

多轮对话综述分享

2019年秋季学期自然语言处理课程

曾四为



中国科学院网络数据科学与技术重点实验室
Key Laboratory of Network Data Science & Technology, CAS

概述

■ 问题背景

Turing在1950年定义了提出正式的人工智能测验后，创建一个能够和人自然、流畅交流的对话系统就成为自然语言处理领域的一个重要目标。

■ 什么是对轮对话系统需要解决的问题？

用户给出一个Query语句，系统需要根据这个Q（以及前面对话中的历史Q）来返回一个合适的Answer语句。

■ Example

Q: 怎么退货

A: 本平台7天之内都可以无理由退货哦，退货可以在XX页面进行XX操作。

对话系统分类

- 对于对话系统，根据不同的分类标准，有不同的划分方式。
- 根据其应用领域，可以分为问答、闲聊和任务对话系统；
- 根据其涉及的领域，可以分为开放领域和垂直领域对话系统；
- 根据数据的来源，可以分为基于语料库的和基于知识图谱的；
- 根据其是否使用历史对话信息，可以分为单轮和多轮对话系统；
- 根据研究的发展历史和系统构建方式，可以分为第一代基于符号规则和模版的对话系统、第二代基于管道的对话系统和第三代大数据驱动的对话系统。

Example:

任务型对话系统=垂直领域+多轮+第二代模型

问答型对话系统=垂直领域+多轮+基于语料+基于知识图谱

闲聊型对话系统=开放领域+多轮+基于语料+第三代模型

Answer的获取

- **简单的想法——基于规则** (1966-2000)
 - if else + 正则
- **给定语料库，利用信息检索的方式——检索式** (2013-至今)
 - 给Q找A，不就是匹配问题？
- **使用Seq2Seq模型——生成式** (2015-至今)
 - 端对端，好像挺不错？
- **复杂但是有效的方法——基于统计的管道模型** (1999-至今)
 - 很多需要解决的问题
 - 如何理解语义
 - 如何维护多轮的语义信息
 - 如何生成

基于规则的模型

■ Example

```
if re.compile("谢谢").search(sentence):  
    return "您太客气了，都是我应该做的，您看还有  
    其他什么可以帮到您的么？"
```

■ 优点

规则解释性好、系统易编写易维护、漏洞易修正

■ 缺点

1. 依赖于专家的人工规则设计，专家水平和规则质量高低严重影响系统质量
2. 难以跨领域迁移，只能针对狭窄的垂直领域
3. 构架成本高，需要消耗大量人力资源和时间成本

基于检索的模型

■ 语义匹配问题

- 问题定义：给定两个句子，计算语义相似度

■ 传统统计模型

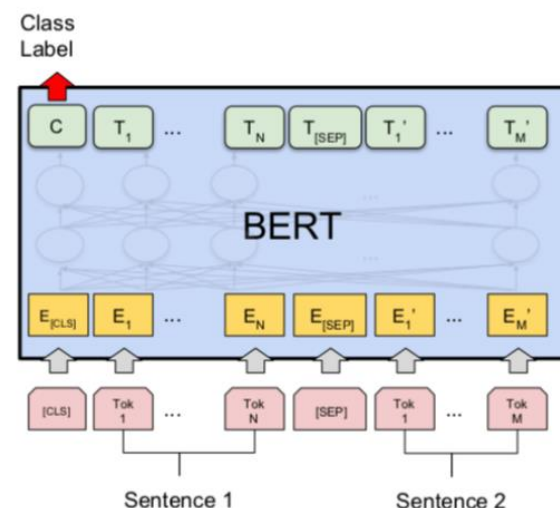
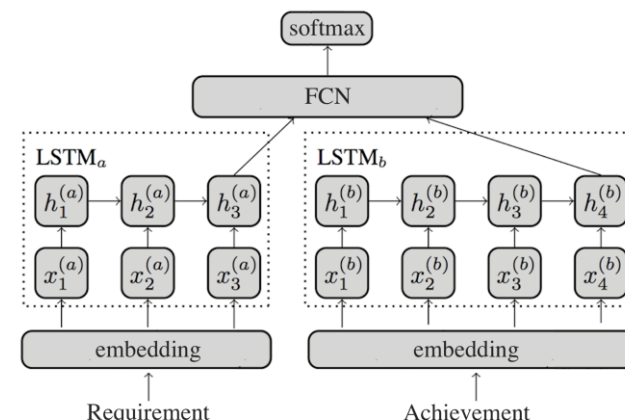
朴素贝叶斯方法、逻辑斯特回归和决策树模型

常见人工特征：

有句的分布式表示、句子长度、实体相似性和句法树结构等。

■ 深度模型

- 表示型：DSSM [1]
- 交互型：Match Pyramid [2]
- SOTA模型：BERT



[1]: Huang, Po-Sen, et al. "Learning Deep Structured Semantic Models for Web Search Using Clickthrough Data." ICKM 2013

[2]: Pang, Liang, et al. "Text Matching as Image Recognition." AAAI 2016.

基于检索的模型

■ 对话系统的匹配难点

- 相同点：都是使用两个语句进行匹配程度（或者称关联程度）的计算
- 不同点：相比于语义的相同程度，更注重语意的连贯。

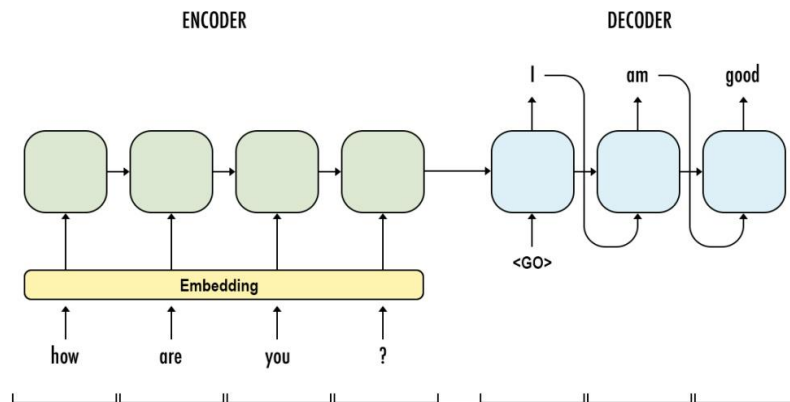
例如：“明天你打算和我去逛街吗” 对应回复为“好的，明天天气不错”

表 1: 对话和其他文本匹配任务对比表

| Tasks | Text 1 | Text 2 | Objective | Totally match |
|---------------------------|----------|------------|----------------|---------------|
| Information Retrieval | query | document | ranking | ✓ |
| Paraphrase Identification | string 1 | string 2 | classification | ✓ |
| Textual Entailment | text | hypothesis | classification | × |
| Question Answer | question | answer | ranking | ✓ |
| Conversation | dialog | response | ranking | × |

基于生成的模型

■ 简洁的Seq2Seq模型



■ Seq2Seq存在的困难

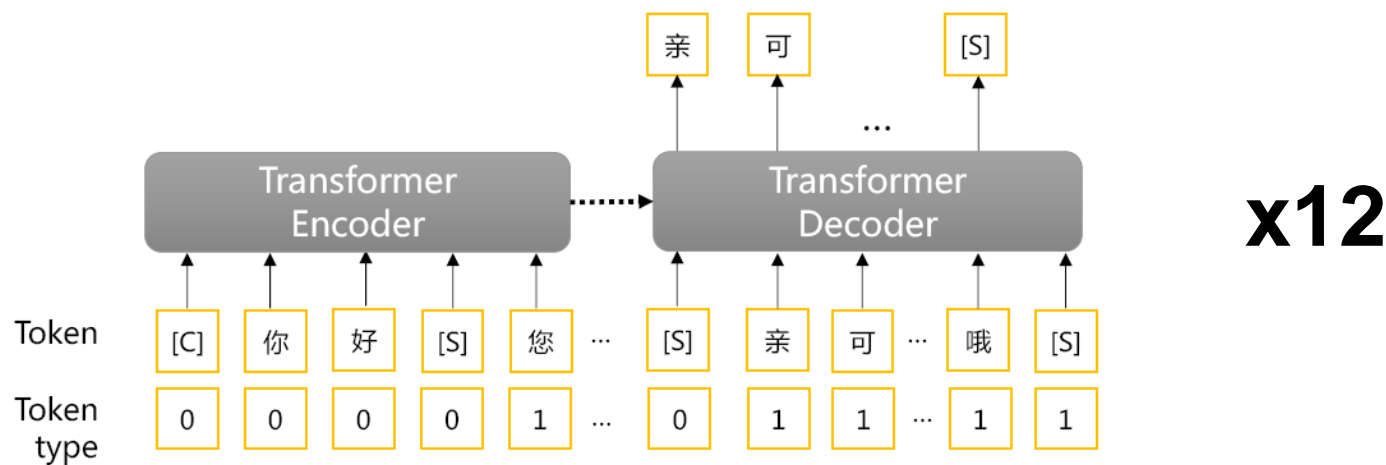
- 模型不能参考历史对话来进行生成回复。
- 模型容易生成大量无意义、通用的语句作为回复。
- 模型没有结合世界知识，很难理解深层次语义。
- 对话系统应该根据对象的不同而采用不同的说话方式。

基于生成的模型

■ 解决方法

存在非常多的论文来解决历史对话和无意义问题，最简单的方法就是：

1. 将对话历史和当前语句一起拼接——全部塞进模型
2. 堆若干个transformer——多训练训练



端对端模型简要回顾

■ 基于检索的模型

- 优点：语句流畅度有保证
- 缺点：在语料库中可能不存在能够完美匹配的回复

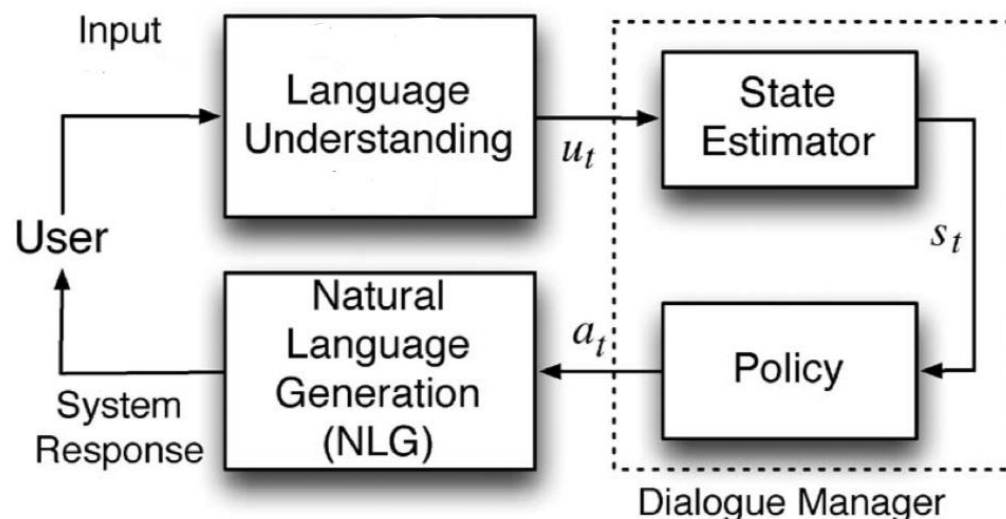
■ 基于生成的模型

- 优点：能够真正根据用户的查询语句与对话历史生成回复
- 缺点：模型尚且不能真正理解从查询到回复的语义差距

基于管道的模型

■ 主要模块

- 自然语言理解模块
- 对话状态跟踪模块
- 策略学习模块
- 自然语言生成模块



■ 主要关心

- 如何理解语义
- 如何维护多轮的语义信息
- 如何生成

■ 从统计模型到深度模型

未来展望

- **如何让生成式模型理解真实世界**
 - 逻辑推理建模
 - 多模态
 - 知识图谱
 - 更合理的对话评价机制



Thank you!