

基于图论的关系抽取任务

网络安全 赵正一

摘要

关系抽取是从一段文本中，在两个命名实体之间，抽取出某种预定义好的关系类别的一种工作。关系的主客体双方必须为出现在同一空间内的命名实体，这一空间可以是句子级、段落级甚至是文档级。当给定的文本很长的时候，从这一文本中提取出特定的有关联的主客体以及确定他们之间的关系就成为了这一任务的一大难点。近些年来，关系抽取工作主要聚焦于使用预训练的语言模型如 BERT 模型。这样的工作往往是序列的标注工作，讲文本当作长序列，顺序学习提取出命名实体和他们之间的关系。本文在图论的基础上，提出了一个重构的多视图图模型用来捕捉命名实体之间的关系。之后通过调整图结构来选择出对于关系抽取工作来说最重要的字词，最后通过在序列上使用 BERT 模型，在结构上使用基于图论的图模型来共同抽取关系。通过实验验证，本文提出的方法在对话级别的关系抽取中有比较好的效果。

关键字：关系抽取；图论；对话级；BERT 模型

Abstract

Relation extraction is a work of extracting a certain predefined relationship category between two named entities from a piece of text. The subject and object of the relationship must be named entities that appear in the same space, which can be sentence-level, paragraph-level, or even document-level. When the given text is very long, extracting specific related subjects and objects from this text and determining the relationship between them becomes a major difficulty in this task. In recent years, relation extraction work has mainly focused on the use of pre-trained language models such as the BERT model. Such work is often sequence labeling work. The text is treated as a long sequence, and sequential learning is used to extract named entities and their relationships. Based on graph theory, this paper proposes a reconstructed multi-view graph model to capture the relationship

between named entities. Afterwards, the graph structure is adjusted to select the most important words for the relationship extraction work, and finally the relationship is jointly extracted by using the BERT model in the sequence and the graph model based on graph theory in the structure. Through experimental verification, the method proposed in this paper has a relatively good effect in the extraction of relations at the dialogue level.

Keywords: Relation Extraction; Graph Theory; Dialog-Level; BERT-based model

目录

1	背景介绍	1
2	相关工作	2
2.1	基于循环神经网络的关系抽取	2
2.2	基于图结构的关系抽取	2
2.3	基于 BERT 模型的关系抽取	3
3	模型介绍	3
3.1	预备知识	3
3.1.1	形式化表示	3
3.1.2	多视图图模型	3
3.2	BERT 模块	4
3.3	高斯图模块	4
3.4	图卷积模块	5
4	实验验证	5
4.1	实验设计	5
4.2	结果分析	6
5	总结和展望	7

1 背景介绍

给定一段文本和两个命名实体，通过理解文本含义推断出两个命名实体之间的语义关系的工作就被称作是关系抽取。文本段落可以是一句话、一段话或者是一篇文章，甚至也可以是一段对话。

这里使用 TACRED[1] 数据集中的一个例子来解释关系抽取任务。值得一提的是，在关系抽取任务中，我们关心两个命名实体之间的关系类别，这个类别往往是预定义好的。比如说下面这个例子中：(*Cathleen P. Black, chairwoman*) 这两个实体的关系应当为 (per:title)。原句为：*Carey will succeed Cathleen P. Black, who held the position for 15 years and will take on a new role as chairwoman of Hearst Magazines, the company said.* 如果两个命名实体之间的关系不在预定义好的类别中，那么判定他们之间的关系为无关。

再从 DialogRE 数据集中举例进行说明。DialogRE 数据集是第一个人工标注的适用于对话级的关系抽取数据集，如表 1 所示。对话级关系抽取工作的目标旨在从对话中抽取出多轮对话的实体关联信息。

S1:	Hey Pheebs.	Argument Pair	Trigger	Relation type	
S2:	Hey!	R1	(Frank, S2)	brother	per:siblings
S1:	Any sign of your brother ?	R2	(S2, Pheebs)	none	per:alternate names
S2:	No, but he is always late.				
S1:	I thought you only met him once?				
S2:	Yeah, I did. I think it sounds y'know big sistery, y'know, 'Frank's always late.'				
S1:	Well relax, he'll be here.				

表 1: DialogRE 数据集示例 [2]

随着技术的发展，图结构的神经网络模型因为它优异的性能而被广泛使用。通常，图中的每一个节点都表示给定文本中的一个命名实体。边的构造方法有很多，通常研究依赖于外部的句法解析树，将文本序列转换成从属的树结构来初始化图结构。也正是因为这个原因，由于句法树的错误提取，可能会导致错误传播的问题发生。

本文提出了一个基于图论的关系抽取模型用来抽取对话级别的实体关系。在没有外部知识的情况下，构造一个多视图的图结构模型高度依赖人工提取的特征的有效性。这个图结构以文本中所有的命名实体为节点，基于 BERT 模型求出节点的表征向量，再根据某种规则对这些向量进行有向连接形成边。最后通过图卷积和图池化来逐一剪枝，最

后确定本文中的关系。

在学习潜在图的时候，两个命名实体之间的边往往记录了他们从文本序列中抽取出来时涵盖着的关系。这份关系往往不是对称的，所以我们需要生成有向图来初始化我们的图结构。具体来说，本文使用了高斯图生成器 (GGG) [3][4] 来初始化多视图图结构的边。然后通过测量不同节点的高斯分布之间的 Kullback-Leibler (KL) 散度 [5] 来计算边节点的表征向量。由于 KL 散度的不对称性，GGG 生成的图当然是有向图。在序列较长的情况下，初始化后的潜在图往往非常大，所以需要通过迭代图卷积和图池化来剪枝缩小规模。从而得到一个具有自适应能力的任务特定的图结构来进行本文关系抽取工作。

2 相关工作

本文从基于长短期记忆模块的循环神经网络模型 (LSTM-RNN)、基于图结构模型和基于 BERT 模型的关系抽取方法对相关工作进行了简要综述。

2.1 基于循环神经网络的关系抽取

早期关系抽取的工作高度依赖于人工执行的特征抽取工作 [6][7]，通过摘取不同的特征来表示命名实体。模型高度依赖人工提取的特征的有效性。当前工作主要聚集于学习模型的表示，比如将循环神经网络 (RNN) 应用到关系抽取工作中去。前人提出的双向长短期记忆模块 (LSTM) [8] 极大的增强了 RNN 对于长距离依赖问题难以解决的困扰。在此基础上，又有人提出了 PA-LSTM[1] 来额外编码序列中的位置信息，增强了关系抽取模型的性能。

2.2 基于图结构的关系抽取

基于图结构的模型由于在关系推理中的有效性优势，在关系抽取工作中得到了广泛应用。有人利用图卷积神经网络 (GCN) [9] 来捕捉远距离依赖的结构信息。有人提出了一个注意力导向的 GCN (AGGCN) 模型 [10]，通过增加自注意力机制来增强图结构表示学习的能力。AGGCN 在句子级的关系抽取工作中表现优异，但是仍有不少缺点。比如它高度依赖于外部的句法知识信息。错误的句法信息在后续的图表示学习中可能引起错误传播。为了消除这一问题的影响，有人提出了隐性图学习模型 [11][12]。使用一个端到端的学习方式来学习图结构中的隐藏信息，这样的学习方式不再需要外部的句法树作为辅助工具。除此之外，有人还尝试把这一结构看作是一个整体变量 [13]，通过对这一变量的学习调整来获取图的特征表示。

本文基于这一思想，尝试生成多视图的图结构，用来模拟不同节点之间所有的结构可能。除此之外，本文还尝试提取出每一个图中的重要信息，来验证图结构的合理性。

2.3 基于 BERT 模型的关系抽取

近年来，大规模的预训练语言模型在很多任务上都达到了最佳的表现，其中比较典型的预训练语言模型就是 BERT[17]。基于 BERT 的模型比起基于 RNN 的模型和基于图结构的模型都在不同程度上表现优异 [14][15][2]。基于此，有人提出了 SpanBERT 模型来学习更好的表示信息 [15]，并且在 TACRED 数据集上表现达到了最优 [1]。TACRED 数据集是一个句子级的关系抽取数据集。对于对话级的关系抽取任务来说，有人认为将对话人的个人信息融入到对话本身可以增强 BERT 的表现能力，并提出了 BERTs 模型 [2]。本文也使用了 BERT 模型来提取多视图图结构的节点特征表示。

3 模型介绍

3.1 预备知识

3.1.1 形式化表示

假设 $X = \{x_1, x_2, \dots, x_T\}$ 是一个输入序列，其中 x_t 是序列中的第 t^{th} 个字/词， T 是句子中字词的总数，也是句子的长度。对于句子级别的关系抽取工作而言， X 是输入的句子；对于对话级别的关系抽取工作而言， X 是整段对话。也就是说，对于本文提出的模型而言，不明显区分句子和对话。为了预测两实体之间的关系，本文还定义了两个实体，主体 X_s 和客体 X_o 。主体和客体都是序列中的子序列。一个实体可能含有多个字/词。给定主体、客体和全体输入序列，我们的任务是提取出主体客体之间的关系 $r \in R$ ，其中 R 是所有与定义好的关系类别。

3.1.2 多视图图模型

对于一个多视图的图结构来说，对于同一对节点，存在多个不同的边，这些边并不是在同一张图内形成的。形式化来说，就是对于一个多视图的图 $G = \{V, A^1, A^2, \dots, A^N\}$ 来说， V 是全体节点的集合， A 是一个邻接矩阵， N 是视图的数目。

在本文中，多视图图模型用来构建给定文本中复杂的实体间关系。图中的每个节点都对应着一个字/词，这些字词有多种组合的方式，不同的组合方式构建成为不同的视图。通过对不同的视图进行分析编码，就可以最大程度上还原原文本中的语义关系。

3.2 BERT 模块

本文使用 BERT 模型作为特征的编码器，用 BERT 的编码向量来进行后续的抽取工作。选用 BERT 是因为 BERT 在表示学习上的高效性 [15][2]。给定含有 T 个字/词的序列 X ，我们将序列 X 转换成 BERT 模型需要的输入结构 $X_{input} = \{x_0, x_1, x_2, \dots, x_T, x_{T+1}\}$ 。这里的 x_0 表示 $[CLS]$ ，代表着一句话的开始。这里我们所说的一句话并不是狭义上的一个句子，而是送入模型的一个批次（Batch）。因为模型是一个批次一个批次来进行学习训练的，所以这里为了表述方便写作一句话。这里的 x_{T+1} 代表 $[SEP]$ ，代表着一句话的结束。与输入序列相对应的隐层结果表示为 $H = \{h_0, h_1, h_2, \dots, h_T, h_{T+1}\}$ 。其中， h_0 被当作是每句话的句子级特征向量，被直接送入 SoftMax 分类模型中进行预测类别。除此之外，本文将剩下的隐层向量送入图结构中，通过图节点和边的关系来建立更抽象层次的特征表示。最后，从图中产生的结果会被送入 SoftMax 层和 h_0 一起进行分类。

3.3 高斯图模块

从 BERT 模型中产生的输出表示会被分为两部分，其中 h_0 作为一句话的句子级特征，代表了这句话归属于什么样的任务，这是第一部分。另一部分是剩下的隐层表示 $H' = \{h_1, h_2, \dots, h_T, h_{T+1}\}$ 。在图结构部分，本文只用第二部分生成多视图的图模型。

首先，将最初始的节点向量表示为 $V^0 = \{v_1^0, v_2^0, \dots, v_{T+1}^0\}$ ，每一个节点都对应着原文本序列中的一个字/词。之后，使用高斯图生成器对多视图的图模型在给定节点集合的情况下进行边初始化。对于同一批节点来说，高斯分布的参数 μ 和 σ 分别表示为：

$$\{\mu_i^1, \mu_i^2, \dots, \mu_i^N\} = g_\theta(v_i^0) \quad (1)$$

$$\{\sigma_i^1, \sigma_i^2, \dots, \sigma_i^N\} = \phi(g'_\theta(v_i^0)) \quad (2)$$

其中， g_θ 和 g'_θ 是两个神经网络中需要训练的参数， ϕ 是一个激活函数，激活函数是用来将神经网络中线性的变换转换为非线性的一种特殊机制。 N 代表了多视图图模型中视图的数目。本文将激活函数 ϕ 设置为 SoftPlus 函数，因为高斯分布的标准差的值域是 $(0, \infty)$ 。因此，我们可以得到对于第 n 个视图的一系列高斯分布 $\{\mathcal{N}_1^n, \mathcal{N}_2^n, \dots, \mathcal{N}_{T+1}^n\}$ 。每一个高斯分布都是期望为 μ_i^n ，方差为 σ_i^{n2} 的分布函数，表示为 $\mathcal{N}_i^n(\mu_i^n, \sigma_i^{n2})$ 。

多视图图模型是用来捕捉全部的关系可能的，所以本文在每一个字/词之间都使用了信息传播方法。本文使用了 KL 散度来计算每两个字/词之间的高斯分布差异度，根据这个差异度来生成边的权重关系。对于第 i^{th} 和第 j^{th} 个节点来说，在第 n^{th} 个视图中，KL 散度的计算公式如下：

$$e_{ij}^n = \text{KL}(\mathcal{N}_i^n(\mu_i^n, \sigma_i^{n2}) || \mathcal{N}_j^n(\mu_j^n, \sigma_j^{n2})) \quad (3)$$

在计算完每个视图的每条边之后，我们就得到了一组临边矩阵 $\{A^1, A^2, \dots, A^N\}$ ，每一个临边矩阵都代表了一个视图。因此，通过高斯图模块生成的图可以表示为 $G = (V^0, A^1, A^2, \dots, A^N)$ 。

3.4 图卷积模块

本文在前人工作的基础上 [10][16]，探索了多视图训练工作的方法。使用稠密的图卷积神经网络来捕捉真实的图上结构信息。通过稠密的全连接层，我们可以训练出一个更深的模型来捕捉局部和全局的信息。多视图的图卷积网络可以写作：

$$v_{n_i}^{(l)} = \rho \left(\sum_{j=1}^T A_{ij}^n W_n^{(l)} k_j^{(l)} + b_n^{(l)} \right) \quad (4)$$

其中， $W_n^{(l)}$ 和 $b_n^{(l)}$ 是训练参数（权重和偏置值）。 ρ 是一个激活函数， $k_j^{(l)}$ 是全连接过程的初始节点表示向量，这个向量包含了之前序列所有节点的信息。第一次图卷积工作输出的结果为 $V^1 = \{v_1^1, v_2^1, \dots, v_{T+1}^1\} = \{k_1^{(1)}, k_2^{(1)}, \dots, k_{T+1}^{(1)}\}$ 。

卷积层和池化层在神经网络中往往成对出现并不断循环，本文选用的是 SAGPool 池化模型 [19]，计算公式如下：

$$s_{n_i} = \alpha \left(\sum_{j=1}^T A_{ij}^n W_{pool} v_j + b_{pool} \right) \quad (5)$$

4 实验验证

本文提出的基于图论的关系抽取模型可以应用于句子级和对话级的关系抽取任务。由于选用数据集的数据格式以及适用的基线模型是不同的，本文进行了两组实验来对模型的有效性进行了验证。以及对模型可能存在的修改空间进行了描述，以应对不同的任务类别。

4.1 实验设计

DialogRE 数据集是第一个人工标注的适用于对话级的关系抽取数据集 [2]，它包含有 1788 段《老友记》中的对话文本。DialogRE 数据集在此基础上人工定义了 36 种关系类型。表一给出了 DialogRE 中的一个示例。

本文根据最近大火的 BERT 模型 [17]，对本文提出的基于图论的关系抽取模型进行了改进。BERT 是一个预训练的语言模型，能够感知对话中说话人的信息，并可以根据将说话人的信息融入到对话本身来提取特征。在对话级的关系抽取任务上有不错的表

现。文本同时还针对 CNN 模型，基于 LSTM 的 RNN 模型分别进行了实验，以保证实验的完整性。

在试验参数的设置上，本文采用了相同的输入数据。值得一提的是，给定的 X , X_s 和 X_o 根据 BERT 的成文规则与 CLS 拼接在一起，并在每句话的结尾添加 SEP 标签。最终我们得到的序列为：

$$[CLS] - X - [SEP] - X_s - [SEP] - X_o - [SEP]$$

抽取对话的触发词不作特别标注，与原序列一起进行训练。

除此之外，优化器我们选择了 Adam[18]。学习率设置为 $3e - 5$ ，池化率设置为 0.7。我们使用了常见的 SAGPool 池化模型 [19] 作为本文的池化层。由于池化层的不断累加，最终图中的节点数量很少，采用传统的 $F1$ 值作为评价指标不具备普适性，所以我们采用了 $F1c$ [2] 指标作为模型试验的评价指标。输入上本文每批次仅将一组对话作为输入，而不是全部对话记录。

4.2 结果分析

表 2 展示了在 DialogRE 数据集上实验的结果。通过这个结果我们可以看出，基于 BERT 模型的方法比基于 CNN 模型的和基于 LSTM 的 RNN 模型的效果有明显的提高。融入了对话人身份信息的 BERT 模型比标准的 BERT 模型效果更好。基于这个想法，本文提出的基于图论的关系抽取模型是在融入了对话人身份信息的 BERT 模型的基础上进行的扩展。实验结果表明，本文提出的方法比 BERTs 的效果，在 $F1$ 值上高 5.9%，在 $F1c$ 值上高 1.7%。

Model	$F1$	$F1c$
CNN[20]	48.1%	46.7%
LSTM-RNN[21]	49.7%	46.2%
BERT[17]	66.8%	60.4%
BERTs[2]	73.0%	65.8%
Our Model	78.9%	67.5%

表 2：针对 DialogRE 数据集的对比实验

-
- S1:** Hey, you guys! Look **what** I found! Look at this! **That's my** Mom's writing!
Look.
- S2:** Me **and Frank** and Phoebe, Graduation 1965.
- S1:** Y'know what that **means**?
- S2:** That you're actually 50?
- S1:** No-no, that's **not**, that's not me Phoebe, **that's** her pal Phoebe. According to
her high school yearbook, they were like B.F.F. Best **Friends** Forever.
-

	Argument Pair	Trigger	Relation type
R1	(S1, Frank)	high school	per:alumni

表 3: DialogRE 测试数据集示例

同时，根据表 3 我们可以发现，通过本文提出的模型，在 DialogRE 数据集上选取的这一对话，我们可以选择出对关系抽取有用的触发词，如：high school。通过逐步减少对关系抽取无用的实体，使得模型在提取关系时有了更高的准确率。

5 总结和展望

本文提出了一个基于图论的关系抽取模型用来抽取对话级别的实体关系。模型很好地解决了在长序列文本中难以抽取出特定的实体以及实体之间的关系的问题。在对话级别的关系抽取工作中，通过实验验证，能够达到不错的实验效果。特别的，本文提出的模型包含有预训练语言模块——BERT 模块、图结构模块——高斯图模块和图卷积模块。通过高斯图模块和图卷积模块，模型可以很好的模拟长序列中不同实体之间的可能的关系并做卷积学习。这种机制在处理长序列方面有很大的潜力。

在将来，应当更关注于这一模型在其他级别的数据上的泛化能力以及讨论模型的可解释性。除了关系抽取工作，命名实体识别、事件抽取以及共指解析工作同样是知识抽取工作的其他研究重点，讨论如何将关系抽取模型泛化到其他三个子任务中也是将来值得探索的重点。

参考文献

- [1] Zhang, Y.; Zhong, V.; Chen, D.; Angeli, G.; and Manning, C. D. 2017. *Position-aware Attention and Supervised Data Improve Slot Filling*. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 35–45. Copenhagen, Denmark: Association for Computational Linguistics.
- [2] Yu, D.; Sun, K.; Cardie, C.; and Yu, D. 2020. *Dialogue-Based Relation Extraction*. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 4927–4940. Online: Association for Computational Linguistics.
- [3] Tarjan R E. *Graph theory and Gaussian elimination*[J]. Sparse Matrix Computations, 1976: 3-22.
- [4] Bagdanov A D, Worring M. *First order Gaussian graphs for efficient structure classification*[J]. Pattern Recognition, 2003, 36(6): 1311-1324.
- [5] Johnson D, Sinanovic S. *Symmetrizing the kullback-leibler distance*[J]. IEEE Transactions on Information Theory, 2001.
- [6] Miwa, M.; and Sasaki, Y. 2014. *Modeling Joint Entityand Relation Extraction with Table Representation*. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1858–1869. Doha, Qatar: Association for Computational Linguistics.
- [7] Gormley, M. R.; Yu, M.; and Dredze, M. 2015. *Improved Relation Extraction with Feature-Rich Compositional Embedding Models*. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 1774–1784. Lisbon, Portugal: Association for Computational Linguistics.
- [8] Zhou, P.; Shi, W.; Tian, J.; Qi, Z.; Li, B.; Hao, H.; and Xu, B. 2016. *Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification*. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 207–212. Berlin, Germany: Association for Computational Linguistics.
- [9] Zhang, Y.; Qi, P.; and Manning, C. D. 2018. *Graph Convolution over Pruned Dependency Trees Improves Relation Extraction*. In Proceedings of the 2018 Conference on Empirical

- Methods in Natural Language Processing, 2205–2215. Association for Computational Linguistics.
- [10] Guo, Z.; Zhang, Y.; and Lu, W. 2019. *Attention Guided Graph Convolutional Networks for Relation Extraction*. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 241–251. Florence, Italy: Association for Computational Linguistics.
- [11] Christopoulou, F.; Miwa, M.; and Ananiadou, S. 2018. *A Walk-based Model on Entity Graphs for Relation Extraction*. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 81–88. Melbourne, Australia: Association for Computational Linguistics.
- [12] Nan, G.; Guo, Z.; Sekulic, I.; and Lu, W. 2020. *Reasoningwith Latent Structure Refinement for Document-Level Relation Extraction*. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 1546–1557. Online: Association for Computational Linguistics.
- [13] Guo, Z.; Nan, G.; LU, W.; and Cohen, S. B. 2020. *Learning Latent Forests for Medical Relation Extraction*. In Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20, 3651–3657. International Joint Conferences on Artificial Intelligence Organization.
- [14] Wu, S.; and He, Y. 2019. *Enriching Pre-Trained Language Model with Entity Information for Relation Classification*. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, CIKM ’ 19, 2361–2364. New York, NY, USA: Association for Computing Machinery.
- [15] Joshi, M.; Chen, D.; Liu, Y.; Weld, D. S.; Zettlemoyer, L.;and Levy, O. 2020. *SpanBERT: Improving Pre-training by Representing and Predicting Spans*. Transactions of the Association for Computational Linguistics8: 64–77.
- [16] Guo, Z.; Zhang, Y.; Teng, Z.; and Lu, W. 2019. *Densely Connected Graph Convolutional Networks for Graph-to-Sequence Learning*. Transactions of the Association for Computational Linguistics7: 297–312.

- [17] Devlin J, Chang M W, Lee K, et al. *Bert: Pre-training of deep bidirectional transformers for language understanding*[J]. arXiv preprint arXiv:1810.04805, 2018.
- [18] Kingma, D. P.; and Ba, J. 2015. *Adam: A Method for Stochastic Optimization*. In International Conference on Learning Representations
- [19] Lee, J.; Lee, I.; and Kang, J. 2019. *Self-Attention Graph Pooling*. In Proceedings of the 36th International Conference on Machine Learning, volume 97 of Proceedings of Machine Learning Research, 3734–3743. Long Beach, California, USA: PMLR.
- [20] Lawrence, S.; Giles, C. L.; Ah Chung Tsoi; and Back, A. D. 1997. *Face recognition: a convolutional neural-network approach*. IEEE Transactions on Neural Networks 8(1): 98–113.
- [21] Hochreiter, S.; and Schmidhuber, J. 1997. *Long short-term memory*. Neural computation 9(8): 1735–1780.