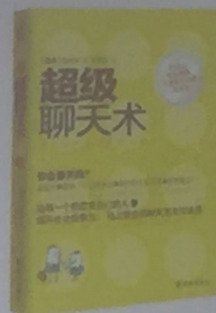


■ 对话系统的问题定义

- 什么构成一个对话流程
 - 给定一个用户输入信息
 - 系统返回一个响应信息
 - 返回信息时的“思考”过程：算法
 - 上下文信息
 - 知识库信息（世界知识）
 - 语义逻辑信息
- 为什么突然走红
 - Timing
 - 大数据驱动 -> 大数据get
- 特别的挑战
 - 源自人类认知直觉的剖析
 - 相关性，趣味性，信息量



■ 对话系统分类

- 按领域分类
 - 开放领域
 - 垂直领域：医疗，金融，法律，等等
- 按回复产生方式分类
 - 检索式对话系统
 - 目前业界主流
 - 生成式对话系统
 - 下一代潮流：深度学习加持
 - 结合式对话系统：新的探索
- 按场景分类
 - 单轮对话/多轮对话
- 按方式分类
 - 被动对话/主动对话

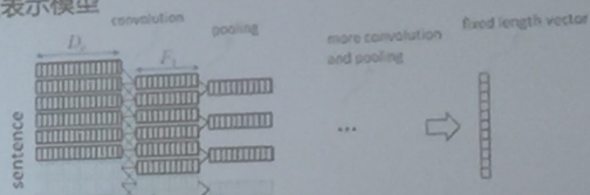
检索式对话背景

- 匹配模型
- 两个特征表示向量的内积

$$\text{match}(x, y) = \langle \Phi_y(x), \Phi_x(y) \rangle > \mathcal{H}$$

$$\text{match}(x, y) = x^T A y = \sum_{m=1}^{D_x} \sum_{n=1}^{D_y} A_{nm} x_m y_n$$

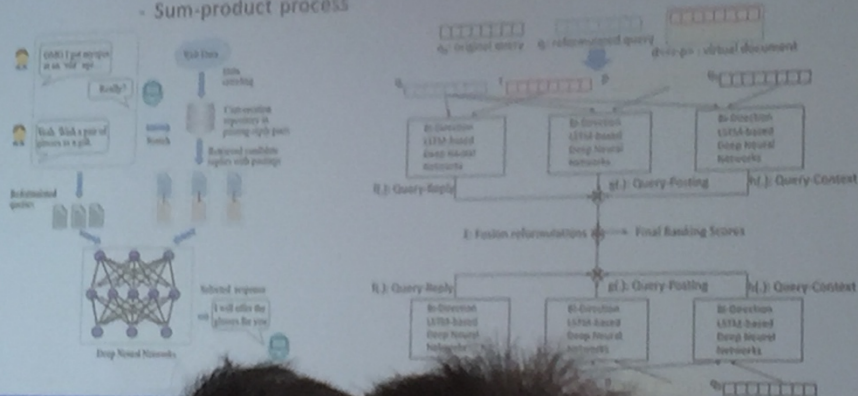
- One-hot表示模型
- 卷积表示模型



检索式对话多轮对话

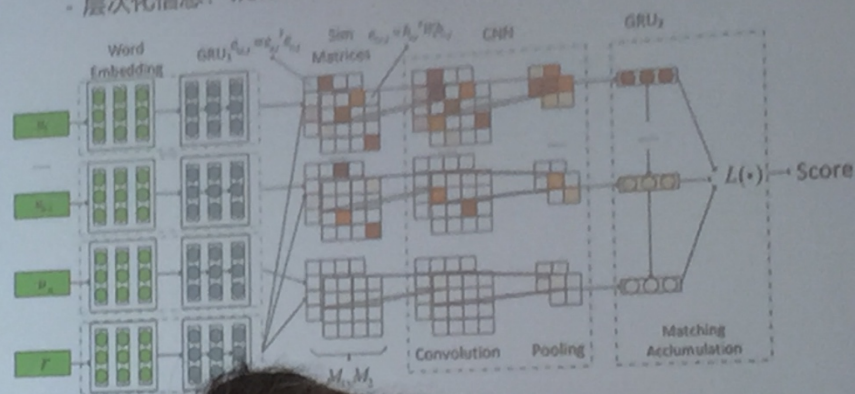
- 百度&北京大学 [SIGIR 2016]
- 改写&整合模式
- Sum-product process

$$F(q, r) = \sum_{p \in \mathcal{P}} (h(q, p) \sum_v (f(v, r) s(v, p)))$$



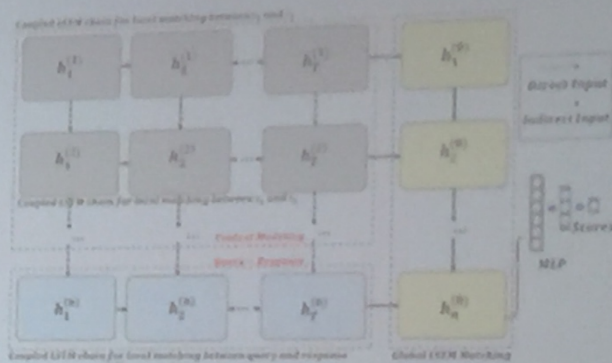
检索式对话多轮对话

- 微软小冰 [ACL 2017]
- 序列化建模: 匹配序列
- 层次化信息: word-level, segment-level, 多通道



检索式对话多轮对话

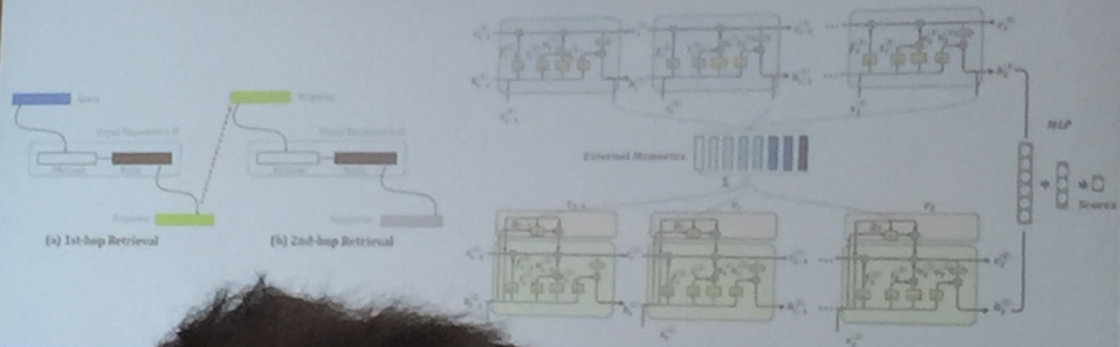
- 北京大学 [KDD 2018]
- 序列化建模: 匹配序列
- 层次化信息: 外部记忆单元+Global/Local-level LSTM序列



检索式对话对话提示

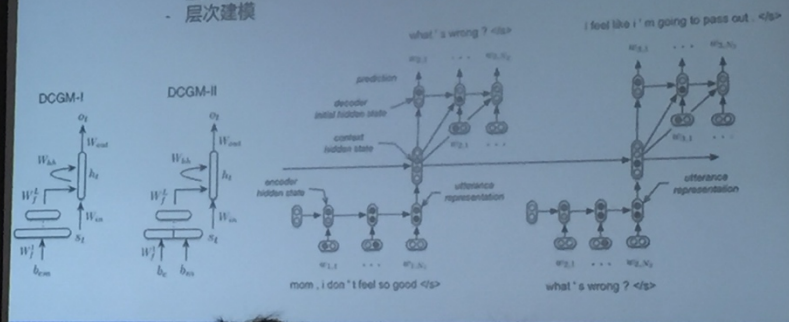
- 北京大学 [SIGIR 2017]

- 查询提示 query suggestion in IR: 很重要
- 是否可以改变一个 well-defined 对话模式
 - 更为 proactive
- 带来潜在的感兴趣内容可聊 (用户持开放态度)



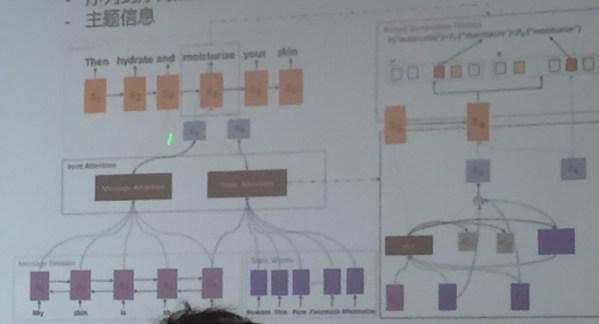
生成式对话多轮对话

- 蒙特利尔大学 [NAACL 2015; AAAI 2016]
- 上下文建模
- 层次建模



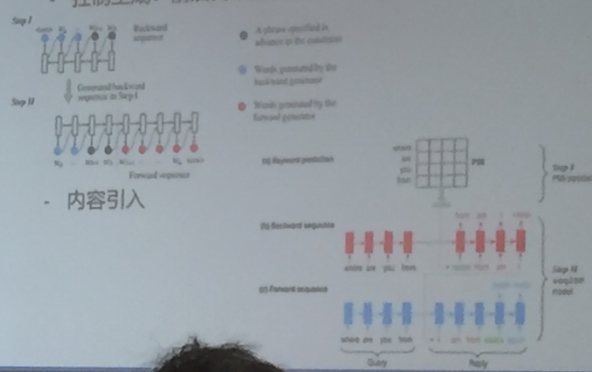
生成式对话话题信息

- 微软小冰 [AAAI 2017]
- 序列到序列建模
- 主题信息



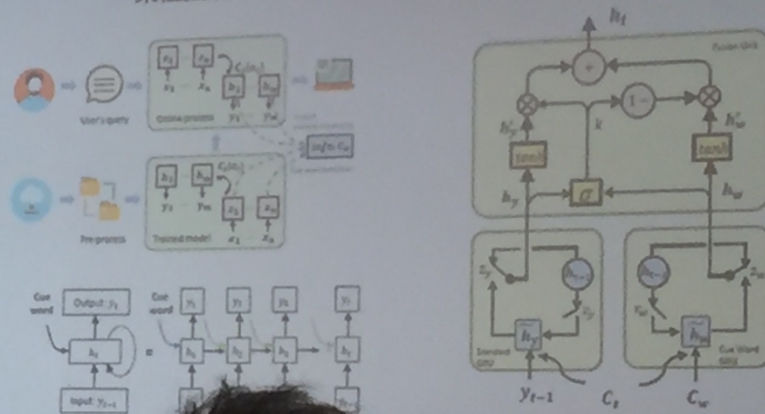
生成式对话主动对话

- 北京大学 [COLING 2016]
- 控制生成：前后向语言模型



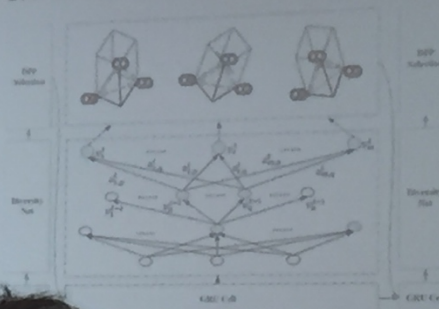
生成式对话主动对话

- 北京大学 [EMNLP 2017]
- 对话模式
- 柔性控制



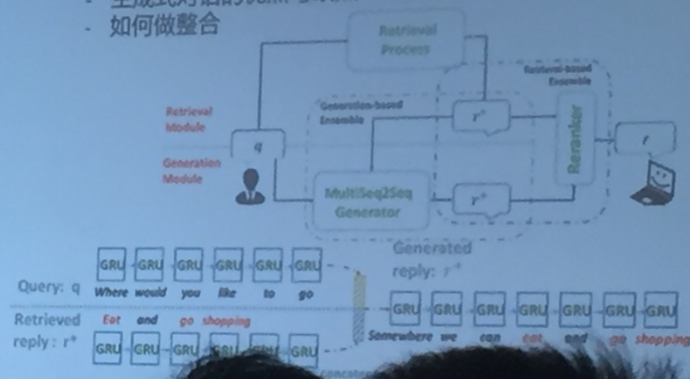
生成式对话多样性

- 北京大学 [AAAI 2018]
- 行列式点过程 (DPP)
- 两种模型:
 - DPP Re-Ranker vs DPP Decoder



结合式对话

- 北京大学 [IJCAI 2018]
- 检索式对话的优点与缺点
- 生成式对话的优点与缺点
- 如何做整合



对话评测

- 自动化评价指标
 - 机器翻译 BLEU, METEOR, NIST
 - 摘要 ROUGE, Pyramid
 - 对话: ?
- 人工指标: pair-wise vs point-wise
- 蒙特利尔大学 [EMNLP 2016]
 - How not to evaluate dialogue systems

