

# 人工智能之NLP

## 意图识别

主讲人: GerryLiu

# 课程要求

- 课上课下“九字”真言
  - 认真听，善摘录，勤思考
  - 多温故，乐实践，再发散
- 四不原则
  - 不懒散惰性，不迟到早退
  - 不请假旷课，不拖延作业
- 一点注意事项
  - 违反“四不原则”，不推荐就业

## 课程内容

- 意图识别应用领域
- 意图识别的难点
- 意图识别主要方法
- 规则模板意图识别
- 深度学习意图识别

## 意图识别的难点

- 用户输入不规范，输入方式多样化，甚至非标准的自然语言
- 表意表现出多意图
- 意图强度，表述不同表现出不同的需求强度
- 意图存在时效性变化，就是随着时间的推移相同问题的意图会发生变化
- 没有固定的评估的标准

## 意图识别应用领域

- 搜索引擎
- 对话系统
- 智能物联网
- 机器人

# 意图识别主要方法

- 意图分类

- (1) 导航型

- 将用户导航到相应的领域或相应的流程中

- (2) 信息型

- 提供给用户想知道的信息

- (3) 事务型

- 给用户提流程中的各个实现环节

## 意图识别主要方法

- 词表穷举法

通过词表直接匹配的方式来获取查询意图，同时，也可以加入比较简单并且查询模式较为集中的类别。

查询词：德国[addr] 爱他美[brand] 奶粉[product] 三段[attr]

查询模式：[brand]+[product];[product]+[attr];[brand]+[product]+[attr]

方式实现较为简单，能够较准确的解决高频词，对长尾query无能为力。

缺点：需要人工参与较多，很难自动化实现。

实用指数：1★

## 意图识别主要方法

- 规则解析法

适用于查询非常符合规则类别，通过规则解析的方式来获取查询的意图。

北京到上海今天的机票价格，可以转换为[地点]到[地点][日期][汽车票/机票/火车票]。

1吨等于多少公斤，可以转换为[数字][计量单位]等于[数字][计量单位]。

对规则性较强的query有较好的识别精度，能够较好的提取准确信息。

缺点：需要人工参与较多，很难自动化实现。

实用指数：4★

场景：搜狗通用搜索、小i智能客服、机器人、智能物联网



## 意图识别主要方法

- 机器学习和深度学习方法

意图识别可以看做是一个分类问题，针对于垂直产品的特点，定义不同的查询意图类别。

对于用户输入的query，根据统计分类模型计算出每一个意图的概率，最终给出查询的意图。

缺点：主要是数据获取和更新较困难，**数据的标注**也需要较准确才能训练出较好地模型。

实用指数：4★

场景：京东JIMI、美团垂直搜索

# 规则模板意图识别

- 规则意图识别的难点

1. 问题不规范

2. 多意图

仙剑奇侠传

游戏? --> 游戏软件? .....

电视剧? --> 电视剧下载? 相关新闻? .....

3. 意图强度

天气很好啊-->弱意图

今天天气怎么样-->强意图

# 规则模板意图识别

- 规则意图识别的一般技术
  - 采用实体-主域-模板的整体框架来判别识别领域
  - 命中主域后，模板采用意图动词(下载、查询等)或意图疑问词(怎么样、为什么等)判别意图
  - 区分弱意图和强意图，针对性提供解决方案；判断实体强度

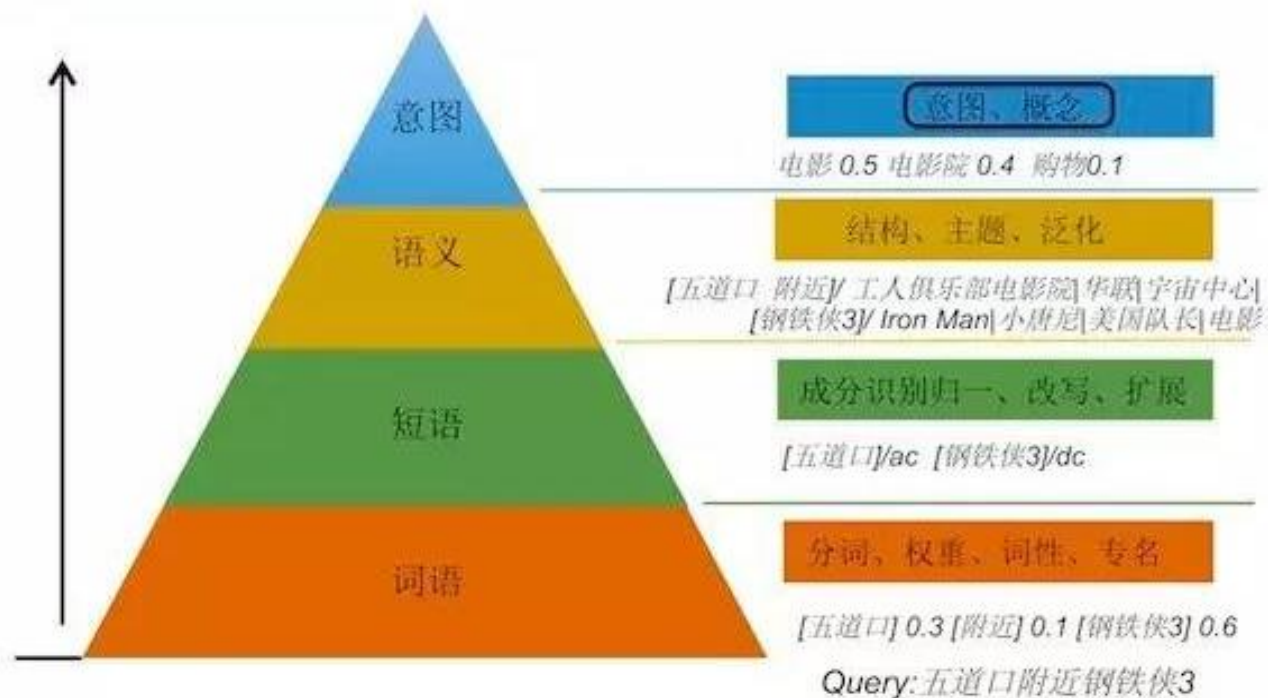
搜索引擎意图识别领域判别表：仙剑奇侠传

意图	强度	实体信息
视频	0.8	(0, 2, 影视剧)
小说	0.3	(0, 2, 小说)
游戏	0.3	(0, 2, 游戏)

# 机器学习意图识别

- 美团意图识别

## 案例1-Query意图识别



# 机器学习意图识别

- 美团意图识别

## 案例1-Query意图识别

场景

- Q:动物园门票 → 意图:旅游
- Q:全聚德望京 → 意图:美食

Query



# 机器学习意图识别

- 美团意图识别

## 案例1-Query意图识别

### 思路

#### 1 训练样本获取

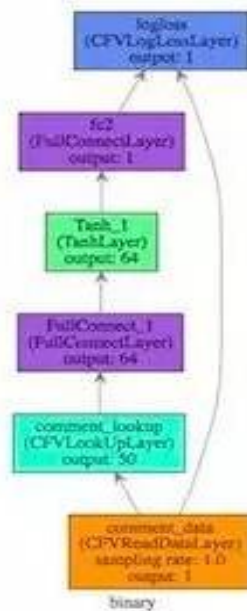
- 自动标注：统计值，主动学习，Bootstrapping,
- 人工标注

#### 2 特征设计特点

- Query语义表示
- Query的信息扩展
- 用户地域：eg 黄鹤楼

#### 3 模型尝试

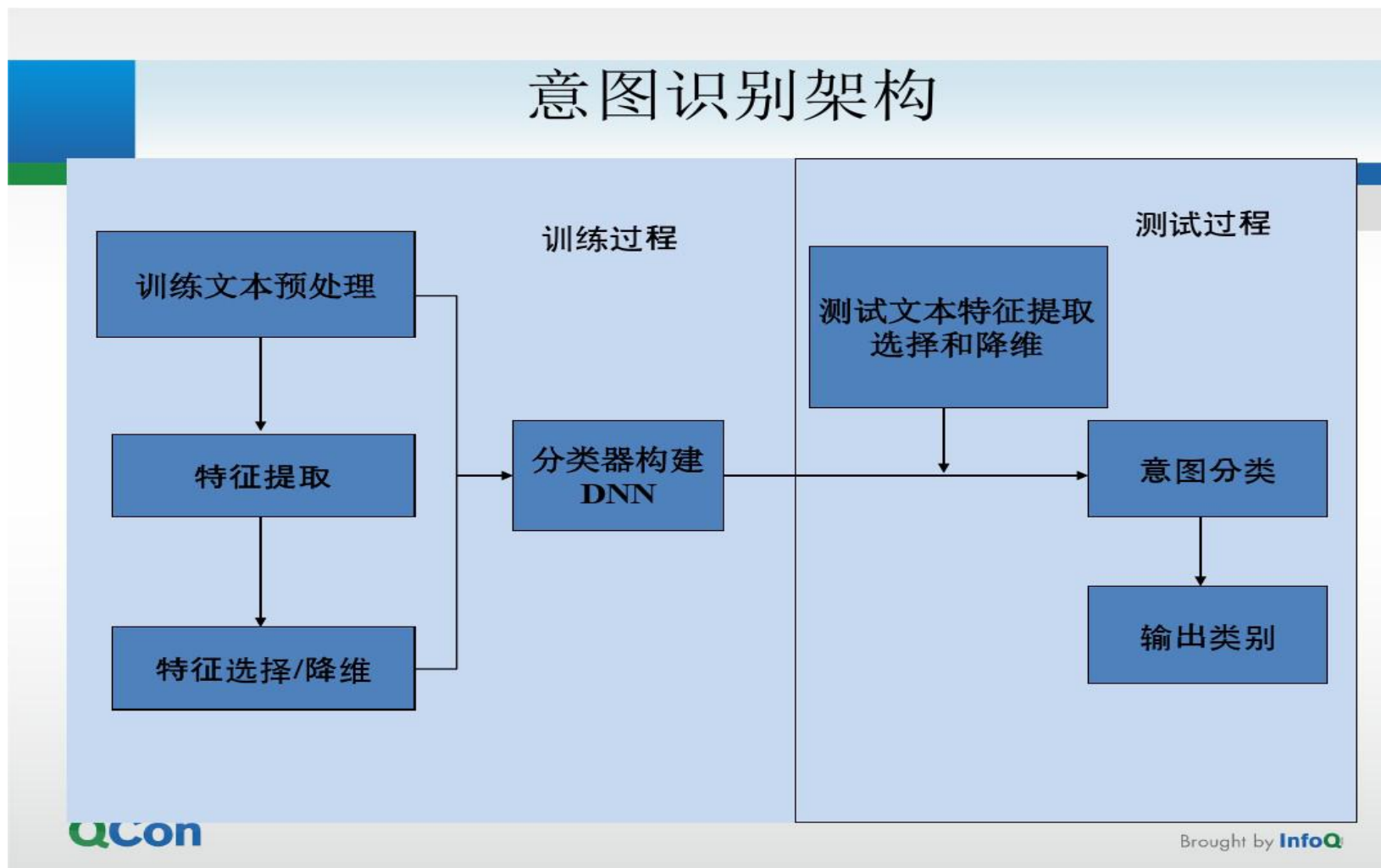
- Xgboost
- DNN on Ginger



DNN on Ginger 文本分类模型

# 深度学习意图识别

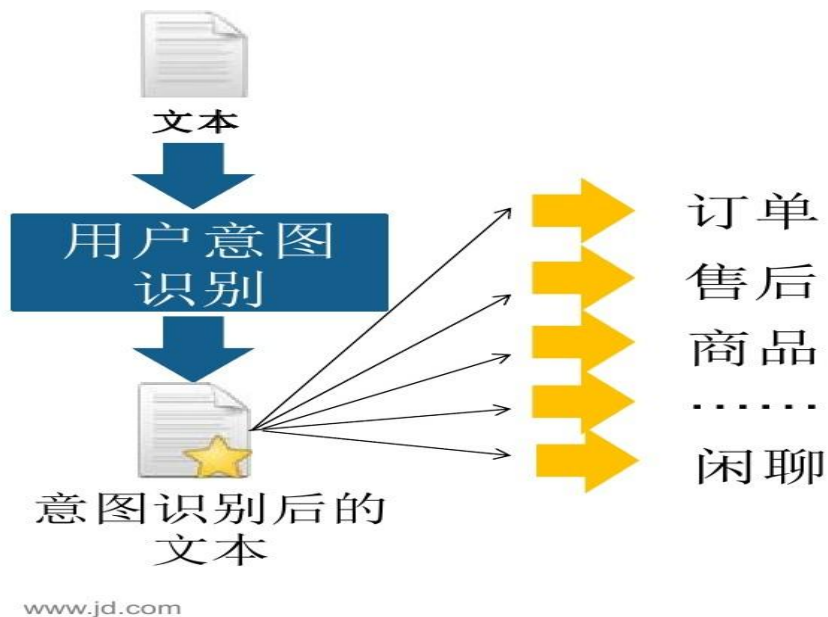
- 京东JIMI意图构架(改进版)



# 深度学习意图识别

- 京东JIMI意图识别

## 用户意图识别



意图识别是JIMI系统里面非常重要的一部分,它首先要通过用户的问话,知道用户想要做什么,只有意图识别对了,才能在相应的类别里面反馈用户的答案

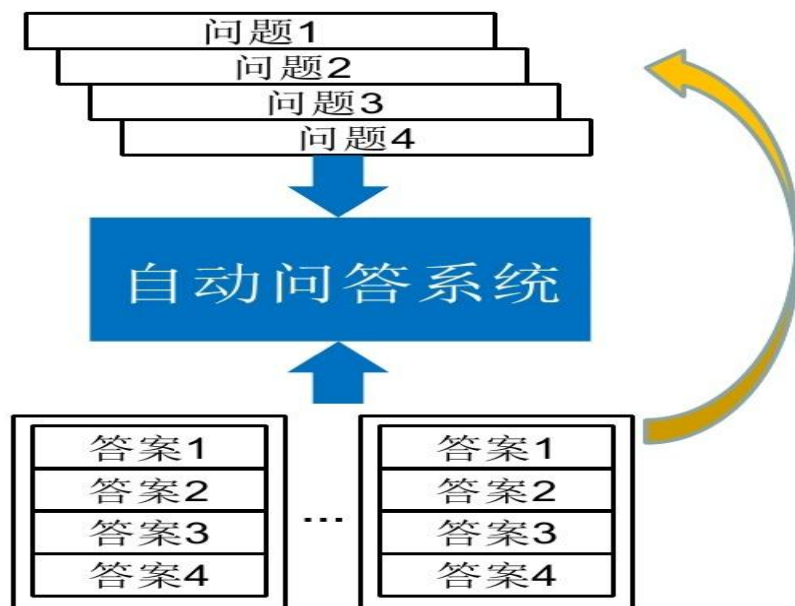




# 深度学习意图识别

- 京东JIMI意图识别

## 自动问答



答案匹配，候选  
答案抽取和排序  
www.jd.com

京东技术 开放日 JD.COM 京东

- ✓ 对用户的问题进行意图识别
- ✓ 确定问题的分类
- ✓ 答案匹配，抽取和排序
- ✓ 给用户反馈最佳答案和建议



## 深度学习意图识别

- 京东JIMI意图识别(baseline)

深度学习CNN模型，可用于求解一个分类问题，将用户的问题映射到一个具体的分类。最终在算法选型上，采用深度学习CNN模型，其中模型参数：

词向量采用100维

每个样本限定30个字以内，超出30截断，不足30补充随机向量

单层CNN网络，第一层卷积核大小3\*50

## 深度学习意图识别

- 京东JIMI意图识别(改进版)
- 各种模型意图识别各种问题分类效果对比

模型	点击率	准确率
Baseline	52.4%	62.4%
LSTM	68.7%	75.5%
LSTM + Attention	71.2%	78.2%

- 寄语：你缺少的只是'打了标记'的样本和机会

