

Attention Guided Graph Convolutional Networks for Relation Extraction

Zhijiang Guo*, Yan Zhang* and Wei Lu

StatNLP Research Group

Singapore University of Technology and Design

{zhijiang.guo, yan.zhang}@mymail.sutd.edu.sg, luwei@sutd.edu.sg

收录会议: ACL2019

任务: 基于句子和跨句子的关系抽取, 使用 TACRED 数据集提供的依存树结构信息。

动机: 大家通常会使用依存树来辅助关系抽取任务, 但是依存树中通常包含有一些无关信息需要剔除, 现有方法使用基于规则的硬剪枝策略, 但有时候会将一些重要的信息给剪掉。因此本文提出一个使用 Attention 来引导的 GCN 网络, 让模型自己学会剪枝, 即软剪枝, 来得到最优的剪枝。

模型架构:

- (1) 首先将句子变成依存树 (这个由数据集给定), 然后再通过加入反向连接和自环连接依存树变成图结构,
- (2) 然后使用 multi-head attention 将这个图变成 N 个全连接的邻接矩阵, 矩阵中的元素是两个词之间的关系权重

$$\tilde{A}^{(t)} = \text{softmax}\left(\frac{QW_i^Q \times (KW_i^K)^T}{\sqrt{d}}\right)V \quad (2)$$

- (3) 然后得到 N 个邻接矩阵之后, 使用本作者在 TACL 2019 Q3 提出来的 DCGCN (Densely Connected Graph Convolutional Network, 类似于 CV 里面用的 Densenet, 每一层的输出都需要加上前面层的输出) 来进行信息聚合操作

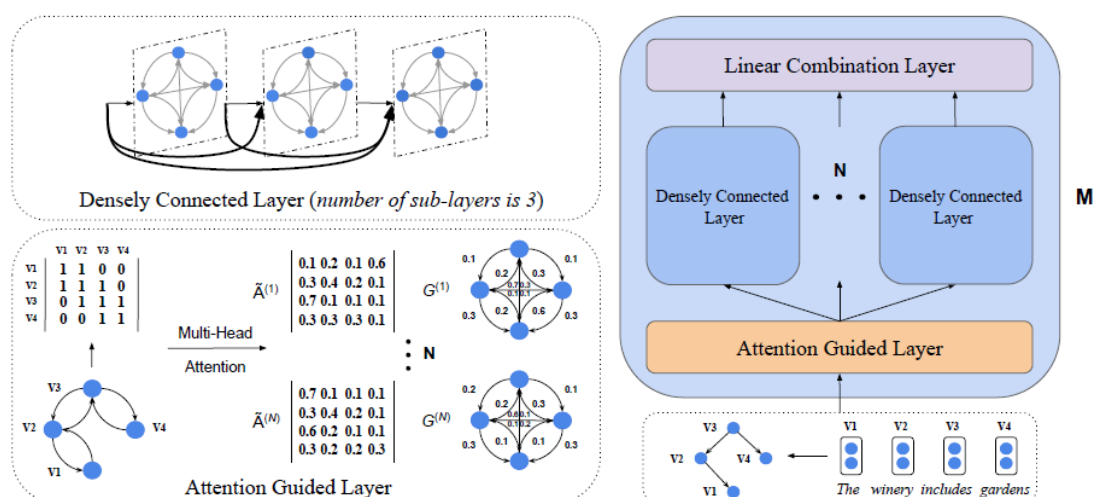
$$\mathbf{g}_j^{(l)} = [\mathbf{x}_j; \mathbf{h}_j^{(1)}; \dots; \mathbf{h}_j^{(l-1)}].$$

$$\mathbf{h}_{t_i}^{(l)} = \rho\left(\sum_{j=1}^n \tilde{A}_{ij}^{(t)} \mathbf{W}_t^{(l)} \mathbf{g}_j^{(l)} + \mathbf{b}_t^{(l)}\right) \quad (4)$$

- (4) 最后将 N 个 DCGCN 得到的图使用 Linear Combination 进行聚合

$$\mathbf{h}_{comb} = \mathbf{W}_{comb} \mathbf{h}_{out} + \mathbf{b}_{comb}$$

$$\mathbf{h}_{out} = [\mathbf{h}^{(1)}; \dots; \mathbf{h}^{(N)}] \in \mathbb{R}^{d \times N}.$$



实验结果:

首先,肯定超越了前人用硬剪枝树做 GCN 的方法,也验证了他们自己提出的 DCGCN 同时捕捉 local 信息和 global 信息的优势

Model	Binary-class				Multi-class	
	T		B		T	B
	Single	Cross	Single	Cross	Cross	Cross
Feature-Based (Quirk and Poon, 2017)	74.7	77.7	73.9	75.2	-	-
SPTree (Miwa and Bansal, 2016)	-	-	75.9	75.9	-	-
Graph LSTM-EMBED (Peng et al., 2017)	76.5	80.6	74.3	76.5	-	-
Graph LSTM-FULL (Peng et al., 2017)	77.9	80.7	75.6	76.7	-	-
+ multi-task	-	82.0	-	78.5	-	-
Bidir DAG LSTM (Song et al., 2018b)	75.6	77.3	76.9	76.4	51.7	50.7
GS GLSTM (Song et al., 2018b)	80.3	83.2	83.5	83.6	71.7	71.7
GCN (Full Tree) (Zhang et al., 2018)	84.3	84.8	84.2	83.6	77.5	74.3
GCN ($K=0$) (Zhang et al., 2018)	85.8	85.8	82.8	82.7	75.6	72.3
GCN ($K=1$) (Zhang et al., 2018)	85.4	85.7	83.5	83.4	78.1	73.6
GCN ($K=2$) (Zhang et al., 2018)	84.7	85.0	83.8	83.7	77.9	73.1
AGGCN (ours)	87.1	87.0	85.2	85.6	79.7	77.4

总结:

这篇 paper 是典型的表示学习模型,首先利用依存树确实能够得到很多语义信息,但是前人的方法由于想要去噪,使用最低公共祖先的方式去做硬剪枝,会损失很多信息,正是基于这个观察,作者使用了自己 TACL2019 Q3 提出来的 DCGCN 来去做尝试,企图能够直接从整个树上学到一种软剪枝(也就是稳重的 attention guided layer),因此这篇 paper 也就是套用了 DCGCN,解决剪枝的问题。

问题:

attention guided layer 其实从含以上是类似于 GAT 的,只不过这里是讲计算 attention 系数分隔开来,取了个好听的名字叫 self-attention,其实就是 DCGAT,所以感觉可以验证 DCGAT 是否更好?