

Effective Deep Memory Networks for Distant Supervised Relation Extraction

Xiaocheng Feng, Jiang Guo, Bing Qin, Ting Liu, Yongjie Liu

SCIR, Harbin Institute of Technology, Harbin, China

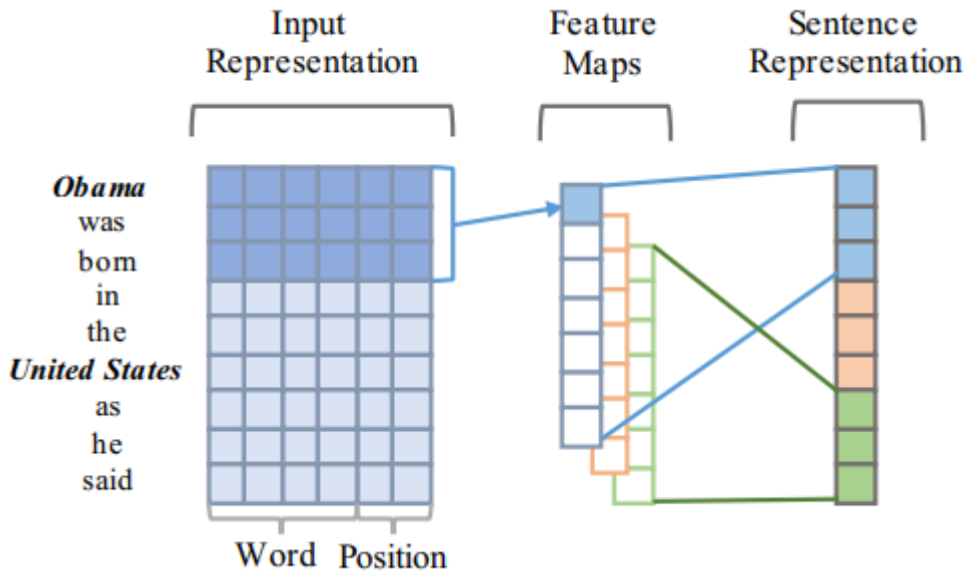
{xcfeng, jguo, qinb, tliu, yongjieliu}@ir.hit.edu.cn

收录会议: IJCAI2017

动机: DS RE 通常用多实例多标签学习的方法来解决, 本文针对多实例的 attention 机制, 提出两个基于 attention 机制的 memory 组件来同时显式地捕捉 context word 的信息 (比如 instance 中的每个词对分类的关系贡献程度不一样) 和关系类别之间的内在依存信息 (比如两个关系类别可能有对称关系或者蕴含关系)

模型:

1. 先使用 PCNN 对 bag 中的每个 instance 进行编码



2. 然后使用 word-level attention-based memory network 对 context-word 进行内存化, 并使用实体对来 query 内存, 得到该 instance 中实体对对 context-word 的表示, 然后 concat 到 PCNN 得到的 instance 的表示后面

$$x = \sum_{i=1}^w \alpha_i m_i \quad (1)$$

$$g_i = \tanh(W_{word-att}[m_i; w_{eh}; w_{et}] + b_{word-att}) \quad (2)$$

$$\alpha_i = \frac{\exp(g_i)}{\sum_{j=1}^w \exp(g_j)} \quad (3)$$

- 使用两层 sentence-level 的 relation memory，初始的 relation 表示是随机初始化的，然后每个 relation 先去从得到的 instance 的表示中做选择性的 attention，然后再使用每个 relation 对其他 relation 进行 attention，得到最终的 relation 的表示

Selective attention over instances:

$$R_j = \sum_{i=1}^n \beta_i x_i \quad \beta_i = \frac{\exp(z_i)}{\sum_{p=1}^n \exp(z_p)} \quad z_i = x_i A v_{r_j}$$

Selective attention over relations

$$R_j^* = \sum_{i=1}^k \gamma_{ji} R_i \quad \gamma_i = \frac{\exp(h_i)}{\sum_{q=1}^k \exp(h_q)} \quad h_i = R_i B R_j$$

最终的架构如下：

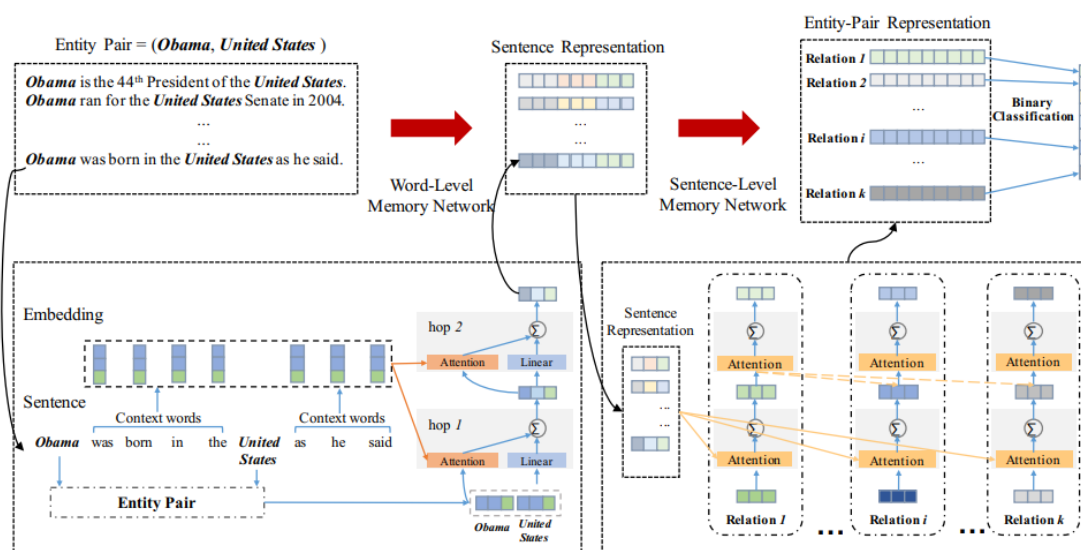


Figure 3: Illustration of our deep memory networks with two computational layers (hops) for distant supervised RE.

总结：针对 attention 机制的细化，提出了 word-level 的 attention 来刻画每个 instance 中的 context-word 对实体对的贡献程度，然后提出 sentence-level 的 attention，首先刻画句子对关系的贡献程度以及关系之间的依赖程度。动机很好，结果也有很大的提升

缺点：虽然动机不错，结果也挺好，但是没有做 attention 的可视化，关系的表示是随机初始化的，可能受初始化的影响特别大。

深层 hop 的 word-level memory，作者认为是随着 memory 的加深，梯度回传消失了，而且 8 跳的效果还特别低，感觉是不是可以引入一些 residual 或者 layerNorm 的机制来解决呢