

Graph Neural Networks with Generated Parameters for Relation Extraction

Hao Zhu¹ Yankai Lin¹ Zhiyuan Liu¹ Jie Fu² Tat-seng Chua³ Maosong Sun¹

¹ State Key Lab on Intelligent Technology and Systems

Department of Computer Science and Technology

Institute for Artificial Intelligence, Tsinghua University, Beijing, China

² National University of Singapore, Singapore

³ Université de Montréal, Montréal, Québec, Canada

{zhuhaol15, linyk14}@mails, {liuzy, sms}@tsinghua.edu.cn

jie.fu@polymtl.ca, chuats@comp.nus.edu.sg

收录会议: ACL2019

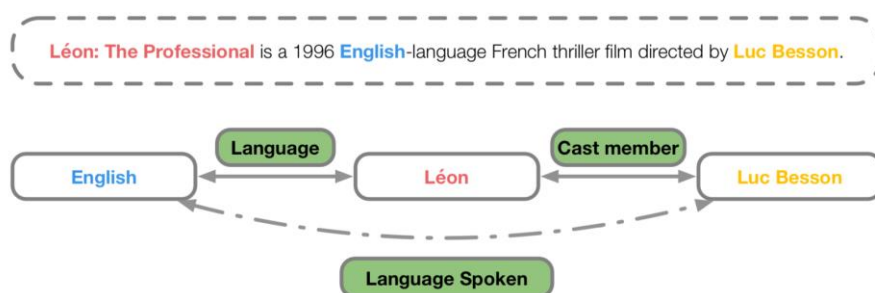


Figure 1: An example of relation extraction from plain text. Given a sentence with several entities marked, we model the interaction between these entities by generating the weights of graph neural networks. Modeling the relationship between “Léon” and “English” as well as “Luc Besson” helps discover the relationship between “Luc Besson” and “English”.

动机:

- 1、通过这个例子我们可以看出，现有的 RE 方法可以轻松地识别出 Luc 这个人导演了 Leon 这部电影，并且这部电影是英文的，但是在没有多跳推理的情况下很难推断 Luc 这个人 和 English 的关系（luc 说英语），因此多跳推理对于 RE 这个任务来说很重要，现在 GNN 是解决多跳推理的一个好方法
- 2、但是现有的 GNN 只能在预先定义好的图上处理多跳关系推理，而不能直接应用在自然语言文本上进行关系推理，所以作者们提出用生成 GNN 的邻接矩阵的方式，来将文本和图进行联系

模型架构主要分为三个组件

首先 GP-GNNs 创建一个全连接的实体图

- 1、编码组件。使用编码器（BiLSTM, CNN 都可以，本文用的 BiLSTM）对文本进行编码，来生成 GCN 中的边的信息。其中每个词的编码包括词嵌入和位置嵌入（每个词相对于实体对之间的距离）。 $f(\cdot)$ 是每个词过了 BiLSTM 后再过一个带非线性激活的 MLP。得到的表示再 reshape 成邻接矩阵

parameters of the propagation module, by

$$\mathcal{A}_{i,j}^{(n)} = f(E(x_0^{i,j}), E(x_1^{i,j}), \dots, E(x_{l-1}^{i,j}); \theta_e^n), \quad (1)$$

where $f(\cdot)$ could be any model that could encode sequential data, such as LSTMs, GRUs, CNNs, $E(\cdot)$ indicates an embedding function, and θ_e^n denotes the parameters of the encoding module of n -th layer.

- 2、信息传播组件即 GCN。使用 GCN 来传播节点之间的关系信息。
- 3、使用 GCN 所有层得到的实体对表示进行 concat，来进行实体对的关系预测

总结：从文本中学习图的边的信息，将没有关系或者有非常有限关系图上的推理扩展到非结构化的文本上。Enable message-passing with rich text information

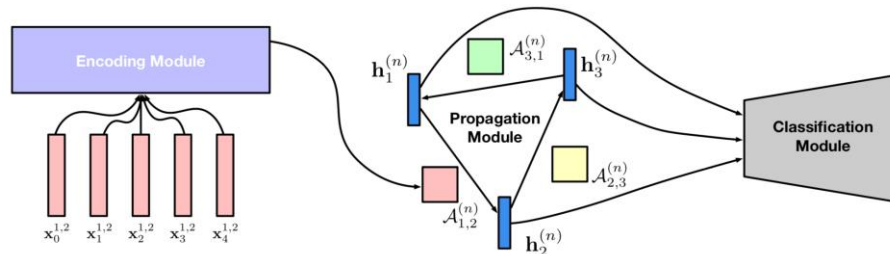


Figure 2: Overall architecture: an encoding module takes a sequence of vector representations as inputs, and output a transition matrix as output; a propagation module propagates the hidden states from nodes to its neighbours with the generated transition matrix; a classification module provides task-related predictions according to nodes representations.

问题：

- 1、刚开始是全连接的图，会不会有噪声？
- 2、分类端，仅仅使用节点的信息，而不适用边的信息，是不是太粗糙？
- 3、Multi-hop 层之间没有交互，仅存在于节点之间的交互
- 4、Position embedding 和 Zeng 不一样，他是相当于对每个词划分成 3 类：要么属于头实体，要么属于尾实体，要么都不属于，然后用着三类的 embedding 层，这样做更好吗？
- 5、GCN 每层都生成边的参数？可不可以合到一层？