



怀宝兴

华为云  
EI 语音语义团队算法总监



全球技术领导力峰会

Geekbang> | TGO 鲲鹏会  
极客邦科技

# 500+ 高端科技领导者与你一起探讨 技术、管理与商业那些事儿



🕒 2019年6月14-15日 | 📍 上海圣诺亚皇冠假日酒店

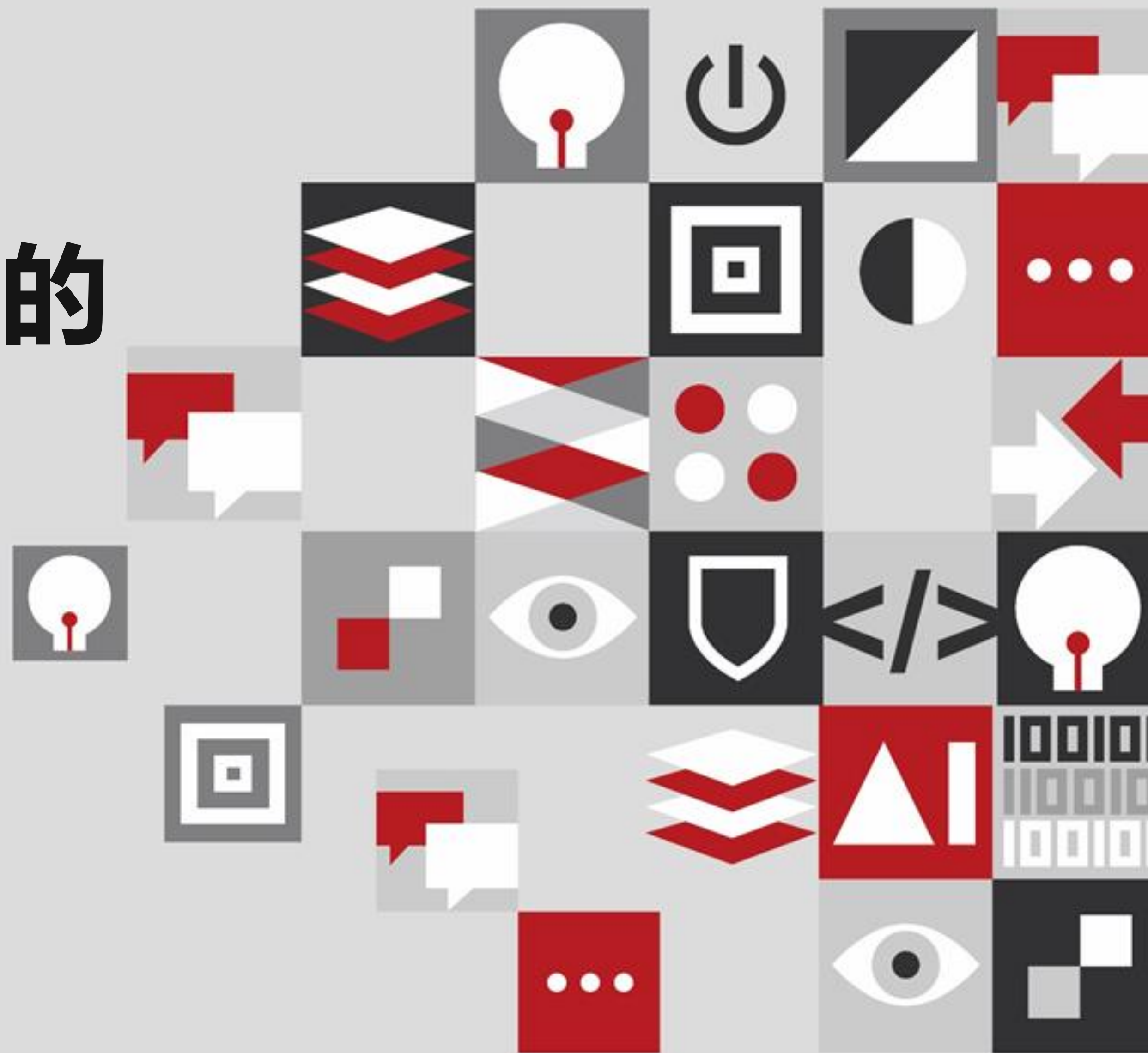


扫码了解更多信息

# NLP在面向企业场景中的 算法优化与落地

怀宝兴

华为云语音语义域算法总监





# 自我介绍

怀宝兴

毕业于中国科学技术大学

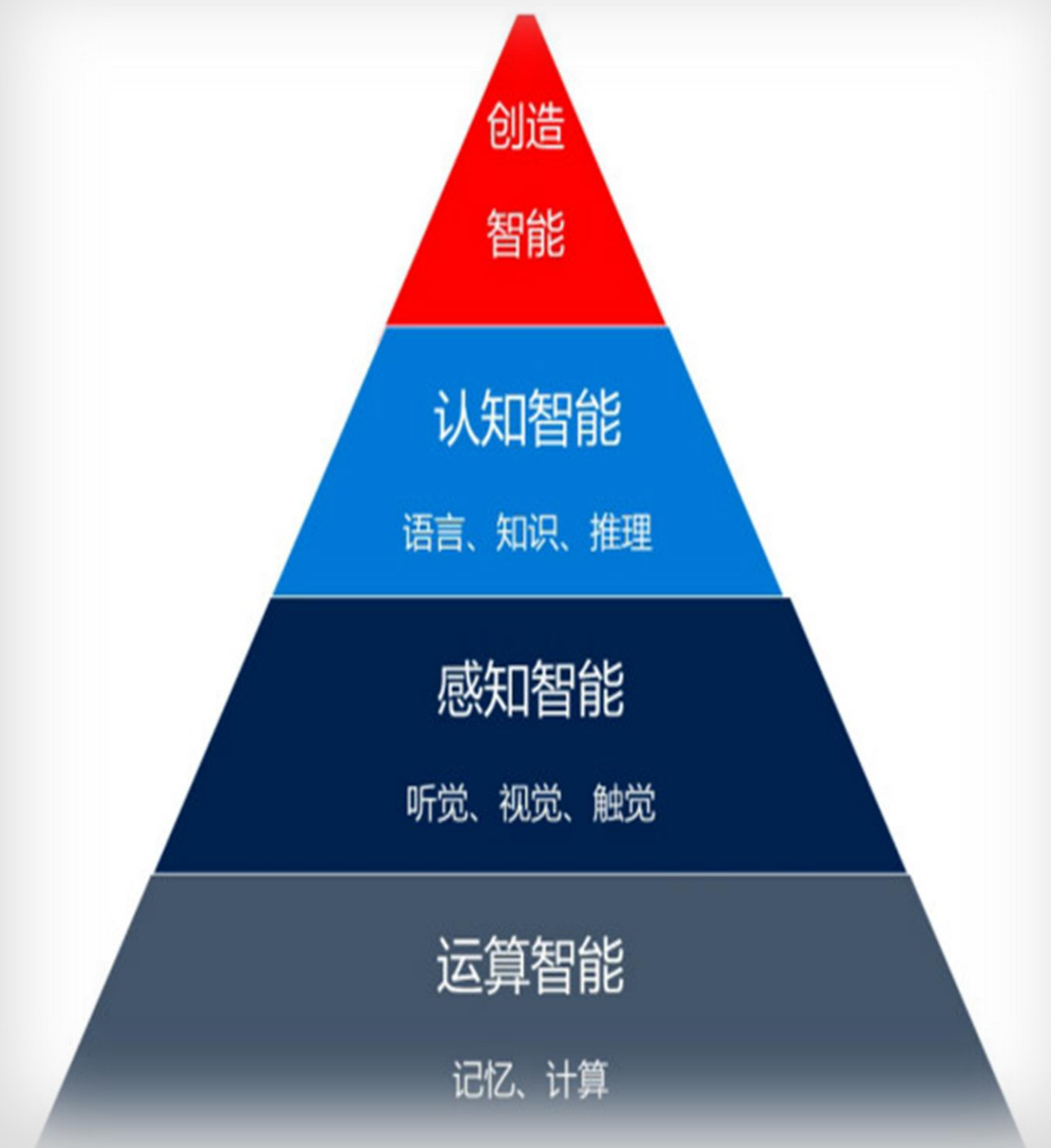
研究方向为文本挖掘、机器学习、数据挖掘

2014~至今：华为，机器学习、文本挖掘等算法工作

- ◆ 自然语言是什么
- ◆ 智能对话系统中的NLP能力
- ◆ 搭建一个智能对话系统
- ◆ 对话系统在企业智能中如何发挥作用

# 自然语言是什么？

- 最不“自然”的自然语言
  - 离散的符号系统
  - 社会性
- 语言 VS 图像、语音
  - 语音、图像——自然界存在的连续信号
  - 意义都是人赋予的



人要是行，干一行行一行，一行行行行行；  
要是不行，干一行不行一行，一行不行行行不行

## ➤ 歧义

- 乒乓球拍卖完了（分词歧义）
- [咬死猎人]的狗 VS. 咬死[猎人的狗]（短语歧义）
- 你真讨厌！（语用歧义）
- 冬天能穿多少穿多少，夏天能穿多少穿多少（世界知识）

## ➤ 病构

- 不合乎语法的句子
  - 例子：他非常男人。（名词不能受程度副词修饰）
- 不合乎语义约束的搭配
  - 例子：My car drinks gasoline like water

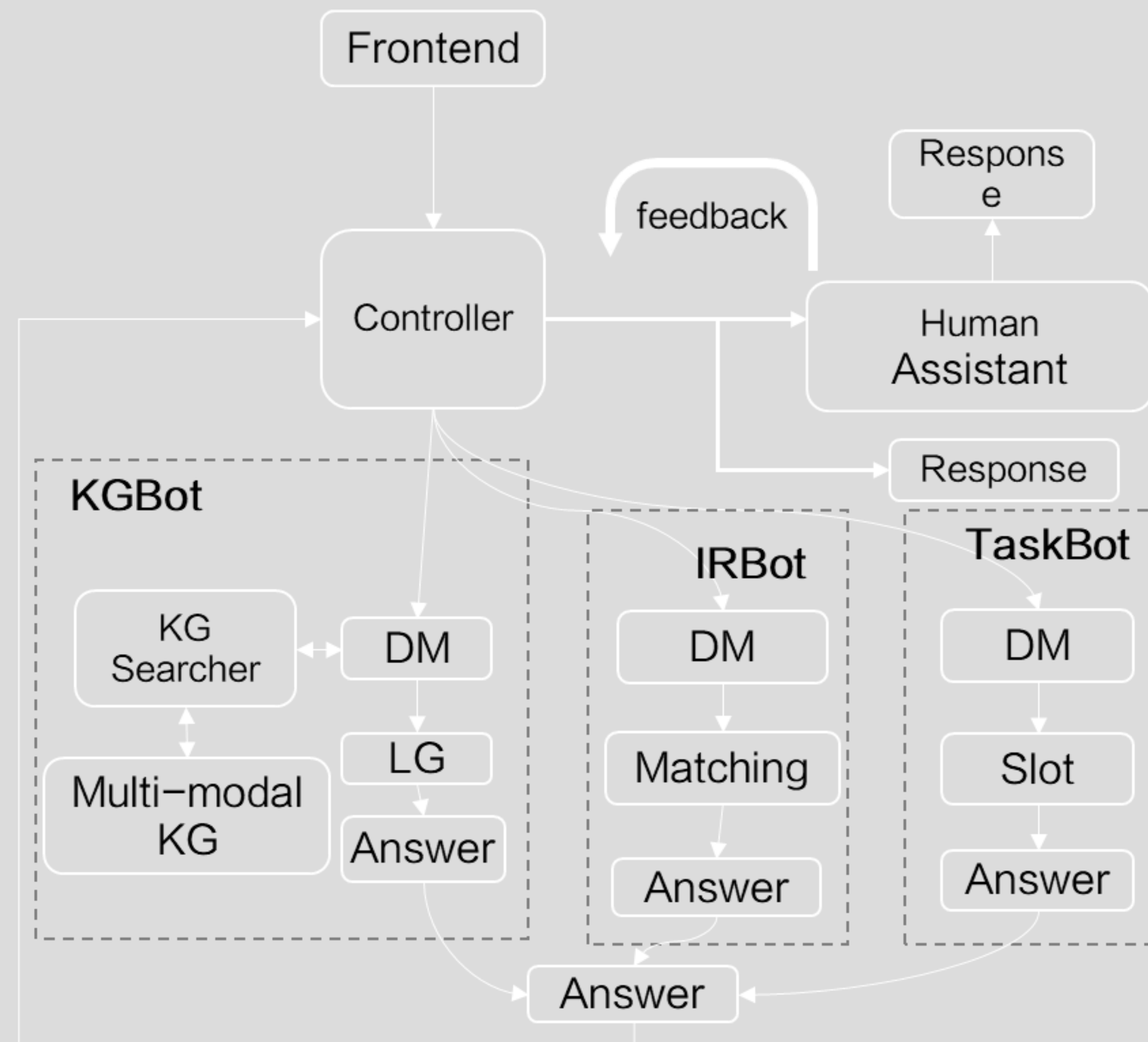
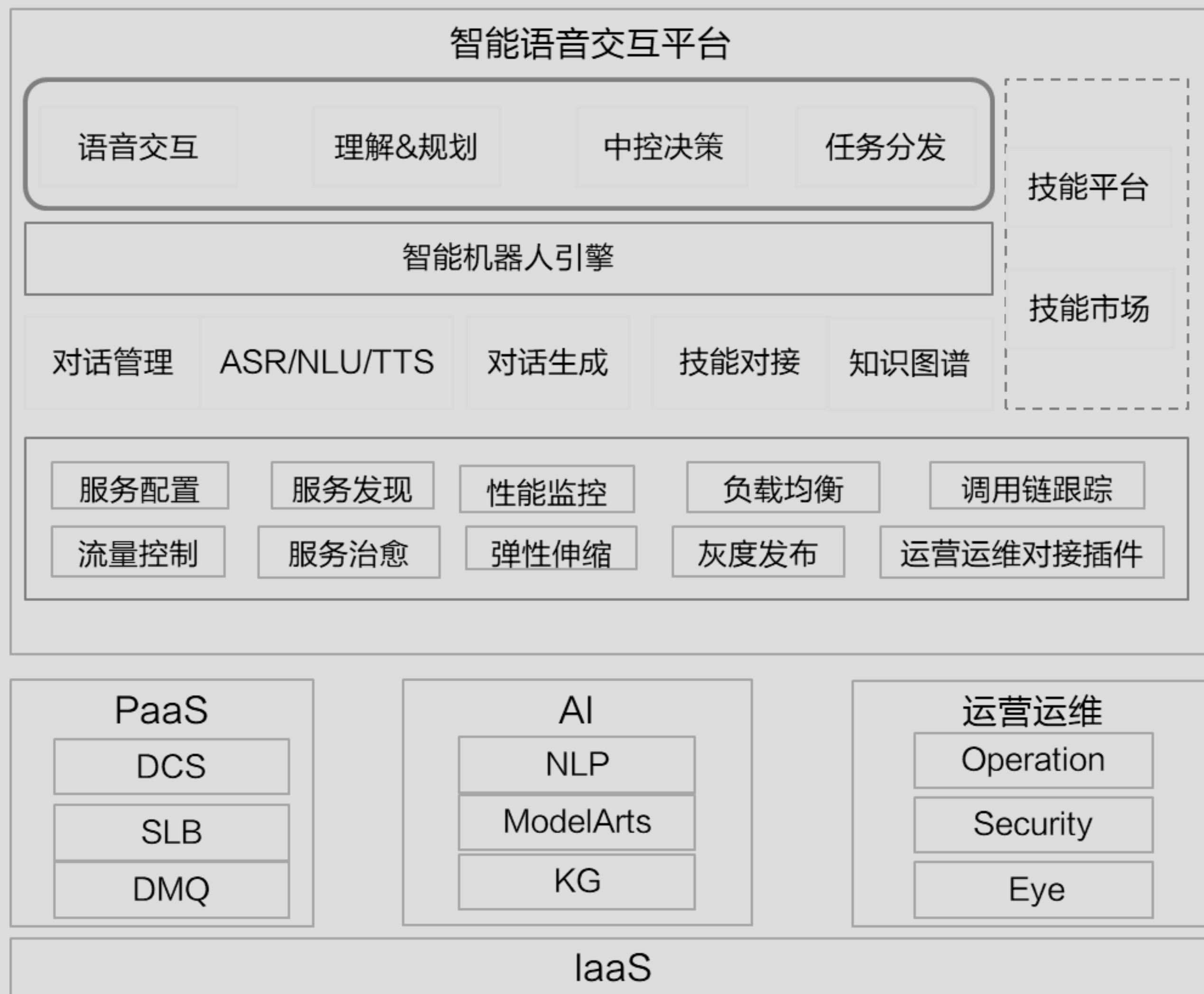


# NLP-应用场景





# NLP-对话系统



# NLP算法-分词

## 中文分词的难点

歧义：严守一/把/手机/关了

世界知识：夏天能穿多少穿多少，冬天能穿多少穿多少

## 企业应用的难点

适配多种应用的分词粒度

检索：细粒度=> 高召回率

意图识别：粗粒度=>实体

领域适配

封闭域=>开放域

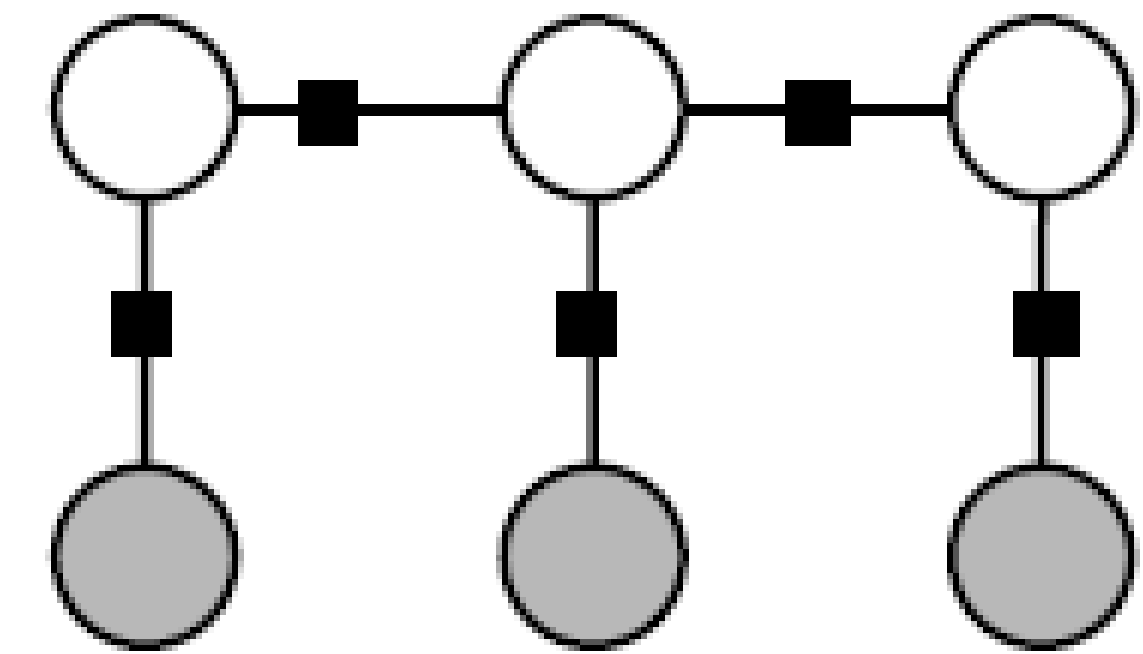
通用域=>垂直领域

## 实战经验

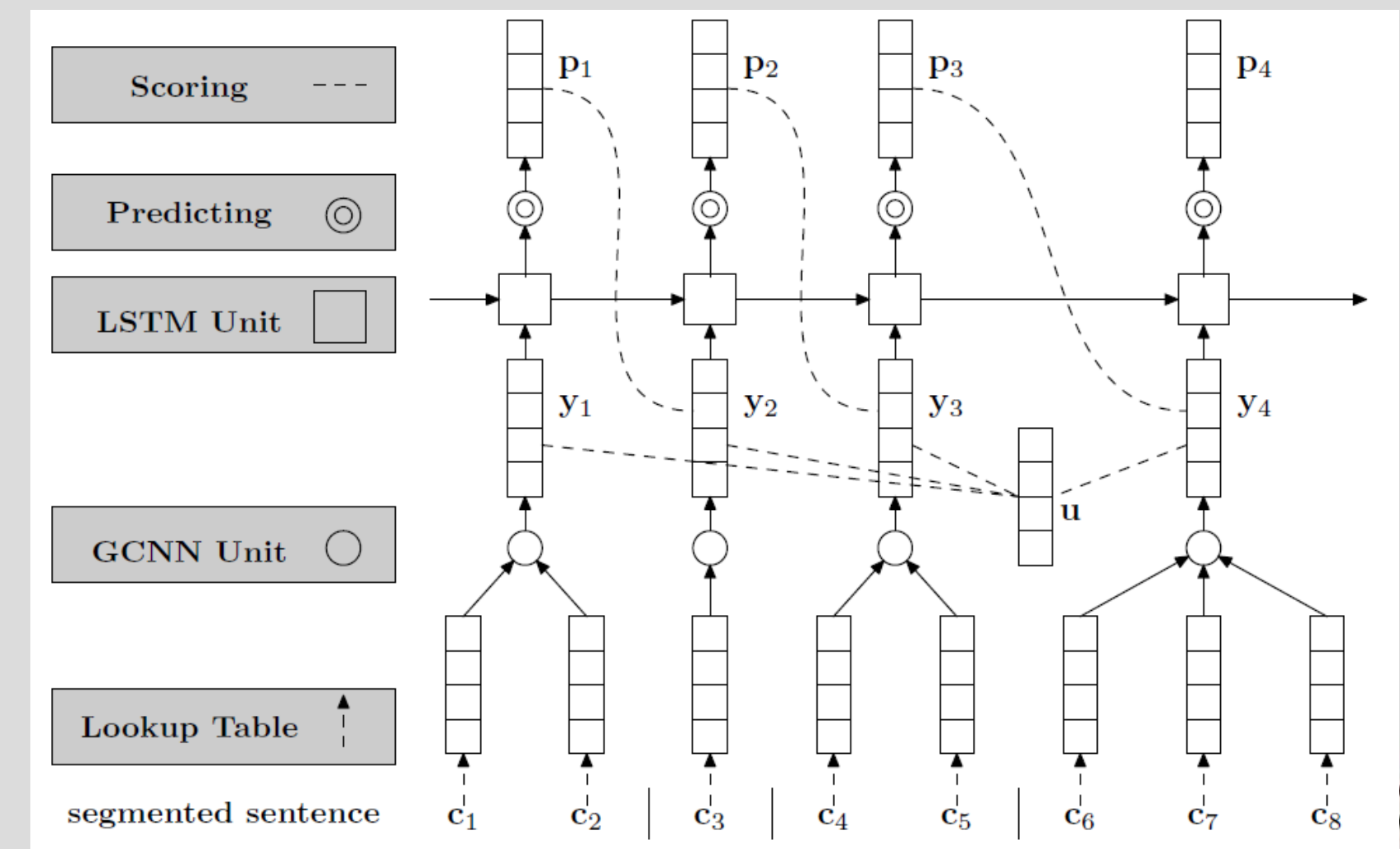
多粒度分词

领域词典构建+模型融合词典

深度学习



Linear-chain CRFs



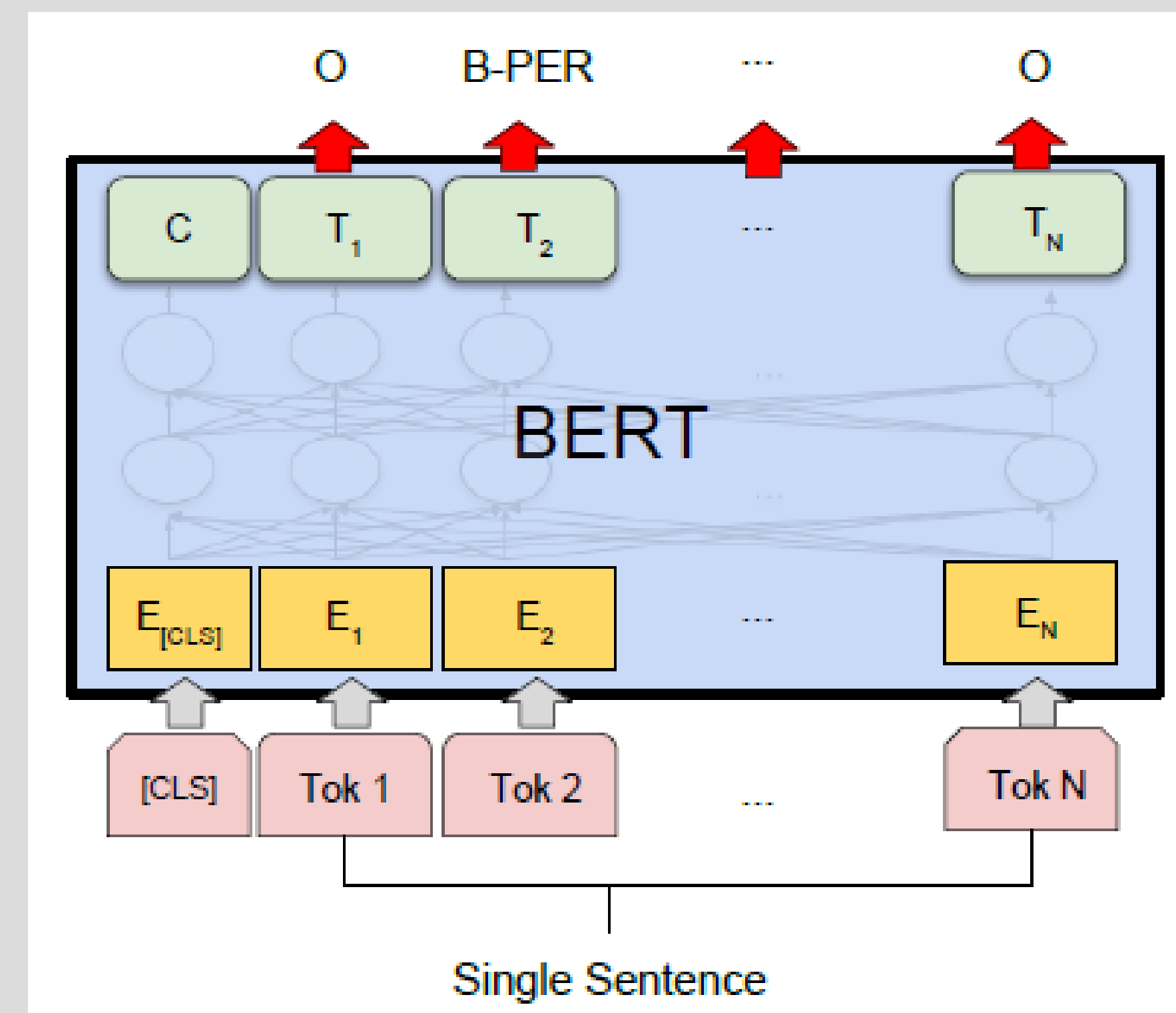
# NLP算法-NER

## 企业场景的难点

- 不同领域都有不同类别的实体  
娱乐=>歌名、电影、电视剧  
医疗=>药名、疾病、症状
- 不同任务需要识别不同类别的实体  
安防/舆情=>人名、证件号、手机号、Email  
对话=>各种槽填充（餐馆、航司、地名）
- 标注数据少、迁移困难

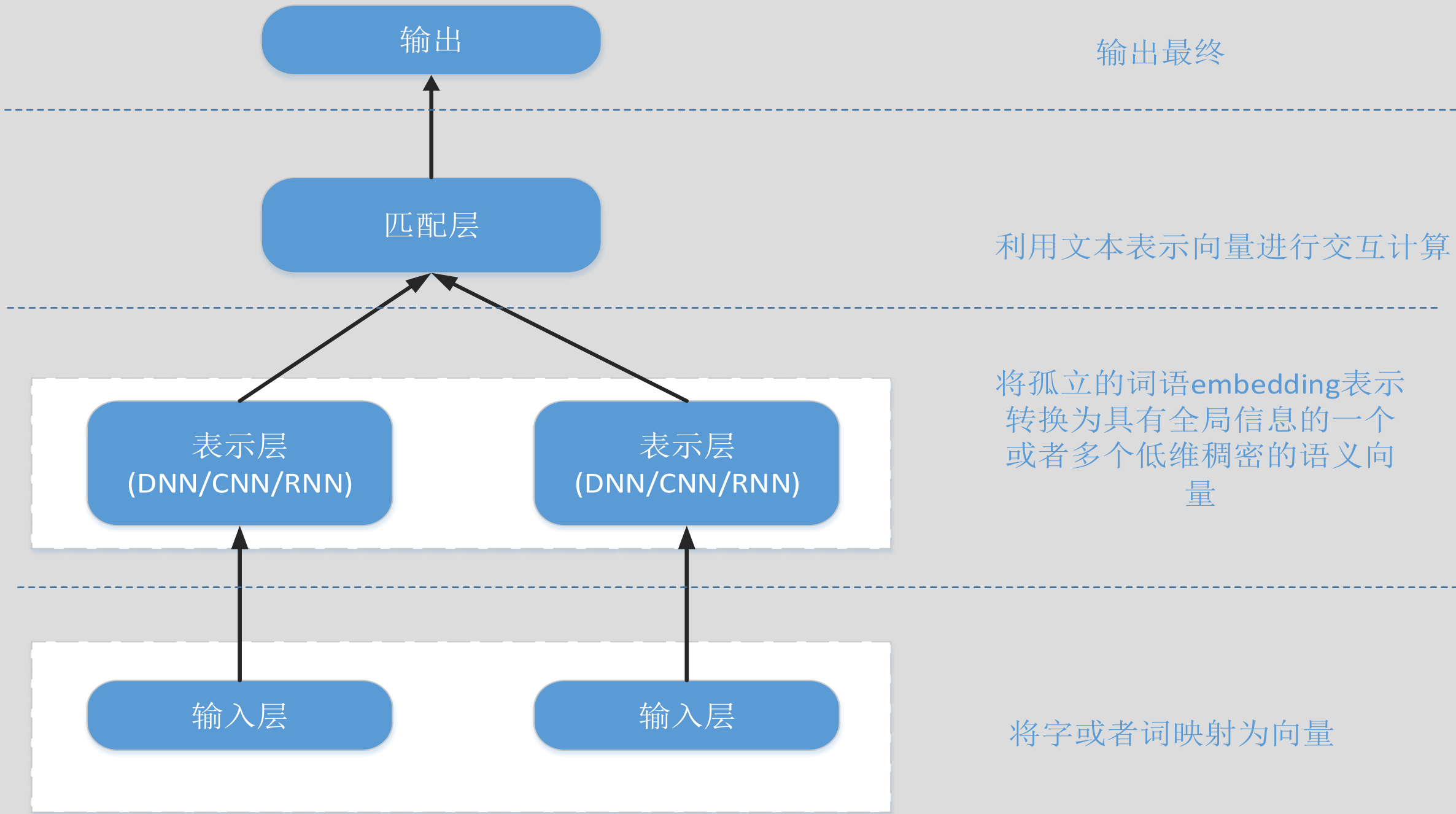
## 实战经验

没有一招鲜吃遍天的万能药（即使是BERT）  
多种方式的混合解决方案（正则、词典、深度学习）



# NLP算法-文本相似度

任务	Texts类型	文本间关系
信息检索	A=query, B=document	相关
自动问答	A=question, B=answer	答案对应问题
机器翻译	A=text, B=text	相关
聊天系统	A=text, B=text	回复和评论
相似文档检测	A=text, B=text	相似
在线广告	A=query, B=Ads	相关
文本蕴含	A=text, B=text	蕴含



## 语义匹配的难点

- 汉语的一词多义和同义词问题
  - 苹果=>水果? iPhone ? Apple inc?
  - 妻子=>老婆, 媳妇儿
- 语言的组合性问题
  - 北京到上海的航班和上海到北京的航班词语虽然完全相同, 但语义完全不同
  - 广州恒大打败了上海上港和上海上港被广州恒大打败了语义完全相同

## 实战经验

基于表示的模型: Siamese Network  
基于交互的模型: tensor、attention  
多粒度多层次模型: 词汇级、短语级  
预训练模型: ELMo、BERT



# NLP算法-意图识别

常见的方法：

- 1、关键词、规则
- 2、ML: svm、lr等
- 3、DL:如cnn、rnn、transformer、bert等。

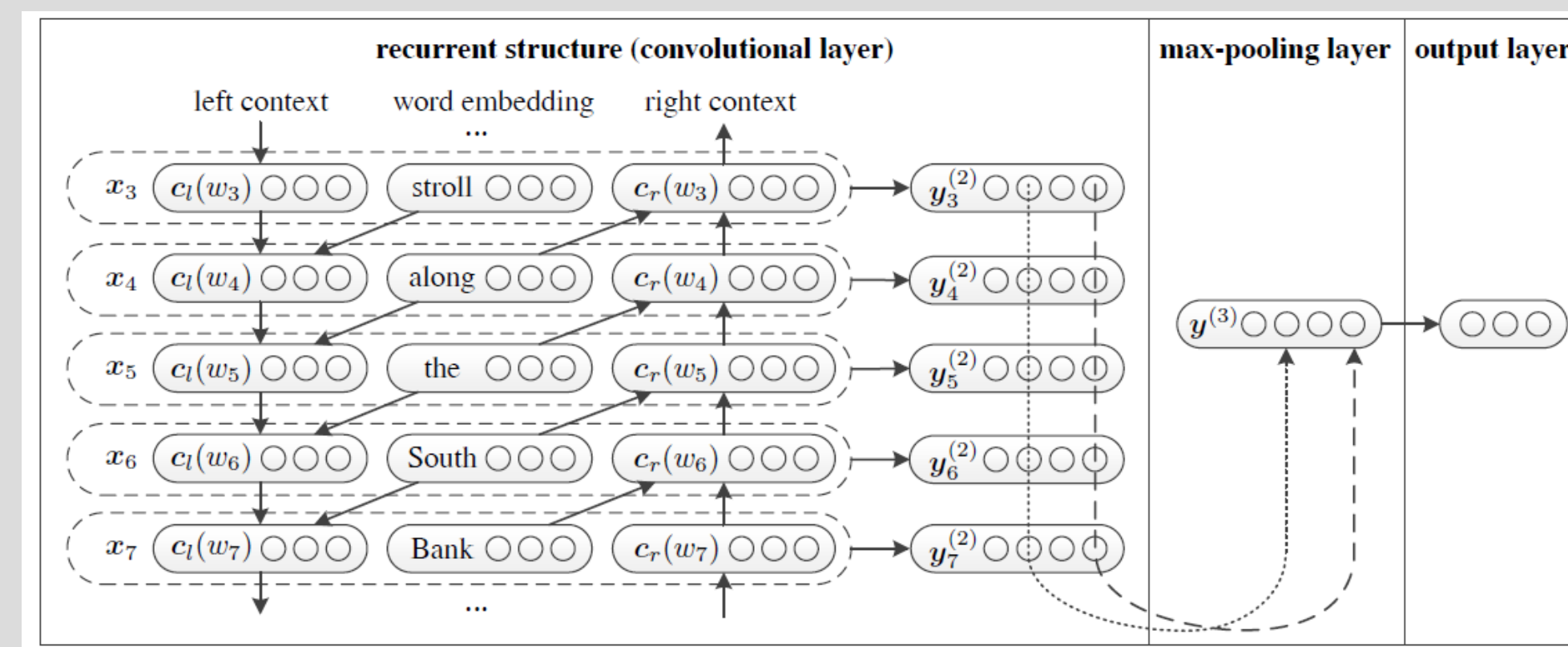
企业场景的难点

- 数据少、质量差、冷启动
- 场景不同类别不同，类别数量决定难度

实战经验

- ✓ 根据不同阶段，设定不同模型。比如：前期冷启动中使用一些正则来匹配，以及无监督模型，比如利用预训练的词向量。
- ✓ 在数据量充足的情况下，采用深度学习模型以及利用知识图谱对词汇进行泛化抽象。

## RCNN



# NLP-对话系统-算法-Entity linking



- 实体链接可以同时解决歧义和多样性问题
  - 对话系统：意图理解、槽填充
  - 搜索系统：查询分析、改写、扩展
- 实体链接的难点
  - 如何识别实体边界
  - 如何链接实体与知识库
  - 如何克服训练样本稀疏
- 企业场景的实战
  - ✓ 通用域=>实体标注问题，百科图谱作为知识库
  - ✓ 垂直域=>实体识别+对齐，依赖领域知识库

苹果就“降速门”致歉：iphone换电池降价390元

苹果 苹果公司  
iPhone iPhone

红富士苹果

红富士 红富士  
苹果 苹果（蔷薇科苹果属果实）

《李白》是李荣浩作词作曲并演唱的歌曲

李白 李白（李荣浩演唱歌曲）  
李荣浩 李荣浩（中国流行男歌手、音乐人、演员）

唐朝诗仙李白

唐朝 唐朝（中国历史朝代）  
诗仙 李白（唐代著名浪漫主义诗人）  
李白 李白（唐代著名浪漫主义诗人）

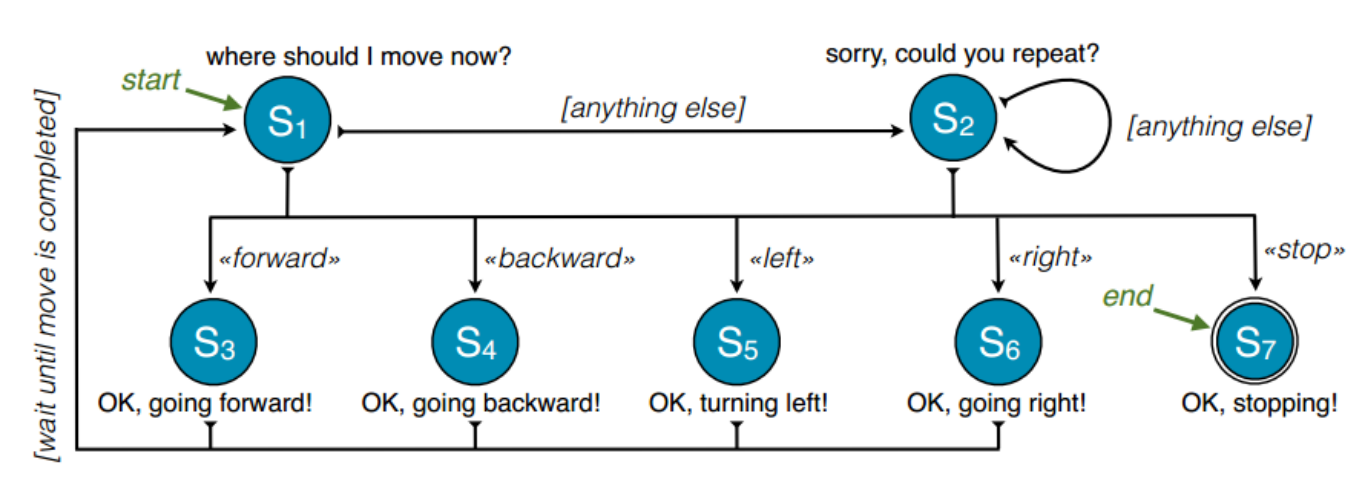


# NLP算法-多轮对话管理

➤ 难点:

- 语音识别/语言理解错误,
- 不确定性大,
- 真实场景状态空间大

状态机



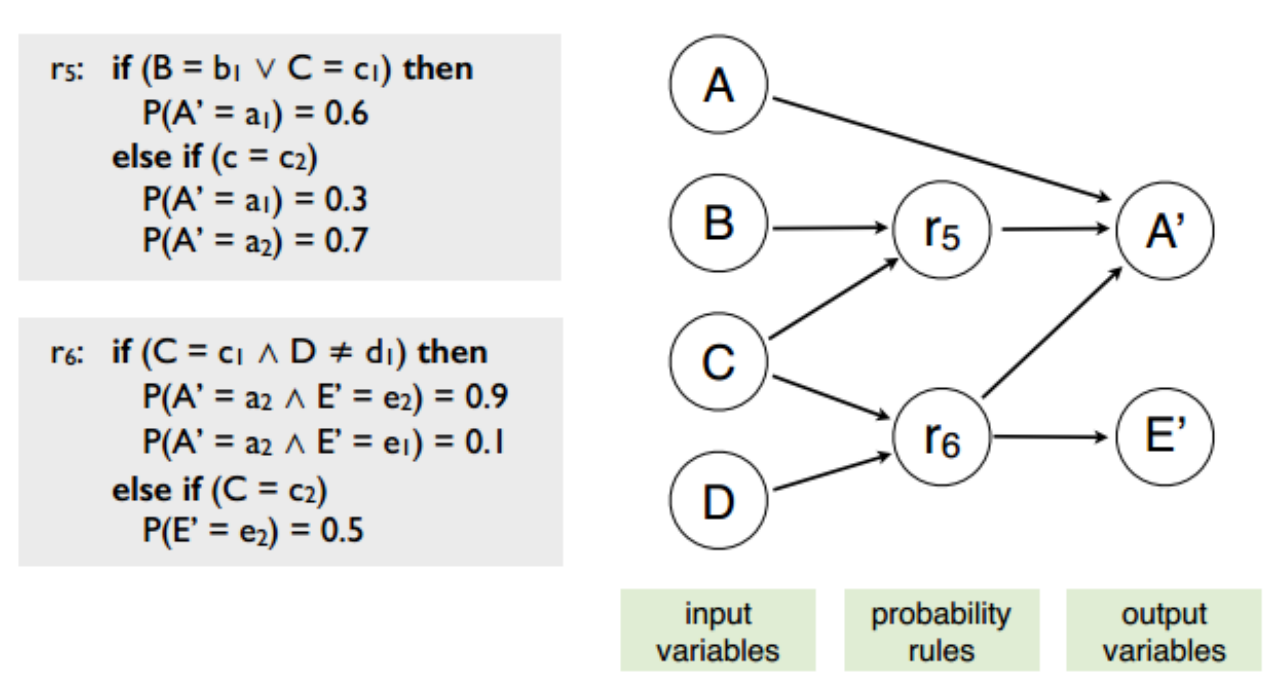
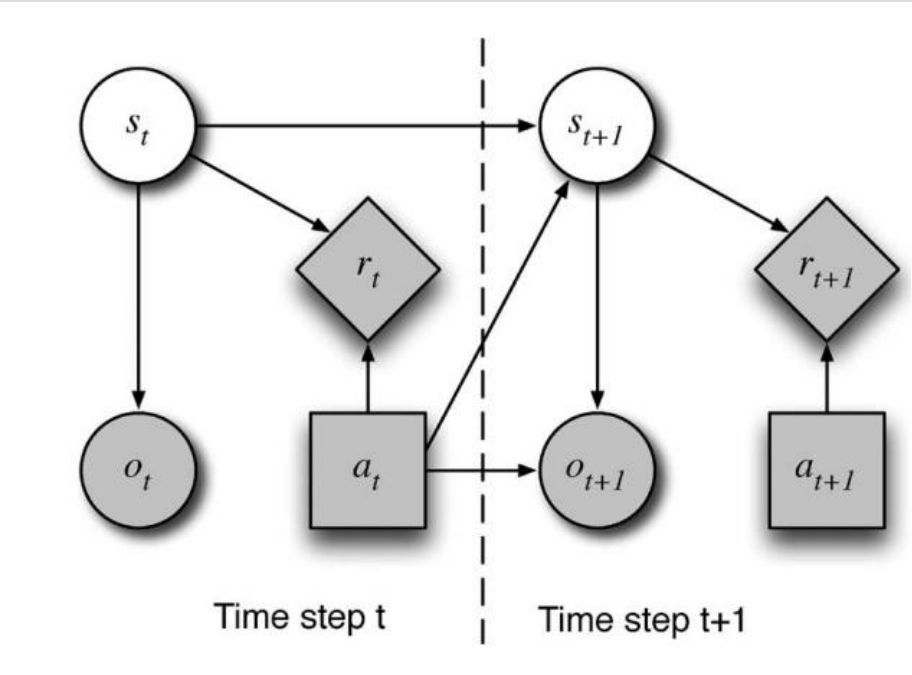
➤ 常用方法:

- 有限状态机
- 槽填充
- Pomdp、prob-rule-based
- 深度学习端到端

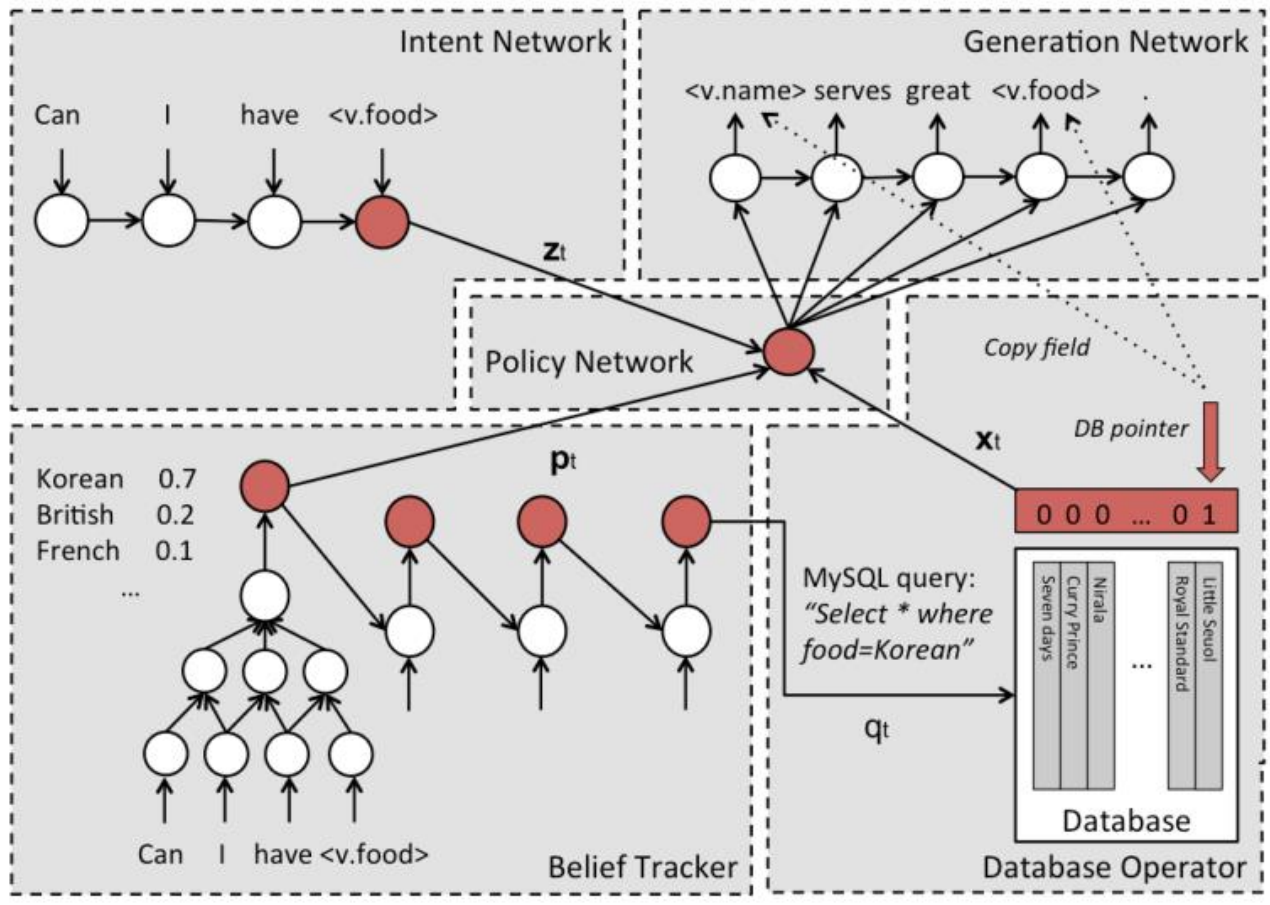
槽填充框架

!source	-->	prompt_source
!destination	-->	prompt_destination
!airline	-->	prompt_airline
!date	-->	prompt_date
nprompts > 1	-->	mark_multiple
nprompts = 0	-->	retrieve_flights

POMDP



概率规则



深度端到端

IT服务中心  
IT SERVICE CENTER

专业、热忱、有温度、值得信赖！

中文 | English

桌面云登不上去了



桌面云登录不上如何处理

您好，桌面云登录不上时，请参考以下方法解决：

方法/步骤

展开

其他相关问题

桌面云无法进入自主维护台的解决方法

如何使用自助维护台修复桌面云连接问题

如何申请桌面云


👍 满意

👎 不满意

请描述您的问题，回车或点击发送按钮提交问题

发送

奥迪a6l和q5对比怎么样





价格上来看的话，奥迪A6L在40.28万元到69.48万元之间，有7款车型，Q5有8款车型， 价格从39.64万元到66.8万元不等。奥迪A6L属于中大型车，Q5属于中型SUV，两者属于不同级别的车，空间，动力都有较大的差别。 您想比较哪两款车型啊？

车型	厂商指导价(万元)
奥迪A6L 2018款 30周年年型 TFSI 进取型	40.28
奥迪A6L 2018款 30周年年型 35 TFSI 时尚型	41.78
奥迪A6L 2018款 30周年年型 35 TFSI 运动型	44.5
奥迪A6L 2018款 30周年年型 30 FSI 风尚型	46.38
奥迪A6L 2018款 30周年年型 45 TFSI quattro 运动型	50.66
奥迪A6L 2018款 30周年年型 50 TFSI quattro 尊享型	69.48

请输入要查询的内容

发送



# 快速搭建一个智能客服机器人



## ➤ 开通机器人

问答机器人名称/ID	状态	操作
IT问答 f2bcef3b-9ad2-40e0-9a76-02a8802812a6	创建中	机器人管理 删除
qabot-lk 24e554c1-25e7-450e-a61e-9b4663d886c7	可用	机器人管理 删除
qa_test 33637ec8-6f3f-4fac-96a8-02f0192ef638	可用	机器人管理 删除

## ➤ 领域知识

新建

删除

导入

发布词典

请输入标准词

Q

C

<input type="checkbox"/> 标准词	同义词	更新时间	操作
<input type="checkbox"/> 粘贴	<div>paste</div> <div>X</div> <div>+</div>	2019-04-22 16:45:12	<a href="#">删除</a>
<input type="checkbox"/> 测试	<div>测试test</div> <div>X</div> <div>+</div>	2019-04-25 15:19:35	<a href="#">删除</a>
<input type="checkbox"/> 复制	<div>copy</div> <div>X</div> <div>+</div>	2019-02-18 17:20:43	<a href="#">删除</a>
<input type="checkbox"/> 桌面云	<div>云桌面</div> <div>X</div> <div>+</div>	2019-04-25 15:19:53	<a href="#">删除</a>
<input type="checkbox"/> huawei	<div>华为</div> <div>X</div> <div>+</div>	2019-04-25 15:19:57	<a href="#">删除</a>
<input type="checkbox"/> 明白	<div>清楚</div> <div>X</div> <div>+</div>	2019-04-25 15:20:13	<a href="#">删除</a>
<input type="checkbox"/> last	<div>最后</div> <div>X</div> <div>最末</div> <div>+</div>	2019-02-18 17:20:29	<a href="#">删除</a>
<input type="checkbox"/> whisper	<div>轻语</div> <div>X</div> <div>+</div>	2019-04-25 15:20:19	<a href="#">删除</a>

## ➤ 新建或导入语料

知识库

未解决问题

问答日志

问题诊断

请输入问题类别名称

test0422

机场安检

Email

Adobe

Office

办公电脑

金山词霸

新建

导入

删除

请输入问题关键字

<input type="checkbox"/> 标准问	答案	问题类别	更新时间	操作
<input type="checkbox"/> 桌面云怎么打开	福娃uyOJHVDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD	test0422	2019-04-22 16:46:54	<a href="#">编辑</a> <a href="#">删除</a>
<input type="checkbox"/> Word2013中将阿拉伯数字转换为中文数字	已为您找到解决方案，请点击：<a class="aUriClass" tar...	Office	2019-04-22 16:07:08	<a href="#">编辑</a> <a href="#">删除</a>
<input type="checkbox"/> Word2013中插入幕函数的输入方法	已为您找到解决方案，请点击：<a class="aUriClass" tar...	Office	2019-04-22 16:07:08	<a href="#">编辑</a> <a href="#">删除</a>
<input type="checkbox"/> Word2013中插入的表格如何设置标题行重复	已为您找到解决方案，请点击：<a class="aUriClass" tar...	Office	2019-04-22 16:07:08	<a href="#">编辑</a> <a href="#">删除</a>
<input type="checkbox"/> Word2013中水平或垂直滚动条消失的解决方法	已为您找到解决方案，请点击：<a class="aUriClass" tar...	Office	2019-04-22 16:07:08	<a href="#">编辑</a> <a href="#">删除</a>
<input type="checkbox"/> Word2013中删除文档中空白的解决方法	已为您找到解决方案，请点击：<a class="aUriClass" tar...	Office	2019-04-22 16:07:08	<a href="#">编辑</a> <a href="#">删除</a>
<input type="checkbox"/> Word2013中英文单词可以换行的设置方法	已为您找到解决方案，请点击：<a class="aUriClass" tar...	Office	2019-04-22 16:07:08	<a href="#">编辑</a> <a href="#">删除</a>
<input type="checkbox"/> Word2013中表格和文字之间的间隔无法调整的解决方法	已为您找到解决方案，请点击：<a class="aUriClass" tar...	Office	2019-04-22 16:07:08	<a href="#">编辑</a> <a href="#">删除</a>
<input type="checkbox"/> Word2013中设置表格跨页显示的方法	已为您找到解决方案，请点击：<a class="aUriClass" tar...	Office	2019-04-22 16:07:08	<a href="#">编辑</a> <a href="#">删除</a>
<input type="checkbox"/> Word2013中输入第一个字母大小自动改变的解决方法	已为您找到解决方案，请点击：<a class="aUriClass" tar...	Office	2019-04-22 16:07:08	<a href="#">编辑</a> <a href="#">删除</a>

10

总条数: 1,593

< 1 2 3 4 5 ... 160 >

## ➤ 模型管理、训练

线上版本

p@1

35.42%

p@3

68.75%

p@5

91.67%

模型 ID：5fa3d799-0ea0-4e06-826f-394b1e974201

调整调值

下线

历史版本

新建

您还可以创建7个模型。

模型 ID

创建时间

p@1

p@3

p@5

操作

213dede8-7628-41b6-bdc9-3f7e650d360f

2019-04-22 16:46:01

35.42%

68.75%

91.67%

调整调值

上线

删除

5fa3d799-0ea0-4e06-826f-394b1e974201

2019-04-22 16:45:50

35.42%

68.75%

91.67%

调整调值

上线

删除

d34309cb-281b-43ea-88fa-c39dd5596040

2019-04-22 16:08:21

35.42%

68.75%

91.67%

调整调值

上线

删除



# 快速搭建一个qabot：问题闭环



## ➤ 未解决问题学习

知识库🔗 未解决问题🔗 问答日志🔗 问题诊断🔗						
<div><div></div><div>全部状态</div><div>全部类型</div><div>请输入未解决问题</div></div>						
未解决问题	重复次数	类型	创建时间	更新时间	状态	操作
时代	13	无答案	2019-02-03 14:46:59	2019-02-03 14:47:23	🔴 未处理	<a href="#">添加新问答语料</a> <a href="#">更多</a>
知足	11	无答案	2019-02-11 17:25:03	2019-02-11 17:25:15	🔴 未处理	<a href="#">添加新问答语料</a> <a href="#">更多</a>
当兵	3	无答案	2019-02-03 15:15:15	2019-02-03 15:22:21	🔴 未处理	<a href="#">添加新问答语料</a> <a href="#">更多</a>
nihao	3	无答案	2019-02-14 16:45:34	2019-02-14 17:01:59	🔴 未处理	<a href="#">添加新问答语料</a> <a href="#">更多</a>
玩机	2	无答案	2019-01-31 17:17:50	2019-01-31 17:17:52	🔴 未处理	<a href="#">添加新问答语料</a> <a href="#">更多</a>
充实自我	1	无答案	2019-02-11 17:23:26	2019-02-11 17:23:26	🔴 未处理	<a href="#">添加新问答语料</a> <a href="#">更多</a>
苹果	1	无答案	2019-02-25 17:18:58	2019-02-25 17:18:58	🔴 未处理	<a href="#">添加新问答语料</a> <a href="#">更多</a>
我好开心	1	无答案	2019-04-25 15:25:35	2019-04-25 15:25:35	🔴 未处理	<a href="#">添加新问答语料</a> <a href="#">更多</a>

## ➤ 问答日志

知识库🔗

未解决问题🔗

问答日志🔗

问题诊断🔗

—

📅

请输入用户问题

用户问题	创建时间 ⬆	答案	满意度 ⬆	问题来源	转人工
一加一	2019-04-25 15:25:58	这个太简单了，是2.000	... 满意	用户输入	否
1+1	2019-04-25 15:25:52	No answer!	满意	用户输入	否
1+！	2019-04-25 15:25:49	No answer!	满意	用户输入	否
我好开心	2019-04-25 15:25:35	No answer!	满意	用户输入	否
你好	2019-04-25 15:25:20	您好，请问您有什么需要	... 满意	用户输入	否
你叫什么	2019-03-28 12:40:45	我叫%s	... 满意	用户输入	否
苹果	2019-02-25 17:18:58	No answer!	满意	用户输入	否
你好	2019-02-14 17:02:02	你好啊！	... 满意	用户输入	否

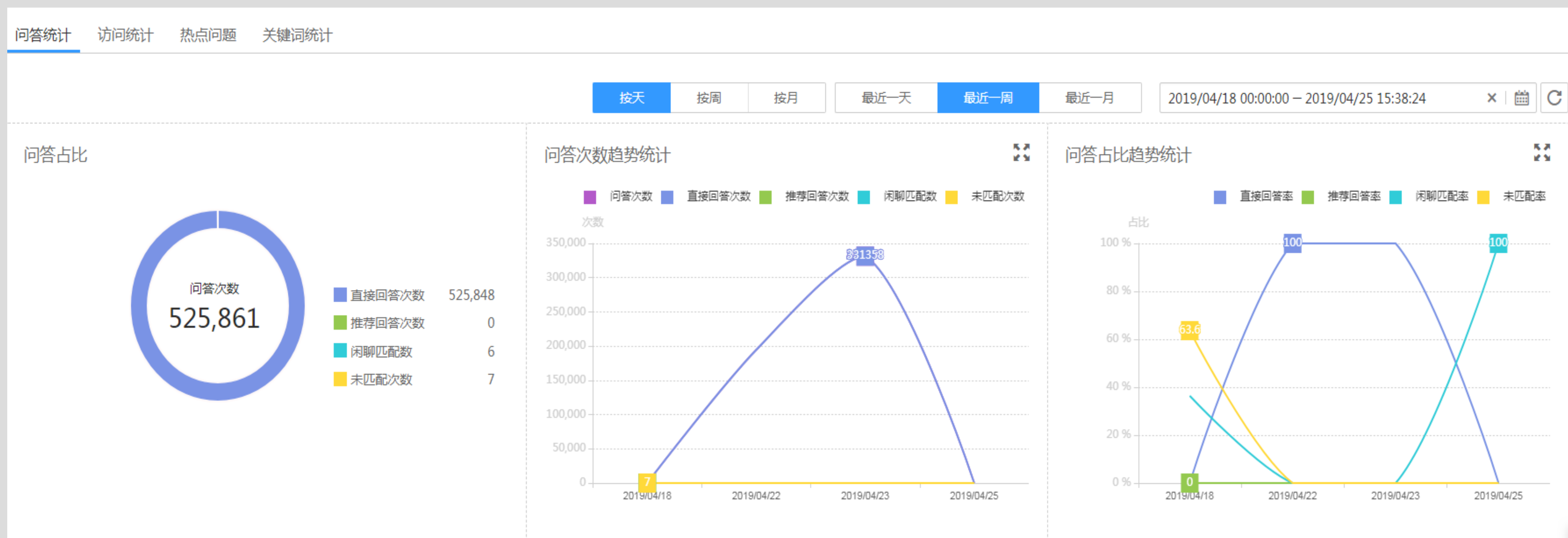
## ➤ 问题诊断

知识库🔗 未解决问题🔗 问答日志🔗 问题诊断🔗	
用户问：	<div>word目录插入</div> <div>问题类别：</div> <div><div>✕</div><div>+</div></div> <div>Debug</div>
Debug	
分词结果：	<div>word，目录，插入</div>
短语匹配：	<div>{}</div>
初筛结果：	<div><pre>{   "qa_pairs": [     {       "score": 11.091,       "qa_pair_id": "8b98edf4-bd2c-46a2-9fc2-3cc5929142b8",       "st_question": "Word中插入目录无法显示页码？",       "ex_questions": [         {           "content": "Word中插入目录无法显示页码？"         }       ]     }   ],   1. }</pre></div>
重排序结果：	<div><pre>{   "qa_pairs": [     {       "score": 0.446,       "topExquesionTerms": "， word， 中， 插入， 目录， 无法， 显示， 页码",       "qa_pair_id": "8b98edf4-bd2c-46a2-9fc2-3cc5929142b8",       "st_question": "Word中插入目录无法显示页码？",       "ex_questions": [         {           "content": "Word中插入目录无法显示页码？"         }       ]     }   ],   1. }</pre></div>

# 快速搭建一个qabot：运营统计



## ➤ 运营统计



# 快速搭建一个qabot：接口调用



## ➤ Token认证

```
{
  "auth": {
    "identity": {
      "methods": [
        "password"
      ],
      "password": {
        "user": {
          "name": "username",
          "password": "password",
          "domain": {
            "name": "domainname"
          }
        }
      }
    },
    "scope": {
      "project": {
        "name": "aaa" //假设区域名称是"aaa"
      }
    }
  }
}
```

## ➤ 搜索提示接口

### • 请求示例

```
POST https://cbs-ext.cn-north-1.myhuaweicloud.com/v1/41551a23c2be484cb0070631e323ee52/qabots/b36a01da-7fc2-43f8-8990-218e57a
Request Header:
Content-Type: application/json
X-Auth-Token: MIINRwYJKoZIhvcNAQcCoIINODCCDTQCAQExDTALBglghkgBZQMEAgEwgguVBgkqhkiG...
Request Body:
{
  "question": "桌面云",
  "top": 2
}
```

### • 成功响应示例

```
{
  "questions": [
    "桌面云打不开",
    "桌面云无法登陆"
  ]
}
```



# 快速搭建一个qabot：接口调用



## ➤ 问答请求接口

### • 请求示例。

```
POST https://cbs.cn-north-1.myhuaweicloud.com/v1/1604d8170cd84440a81281e5a3a04f8b/qabots/303a0a00-c88a-43e3-aa2f-d5b8b9832b0

Request Header:

Content-Type: application/json

X-Auth-Token: MIINRwYJKoZIhvcNAQcCoIINODCCDTQCAQEExDTALBgIghkgBZQMEAgEwgggVBgkqhkiG...

Request Body:
{
  "user_id": "z00123456",
  "session_id": "xxxxxxxxxxxxxxxx",
  "operate_type": 3,
  "question": "桌面云无法登陆",
  "extends": {
    "domains": [
      "桌面云",
      "公有云"
    ],
    "top": 3
  }
}
```

### • 响应示例

示例：桌面云打不开。

#### • 成功响应示例

```
{
  "request_id": "303a0a00-c88a",
  "question": "桌面云打不开",
  "answers": [
    {
      "qa_pair_id": "123",
      "st_question": "桌面云打开不了",
      "answer": "用户账号不对",
      "link": " http://www.hwcloud.com/help/123.html ",
      "score": 0.952,
      "domain": "桌面云",
      "relevance_details": [
        {
          "st_question": "桌面云登录不了",
          "link": " http://www.hwcloud.com/help/124.html "
        },
        {
          "st_question": "桌面云无法进入",
          "link": " http://www.hwcloud.com/help/125.html "
        }
      ]
    }
  ]
}
```

# 应用案例-售后智能客服



# 应用案例-智能家居





# 应用案例-智能呼叫中心





# 应用案例-话务机器人



# EI体验空间



扫描小程序  
为讲师及分享内容评分





# TGO 鲲鹏会

## 汇聚全球科技领导者的高端社群

🏢 全球12大城市

👤 850+ 高端科技领导者

使命

Mission

为社会输送更多优秀的  
科技领导者

愿景

Vision

构建全球领先的有技术背景  
优秀人才的学习成长平台



扫描二维码，了解更多内容





# AI与云原生实践

2019.5.6 北京国际会议中心

