第8周汇报

——张溢炉

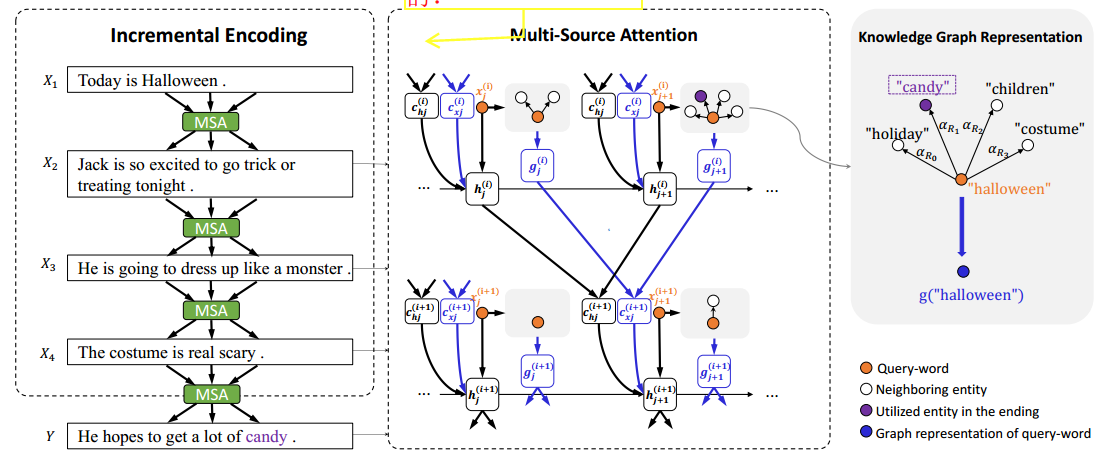
1. 《Story Ending Generation with Incremental Encoding and Commonsense

Knowledge》总结

1. 内容摘要

本文提出了一种自动生成故事结尾的模型。

1. 总结及理解

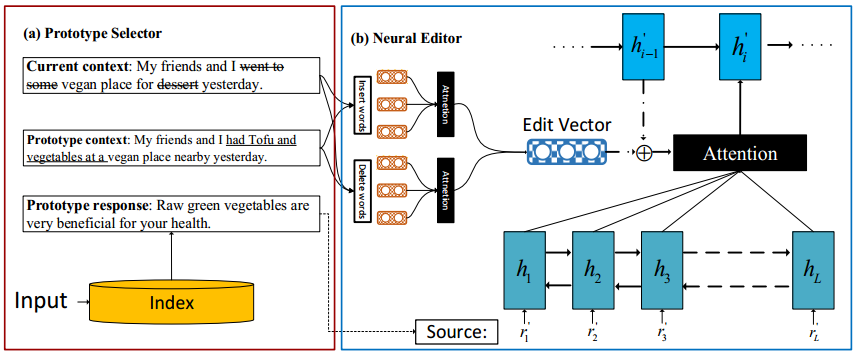
* 动机：通过分析，作者认为文章的结尾主要与文章中的线索和常识有关，线索指各时间和实体之间的上下文逻辑、因果关系。于是，作者分别设计了增量式编码方案和多头注意力机制来利用线索和常识。
* 模型如下图所示，故事和结尾的句子数是固定，分别是4句和1句。模型以句为单位进行编码，注意h由(当前词的词向量)使用LSTM编码而成，使用注意力机制存储上一层隐层h信息，使用注意力机制存储上一层隐层三元组g信息，三元组信息是通过当前词搜索知识库所得。
* 数据集
* 语料集：4句话的故事和1句话的结尾
* 知识库ConceptNet：16652条三元组关系，45种关系，关系种类有点少
* 词向量：6位小数的200维Glove词向量
* 复现代码总结
* 读代码时发现：其不能体现2个相同h、t实体间的多种关系，只会选取数据集中的第一种关系
* 语料集：训练集有90000个故事，验证集和测试集都是4081个故事
* 词向量（GloVe.6B）：下载网址: <https://github.com/stanfordnlp/GloVe>。网址中提供训练词向量的源码，可根据需要训练词向量。本文使用的是200维向量，742个词。
* CPU跑完一个epoach用了4个小时左右，GTX 1070花了不到9分钟左右
* 需要对句子进行填充，统一句子长度，可引入掩膜使填充位置乘0消除其对输出的影响
* 训练PPL（6）与测试PPL（9）相差较大，我觉得这应该是一个欠拟合的问题，要减小两者差可以尝试增加模型的复杂度（深度和宽度等）、增大训练集。其中可能原因是：任务的输出关于输出的函数极其复杂，或者说不存在能准确表达输出与输入关系的数学表达式，这与人类语言表达的多样性、随机性、主观性等有关，故不可能训练出能精确捕获复杂语言任务关系的模型
* 通过观察生成结尾及与参考结尾对比发现：生成结尾语句通顺，语法错误较少，与上下文相关，但是抓不住重点；生成结尾与参考结尾相差较大，可能因为人的思维和生成都具有随机性，还可能和数据集有关，数据集是由多个人一起收集的，这使得模型更难捕获数据集的思维（规律）
* 首先要设计好代码的整体框架；然后，要注意数据的预处理，要将文本转换成词向量，并且要与模型的输入格式对应，这其中有很多细节性问题，很可能出现一些不易察觉的逻辑错误，如将词与词向量的映射搞错。编码过程中最好分段调试，将数据打印出来观察是否出错；最后就是模型的搭建，模型常使用类来表示，复杂的模型包含很多公式，也很容易出错，这个可以利用tensorflow可视化功能，将模型以图的形式展现出来。

1. 《Response Generation by Context-aware Prototype Editing》总结
2. 内容摘要

本文提出一种可以利用检索到的原型上下文及其与输入上下文差异、原型回答来生成回答的模型

1. 总结及理解

* 动机：利用检索到的原型具有良好的语法和信息量为生成提供了良好的起点，后期编辑过程进一步提高了原型的相关性和连贯性
* 生成句子追求目标：与上下文相关性强、流畅、语法错误少、多样性、原创性，原创性是本文提出的
* 模型如下图，模型的大致思路是首先检索出与当前输入上下文相关的原型上下文，然后根据上下文差异（当前上下文没有而原型有的词（insert words）、当前上下文有而原型没有的词（delete words））通过注意力机制得到edit向量，最后将edit向量和原型回答串联作为解码器输入，输出回答。其中，测试时的检索方式与训练时的检索方式是不同的，测试时是利用Lucene方法检索上下文相似的原型；而训练时是利用回答的相似性来检索的。



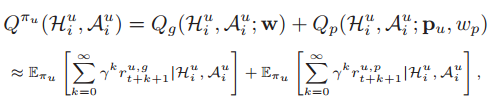
* 模型的亮点是利用了已有的数据集作为参考，提高了生成句子的质量

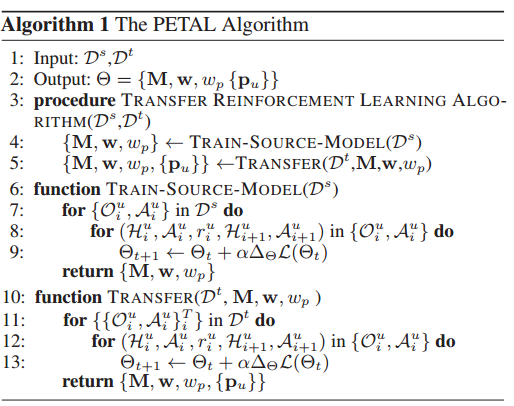
1. 《Personalizing a Dialogue System with Transfer Reinforcement Learning》总结
2. 内容摘要

本文提出了迁移强化学习的方法使对话agent的更具个性化，可以对常客做到私人订制，减少了对话轮次，有利于提升客户体验。

1. 总结及理解

* 动机：利用迁移学习可以解决关于某个人的数据少的问题，同时也解决了之前迁移忽略个体差异带来的负面影响；利用端到端的方法解决基于规则学习需手动设置规则和特征的麻烦
* 模型是基于POMDP的思想创建的，包含7个元素{S,A,O,P,R,Z,}。S为隐层不可观测的状态；A为agent生成的回答；O为顾客所说；P为状态转移矩阵；R为奖励；Z为观察函数；为贴现因子，使对话更短。个性化的Q-function如下式：其中，第一项代表所有顾客的偏好，第二项代表个人偏好，所有b为由历史信息{A，O}经矩阵M仿射而得信念状态。参数M，W，由所有顾客的数据训练而得，在本文实验中参数个数为85000；参数由某个顾客数据训练而得，在本文实验中参数个数为100。利用迁移学习解决了个人数据匮乏的问题，也利用了个体之间的差异



* 本文采用State-Action-Reward-State-Action(SARSA)方法来优化参数，步骤如下图，先用所有顾客数据训练M，W，，再用某个人数据来微调M，W，，。
* 本文实验real-world dataset，实验结果明显优于其他baseline
* 疑问：
* 如何编码实现？