游详情：<http://you.ctrip.com/place/neimenggu100062.html> 参考一下

[1] Lowe R, Pow N, Serban I, et al. The ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems[J]. arXiv preprint arXiv:1506.08909, 2015.

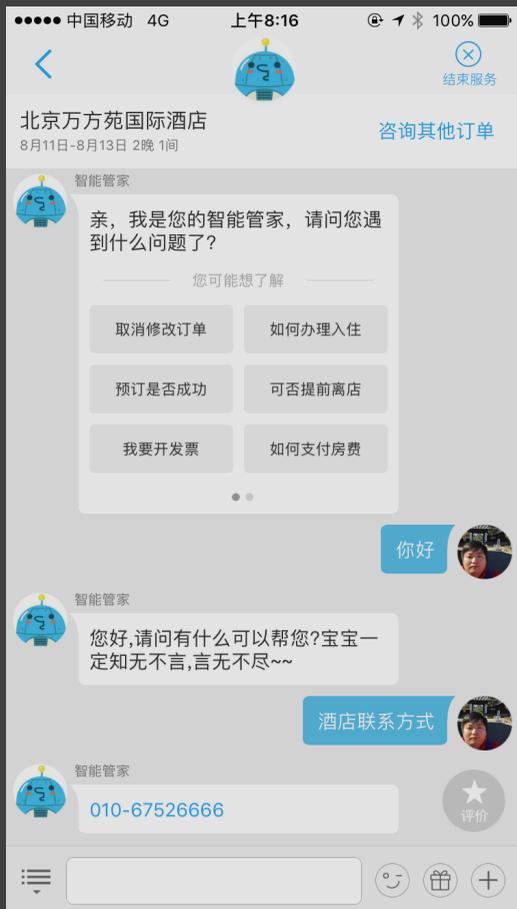
[2] Yang Z, Yang D, Dyer C, et al. Hierarchical Attention Networks for Document Classification[C]//HLT-NAACL. 2016: 1480-1489.



AI云化的实践

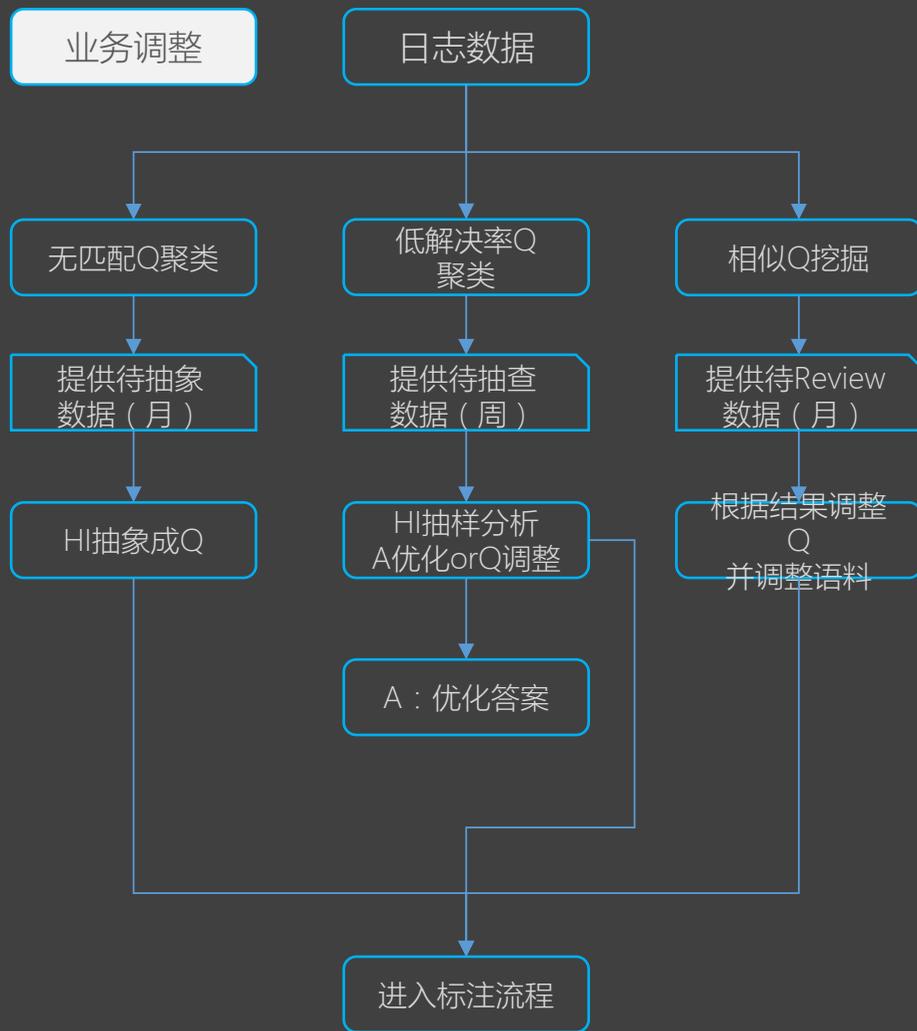
客服机器人、AI云接入

智能客服（整体解决率达到80%左右）



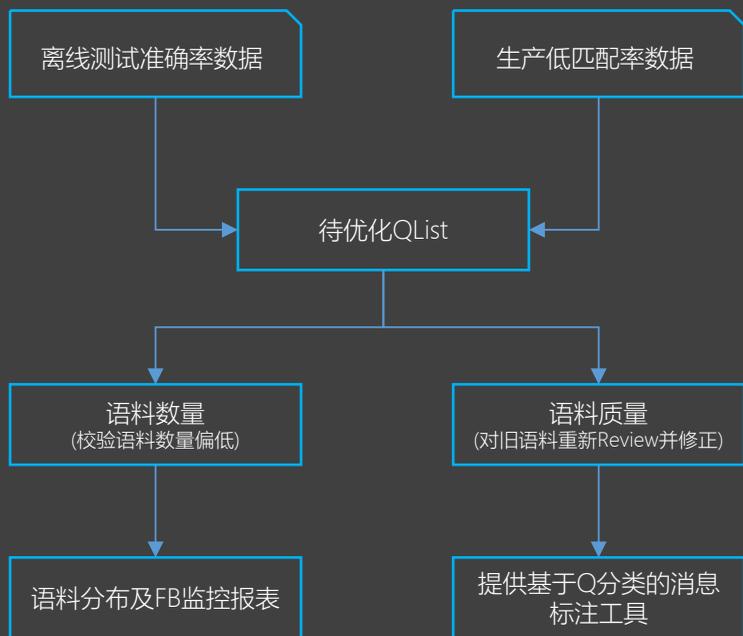
如何支持集团内60余条业务线?

SQ运营云化

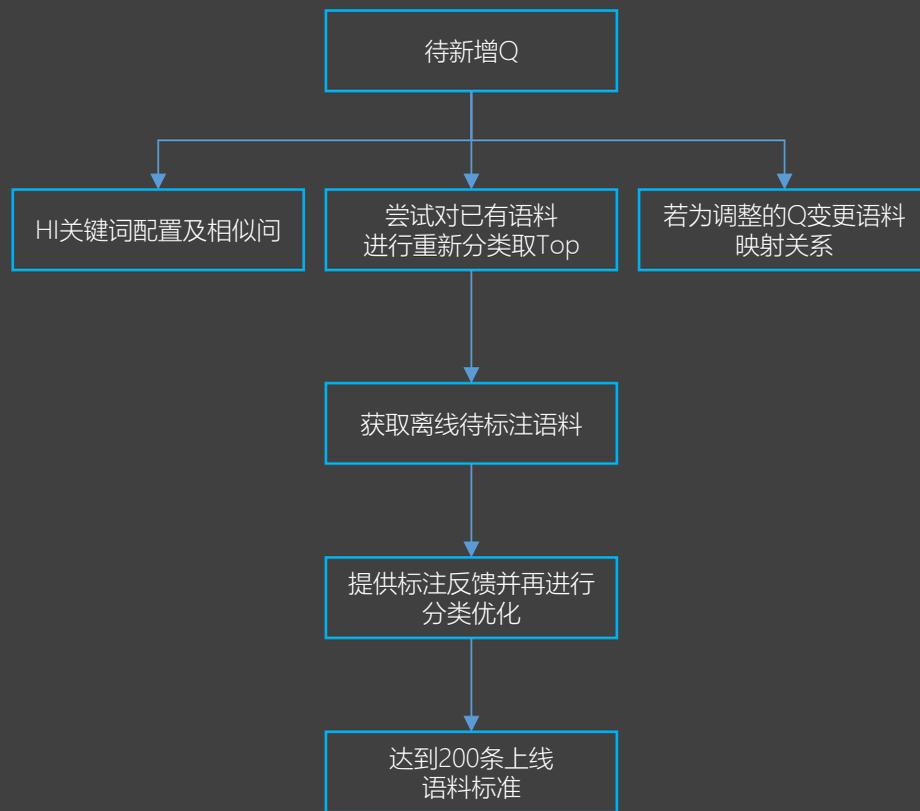


语料标注流程云化

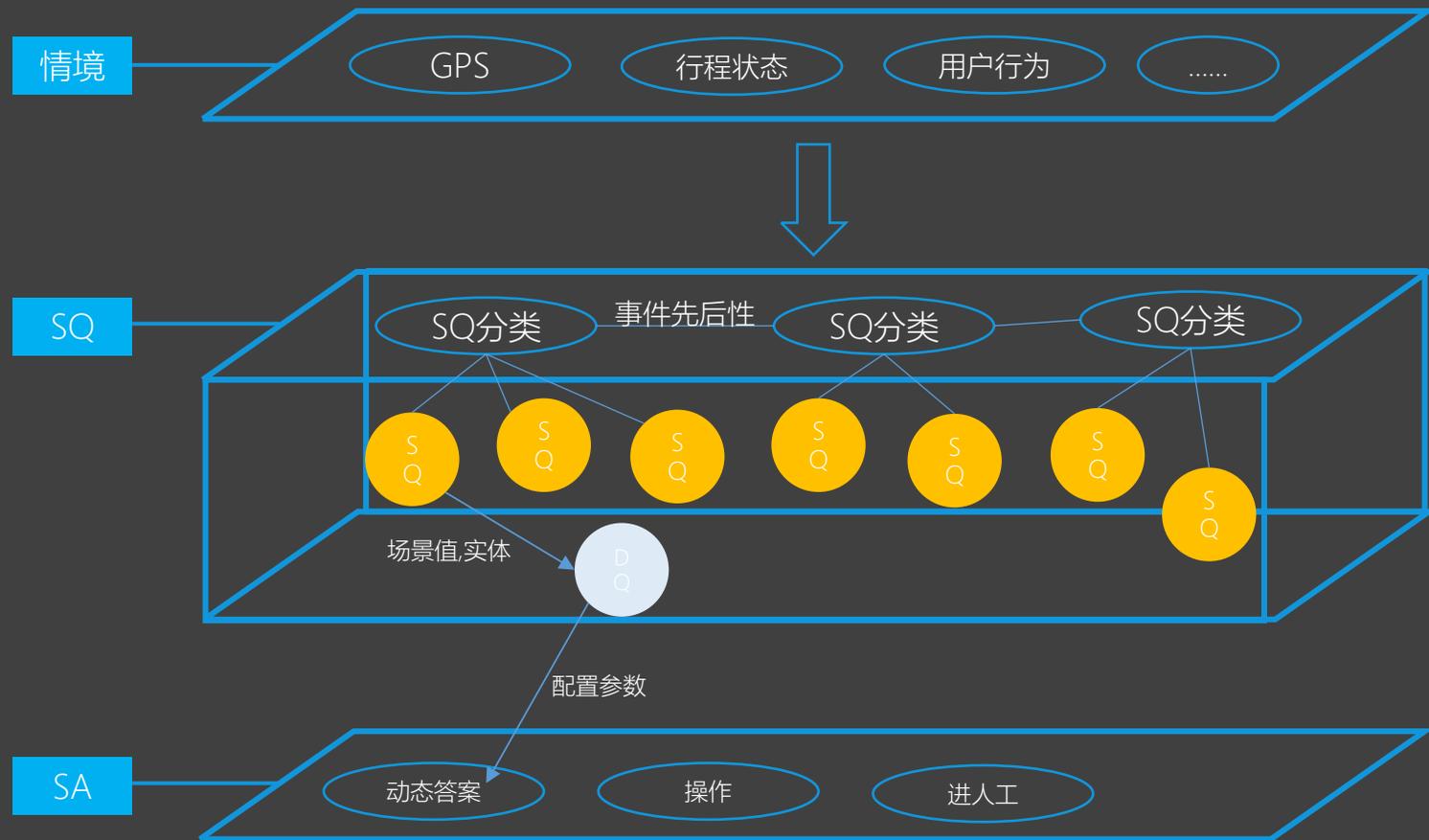
1 已有Q的语料标注



2 新增Q的语料标注



对话管理业务逻辑的云化



聚类用户情境
建立连接

事件的流转
通过场景、实体等对SQ的推理

后续提供服务的Action基于用户详情需求深度定制覆盖

Democratize AI

合作、互补、共享 → 共赢、高速创新 → 下一代的OTA用户体验





动物识别

植物识别

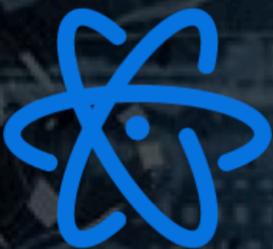
个性化智能导购

图像主检测

相同图检索

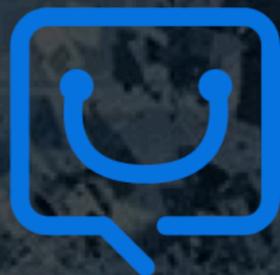
相同图入库

图像审核



用户意图

行中
行前
决策期
站外



用户画像

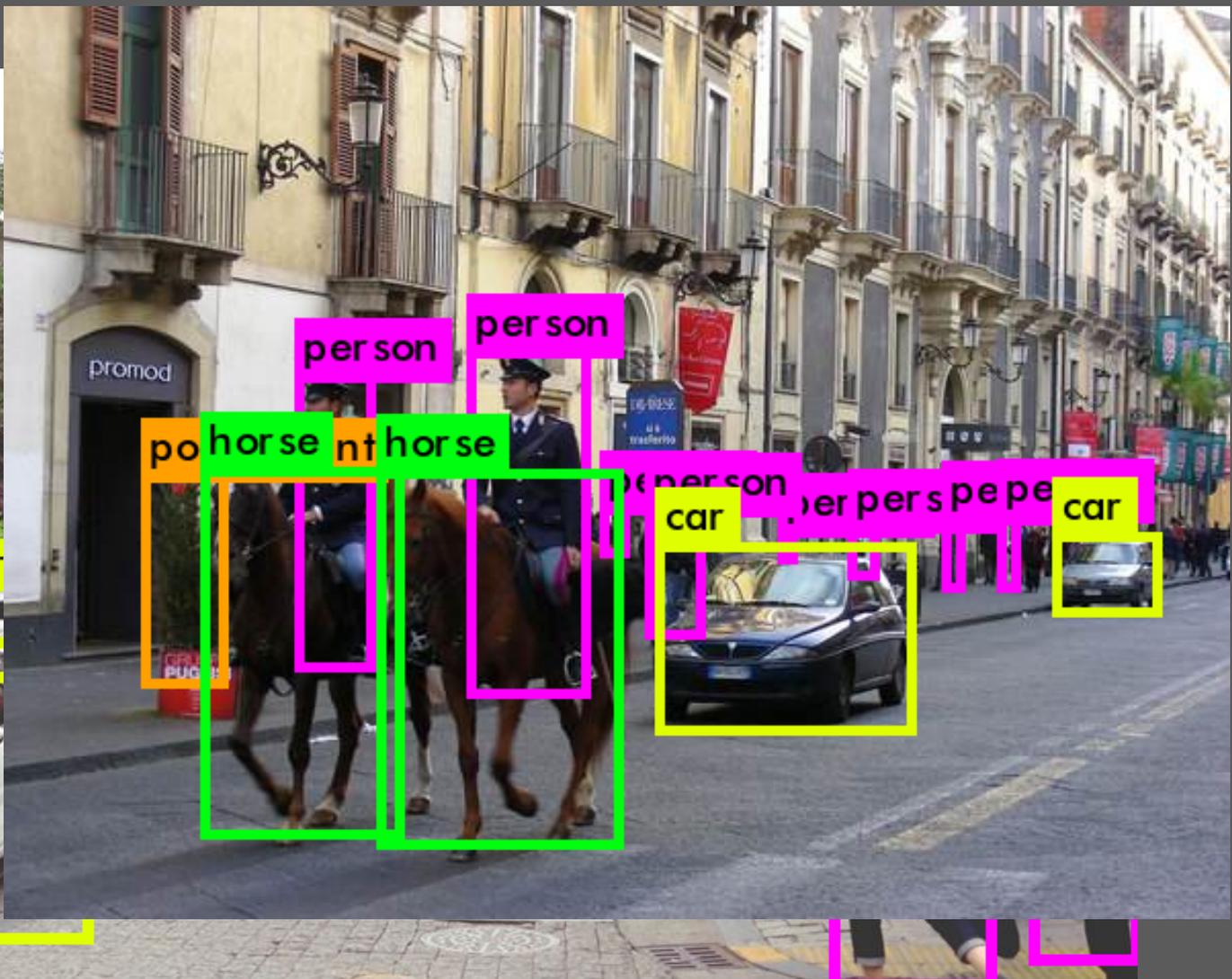
人口属性
外貌属性
消费偏好



计算广告

DSP出价AI
赛马监督

图片检测(Image Detection)





集团个性化云平台

面向60余条业务线

行前 个性化与智能目的地推荐

PERSONALIZED RECOMMENDATION



行前 个性化与智能目的地导购

PERSONALIZED RECOMMENDATION



- 全业务线，落地**100**余个场景
- 年化营收（毛利）提升**2.1**亿元
- 架构：可配置模块化，高速迭代
- 算法：深度神经网络、迁移学习的引入（AAAI-17）

想好去哪玩？并不容易

- 有多少候选目的地？
国内游：21天，232次访问操作
东南亚：40天，459次访问操作
- 每个目的地，能玩什么？
- 机票、酒店价格？
- 该什么时间去？





用户粘性 (使用UV / 曝光UV)



人均毛利 (总毛利 / 曝光UV)

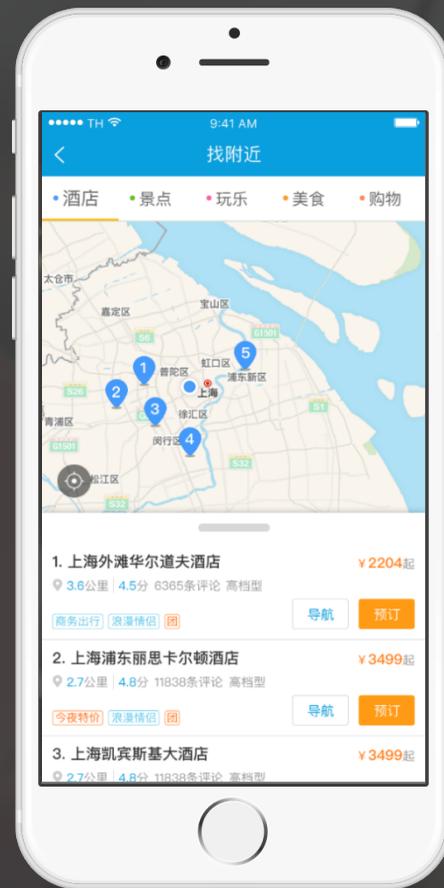


用户离开“常住地”



聚合不可缺的：吃住行娱购游

- 出行人的属性
- 出行人的群体兴趣
- 行程特征（差旅/旅游/亲子/消费档次）



行前 个性化与智能目的地推荐

PERSONALIZED RECOMMENDATION



架构 组件式可配置化架构

PERSONALIZED RECOMMENDATION

1.4亿次/天的高并发访问
场景接入/更新：36个人日->1.3个人日

行业内突破：
场景上线配置化
场景迭代拆分化

Big Picture of Arch.



配置推荐场景

场景名称	<input type="text" value="详情页推荐"/>	公共配置
请求量(kps)	<input type="text" value="0"/>	
场景Code	<input type="text" value="572a004f928b9ff07177316"/>	

SOA 请求列表

name	url	request	response	
订单历史	http://ws.myctrip.members.ctripcorp.com/orderindex-service/api/getbrieforderinfosbyuid	{@ ...}	0	
收藏历史	http://ws.content.members.ctripcorp.com/member-favoritews/GetMyFavorites	{@ ...}	0	
浏览历史	http://10.15.100.68:8080/wsruserbehavior/rtuserbehaviorquery.json	{@ ...}	0	
发现状态服务	http://10.8.89.249:8080/intelligentrec/getDiscoveryTourStatus.json	{@ ...}	0	
发现首页V2	http://10.8.89.249:8080/intelligentrec/getDiscoveryRecommendV2	{@ ...}	3	
周边游	http://10.15.146.237:8080/intelligentrec/gethandinhandtourrecommend.json	{@ ...}	2	
站外(今日头条)	http://10.15.146.237:8080/intelligentrec/getOutsideADRecommend.bjjson	{@ ...}	0	
站内在线(Landing)	http://10.15.146.237:8080/intelligentrec/getMarketOnlineRecommend.json	{@ ...}	1, recommendations, travels	
特惠酒店	http://10.15.146.237:8080/intelligentrec/getHTLUserPreference.json	{@ ...}	0	
站内在线(EDM)	http://10.15.146.237:8080/intelligentrec/getMarketOnlineRecommend.json	{@ ...}	1, recommendations, travels	
站内市场离线	http://10.15.146.237:8080/intelligentrec/getMarketBatchRecommend.json	{@ ...}	1, recommendations, travels	
微领队	http://10.15.146.237:8080/intelligentrec/getMicroGuideRecommend.json	{@ ...}	0	
牵手游	http://10.15.146.237:8080/intelligentrec/gethandinhandtourrecommend.json	{@ ...}	2	
发现首页	http://10.15.146.237:8080/intelligentrec/getDiscoveryRecommend	{@ ...}	3	

- 董锐
- 产品与服务
- 个性化推荐
- 精准化营销
- 推荐营销 DEMO
- 对话标签
- 用户中心

- 董锐
- 董锐
- 产品与服务
- 个性化推荐
- 精准化营销
- 推荐营销 DEMO
- 对话标签
- 用户中心
- 用户账号
- 维表管理
- SOA 请求列表
- bu 维表管理
- 约束条件管理
- 策略维表管理
- 策略可配置项管理
- 契约参数类型维表
- 契约可配置参数维表
- 推荐管理
- 推荐配置表

场景名称
请求量(kps)
场景Code

完成 返



旅游科技的未来

智能化



移动互联网10年发展趋势



截止2016年，移动网民占比已达**95.1%**，后续增长困难，用户红利已离我们远去

下一代互联网的探索

网页互联网



移动互联网



AI互联网

Penetrate all OTA-related Scenarios

Shopping on APP

On-travel

At-Home

Hotel Room

Airport

Road Trip

Museum

Embedded in Various AI Devices with Low Cost



Voice

Vision/Text

Focus on Ctrip Dialogue Robot Engine

After-sales Robot



Product Consulting Robot



Shopping Guide Robot



Tour Guide Robot

1. Construct Ctrip Product/info Knowledge-map
2. Product Info Reinforcement in a short word
3. Personalization

Infra Tech Reserves

1. Construct Novel Product Design via Voice-vision
2. Design Shopping Guide Flow based on AI-UI

Product PoC

1. Collect & Analysis UE at each step
2. Collect Training Corpus to improve intelligence

User Feedback Collecting

Focus on **Ctrip Dialogue Robot Engine**

After-sales Robot



Pre-sales Robot



Shop Guide Robot



Tour Guide Robot



THANKS