

依存约束的图网络实体关系联合抽取^①



任鹏程, 于 强, 侯召祥

(中国石油大学(华东)计算机与技术科学学院, 青岛 266580)

通讯作者: 任鹏程, E-mail: rsanshierli@163.com

摘 要: 实体关系抽取是信息抽取的关键任务之一, 是一种包含实体抽取和关系抽取的级联任务. 传统的实体关系抽取方式是将实体与关系抽取任务分离的 Pipeline 方式, 忽略了两个任务的内在联系, 导致关系抽取的效果严重依赖实体抽取, 容易引起误差的累积. 为了规避这种问题, 我们提出一种端到端的实体关系联合抽取模型, 通过自注意力机制学习单词特征, 基于句法依存图蕴含的依赖信息构建依存约束, 然后将约束信息融入图注意力网络来实现实体与关系的抽取. 通过在公共数据集 NYT 上进行实验证明了我们工作的先进性和显著性, 我们的模型在保持高精度的情况下, 召回率有了显著的提升, 比以往工作中的方法具有更好的抽取性能.

关键词: 实体关系联合抽取; 依存约束; 图注意力网络; 自注意力机制

引用格式: 任鹏程, 于强, 侯召祥. 依存约束的图网络实体关系联合抽取. 计算机系统应用, 2021, 30(3):24-32. <http://www.c-s-a.org.cn/1003-3254/7825.html>

Graph Network with Dependency Constraints for Joint Entity and Relationship Extraction

REN Peng-Cheng, YU Qiang, HOU Zhao-Xiang

(College of Computer Science and Technology, China University of Petroleum, Qingdao 266580, China)

Abstract: Entity relationship extraction is one of the key tasks of information extraction, which involves a multi-task cascade including entity extraction and relationship extraction. Traditional methods of entity relationship extraction follow a mode of Pipeline which separates entity extraction from relationship extraction, ignoring the internal connection between the two. As a result, the effect of relationship extraction depends heavily on entity extraction, and it is prone to error accumulation. To avoid this problem, we propose an end-to-end joint entity and relationship extraction model, which relies on the self-attention mechanism to learn word features, constructs dependency constraints based on dependency information contained in syntactic dependency graphs, and then integrates constraint information into a graph attention network for entity and relationship extraction. Experiments on the public data set NYT demonstrate the advance and significance of our model which has a high recall rate and better extraction performance than previous methods.

Key words: joint entity and relationship extraction; dependency constraint; graph attention network; self-attention mechanism

信息抽取是从自然语言中提取出特定需求信息的过程, 例如从医疗文档中提取疾病名称、疾病原因、药物名称信息, 或者从新闻文本中提取时间、地点、人物信息等, 从而对文本数据进行分类、提取、重构. 其

中关键任务之一就是提取形为 SPO (Subject, Prediction, Object) 的实体关系三元组, 从而将无结构数据转换为结构化的可用形式.

传统上实体关系抽取使用 Pipeline 方式, 即首先通

① 收稿时间: 2020-07-16; 修改时间: 2020-08-13; 采用时间: 2020-08-21; csa 在线出版时间: 2021-03-03

过命名实体识别模型提取实体, 然后对提取的实体进行关系分类。Pipeline 形式简单、操作灵活, 但是任务的级联特性会导致模型误差的累积, 命名实体识别模型的错误预测会传播到关系抽取模型, 影响关系分类任务的性能。同时在关系分类任务前, 需要对命名实体识别的结果进行实体对的两两配对, 如果匹配的实体对本身不存在关系则产生“噪声”, 引起模型错误率及计算复杂度的增加。联合抽取即将命名实体识别模型和关系分类模型合并为一个模型进行统一建模, 考虑两个任务之前潜在关联, 以此削弱级联误差传播影响, 文本中所有实体、实体与关系、关系与关系的交互信息学习的越多, 联合模型的效果就会越好。

关系抽取还面临重叠关系的挑战, 即实体间存在多种对应方式, 一对多或者多对多等情况, 例如“Quebec, Canada’s second most populous province, after Ontario, has not decided to go that far.”, 包含 3 个实体关系三元组 (Ontario, country, Canada)、(Canada, Contains, Ontario) 以及 (Canada, administrative_divisions, Ontario), 其中 (Canada, Ontario) 实体对对应两个关系, 预测时就会面临只预测出一种关系的情况。重叠关系的问题在基于知识库的抽取方式中相对容易解决, 但是通过模型进行联合抽取依旧面临挑战。

语句的依存分析是 NLP 中处理文本的关键技术之一, 其目标是确定句子的句法结构和各单词间的依存关系。图神经网络是图结构数据建模的有效模型, 图注意力网络 (Graph Attention Network, GAT)^[1] 是一种空域图神经网络, 适用于动态图的处理, 有利于将顶点特征之间的关联性更好得融合进模型。

为优化传统 Pipeline 方式并实现高效的实体关系联合抽取模型, 本文综合依存分析和图注意力网络特性建立依存约束的图注意力网络 DCGAT (Graph Attention Network with Dependency Constraints) 联合抽取模型。将句法结构拓扑到句子中各个单词间的内在联系, 通过图注意力网络进行约束依赖的学习, 并采用自注意力机制进行单词底层特征学习, 实现实体与关系的联合抽取。

本文主要贡献为以下 3 个方面:

1) 通过对句子中各字词的依存分析动态建立依存分析图, 图顶点表示文本中的单词, 图的边表示各单词间的依赖关系, 建模句子中所有实体、实体与关系、关系与关系间的依赖, 形成依存约束。

2) 提出一种端到端的实体关系联合抽取模型 DCGAT, 通过多头自注意力机制学习单词多层次表示, 通过图注意力网络融合依存约束矩阵进行句法依存图节点的权重计算, 动态学习实体关系的特征表示, 实现实体与关系的联合抽取。

3) 将依存分析与图注意力网络进行结合, 深入学习单词特征表示, 充分考虑句子中各个单词依赖关系, 显著提升关系预测的 Recall, 从而改善重叠关系的预测。

在公共关系抽取 NTY 数据集上进行多次实验, 实验结果表明, 本文提出的方法显著提升了模型的预测效果。

1 相关工作

1.1 实体关系联合抽取

现有的实体关系联合抽取模型中, 实体与关系的抽取顺序依据模型特点各有差异。Miwa 和 Bansal^[2] 提出 BiLSTM-TreeLSTM 模型进行实体和关系的共同建模, 模型通过 Bi-LSTM 学习的字词特征由实体标记器和最短依赖路径关系分类器共享, 但是模型在抽取关系时需通过 Pipeline 传递标记器抽取的实体, 形成实体对再抽取关系, 并且未解决关系重叠问题。Dixit 等^[3] 提出片段排列的方式解决实体重叠问题, 模型在 Span 范围内通过注意力机制与原始 LSTM 编码进行特征学习, 然后将所有的实体 Span 并行输入 Softmax 进行关系解码, 不过模型计算复杂度极高, 关系解码时也存在冗余发生。Bekoulis 等^[4] 提出多头选择机制, 将关系抽取当作多头选择任务执行, 该机制提高了多重关系的抽取能力, 实体抽取则通过 CRF 进行线性建模, 所以模型对于被其他实体或者非实体词间隔的实体无法进行处理。

不同于上述模型先实体后关系的抽取顺序, Zeng 等^[5] 提出一种联合抽取的 Seq2Seq 框架和 Copy 机制, 依次抽取关系、头实体和尾实体。文献中使用 Bi-LSTM 进行句子编码, 使用 LSTM 解码关系并从文本中复制两个单词进行三元组动态解码。然而, 模型在复制实体时忽略了多个单词构成的实体组合情况。Katiyar 与 Cardie^[6] 提出一种同时抽取实体与关系的指针网络, 识别出当前位置的实体之后进行前向回溯, 查询已识别实体并计算注意力得分并进行关系分类, 但查询回溯只能判断已知实体, 存在实体遗漏问题。

此外, Li 等^[7] 放弃三元组抽取结构, 将实体关系抽

取问题转换为多轮对话问题,将各实体和关系通过人工问答模板进行重构,并应用强化学习优化多轮对话惩罚奖励实现联合抽取,然而通常情况下问答模板的重构编码需要更高的时间复杂度。

1.2 依存分析与图神经网络

依存分析任务的方式一般有3种:基于规则的方法,基于统计的方法和基于深度学习的方法。目前多数工作将依存分析作为其他任务的子任务来进行。Jia等^[8]在他们的研究工作中提出了一种基于依存语义范式(DSNFs)的无监督中文开放式关系抽取模型,在不考虑实体关系位置因素的情况下通过提取动名词之间的关系来实现关系识别。

自从图神经网络发展起来以后,其在自然语言处理方面的研究多样化。Zhang等^[9]利用GCN编码依存树信息,并提出path-centric pruning方法对依存树中无关path进行剪枝,优化实体、关系冗余问题。Zhu等^[10]提出具有生成参数的图神经网络进行关系抽取,使GNN能够处理非结构化文本输入上的关系推理。Fu等^[11]的研究工作中将实体关系的联合抽取分成两阶段,第一阶段通过Bi-LSTM和GCN对文本字词和依赖特征进行学习,抽取普通三元组,第二阶段重构隐状态输出为全连接图,通过加权GCN模型实现重叠关系的再次学习,但其使用Bi-LSTM进行特征抽取,提取效益相较自注意力机制不足。文本语句的依存分析图天然是一种动态的图结构,相比GCN的研究工作^[9-13],GAT自身更适用于动态图的任务,例如许晶航等^[14]的工作中将GAT应用于因果关系的抽取。

1.3 自注意力机制

注意力机制最早应用于图像领域,Bahdanau等^[15]的工作中将类似Attention的机制应用在机器翻译任务上,开创了注意力机制在NLP使用的开端。随后Google的“Attention is All You Need”^[16]工作提出的自注意力机制,相比传统RNN、CNN在自然语言处理中的应用效果有极大的提升,Tang等^[17]、凡子威等^[18]以及李明扬等^[19]的工作也论证了自注意力机制的语义抽取能力的先进性。

本文依托自注意力机制的强大特征学习能力,融合依存分析与图注意力网络的结构优势,编码依存约束矩阵充分建模文本中实体与关系的相互作用,建立DCGAT模型实现实体与关系的联合抽取。

2 实体关系联合抽取模型

本文提出的DCGAT模型总体架构如图1所示,原始文本(original text)编码成词向量(word)和位置向量(pos),由多头自注意力机制(multi self-attention)计算单词特征与语义特征的注意力得分(score),通过依存分析图(semantic dependency graph)构建依存约束矩阵(dependency constrain matrix),将约束矩阵应用于图注意力网络(GAT)进一步学习句子中各单词的潜在关系,实现实体与关系的联合抽取。

2.1 多头自注意力机制

模型采用多头自注意力机制自动学习句子的单词特征及语义特征,并堆叠 N 个多头自注意力块进行深入编码。首先将词向量与位置向量组合嵌入为初始输入特征,词向量采用Glove预训练词向量,位置向量由positional encoding进行编码,具体方式为采用sine和cosine函数进行直接计算,

$$PE_{(pos,2i)} = \sin(pos/10\,000^{2i/d_{model}}) \quad (1)$$

$$PE_{(pos,2i+1)} = \cos(pos/10\,000^{2i/d_{model}}) \quad (2)$$

式中, i 表示词向量的位置, pos 表示单词在句子中的位置, d_{model} 表示词向量维度。线性变换层映射得到Query、Key、Value向量,分别由 Q 、 K 、 V 表示:

$$Q = W_Q X, K = W_K X, V = W_V X \quad (3)$$

然后通过Dot-Product Attention进行权重计算。Dot-Product Attention计算公式为:

$$DotProductAtt(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

其中, d_k 为词向量维度,约束内积值不宜过大。

为保证模型可以在不同的子表示空间学习到不同的空间信息,将注意力机制拓展到多头计算方式。

$$head_i = DotProductAtt(QW_i^Q, KW_i^K, VW_i^V) \quad (5)$$

$$\begin{aligned} MultiHeadAtt(Q, K, V) \\ = Concat(head_1, head_2, \dots, head_h)W \end{aligned} \quad (6)$$

通过多头计算得到句子中单词和语义表示的注意力得分。

2.2 依存约束

句子的依存分析是对句子各成分进行依赖关系的分析来解释句法结构,可通过依存分析图进行可视化。

如图2所示,对句子进行依存分析,默认创建“Root”,

通过“Root”有向弧指向“Beijing”, “capital”和“China” 分别有多条指向其他单词的有向弧, 表示“capital”和

“China”与其他词的依赖关系, 如图中显示的“SCO”、“FEAT”、“mRELA”等.

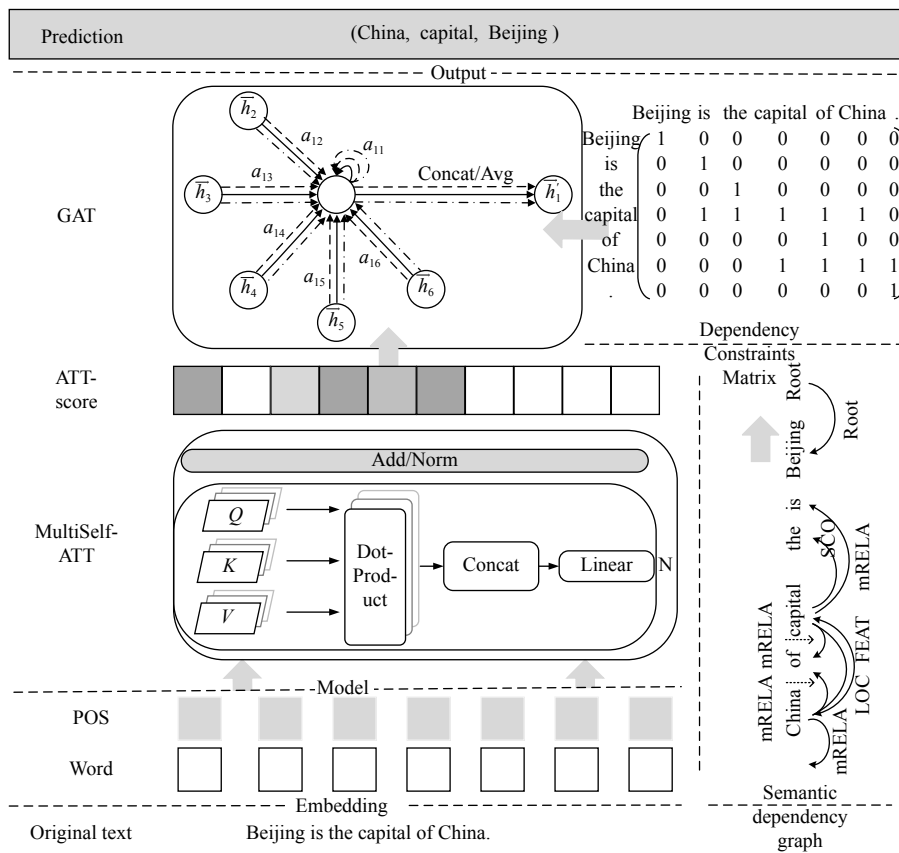


图1 DCGAT 总体架构图

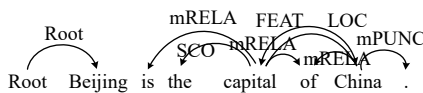


图2 依存分析图

根据依存分析图的结果, 构建依存约束矩阵, 如图3所示. 矩阵构建的规则为:

1) 矩阵维度为句子元素长度, 行表示目标节点(单词), 列表示邻域节点(单词), 矩阵值表示目标节点与邻域节点的几何关系, 即单词间的依赖情况, 对于输入的

各个句子动态构建约束矩阵.

2) 忽略“Root”有向弧, 其他句法依存图中的边依赖情况在矩阵中相应位置以“1”表示连接关系, 无有向弧连接以“0”填充. 因重点关注是否有依赖关系, 所以不考虑相关依赖关系类型, 即依存矩阵中不存储依存分析图中的有向弧标签信息(如“FEAT”等). 例如, “capital”与“is”之间存有有向弧连接, 则矩阵中(capital, is)位置由“1”表示, 而“is”没有指向“capital”的边存在, 则相应位置(is, capital)处赋值为“0”.

	Beijing	is	the	capital	of	China	.
Beijing	1	0	0	0	0	0	0
is	0	1	0	0	0	0	0
the	0	0	1	0	0	0	0
capital	0	1	1	1	1	1	0
of	0	0	0	0	1	0	0
China	0	0	0	1	1	1	1
.	0	0	0	0	0	0	1

(a) 考虑自身依赖

	Beijing	is	the	capital	of	China	.
Beijing	0	0	0	0	0	0	0
is	0	0	0	0	0	0	0
the	0	0	0	0	0	0	0
capital	0	1	1	0	1	1	0
of	0	0	0	0	0	0	0
China	0	0	0	1	1	0	1
.	0	0	0	0	0	0	0

(b) 不考虑自身依赖

图3 依存约束矩阵

3) 对于单词自身与自身的依赖情况, 若依存分析图中不存在有向弧连接, 在构建依存约束矩阵时, 分为两种方式: 一是考虑单词自身约束, 矩阵中 ($word_i, word_i$) 位置赋值为“1”, 二是不考虑自身依赖, 对应位置赋值为“0”。

2.3 图注意力网络

图注意力网络是空域图神经网络的代表模型, 从对应空间结构学习目标节点与邻域节点的几何关系. GAT 对输入图信息进行注意力计算, 对目标节点与邻域节点计算注意力得分以分配不同的权重, 从而提升图结构中影响较大节点的关注度并降低影响较小节点的关注度。

图注意力网络的计算过程分为两步: 第一步计算注意力系数. 将多头自注意力机制提取的单词特征通过权重矩阵 W 进行线性映射, 将映射后的顶点 i 与邻域节点 j 的特征向量进行拼接 (由“||”表示), 通过单前馈网络层 $f(\cdot)$ 进行转换, 完成目标顶点与其邻域节点的相似系数 e_{ij} 的计算。

$$e_{ij} = f\left(\left[Wh_i \parallel Wh_j\right]\right) \quad (7)$$

将相似系数通过 $LeakyReLU(\cdot)$ 非线性函数进行激活, 将激活后的矩阵通过依存矩阵进行 $Mask(\cdot)$ 约束, 强化单词之间依赖关系的影响, 并且可以稀疏化矩阵加速运算. 最后通过 $Softmax$ 进行归一化, 完成注意力系数的计算. 注意力系数 (α_{ij}) 的计算过程如图 4 所示。

$$\alpha_{ij} = Softmax\left(Mask\left(LeakyReLU\left(e_{ij}\right)\right)\right) \quad (8)$$

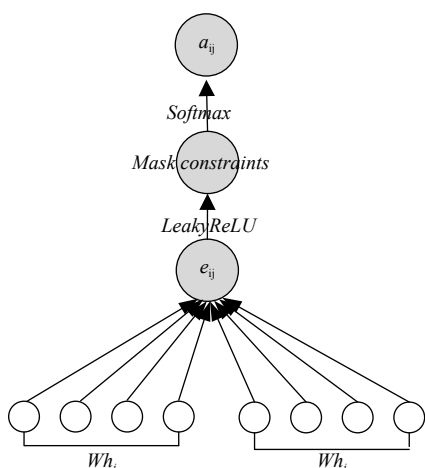


图 4 注意力系数计算

第二步计算为注意力系数的聚合. 将计算完成的注意力系数进行加权、激活得到顶点特征 h_i , 当前对于

每个顶点 i 都聚合了其依赖特征信息。

$$h_i = \sigma\left(\sum \alpha_{ij} Wh_j\right) \quad (9)$$

为提高模型的健壮性, 图注意力网络中的注意力计算同样采用多头计算的方式, 由式 (10) 将 K 个相互独立的注意力系数进行 Concat 和 Average 计算, 得到最终的输出结果. 多头计算过程如图 1 中“GAT”部分所示, 线条数代表 K 值, 例如图中表示 $K=3$ 的多头计算。

$$h_i = \parallel_{k=1}^K \sigma\left(\sum \alpha_{ij}^k\right) W_k h_j = \sigma\left(\frac{1}{K} \sum_{k=1}^K \alpha_{ij}^k W_k h_j\right) \quad (10)$$

[\left\{ \right\} \right]

2.4 模型训练与推理

DCGAT 模型整体的训练损失包括两部分: 命名实体损失 $loss_e$ 和关系分类损失 $loss_r$. 两者都由分类损失 CrossEntropy 函数进行计算, 总体损失 $Loss_{model}$ 为命名实体识别损失与关系分类损失之和, 二者损失权重系数比为 1:1. 实体标签由“BIESO”方法进行标注, 损失计算为对应标注的判断. 关系标签由数据集中所有类别通过 LabelEncoder 进行映射, 关系损失基于预测的实体对 (Subject, Object) 进行计算。

$$Loss_{model} = loss_e + loss_r \quad (11)$$

模型的 Inference 解码采用阈值方法. 将所有单词对都作为候选分别进行判断, 当模型中的实体、关系类别解码概率大于阈值时, 选取为 Inference 结果. 阈值可自定义初始化, 根据需要进行权衡调整。

模型效果的评估采用标准的 $F1$ 分数对预测的三元组 SPO 进行判别, $Precision$ 代表模型预测效果, $Recall$ 代表模型的预测精度, 三者计算公式如式 (12)~式 (14) 所示。

$$Precision = \frac{\sum_{i=1}^n sent_i \text{中预测正确的SPO个数}}{\sum_{i=1}^n sent_i \text{所有预测到的SPO个数}} \quad (12)$$

$$Recall = \frac{\sum_{i=1}^n sent_i \text{中预测正确的SPO个数}}{\sum_{i=1}^n sent_i \text{应包含的SPO个数}} \quad (13)$$

$$F1-Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (14)$$

3 实验分析

3.1 实验数据及参数

实验数据来自 Zeng 等^[5] 工作中使用的 NYT 关系抽取数据集, 关系类别数为 25 类, 其中包括 24 种实体

关系类型及“None”关系类型。训练集大小为 56 196, 验证集大小为 5000, 测试集大小为 5000。训练集中部分关系类型数据量如表 1 所示, 样本格式如图 5 所示, 包含句子信息“sentText”、关系信息“relationMentions”和实体信息“entityMentions”等实体通过“BIESO”标注法与实体类型组合标注, 依存分析通过 spaCy^[20] 获取依赖关系, 并对数据集中的单词、关系、依赖类型进行向量化编码。

```
{'sentText': 'Of Boca Raton , Florida , peacefully on April 30 .',
'articleId': '/m/vinci8/data1/riedel/projects/relation/kb/nyt1/docstore/nyt-2005-2006.backup/1758
823.xml.pb',
'relationMentions': [{'em1Text': 'Florida',
'em2Text': 'Boca Raton',
'label': '/location/location/contains'}],
'entityMentions': [{'start': 0, 'label': 'LOCATION', 'text': 'Boca Raton'},
{'start': 1, 'label': 'LOCATION', 'text': 'Florida'}],
'sentId': '1'}
```

图 5 数据集字段

模型词向量使用预训练的 GloVe(300d) 作为固定的词向量表示, POS 位置向量通过式 (1)、式 (2) 进行编码, 两者拼接作为每个单词的最终模型输入表示。多头自注意力机制中 N 设置为 3, 表示三层多头自注意力的堆叠, 在构建依存约束矩阵时考虑自身依赖, 即图 3(a) 所示矩阵构建方式, 图注意力网络中多头 K 设置为 6。初始学习率设置为 $5e-5$, 使用 Adam 优化器进行梯度优化。

3.2 模型效果评估

将 DCGAT 模型的效果与以往相关工作^[5,11,21] 进行了比较, 其中 Zheng 等^[21] 的研究提出 Tagging Scheme, 将命名实体识别的序列标注任务和关系抽取的分类任务合并成了一个序列标注问题, 把关系类型集成在序列中一同标注; Zeng 等^[5] 的工作在 Seq2Seq 框架中采用 Copy Mechanism, 提出 OneDecoder 和 MultiDecoder 模型动态解码提取关系三元组; Fu 等^[11] 的工作中 GraphRel 在两阶段中应用图卷积神经网络到关系提取中。对比模型的数据来源于原始论文, 抽取结果以三元组 SPO 表示, 当且仅当两个对应实体的及关系与真实标签相同则判断为正确。

由表 2 各个模型的实体关系抽取效果可以看出, DCGAT 模型的 $F1-Score$ 相比以往的工作有了明显的提升。相比以往工作中的最优模型 GraphRel-2, DCGAT 模型的 $F1-Score$ 提升了 2.5%, $Precision$ 上提升 1.2%, $Recall$ 改变最大, 提升 3.9%。源于自注意力机制的强大

表 1 部分关系类型数据量

Relation	Count	Relation	Count
/location/location/ contains	45 609	/people/person/ place_of_birth	2814
/business/person/ company	4951	/people/person/ place_lived	6363
/sports/sports_team_ location/teams	187	/business/company/ place_founded	361
/business/company/ advisors	40	/people/person/ ethnicity	21

特征提取能力, 以及融合依存关系的图注意力网络的约束性学习, 模型对于实体边界及重叠关系的判断更加准确, 使得 $Recall$ 提升尤其明显, 这也对整体效果 $F1-Score$ 的提升产生了关键性的影响。

表 2 模型实体关系抽取效果

模型	$Precision$	$Recall$	$F1-Score$
Tagging Scheme ^[21]	0.624	0.317	0.420
Copy Mechanism-One ^[5]	0.594	0.531	0.560
Copy Mechanism-Multi ^[5]	0.610	0.566	0.587
GraphRel-1 ^[11]	0.629	0.573	0.600
GraphRel-2 ^[11]	0.639	0.600	0.619
DCGAT (本文)	0.651	0.639	0.644

3.3 模型实验分析

(1) 约束矩阵构建方式分析

在表 2 使用的 DCGAT 模型中, 依存约束矩阵的创建考虑每个单词的自我约束, 这一部分工作中我们对比了不同约束矩阵的构建方式以及无关系约束 (即不考虑依存关系, 对应关系矩阵全部取值为 1, 不产生 Mask 约束作用) 产生的影响及模型效果。另外, 将依存图中的有向边依赖拓展至无向边, 在只考虑是否连接的情况下构建无向约束矩阵一同进行实验对比, 无向约束矩阵如图 6 所示。

由图 7 可以看出, 考虑自身依赖的约束矩阵的 $F1-Score$ 整体效果最好, 不考虑自身依赖的情况次之, 原因是不考虑自身约束的情况下, 节点丢失了相应自身信息。对于转换而成的无向约束矩阵, 我们猜测, 依赖

信息的扩增从而获得了更多节点特征,使得模型对于实体识别的判断更加敏感, *Precision* 相比考虑自身依赖的情况还有高出 0.002,但是这种扩增信息丢失了有向的依赖指向关系,对于三元组中含有多个关系时弱化了对应差别,所以 *Recall* 效果显著降低。

	Beijing	is	the	capital	of	China	.
Beijing	1	0	0	0	0	0	0
is	0	1	0	1	0	0	0
the	0	0	1	1	0	0	0
capital	0	1	1	1	1	1	0
of	0	0	0	1	1	1	0
China	0	0	0	1	1	1	1
.	0	0	0	0	0	1	1

图6 无向约束矩阵

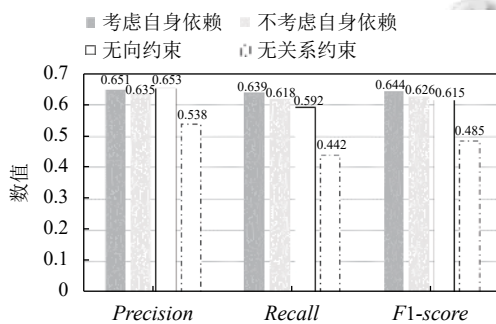


图7 不同约束矩阵模型对比

(2) 注意力多头数分析

为了验证注意力网络中多头的计算方式给模型带来的影响,实验分别对比了自注意力多头 N 和图注意力多头 K 的不同取值产生的模型效果. N 、 K 分别取值 1、3、6,实验结果如表 3 所示。

表3 不同多头取值模型效果

指标	$N=1, K=1$	$N=1, K=3$	$N=1, K=6$	$N=3, K=1$	$N=3, K=3$	$N=3, K=6$	$N=6, K=1$	$N=6, K=3$	$N=6, K=6$
<i>Precision</i>	0.620	0.623	0.623	0.623	0.642	0.651	0.625	0.644	0.635
<i>Recall</i>	0.607	0.615	0.618	0.610	0.629	0.639	0.609	0.632	0.628
<i>F1-Score</i>	0.613	0.618	0.620	0.616	0.635	0.644	0.617	0.637	0.631

由表 3 看出,不同的头取值对于模型的效果都会产生不同的影响. Multi Self-Attention 中 $N=3$ 时相对效果最好, GAT 中 $K=6$ 时相对效果也最好. 实验结果表明, K 、 N 的最优值属于一定的范围,取值较小为 1 时模型的注意力的提取能力相对较弱,不同空间的不同特征信息考虑相对不足;取值变大时,可以捕捉到更多位于不同分布空间的特征信息,随之模型的效果得到

提升.但是, K 、 N 取值都取最大为 6 时效果不如 $K=3$, $N=6$ 的情况,这表明更多的头也可能带来信息的冗余。

(3) 阈值取值分析

模型解码阶段采用阈值判别方法,实验对比阈值为 0.3 到 0.9 之间的不同效果,实验效果如图 8 所示.由图可以看出,随着阈值不断增大, *Precision* 也增大,模型对于实体关系边界的正确选择概率扩大,在 0.5 和 0.6 左右取值达到最优,伴随阈值的继续增大,模型对于边界判断强严格,模型的 *Precision* 效果随之降低。

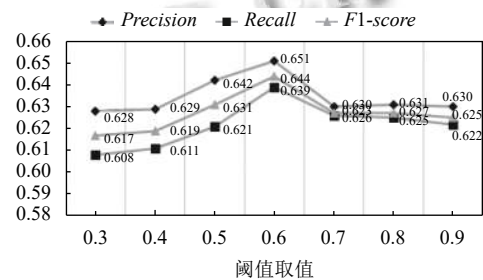


图8 不同阈值取值对模型效果的影响

对于 *Recall*,随着阈值的逐渐增加并趋于相对稳定状态,较大阈值控制模型的判断准确性,保证句子中应包含的关系被正确识别,相对提升重叠关系的预测效果. 阈值法有效地调整了不同阈值选择下 *Precision* 和 *Recall* 之间的平衡,保证最终的 *F1-Score* 达到最优。

(4) 重叠关系分析

我们还实验对比了 DCGAT 模型与文献 [5,11] 工作对于重叠关系的预测效果.表 4 是挑选的两个典型的句子样例及真实标签,其中包含多个重叠关系。

表4 原始句子及真实标签

样例	Original Text	True Label
Sent1	The bungalows at Fort Aguada, in Goa, India.	(India, contains, Fort Aguada)
		(India, contains, Goa)
Sent2	PAGE 10 The Enigma of Damascus President Bashar al-Assad and his Western-educated wife say they are transforming Syria into an open society.	(India, administrative_divisions, Goa)
		(Goa, country, India)
		(Goa, contains, Fort Aguada)
		(Bashar al-Assad, place_of_birth, Damascus)
		(Bashar al-Assad, nationality, Syria)
		(Syria, administrative_division, Damascus)
		(Syria, contains, Damascus)
		(Syria, capital, Damascus)
		(Damascus, country, Syria)

表 5 为 3 个模型的实验对比, 后者分析结果源于其工作开源的官方代码, 参数默认使用官方设置, 文

献 [21] 中未开源官方代码暂不进行实验对比。

由表 5 的对比结果分析可知, DCGAT 模型的预测效果最好, 其对于三元组的预测更加接近真实标签。在样例“The bungalows at Fort Aguada, in Goa, India.”中, Copy Mechanism-Multi 模型仅仅预测到两个关系, “India”与“Goa”同样具有“contains”关系, 却没有得到

充分学习; GraphRel-2p 引入 GCN 及依存树的关系提升了效果, 预测出了“India”与“Goa”的关系, 但是未预测到同实体对的“administrative_division”关系类型; 本文中的 DCGAT 预测效果最好, 对于“India”与“Goa”实体对成功预测到“administrative_division”的关系类型。

表 5 模型预测结果

样例	Copy Mechanism-Multi	GraphRel-2p	DCGAT
Sent1	(India, contains, Fort Aguada)	(India, contains, Fort Aguada)	(India, contains, Fort Aguada)
	(Goa, contains, Fort Aguada)	(Goa, contains, Fort Aguada)	(India, contains, Goa)
		(India, contains, Goa)	(India, administrative_divisions, Goa)
Sent2	(Bashar al-Assad, place_of_birth, Damascus)	(Bashar al-Assad, place_of_birth, Damascus)	(Bashar al-Assad, place_of_birth, Damascus)
	(Syria, contains, Damascus)	(Syria.administrative_division, Damascus)	(Syria administrative_division, Damascus)
	(Syria, capital, Damascus)	(Syria, contains, Damascus)	(Syria, contains, Damascus)
			(Syria, capital, Damascus)
			(Damascus, country, Syria)

4 结论

本文提出了依存约束的图注意力网络实体关系联合抽取模型 (DCGAT), 通过多头自注意力机制学习单词表示及语义关系, 借助语义分析中的依存分析图构建约束矩阵, 并通过图注意力网络实现约束矩阵对隐状态的约束优化, 提升模型对于实体、关系单词特征的捕获能力并实现了实体关系的联合抽取。通过公共数据集 NYT 的实验论证, DCGAT 模型相比以往工作实现了更优的性能。在以后的工作中, 我们会继续针对 DCGAT 的抽取性能进行优化, 进一步提升预测精度和重叠关系的预测效果。

参考文献

- 1 Veličković P, Cucurull G, Casanova A, *et al.* Graph attention networks. arXiv: 1710.10903, 2017.
- 2 Miwa M, Bansal M. End-to-end relation extraction using LSTMs on sequences and tree structures. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany. 2016. 1105–1116.
- 3 Dixit K, Al-Onaizan Y. Span-level model for relation extraction. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy. 2019. 5308–5314.
- 4 Bekoulis G, Deleu J, Demeester T, *et al.* Joint entity recognition and relation extraction as a multi-head selection problem. Expert Systems with Applications, 2018, 114: 34–45. [doi: 10.1016/j.eswa.2018.07.032]

- 5 Zeng XR, Zeng DJ, He SZ, *et al.* Extracting relational facts by an end-to-end neural model with copy mechanism. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia. 2018. 506–514.
- 6 Katiyar A, Cardie C. Going out on a limb: Joint extraction of entity mentions and relations without dependency trees. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada. 2017. 917–928.
- 7 Li XY, Yin F, Sun ZJ, *et al.* Entity-relation extraction as multi-turn question answering. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy. 2019. 1340–1350.
- 8 Jia SB, E SJ, Li MZ, *et al.* Chinese open relation extraction and knowledge base establishment. ACM Transactions on Asian and Low-Resource Language Information Processing, 2018, 17(3): 15.
- 9 Zhang YH, Qi P, Manning CD. Graph convolution over pruned dependency trees improves relation extraction. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium. 2018. 2205–2215.
- 10 Zhu H, Lin YK, Liu ZY, *et al.* Graph neural networks with generated parameters for relation extraction. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy. 2019. 1331–1339.
- 11 Fu TJ, Li PH, Ma WY. GraphRel: Modeling text as relational

- graphs for joint entity and relation extraction. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy. 2019. 1409–1418.
- 12 Guo ZJ, Zhang Y, Lu W. Attention guided graph convolutional networks for relation extraction. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy. 2019. 241–251.
- 13 Sun CZ, Gong YY, Wu YB, *et al.* Joint type inference on entities and relations via graph convolutional networks. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy. 2019. 1361–1370.
- 14 许晶航, 左万利, 梁世宁, 等. 基于图注意力网络的因果关系抽取. 计算机研究与发展, 2020, 57(1): 159–174. [doi: [10.7544/issn1000-1239.2020.20190042](https://doi.org/10.7544/issn1000-1239.2020.20190042)]
- 15 Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv: 1409.0473, 2014.
- 16 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is All You Need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA, USA. 2017. 5998–6008.
- 17 Tang GB, Müller M, Rios A, *et al.* Why self-attention? A targeted evaluation of neural machine translation architectures. Proceedings of 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium. 2018. 4263–4272.
- 18 凡子威, 张民, 李正华. 基于 BiLSTM 并结合自注意力机制和句法信息的隐式篇章关系分类. 计算机科学, 2019, 46(5): 214–220. [doi: [10.11896/j.issn.1002-137X.2019.05.033](https://doi.org/10.11896/j.issn.1002-137X.2019.05.033)]
- 19 李明扬, 孔芳. 融入自注意力机制的社交媒体命名实体识别. 清华大学学报(自然科学版), 2019, 59(6): 461–467.
- 20 Honnibal M, Johnson M. An improved non-monotonic transition system for dependency parsing. Proceedings of 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal. 2015. 1373–1378.
- 21 Zheng SC, Wang F, Bao HY, *et al.* Joint extraction of entities and relations based on a novel tagging scheme. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada. 2017. 1227–1236.