



## 融合依存关系的对话关系抽取

段瑞雪, 刘鑫, 张仰森, 马致远, 张博宣

引用本文:

段瑞雪, 刘鑫, 张仰森, 马致远, 张博宣. 融合依存关系的对话关系抽取[J]. 重庆理工大学学报(自然科学), 2023, 37(7): 217-226.

---

## 相似文章推荐 (请使用火狐或IE浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

### 基于标签校正的端到端实体关系联合抽取

End-to-end Entity Relationship Joint Extraction Based on Label Correction

重庆理工大学学报(自然科学). 2021, 35(10): 163-171 [https://doi.org/10.3969/j.issn.1674-8425\(z\).2021.10.021](https://doi.org/10.3969/j.issn.1674-8425(z).2021.10.021)

### 基于随机森林的N1+N2结构语法关系判定方法研究

Research on Judging Method of N1+N2 Structure Grammatical Relation Based on Random Forest

重庆理工大学学报(自然科学). 2021, 35(7): 125-130 [https://doi.org/10.3969/j.issn.1674-8425\(z\).2021.07.015](https://doi.org/10.3969/j.issn.1674-8425(z).2021.07.015)

### 异柠檬酸脱氢酶1(IDH1)抑制剂的三维定量构效关系和分子对接研究

Three-Dimensional Quantitative Structure-Activity Relationship and Molecular Docking of Isocitrate Dehydrogenase 1 (IDH1) Inhibitors

重庆理工大学学报(自然科学). 2021, 35(12): 234-242 [https://doi.org/10.3969/j.issn.1674-8425\(z\).2021.12.029](https://doi.org/10.3969/j.issn.1674-8425(z).2021.12.029)

### 融合多信息的个性化推荐模型

Personalized Recommendation Model with Multiple Information Fusion

重庆理工大学学报(自然科学). 2021, 35(3): 128-138 [https://doi.org/10.3969/j.issn.1674-8425\(z\).2021.03.017](https://doi.org/10.3969/j.issn.1674-8425(z).2021.03.017)

### IPSO-EKF融合算法的SOC估算研究

Research on SOC prediction Based on Fusion Algorithm of IPSO-EKF

重庆理工大学学报(自然科学). 2021, 35(12): 18-27 [https://doi.org/10.3969/j.issn.1674-8425\(z\).2021.12.003](https://doi.org/10.3969/j.issn.1674-8425(z).2021.12.003)



# 融合依存关系的对话关系抽取

段瑞雪, 刘鑫, 张仰森, 马致远, 张博宣

(北京信息科技大学 智能信息处理研究所, 北京 100101)

**摘要:** 为了提高对话中实体对的关系抽取能力, 将依存关系引入到异构图注意力网络中, 提出了 DEP-GAT 模型。首先, 通过预处理层获取每个词的基本特征, 然后在话语编码层实现上下文特征的抽取, 并加入依存信息进一步掌握话语结构。最后利用特征构建异构图, 设计有效的消息传递机制, 从而使得更新后的对话实体对包含了整个对话的上下文信息和语法特征, 以此提高模型对实体关系抽取的能力。实验结果表明, 在 DialogRE 数据集上, DEP-GAT 模型相比于基线模型, F1 值在开发集上提高了 2.9%, 在测试集上提高了 1.8%。

**关键词:** 实体关系抽取; 依存关系; 异构图; 自然语言处理

**中图分类号:** TP391.41

**文献标识码:** A

**文章编号:** 1674-8425(2023)07-0217-10

## 0 引言

随着互联网的不断发展, 人们越来越倾向于在互联网中进行对话沟通。与此同时, 电子信息的爆炸式增长, 越来越多的有效信息被隐藏在非结构化的信息之中。对话数据是典型的非结构化信息, 与普通文本数据不同, 对话数据具有很强的非结构性。对话关系抽取<sup>[1]</sup>作为信息抽取的子任务, 在理解非结构化文本中发挥着重要作用。

近几年, 人工智能领域的发展取得了显著进步。相关学者基于图神经网络、长短期记忆网络等神经网络模型在实体关系抽取工作上展开研究, 并取得了显著成果。2020年, Yu等<sup>[2]</sup>根据美

国电视情景喜剧《老友记》的剧本进行删减并经过人工标注得到数据集 DialogRE, 标志着基于对话文本的实体关系抽取任务的开始。2021年, Chen等<sup>[3]</sup>首先提出了文档级异构图注意力网络(document-level heterogeneous graph attention networks, DHGAT), 并在 DialogRE 数据集上取得不错的成果。然而 DHGAT 模型中并未充分利用话语信息, 比如依存关系。通过引入依存关系, 能够更好地让模型掌握话语的结构, 深入理解话语的内容, 从而更准确地预测实体对的关系。

本文的贡献如下:

1) 基于现有的模型, 提出了融合依存关系的对话关系抽取模型 DEP-GAT (dependency relation

**收稿日期:** 2023-04-14

**基金项目:** 国家自然科学基金项目(62176023); 北京信息科技大学大学生创新创业训练计划项目-计算机学院(5112310855); 北京市自然科学基金项目(4224090)

**作者简介:** 段瑞雪, 女, 博士, 讲师, 主要从事自然语言处理研究, E-mail: duanruixue@bistu.edu.cn。

**本文引用格式:** 段瑞雪, 刘鑫, 张仰森, 等. 融合依存关系的对话关系抽取[J]. 重庆理工大学学报(自然科学), 2023, 37(7): 217-226.

**Citation format:** DUAN Ruixue, LIU Xin, ZHANG Yangsen, et al. Dialogue relationship extraction with dependency relation [J]. Journal of Chongqing University of Technology (Natural Science), 2023, 37(7): 217-226.

graph attention network),该模型能够有效结合上下文,并充分利用话语信息。通过引入依存关系,模型能够获取更多的长程信息,模型进一步掌握对话中的话语结构,从而提高模型抽取对话实体关系的能力。

2) 提出使用 BERT 模型进行话语编码,使用 BERT 进行话语编码,能够获取更加全面的语义信息,从而提高模型理解对话的能力,最终提升模型抽取对话实体关系的能力。本文的代码已开源在 <https://github.com/lx-bistu/DEP-GAT>。

## 1 相关工作

目前已有的研究大多仅停留在句子级或文档级的关系提取上,面向对话级的关系提取任务的研究还相对较少。实体关系抽取的<sup>[4]</sup>方法主要可分为两类:一类是基于序列的关系抽取方法,另外一类是基于图神经网络的关系抽取方法。

### 1.1 基于序列的关系抽取方法

常见的基于序列的方法采用不同的神经网络编码实体、句子和文档信息,对文档内的句子进行建模,以实现关系抽取。2015年,Nguyen等<sup>[5]</sup>首次将卷积神经网络(convolutional neural network, CNN)应用到关系抽取任务中。在此基础上,Zeng等<sup>[6]</sup>创新性地提出分段卷积神经网络(piecewise convolutional neural network, PCNN),并引入了单句策略,然而该网络使用单句策略,忽视了在关系抽取中句间信息的重要性。为此,Yang等<sup>[7]</sup>提出了一个多级注意力网络,并首次采用门控循环单元来表示语义信息,在 Freebase 数据集的实验结果证明了模型的有效性。此外,Zhang等<sup>[8]</sup>提出了带有位置感知的 LSTM(long short-term memory),并成功地将位置感知引入到了关系抽取中。Zhou等<sup>[9]</sup>提出了一种基于注意力机制的 BiLSTM,能够捕获句子中最重要的语义信息来实现关系抽取。Xue等<sup>[10]</sup>提出 SimpleRE 模型进行句子级关系抽取,并取得了不错的效果。

### 1.2 基于图神经网络的关系抽取方法

基于序列的抽取方法对长距离依赖关系建模能力不足,该方法侧重于词级特征,从而忽略了上下文中其他文本单位之间的关联。相比之下,图神经网络能够充分利用各种信息,并建立更好的

实体之间的联系,从而能够显著提升模型实体关系抽取的能力。例如,Zhang等<sup>[11]</sup>使用一种新颖的剪枝策略得到依存树,并将其引入到图卷积神经网络(graph convolutional network, GCN)中,通过图卷积网络高效地聚合依存结构信息,从而提升了模型关系抽取的能力。Guo等<sup>[12]</sup>则在图卷积网络中引入了注意力机制,提出了 AGGCN(attention guided graph convolution network)模型。该模型能够对依存树进行软修剪,并且能够自动选择对关系抽取任务有用的信息,从而提升实体关系抽取的能力。Christopoulou等<sup>[13]</sup>提出了一种面向边的图神经网络模型,用来解决文档级的关系抽取任务。该模型利用边进行推理,使得模型能够更好地学习句内和句间的关系,进而提高了关系抽取的能力。Xue等<sup>[14]</sup>利用多视图结构来捕获词之间的各种可能的关系,并使用 BERT 模型选择可能用于关系抽取的重要词语,并将这些表示用于关系抽取。Xu等<sup>[15]</sup>提出了 SIEF(sentence importance estimation and focusing)框架,并将其引入到了图网络模型中,以提升模型的关系抽取能力。Bai等<sup>[16]</sup>则将抽象意义特征引入到图结构中来完成对话建模。Xu等<sup>[17]</sup>则将角色指代信息作为角色引入到了图注意力网络中,发现在 DialogRE 数据集上的实验效果与 DHGAT 模型相比,性能有所提升。Lee等<sup>[18]</sup>则提出了 TUCORE-GCN(turn context aware graph convolutional network)模型,将对话文本中的轮换信息加入到图神经网络中,以提升模型关系抽取的能力。综上所述,图网络模型在关系抽取任务中取得了较好效果。

## 2 DEP-GAT 模型

### 2.1 DEP-GAT 模型概述

与句内关系抽取相比,对话间的关系抽取面临更多的挑战。首先,对话间的序列的长度更长,而 BiLSTM 模型在捕获长距离依赖方面受到了限制。其次,对话中的关系抽取需要考虑更全面的信息,包括句内依赖与句间依赖关系,这正是图神经网络模型在传递句子之间的依赖信息方面表现出色的原因。最后,句法信息,尤其是依存信息在关系抽取任务中起到关键的作用。基线 DHGAT 模型没有考虑语法信息等。因此,为了充分利用

话语信息,本文提出了融合依存关系和图注意力网络的对话关系抽取的总模型 DEP-GAT。针对话语编码层采用了 2 种不同的策略来构建模型,分

别是 DEP-GAT-BERT 和 DEP-GAT-BiLSTM,具体模型的结构如图 1 所示。话语编码层的具体细节将在 2.3 节中介绍。

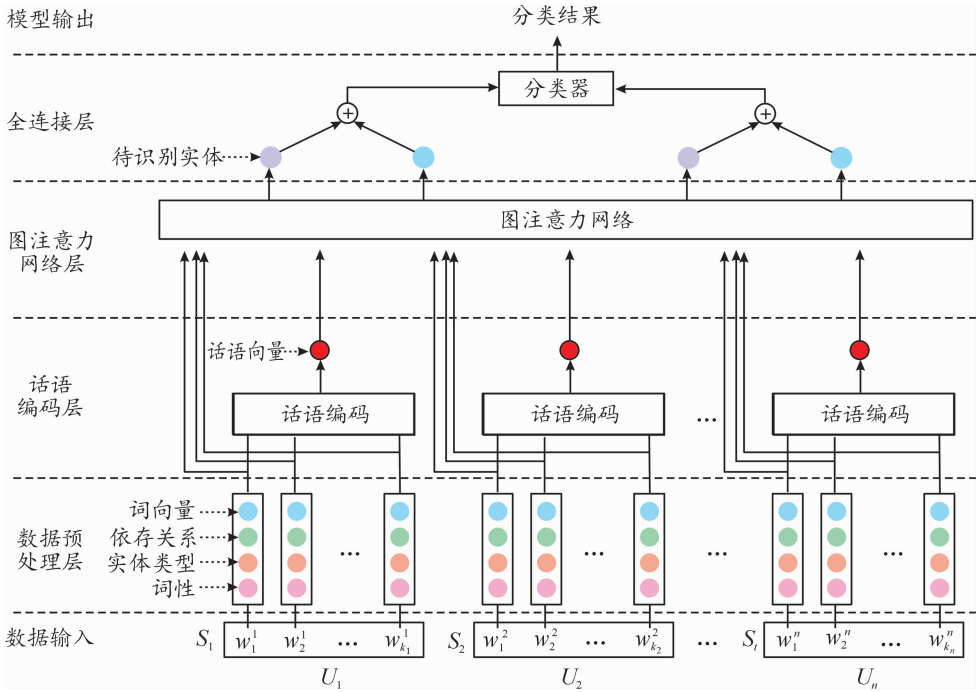


图 1 DEP-GAT 模型结构

该模型包含以下 4 个部分:数据预处理层、话语编码层、图注意力网络层、全连接层。首先对输入的对话进行预处理,提取出相应的特征。再将这些特征传入话语编码层进行编码,得到对话中每个句子的基础向量表示。然后将数据预处理层和话语编码层得到的特征构建成异构图。值得注意的是,在构建异构图时,将由依存关系构建的依存节点直接与对应的词节点进行连接,得到引入依存关系的异构图。接着使用图注意力机制对其进行消息传递,从而使得待识别的实体获得更全面的表示。最后,将从图注意力网络中得到的向量表示传入到全连接层中进行分类。

### 2.2 数据预处理层

利用 spaCy 进行数据预处理,得到对话的分词,以及每个 word 的词性(part-of-speech, POS)、命名实体类型(named entity recognition, NER)、依存关系(dependency, DEP)等,并将得到的表示传入话语编码层和图注意力网络中。

### 2.3 话语编码层

采用 2 种话语编码的策略,第一种是采用

BERT 模型进行编码,另一种是沿用 Chen 等<sup>[3]</sup>的策略,使用双向的 LSTM 进行话语编码。值得注意的是,本文将采用 BERT 模型进行话语编码的模型称为 DEP-GAT-BERT,将采用 LSTM 进行编码的模型称为 DEP-GAT-BiLSTM。

#### 2.3.1 基于 BERT 模型的话语编码

BERT 模型经过大规模无监督训练获取可以获取每个词的向量表示,维度为 768,相近语义的词具有相似的词向量。BERT 模型实质上是多层 Transformer 模型的编码器构成的。由于 Transformer 使用了多头注意力机制,从而使得 BERT 模型获取的话语表示能够理解文本更深层次的语义信息。

对于对话  $C = \{U_1, U_2, \dots, U_n\}$  的每个话语  $U_i = \{w_1^i, w_2^i, \dots, w_k^i\}$ ,使用 BERT 模型可以获取每个语句的话语表示。在话语  $U_i$  的首尾分别添加“[CLS]”和“[SEP]”标志,然后将其传入 BERT 模型中。最后,经过 BERT 模型的编码,可以得到话语中每个 word 的向量表示,以及“[CLS]”和

“[SEP]”2个标志的向量表示。具体可以表示为:

$$V = \text{BERT}([\text{CLS}], w_1, w_2, \dots, w_n, [\text{SEP}]) = (e_{[\text{CLS}]}, e_1, e_2, \dots, e_n, e_{[\text{SEP}]}) \quad (1)$$

其中:  $w_i$  表示输入序列的第  $i$  个 word,  $i$  取值为  $1 \sim n$ ;  $V$  表示 BERT 模型的输出向量。另外, 利用特殊位置 “[CLS]” 的向量表示作为话语  $i$  的向量表示, 进而得到每个话语的向量表示。

基于 BERT 的话语编码如图 2 所示。对于对话中的第  $i$  个话语, 首先在其首尾分别添加 “[CLS]” 和 “[SEP]” 标志, 然后将添加了标志的话语传入到 BERT 模型中, 最后经过 BERT 模型的编码, 可以得到话语中每个 word 的向量表示, 以及 “[CLS]” 和 “[SEP]” 2 个标志的向量表示:

$$V = \text{BERT}(U'_i) = (e_{[\text{CLS}]}, e_1^i, e_2^i, \dots, e_n^i, e_{[\text{SEP}]}) \quad (2)$$

其中,  $U'_i$  表示在话语  $U_i$  的首尾分别添加 “[CLS]” 和 “[SEP]” 标志,  $e$  表示每个 word 的向量表示, 其下标为话语中的第  $e$  个 word 或者特殊标志, 其上标表示对话中的第  $i$  个话语。

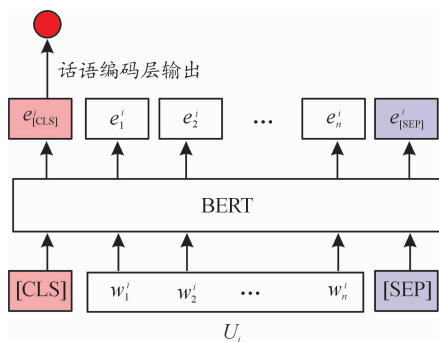


图 2 基于 BERT 的话语编码示意图

使用 BERT 对每个话语进行编码, 并将每个话语的编码结果的  $e_{[\text{CLS}]}$  作为话语的向量表示, 即得到整个对话的话语表示。再将其传入到图注意力网络中构建话语节点。

### 2.3.2 基于 BiLSTM 的话语编码

如图 3 所示, 首先将由 spaCy 得到的 word、POS、NER 的向量表示  $e_{\text{word}}$ 、 $e_{\text{POS}}$  和  $e_{\text{NER}}$ 。其中  $e_{\text{word}}$  是通过 GloVe 得到的每个 word 的词向量,  $e_{\text{POS}}$  和  $e_{\text{NER}}$  分别表示通过随机初始化得到的每个 word 的 POS 和 NER 的向量表示。将得到的向量拼接在一起, 从而得到包含词本身、词性、实体类型的向量表示  $e$ , 具体操作如式(3)所示。

$$e = [e_{\text{word}} : e_{\text{POS}} : e_{\text{NER}}] \quad (3)$$

其中,  $:$  表示拼接操作。

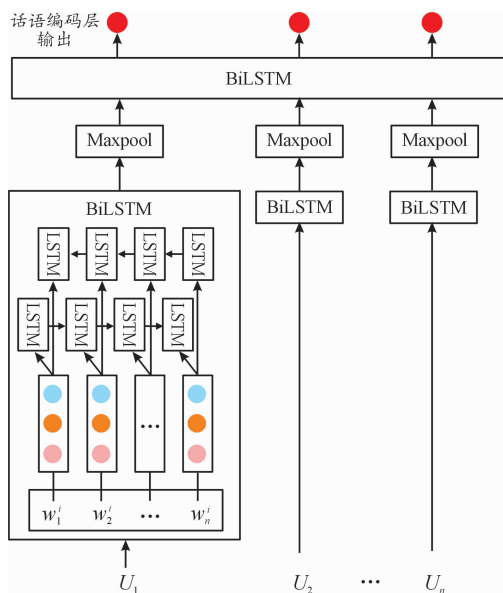


图 3 基于 BiLSTM 的话语编码示意图

将每一个对话  $C = \{U_1, U_2, \dots, U_n\}$  传入话语级的 BiLSTM 中来获取每个语句的话语表示。具体操作如式(4)~(6)所示:

$$\overleftarrow{h}_k^i = \text{LSTM}_l(\overleftarrow{h}_{k+1}^i, e_k^i) \quad (4)$$

$$\overrightarrow{h}_k^i = \text{LSTM}_r(\overrightarrow{h}_{k-1}^i, e_k^i) \quad (5)$$

$$h_k^i = [\overleftarrow{h}_k^i, \overrightarrow{h}_k^i] \quad (6)$$

其中,  $\overleftarrow{h}_k^i$ ,  $\overrightarrow{h}_k^i$  分别表示由左、右 2 个方向的 LSTM 得到的话语  $U_i$  中第  $k$  个 word 的隐藏层表示。  $e_k^i$  表示在话语  $U_i$  中的第  $k$  个 word 的词向量。  $\text{LSTM}_l$  和  $\text{LSTM}_r$  分别表示左、右两个方向的 LSTM。

最后, 将每一层的隐藏层表示进行最大池化, 得到每个语句的话语表示  $e_c = \{c_1, c_2, \dots, c_n\}$ 。另外, 为了得到包含上下文信息的话语表示, 将  $e_c$  传入到对话级的 BiLSTM 进行编码。从而得到包含上下文信息的话语表示, 再将其传入到图注意力网络中构建话语节点。

## 3 图注意力网络层

通过前面两层得到的各种表示, 将对话表示为异构图  $G = (V, E)$ , 其中  $V$  表示各种类型的节点的集合,  $E$  表示各种类型的边集合。下面将对节点以及边的构建进行详细介绍。



### 3.1 节点的构建

Chen 等<sup>[3]</sup>提出的 DHGAT 模型设定 5 类节点,分别为词节点、发言人节点、待识别实体节点、实体类型节点和话语节点。本文在构建异构图时,首次将依存节点引入到异构图中,并构建了引入依存关系的异构图。使用如表 1 所示的方法对各节点进行初始化。同时,沿用 Chen 等<sup>[3]</sup>的策略,把词节点、发言人节点、待识别实体节点统称为基础节点。

表 1 异构图中的各节点的初始化方式

节点类型	初始化方式
词节点(word)	使用 GloVe 进行初始化
话语节点(sent)	话语编码层的输出结果
发言人节点(speaker)	
实体类型节点(NER)	
依存节点(DEP)	随机初始化
待识别实体节点(arg)	

### 3.2 边的构建

DHGAT 模型<sup>[3]</sup>根据节点类型和节点之间的逻辑关系构建了 5 种边,分别是词节点-话语节点、词节点-实体类型节点、话语节点-发言人节点、待识别实体节点-实体类型节点以及话语节点-待识别实体节点。DEP-GAT 模型在此基础上,将依存节点与词节点直接相连,构建了第 6 种边:依存节点-词节点,从而将依存关系引入到了异构图中。上述 6 种边均采用随机初始化。

### 3.3 图注意力机制

使用图注意力机制<sup>[19]</sup>来实现消息传递,具体操作如式(7)~(9)所示:

$$F(h_i, h_j) = \text{LeakyReLU}(a^T(W_i h_i; W_j h_j; E_{ij})) \quad (7)$$

$$\alpha_{ij} = \text{softmax}(F(h_i, h_j)) \quad (8)$$

$$h_i^{\text{in}} = \parallel_{k=1}^K \sigma \left( \sum_j \alpha_{ij}^k W_q^k h_j \right) \quad (9)$$

其中,  $h_i$  和  $h_j$  分别是第  $i$  个节点和第  $j$  个节点的话语表示,  $W_i$ 、 $W_j$ 、 $W_q$  和  $a^T$  是可以学习的参数,  $E_{ij}$  是映射到嵌入空间的边的权重矩阵,  $\alpha_{ij}$  是节点  $i$  和节点  $j$  之间的注意力权重,  $\sigma$  是激活函数,  $\parallel$  是拼接操作。

### 3.4 消息传递

为了充分利用各种语义信息,一共设置 7 层图注意力层,每一层代表一次消息传递。如图 4 所示,其中 Layer A 和 Layer F 是话语节点和基础节点之间的消息传递,Layer B 和 Layer D 是基础节点与实体类型节点之间的消息传递,Layer C 和 Layer E 是基础节点与依存节点之间的消息传递。

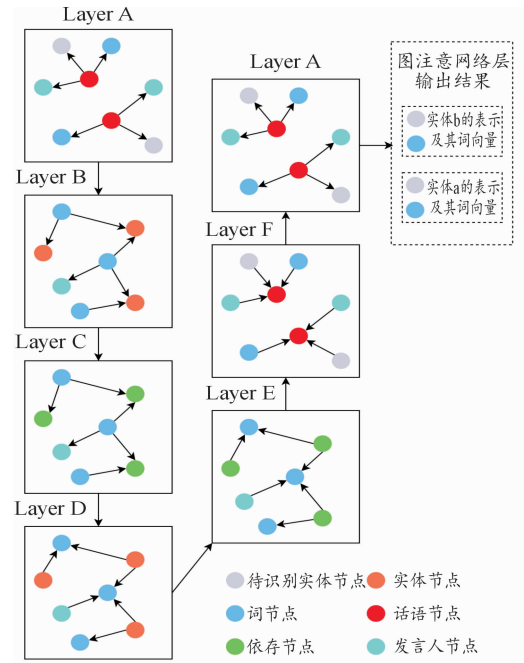


图 4 图注意力网络层的消息传播示意图

不同的消息传播过程会产生不同的效果。本文采用的消息传播过程如下:

首先,使用最初的话语节点对基础节点进行消息传递,得到第 1 次更新的基础节点。

再使用第 1 次更新的基础节点分别对实体类型节点和依存节点进行更新,从而得到更新后的实体类型节点和依存节点。

接着使用更新后的实体类型节点对第 1 次更新的基础节点进行消息传递,得到第 2 次更新的基础节点。

然后使用更新后的依存节点对第 2 次更新的基础节点进行消息传递,得到第 3 次更新的基础节点。

接下来使用第 3 次更新的基础节点对话语节点进行消息传递,得到更新后的话语节点。

最后,使用更新后的话语节点对第 3 次更新的基础节点进行更新,得到第 4 次更新的基础节

点。消息传递过程可以描述成式(10)—(16)。

$$H_b^1 = \text{GAT}(H_b^0, H_u^0) \quad (10)$$

$$H_n^1 = \text{GAT}(H_n^0, H_b^1) \quad (11)$$

$$H_d^1 = \text{GAT}(H_d^0, H_b^1) \quad (12)$$

$$H_b^2 = \text{GAT}(H_b^1, H_n^1) \quad (13)$$

$$H_b^3 = \text{GAT}(H_b^2, H_d^1) \quad (14)$$

$$H_u^1 = \text{GAT}(H_u^0, H_b^3) \quad (15)$$

$$H_b^4 = \text{GAT}(H_b^3, H_u^1) \quad (16)$$

其中, GAT 表示图注意力层,  $H_x^m$  的上标  $m$  表示第  $m$  次更新, 下标  $x$  可表示基础节点  $b$ 、话语节点  $u$ 、依存节点  $d$  和实体类型节点  $n$ 。

此外, 本文继续沿用 Wang 等<sup>[20]</sup>的策略, 使用残差连接来避免在更新的时候出现梯度消失<sup>[21]</sup>, 并在每层消息传递后, 还设置了一个两层的前馈网络, 具体可由式(17)—(18)表示:

$$h_i^{\text{new}} = \text{FFN}(\hat{h}_i) \quad (17)$$

$$\hat{h}_i = h_i^{\text{out}} + h_i^{\text{in}} \quad (18)$$

其中,  $h_i^{\text{out}}$  是第  $i$  次信息传递的输出,  $h_i^{\text{in}}$  是第  $i$  次信息传递的输入。

### 3.5 全连接层

经过 7 层消息传递后, 可以得到每对待识别实体对的表示。将待识别实体节点的表示  $\arg_a$  和  $\arg_b$ , 以及这对待识别实体的词向量  $w_a$  和  $w_b$ , 分别进行最大池化。然后, 将池化后的表示进行拼接, 传入到全连接网络中进行分类。最后, 将分类结果使用激活函数进行激活, 得到每一种关系的概率。具体可以描述成式(19)—(22):

$$w_a^{\text{new}} = [\text{maxpool}(w_a); \text{maxpool}(\arg_a)] \quad (19)$$

$$w_b^{\text{new}} = [\text{maxpool}(w_b); \text{maxpool}(\arg_b)] \quad (20)$$

$$w = [w_a^{\text{new}}; w_b^{\text{new}}] \quad (21)$$

$$P(r|\arg_a, \arg_b) = \sigma(W_w w + b_w), \quad (22)$$

其中,  $P(r|\arg_a, \arg_b)$  是待识别实体对  $(\arg_a, \arg_b)$  的关系类型  $r$  的概率,  $W_w$  和  $b_w$  分别是线性映射的权重和偏移,  $\text{maxpool}$  是最大池化,  $\sigma$  为激活函数。

## 4 实验

### 4.1 数据集

采用 DialogRE 数据集, 该数据集是由 Yu

等<sup>[2]</sup>根据美国电视情景喜剧《老友记》的剧本进行删减并经过人工标注得到。数据集包括 1 788 个对话以及 10 168 个关系三元组。这些三元组一共分为 36 类关系。随机打乱 DialogRE 中的全部对话, 并根据 3:1:1 的比率将数据集划分成 3 部分: 训练集、开发集和测试集。

### 4.2 评价指标

将实体关系抽取任务视为多分类任务, 并使用  $F1$  作为评价指标, 如式(23)所示:

$$F1 = \frac{2P \cdot R}{P + R} \quad (23)$$

其中,  $P$  和  $R$  分别是精确率 (precision) 和召回率 (recall)。

另外, 继续沿用 Yu 等<sup>[2]</sup>设置的  $F1_c$  值,  $F1_c$  值的计算公式与  $F1$  值的一致。区别在于该值是在对话内容不完整的情况下计算得到的。通过对对话内容删减, 验证模型在对话内容不完整的情况下, 是否能够完成实体关系抽取任务。

### 4.3 参数设置

采用深度学习框架 PyTorch 进行代码实现, 并在 Ubuntu16.04 系统上采用 GPU (GeForce RTX 3090) 进行模型的训练和测试, Cuda 的版本为 11.2。BERT 模型使用的是 bert-base-uncased。依存关系 (DEP) 以及发言人 (speaker) 的嵌入向量维度为 30, BiLSTM 的隐藏层设置为 200, batch-size 设置为 16, 优化算法采用 Adam, 学习率 (learning-rate) 设置为 0.0005。边的隐藏层维度设置为 50, 激活函数采用 sigmoid。由于 BERT 模型与 BiLSTM 存在一定的差异, 所以, 设置了不同参数具体如下:

1) 在 DEP-GAT-BERT 模型中, 图注意力网络中的词节点的维度设置为 768, 话语表示的维度设置为 768, 多头注意力的头数设置为 12。

2) 在 DEP-GAT-BiLSTM 模型中, 图注意力网络的词节点的维度设置为 300, 话语级的 BiLSTM 的输出维度为 200, 对话级的 BiLSTM 的输出维度为 300, 多头注意力的头数设置为 10。

### 4.4 结果与分析

#### 4.4.1 对比实验

根据话语编码层的不同, DEP-GAT 模型分别构建 DEP-GAT-BiLSTM 和 DEP-GAT-BERT 模型。前者的话语表示采用 BiLSTM 进行编码, 后者则是

使用 BERT 模型编码层进行编码。

为了验证本文提出的 DEP-GAT 模型的有效性,分别采用 CNN<sup>[6]</sup>、LSTM、BiLSTM<sup>[22]</sup>、AGGCN<sup>[12]</sup>、LSR<sup>[23]</sup> 以及 DHGAT<sup>[3]</sup> 作为基线进行对比。实验结果如表 2 所示,对比 CNN、LSTM 以及 BiLSTM 的实验结果,说明了 BiLSTM 能够很好地解决长句子中的单词依赖问题,从而得到相对较好的关系抽取结果。而 AGGCN 以及 LSR 分别通过 2 种策略来构建依存图,并将依存图传入到图卷积网络中进行消息传递,得到包含丰富信息的待识别节点的表示,从而取得了不错的关系抽取结果。DHGAT 的实验结果可以说明,其能够结合 BiLSTM 和图卷积网络的优点,从而能够在二者的基础上大幅度提升关系抽取能力。

表 2 DialogRE 数据集的实验结果

模型	参数量	Dev/%		Test/%	
		F1	F1 <sub>c</sub>	F1	F1 <sub>c</sub>
CNN	-	46.1	43.7	48.0	45.0
LSTM	-	46.7	44.2	47.4	44.9
BiLSTM	4.1M	48.1	44.3	48.6	45.0
AGGCN	3.7M	46.6	40.5	46.2	39.5
LSR	20.5M	44.5	-	44.4	-
DHGAT	6.0M	56.5	51.8	55.2	49.8
DEP-GAT-BiLSTM (ours)	6.0M	59.4	54.0	57.0	51.2
DEP-GAT-BERT (ours)	5.0M	59.2	53.3	56.0	51.1

基于 DHGAT 模型提出的 DEP-GAT-BiLSTM 模型的 F1 在开发集和测试集分别提升 2.9% 和 1.8%, F1<sub>c</sub> 在开发集和测试集分别提升 2.2% 和 1.4%。该实验结果说明了,由于引入了依存关系,模型能够掌握句子结构,从而加深模型对句子的理解能力,提升了模型在对话中的实体关系抽取能力。与 DHGAT 相比,使用 BERT 进行话语编码的 DEP-GAT-BERT 模型的 F1,在开发集和测试集分别提升了 2.7% 和 0.8%, F1<sub>c</sub> 在开发集和测试集分别提升了 1.5% 和 1.3%。虽然 DEP-GAT-BERT 模型相比于 DHGAT 模型性能有所提升,但是与 DEP-GAT-BiLSTM 模型相比,性能却有所下降。其原因可能是: DEP-GAT-BERT 模型使用

GloVe 模型进行词向量的初始化,导致词向量存在差异,进而影响了模型对话语的理解能力,降低了模型在对话中的实体关系抽取能力。此外,由于 BERT 模型的词向量维度以及多头注意力对维度的要求,必须改变原有模型的多头注意力头数,这也可能对 DEP-GAT-BERT 模型的性能产生影响。

#### 4.4.2 消融实验

为进一步验证模型的有效性,进行了消融实验,实验结果如表 3 所示。其中, w BERT 表示使用 BERT 进行话语编码, w BiLSTM 表示使用 BiLSTM 进行话语编码, head = 10 表示将模型的多头注意力头数设置为 10, head = 12 表示将模型的多头注意力头数设置为 12, w/o DEP 表示在 DEP-GAT-BERT 模型的图注意力网络层中删除第 3 层和第 5 层, w 2-DEP 表示在 DEP-GAT-BERT 模型的图注意力网络层中的第 5 层之后加入 Layer C 和 Layer E 两层。

表 3 DialogRE 数据集上的消融实验结果

实验	模型	Dev/%		Test/%	
		F1	F1 <sub>c</sub>	F1	F1 <sub>c</sub>
1	w BERT	59.2	53.3	56.0	51.1
2	w BiLSTM (head = 10)	59.4	54.0	57.0	51.2
3	w BiLSTM (head = 12)	57.6	53.0	56.2	50.5
4	w/o DEP & w BERT	56.9	51.5	55.7	49.9
5	w 2-DEP & w BiLSTM	56.1	51.3	55.9	50.6
6	w/o DEP & w BiLSTM	56.7	51.7	56.0	51.0

从表 3 中可以得到:

1) 通过对比实验 2 和实验 3,可以说明多头注意力头数的变化确实可能影响模型关系抽取的能力。

2) 对比实验 1 和实验 4,可以看出引入依存关系对模型的性能有明显的提升。这是因为依存关系的引入能够提升模型理解句子的能力,从而提升模型抽取对话实体关系的能力。

3) 实验 5 在图注意力网络中 2 次使用了依存节点进行消息传播,但模型性能却有所下降。其原因可能是:2 次依存节点的消息传播增加了模型图注意力网络的层数,从而使得词节点学到的信息冗余,进而影响模型性能。

4) 实验 6 相比实验 4,模型性能略有提升,可



能是因为在实验 6 中,模型的词向量都是通过 GloVe 模型来初始化的,保证了词向量的一致性。实验 4 中,在话语编译层使用了 BERT 模型,而在图注意力网络层的词节点却使用 GloVe 模型进行初始化,从而导致词向量存在差异,进而影响模型的性能。

## 5 案例分析

与 DHGAT 模型相比,DEP-GAT 模型在图注意力网络层额外获取了依存信息,从而增强了对话语理解,提高了模型性能,如图 5 所示。通过关系中的触发词“brother”,DEP-GAT 模型可以推断“Speaker2”和“Frank”之间的关系是“per: siblings”。

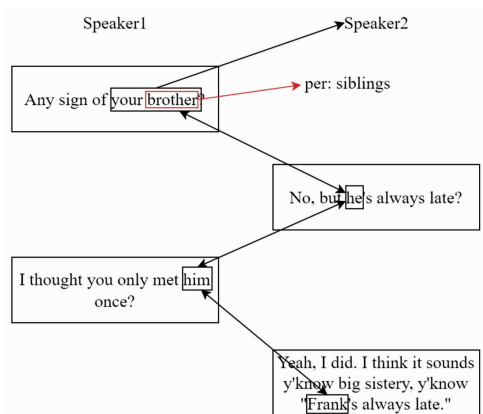


图 5 通过依存关系预测实体关系的示例

首先,通过 spaCy 对句子“Speaker1: any sign of your brother?”进行依存分析,可以得到“your”和“brother”的依存信息。然而,这些依存信息的表示是随机初始化的,因此并未完整包含所有依存信息。而在图注意力网络层进行消息传播之后,模型可以学习更多信息:①“your”和“brother”之间的依存关系是“poss”,即“your”是“brother”的所有者;②“your brother”与“Frank”指代的是同一人。

此外,在图注意力网络层的消息传播过程中,节点“Speaker2”也能够学习到与触发词“brother”相关的各种信息。最终,DEP-GAT 模型能够确定“Speaker2”和“Frank”之间的关系是“per: siblings”。

## 6 结论

本文将依存关系引入到了对话抽取模型中。首先,构建了依存节点,并将其引入到了异构图网络中。然后,提出了融合依存关系的对话抽取模型 DEP-GAT。实验表明,依存关系的引入使得模型可以更加准确地理解对话,从而提升对实体关系抽取的能力。

除此之外,将 BERT 模型引入 DEP-GAT 模型中,提出了 DEP-GAT-BERT 模型。DEP-GAT-BERT 模型在图注意力网络层构建词节点时,仍然使用了 GloVe 模型对词节点进行初始化,从而在一定程度上影响了模型的性能。因此,下一步工作将着重于使用 BERT 对图注意力网络中的词节点进行初始化,以消除词节点的表示不一致对模型性能的影响。

## 参考文献:

- [1] PENG N, POON H, QUIRK C, et al. Cross-sentence nary relation extraction with graph LSTMs [C]//arXiv, Transactions of the Association for Computational Linguistics, Cambridge, 2017: 101 - 105.
- [2] YU D, SUN K, CARDIE C, et al. Dialogue-based relation extraction [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics, 2020: 4927 - 4940.
- [3] CHEN H, HONG P, HAN W, et al. Dialogue relation extraction with document-level heterogeneous graph attention networks [EB/OL]. [2021 - 06 - 20]. <https://arxiv.org/abs/2009.05092>.
- [4] 冯钧, 魏大保, 苏栋, 等. 文档级实体关系抽取方法研究综述 [J]. 计算机科学, 2022, 49(10): 224 - 242.
- [5] NGUYEN T H, GRISHMAN R. Relation extraction: perspective from convolutional neural networks [C]//Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing. Denver, Colorado: Association for Computational Linguistics, 2015: 39 - 48.
- [6] ZENG D, LIU K, CHEN Y, et al. Distant supervision for relation extraction via piecewise convolutional neural networks [C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal: Association for Computational Linguistics, 2015:

- 1753 – 1762.
- [7] YANG L, NG T L J, MOONEY C, et al. Multi-level attention-based neural networks for distant supervised relation extraction [C]//Irish Conference on Artificial Intelligence and Cognitive Science, 2017: 206 – 218.
- [8] ZHANG Y, ZHONG V, CHEN D, et al. Position-aware attention and supervised data improve slot filling [C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, Denmark: Association for Computational Linguistics, 2017: 35 – 45.
- [9] ZHOU P, SHI W, TIAN J, et al. Attention-based bidirectional long short-term memory networks for relation classification [C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Berlin, Germany: Association for Computational Linguistics, 2016: 207 – 212.
- [10] XUE F, SUN A, ZHANG H, et al. An embarrassingly simple model for dialogue relation extraction [M]. ICASSP 2022—2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022: 6707 – 6711.
- [11] ZHANG Y, QI P, MANNING C D. Graph convolution over pruned dependency trees improves relation extraction [C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium: Association for Computational Linguistics, 2018: 2205 – 2215.
- [12] GUO Z, ZHANG Y, LU W. Attention guided graph convolutional networks for relation extraction [C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, 2019: 241 – 251.
- [13] CHRISTOPOULOU F, MIWA M, ANANIADOU S. Connecting the dots: Document-level neural relation extraction with edge-oriented graphs [C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Hong Kong, China: Association for Computational Linguistics, 2019: 4925 – 4936.
- [14] XUE F, SUN A, ZHANG H, et al. GDPNet: Refining latent multi-view graph for relation extraction [C]//Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence, 2021: 14194 – 14202.
- [15] XU W, CHEN K, MOU L, et al. Document-Level relation extraction with sentences importance estimation and focusing [C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019: 764 – 777.
- [16] BAI X, CHEN Y, SONG L, et al. Semantic representation for dialogue modeling [C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021: 4430 – 4445.
- [17] 徐洋, 蒋玉茹, 张禹尧, 等. 融合角色指代的多方对话关系抽取方法研究 [J]. 北京大学学报 (自然科学版), 2022, 58(1): 13 – 20.
- [18] LEE B, CHOI Y S. Graph based network with contextualized representations of turns in dialogue [C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 2021: 443 – 445.
- [19] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph Attention Networks [J]. arXiv preprint arXiv: 1710.10903, 2017.
- [20] WANG D, LIU P, ZHENG Y, et al. Heterogeneous graph neural networks for extractive document summarization [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics, 2020: 1478 – 1489.
- [21] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 770 – 778.
- [22] CAI R, ZHANG X, WANG H. Bidirectional recurrent convolutional neural network for relation classification [C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Berlin, Germany: Association for Computational Linguistics, 2016: 756 – 765.
- [23] NAN G, GUO Z, SEKULIC I, et al. Reasoning with latent structure refinement for document-level relation extraction [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics, 2020: 1546 – 1557.

## Dialogue relationship extraction with dependency relation

DUAN Ruixue, LIU Xin, ZHANG Yangsen, MA Zhiyuan, ZHANG Boxuan

(Institute of Intelligent Information Processing,  
Beijing Information Science and Technology University, Beijing 100101, China)

**Abstract:** In order to enhance the ability to extract entity pair relationship in dialogues, this paper proposes a DEP-GAT model by introducing dependency relation into a heterogeneous graph attention network. Initially, the basic characteristics of each word are obtained through the preprocessing layer. Subsequently, in the discourse coding layer, context features are extracted and dependency information is added to further understand the speech structure. Eventually, a heterogeneous graph is constructed by utilizing the features, and an effective message passing mechanism is designed to enable the updated dialogue entity pairs to contain all the context information and grammatical features of the entire dialogue, thereby further enhancing the ability of the model to extract entity relations. The experimental results reveal that, on the DialogRE data set, the DEP-GAT model performs better than the baseline model does, with an increased  $F1$  value of 2.9% in the development set and 1.8% in the test set respectively.

**Key words:** entity relation extraction; dependency relation; heterogeneous graph; natural language processing

(责任编辑 周媛媛)