

# 北大AI第三讲：何晓东-京东AI研究院常务副院长、IEEE Fellow：自然语言与多模态交互前言技术

## NLP进展：

- 语言理解/语义槽值提取
- 语言理解/意图分类，2016年提出层次化注意力模型（HAN），以更好的在词、句子、段落、等多个层面来理解语言，判断意图，并通过对神经元激活的可视化给出一定程度的可解释性。
- 语言理解/语义的表征：从自然语言中提取出语义并将其投影到语义空间以帮助搜索、推荐、分类、问答等应用；自然语言的描述，通过深度神经网络逐步抽取语义上的不变形（invariance），生成抽象的语义表征。
- 机器阅读理解（MRC）：机器阅读文本，回答问题；BERT模型在SQuAD封闭数据集上的成绩已经超过人类

## 未来

- 多模态智能：综合文字、语音、图像、知识图谱等信息来获取信息
  - 建立多模态语义空间：联结图像和文字
    - 通过深度结构语义模型（DSSM）把图像和文字表征成语义空间内的向量
    - 在此空间中进行语义相似度计算，生成最匹配图像内容的文字表述。
  - 理解场景和知识，用语言表达（image caption）
    - 一个棒球
    - 一个棒球运动员
    - 一个棒球运动员在扔
    - 一个棒球运动员在扔一个球
  - 图像描述机器人：CaptionBot
  - 智能绘画机器人：AI根据语言描述创作绘画
  - AI+Art：更多（艺术化的）创作
  - 综合图像和语言推理，回答问题：eg.那两把蓝色椅子之间是什么？
  - 视觉-语言多模态导航：结合语言理解和对环境的视觉信息建模，机器人按指令从一个地方走到另一个地方。
- 复杂内容创作：比如人工智能写作（长文章）
  - 创作长文的技术挑战：
    - 从简单输入到创作长文需要大量内容的扩充
    - 长文的生成要可控，能满足组合爆炸式需求。模型需要时组合性的，能与训练的。
    - 现有的端到端的模型不适合长文创作：为短文本生成（如机器翻译）而设计，不能抓住长文的高层语义。度量优化，信用分配，维持一致性，目标函数平衡等要重新设计。
  - 长文创作的前沿探索：比如顶层设计和规划
    - 现有的文本生成模型缺乏“规划”，应先产生粗略的高层主题规划，然后再对主题和子主题展开长文
    - 最近的一些工作：多层增强学习模型及其在主题设计和长文生成中的应用。
  - 创作诗歌（控制）
- 情感智能：不只识别人的情感，还能像人一样表达情感和风格
  - 生成带情感的语言：让AI在语言表达中加入情感，提升用户体验
  - 表达情感和风格：让AI用语言表达浪漫或者幽默的风格——StyleNet

- 多轮人机对话：理解语境、常识、语言，生成逻辑严谨的有情感的对话，服务于人
  - 图灵测试：通过人类和机器之间的自然语言对话来判断机器是否具有智能。
  - 主要的人机对话系统框架：任务型对话系统、问答型对话系统、聊天型对话系统、检索性对话系统
  - 我们将成为有史以来第一代与AI共生的人类，《从Eliza到小冰：社交对话机器人的机遇与挑战》

## AI产业化的下一个方向是什么

- 智能服务产业是新蓝海：传统人类密集型产业，有广阔自动化、智能化空间；随着AI技术、IOT技术等的创新，市场在快速成长。
- 服务型对话
  - 服务：生活、娱乐、消费、客服等为人提供的服务
  - 对话：多模态、大规模开放领域，具有常识和情感、能完成复杂任务的智能交互技术。
- 产业界应用：
  - 京东客服机器人：首个大规模商用情感客服机器人：能够检测用户的情感类型，做出道歉、安抚、祝福的动作，提升用户体验。
  - 京东智能服务产品矩阵：JIMI和AlphaSales
  - 京东智能IoT
  - 京东智能市政服务
- 人机融合、多模态智能服务的产业时代
  - 服务即对话：多模态、大规模开放领域、具有常识和情感、能完成复杂任务的对话系统是推动下一代智能产业的核心技术。

分级	目标
低级智能对话	对用户简单意图进行识别并给出预设答案
初级智能对话	能识别复杂意图，联系上下文给出回答
中级智能对话	根据用户问题及情绪完成个性化多轮对话，协助用户完成目标
高级智能对话	能对多模态信息进行推理，自主判断，并组织语言与用户沟通，具备自我学习能力
通用智能对话	能基于一切信息开展自我学习，自我适应，及自我创新。在复杂问题领域达到人类水平。

## 提问环节

- 关于OpenAI的GPT模型暴力解决问题的看法
  - 算法是解决问题的核心，但是光有算法是不够的。GPT2是算法和产业界结合的一个例子，有大量高质量的数据和算力。对科研机构不算一个坏事。算法是灵魂，数据和算力是物质基础
- NLP领域的问题是否可以认为比图像领域的问题要难解决，所有在进度上有所落后？
  - NLP领域的问题是一种认知领域的问题，比图像领域的感知问题要复杂一些，所以图像领域的问题能够比较清晰的被定义，相对而言也会推进的更加深入。
- 如何判断机器是否真正理解了人类的问题？

- 这是一个哲学问题，科学家的作用就是将哲学问题转化成科学问题，可以通过定义一些测试任务来进行一定程度上的判断。比如小冰提出了聊天轮数的metric，作为判断聊天机器人是否能像人类一样进行聊天。
- 智能音响等IoT产品的发展趋势？
  - 每一次交互的革命都能带来一个万亿级别的产业，智能IoT就具有这样的潜力
- 如何看待当前CV行业如火如荼，NLP行业相对比较平静的现象？
  - 当问题已经定义的很清楚的时候，机会相对来说就小了很多。智能服务产业的未来可能要比CV的智能安防产业还要大。