

Relation Extraction with Temporal Reasoning Based on Memory Augmented Distant Supervision

Jianhao Yan¹, Lin He¹, Ruqin Huang², Jian Li^{3*}, and Ying Liu¹

¹Institute for Network Sciences and Cyberspace, Tsinghua University

²Department of Computer Science and Technology, Tsinghua University

³Institute for Interdisciplinary Information Sciences, Tsinghua University

{yjh16, he-l14, hrq16}@mails.tsinghua.edu.cn

lijian83@tsinghua.edu.cn, liuying@cernet.edu.cn

收录会议: NAACL2019

任务: 远程监督关系抽取

动机: 前人 DS RE 的工作忽略了关系实例中的时间信息, 比如实体对在今天和明天可能是不同的关系, 所以本文提出了一个新的任务, 即从时序的角度推理实体对之间的关系, 并提出专门针对时序 RE 的新数据集 WIKI-TIME

TemIME 模型基于远程监督常用的架构 encoding + fusion, 将时序预测关系看作是序列标注任务。

回顾:

1. 前人 encoding 阶段, 对 bag 中每个 instance 使用相同的编码组件, 导致了这些 instance 的表示在时间信息上是相同的, 以至于不能区分他们所在的时序关系。所以作者期望在 encoding 阶段对每个 instance 加入时间信息。

2. 前人 fusion 阶段, 只采用了一步的 fusion (比如用 attention 或者 RL 对 instance 的表示进行聚合或者采样, 这些方法用于降噪), 这导致了没有句子级别的推理。作者认为, bag 中的有些 instance 尽管不是对应关系的正例, 但可以提供支持的证据 (supporting evidence), 并将其称作 remote instance, 作者希望能够从这些看似是负例的 instance 中对正例的推理起到帮助作用。

因此模型的架构如下:

1. 在句子的 Encoding 阶段, 使用 CNN 和 PCNN (15 年曾道建提出), 模型的输入是词向量和 Positional Feature 的合并, 然后送入 CNN 做 1D 的滑动窗口为 3 的卷积, 然后使用 piece-wise 的最大池化, 得到三部分的表示并进行合并, 然后再把所有 channel 的输出进行合并, 得到 instance 最终的表示

下面介绍如何引入时序信息 (应具有两个特点: 1. 时序的编码应该与句子的时间顺序一致, 2. 两个时间点的不同决定了两个时间的编码相似性)

作者不是直接的使用 1,2,3,4 来编码时间, 而是用了类似于 transformer 中的 Positional Encoding 来进行时序的编码

$$PE(j)=\begin{cases} \sin(j/10000^{d/d_m}) & \text{if } d\%2=0 \\ \cos(j/10000^{(d-1)/d_m}) & \text{if } d\%2=1 \end{cases}, \quad (6)$$

where j is the rank of instance s , d is the dimension, and d_m is the dimension of temporal encoding. Obviously, the PE encoding complies with the chronological order and the similarity between

- 得到时间信息后，将其 concat 到 instance 表示的后面，并引入一个放缩的因子 λ
- 在 Fusion 的阶段，将每个 instance 的最终表示看做是 memor 中的一个槽，然后使用一个与时间相关的 query（关系，实体对，时序信息）对 memory 中的 instance 进行迭代的 attention 来对内存进行定址

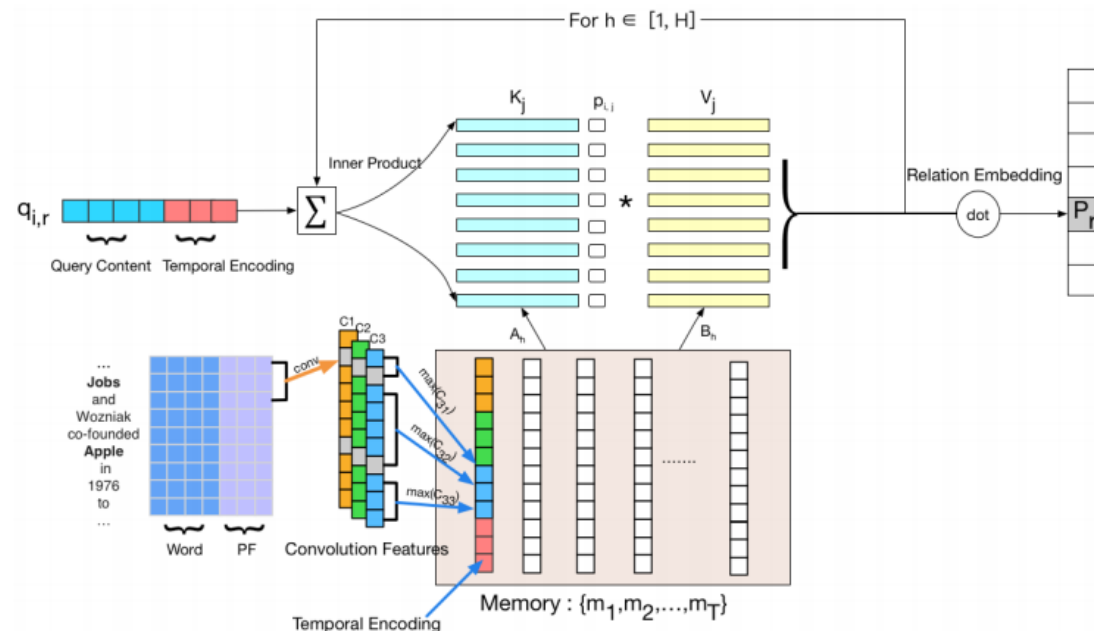


Figure 1: Overall TempMEM architecture

实验数据使用 NYT-10 和 WIKI-TIME

Method	P@N_100	P@N_200	P@N_300
CNN_ATT	67.33	67.66	66.45
CNN_ONE	70.3	68.66	65.78
TempMEM	81.18	82.09	78.41
TempMEM+R	79.21	78.61	75.42
TempMEM+P	81.19	79.1	77.41

Table 2: Comparison with previous models.

P@N_100/200/300 refers to the precision for the highest 100, 200 and 300 predictions in WIKI-TIME.

总结：这篇 paper 从前人工作的问题（缺少时序信息推理）入手，再提出针对性的解决方案（引入缺少的信息）和新的数据集，并使用 **memory network** 的迭代访存来引入推理的概念。