**Reference Type:**  Journal Article

**Record Number:** 34

**Author:** Shuman Liuy;x∗, Hongshen Chenz, Zhaochun Renz, Yang Fengy, Qun Liu}, Dawei Yinz

**Year:** 2018(ACL)

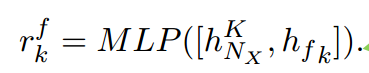
**Title:** <Knowledge Diffusion for Neural Dialogue Generation(ACL2018, Liu et all).pdf>

**Short Title:** <Knowledge Diffusion for Neural Dialogue Generation(ACL2018, Liu et all).pdf>

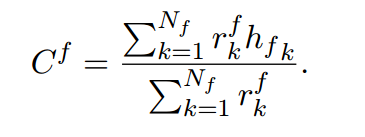
**'File' Attachments:** internal-pdf://0328523247/Knowledge Diffusion for Neural Dialogue Genera.pdf

1. **论文思路及个人分析**

* 论文提出的模型是neural knowledge diffusion（NKD），是在对话生成中嵌入知识（缺乏背景知识和常识的端到端的数据驱动的模型倾向于生成短和无意义的回答，如：哈哈，我不知道），此处的知识主要是电影知识，知识以对象的形式存储和使用。使用语料按照是否有实体分成两类，且均含单轮和多轮对话两种形式，针对不同类型的语料提出了两种方法，分别是facts matching和entity diffusion，即整个NKD的组成成分。作者主张此模型有良好的**收敛**（应该指，对包含实体的对话识别出实体，将回答限制在此实体范围内）和**发散**的能力（应该指，对于不包含实体的对话，单轮则根据语料信息寻找候选实体，即diffusion；多轮则根据上一轮的实体或当前轮的语料信息寻找候选实体）。最终分别在事实型回答系统（factoid-questions，见“六”的解释）和基于知识的闲聊（knowledge grounded chi-chats）两种情境下测试模型。模型有一定的效果，但没有达到state of art。
* Facts matching（三元组形式）：难点是有些语句只有实体，没有关系（如，泰坦尼克号是我最喜欢的电影），则匹配不了；
* Entity diffusion：应用于需要从一个实体推移到另一个实体（如，给我推荐一部跟泰坦尼克号一样类型的电影）
* 模型包含整体流程（包含四个模块层：Encoder RNN，Knowledge Retriever，Context RNN和Decoder RNN）：
* （1）Encoder RNN ：encoder将输入序列编码成向量表征，并产生两个隐藏状态序列，和，用来记录输入序列的信息，它的最后一个状态传递给Context层，也记录了输入序列中实体和关系的信息，最后一个状态传递给Knowledge层；
* （2）Knowledge Retriever：a）通过字符串匹配、实体链接或者命名实体识别等方法在输入序列（或者多轮对话中的历史对话）找到实体，并在识库中找出该实体的相关facts，编码成事实表征（通过求每条fact中实体（subject）向量和关系向量的平均数得到），最后计算输入序列中的【实体-关系】与中所有的【实体-关系】的相关性，相关性系数采用非线性函数或子神经网络来计算，此处采用的是多层感知机：



对于识别出fact（【实体-关系】）的单轮对话，采用上述式子，得到本层次的结果；而对于无法识别fact或者多轮对话（有某些turn没有fact）则无法直接使用这个相关性系数，转而找一个相对值来代替：通过的加权平均数得到相对事实特征，也即在多轮对话中，主题始终围绕此实体：

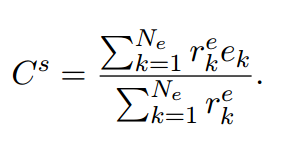


此结果用于本层次的下一环节b），并非本层次的最终结果。

b) 对于无法识别fact或者多轮对话（有某些turn没有fact）这种情况，做法是，基于当前的实体发散挖掘其他实体，通过计算知识库中（除当前实体以外）的实体与当前实体和输入序列中fact【实体-关系】相关性来寻找最佳的实体：（）

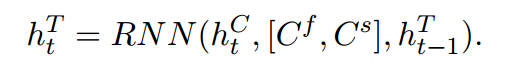


选出相关性最高的K个实体，然后得到相似实体集的表征：

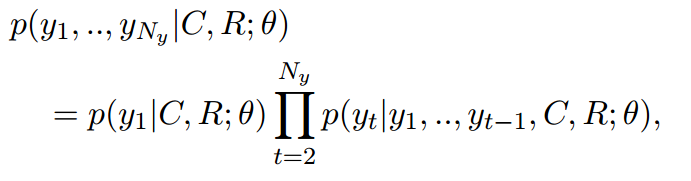


作者观察到在fact matching的turn中，entity diffusion的相关性很低，也即此情况下没必要做entity diffusion，只有在b）的情况下，entity diffusion才能发挥起作用。可以这么理解，当fact matching时，实体和关系是确定的，而entity diffusion是求除掉词实体以外其他实体与输入序列的fact的相关性，所以entity diffusion无效是自然的。

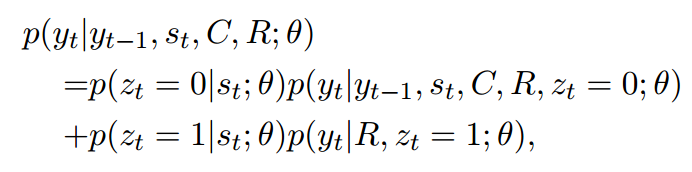
* （3）Context RNN：主要是整合整个输入序列的信息，此RNN的输入为：encoder得到的状态、knowledge层得到的[]和前面词的信息：



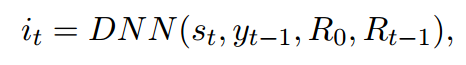
* （4）Decoder RNN：此层次作用是生成回答的词语，提出了两种方法。令为语料信息系数，为知识信息系数。a) **Vanilla decoder：**简单的RNN模型，直接生成responses：



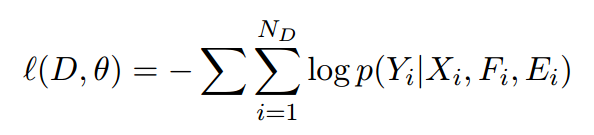
b）**Probabilistic gated decoder：**多加了一个门控，门控的作用是标识词语是否来源于知识库（表示来源于知识库），即是否为知识（entity），产生第t个词的概率是：



住通过logistic线性回归来计算得到，貌似线性二分类问题，多采用logistic。为了防止一个实体在回答中重复出现多次（降低了回答的多样性），对已出现过的实体加以限制，即每个时间都更新知识信息系数，对此也提出两种方法：1）Mask coefficient tracker：直接把已出现过的实体的知识信息系数置为0；2）Coefficient attenuation tracker：通过深度神经网络得到衰减率:



再用去更新。模型的损失函数是log似然函数：



1. **怎么摆出实验结果（如何增强实验说服力）**

* 实验的知识库是从“豆瓣的电影栏目”爬的数据，以对象的形式存储，然后第二层Knowledge extract中检索出后转化为三元组。
* 从“百度知道”中爬取问题及答案（即事实型系统），从“豆瓣的group栏目”中爬取基于知识的闲聊，这些数据的分布特点是：长尾型，即热门的电影（实体）在对话中出现多，冷门的出现少。
* 实验基于了两种类型的对话数据集，某种程度上能够说明在不同数据集上的适用性（能叫做迁移吗？）；同时采用了3种其他作者的模型作为baseline，增加可信度；其次，对于自己提出了方法，组合出三种模型，说明提出的方法是可以随机组合的，且均有效果和不同的优势；最后实验效果的测评采用自动评测（精确度和召回率）和人工（从流利度、知识相关性和正确度三方面来评估）两种方式，一定程度上增加了实验结果的可信度。

1. **问题**

* Encoder层中的和是怎么设置的，差别在哪里，是如何考虑参数的设置。
* 在计算输入序列中的【实体-关系】与中所有是【实体-关系】的相关性系数时，为什么要采用非线性函数或子神经网络来计算，为什么本文挑了多层感知机。

1. **知识拓展**

* 问答系统的分类：
* 事实型：答案一般为命名实体（如，人名，组织名，地名），有些事实型问答系统会固定文本长度，有些返回的是精确答案。
* 列表型：与事实型相近，不同在于列表型的答案可能大于一个，通过答案的覆盖面来评测性能的好坏。
* 描述型：答案通常为一段话，没有固定答案。
* 开放领域型：试图回答各种问题，，如AskJeeves。
* 受限领域型：针对某个特定应用领域，或企业级的知识问答系统，特点是有非常明显的专业特征，所以可以结合该专业的具体知识或领域本体来分析问题和答案，答案的来源可以是结构化数据。

1. **体会及思考**

* 对语句中的知识和普通词加以区分，给予不同的身份标识（即作用不同）。
* Encoder的产生了两个隐藏层状态，分别用于表示不同的信息，一个是知识，一个是语料，这样可以各自发挥各自的作用且有针对性，排除其他内容的干扰。
* 对于一些小块的东西（是否属于知识的二分类问题、两个相关性系数）思考根据其特点适用于什么样的统计方法或者公式或者某种数学方法，及支撑这种方式的理由。
* 本论文的一个优点是：各个步骤（方法）之间联系的非常紧密而且按照人的思维很有道理，有些地方作者并未解释为什么，但是自己深层次的想一下就可以理解到。