



# 计算机科学

COMPUTER SCIENCE

## 面向对话的融入知识的实体关系抽取

陆亮, 孔芳

引用本文

陆亮, 孔芳. 面向对话的融入知识的实体关系抽取[J]. 计算机科学, 2022, 49(5): 200-205.

LU Liang, KONG Fang. Dialogue-based Entity Relation Extraction with Knowledge[J]. Computer Science, 2022, 49(5): 200-205.

---

## 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)

[结合注意力机制的多尺度特征融合图像去雾算法](#)

Multi-scale Feature Fusion Image Dehazing Algorithm Combined with Attention Mechanism

计算机科学, 2022, 49(5): 50-57. <https://doi.org/10.11896/jsjcx.210400093>

[基于改进 CenterNet 的航拍绝缘子缺陷实时检测模型](#)

Real-time Detection Model of Insulator Defect Based on Improved CenterNet

计算机科学, 2022, 49(5): 84-91. <https://doi.org/10.11896/jsjcx.210400142>

[结合注意力机制与几何信息的特征融合框架](#)

Feature Fusion Framework Combining Attention Mechanism and Geometric Information

计算机科学, 2022, 49(5): 129-134. <https://doi.org/10.11896/jsjcx.210300180>

[基于用户关联的立场检测](#)

Stance Detection Based on User Connection

计算机科学, 2022, 49(5): 221-226. <https://doi.org/10.11896/jsjcx.210400135>

[基于时空自适应图卷积神经网络的脑电信号情绪识别](#)

EEG Emotion Recognition Based on Spatiotemporal Self-Adaptive Graph Convolutional Neural Network

计算机科学, 2022, 49(4): 30-36. <https://doi.org/10.11896/jsjcx.210900200>

# 面向对话的融入知识的实体关系抽取

陆亮 孔芳

苏州大学计算机科学与技术学院 江苏 苏州 215006

(20185227045@stu.suda.edu.cn)

**摘要** 实体关系抽取旨在从文本中抽取实体之间的语义关系。该任务在新闻报道、维基百科等规范文本上的研究相对丰富,并取得了一定的成果,但面向对话文本的相关研究还处于起始阶段。目前用于实体关系抽取的对话语料规模较小且信息密度低,有效特征难以捕获;深度学习模型无法像人一样进行知识联想,单纯依靠加大标注数据量和增强计算力难以精细深度地理解对话内容。针对上述问题,提出了一个融入知识的实体关系抽取模型,使用 Star-Transformer 从对话文本中有效捕获特征,同时通过关键词共现的方式构建一个包含关系及其语义关键词的关系集合,将该集合与对话文本进行相关性计算后得到的重要关系特征作为知识融入模型中。在 DialogRE 公开数据集上进行实验,得到  $F_1$  值为 53.6%,  $F_{1c}$  值为 49.5%,证明了所提方法的有效性。

**关键词:** 对话语境;实体关系抽取;Transformer;注意力机制;融入知识

**中图法分类号** TP391

## Dialogue-based Entity Relation Extraction with Knowledge

LU Liang and KONG Fang

School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China

**Abstract** Entity relation extraction aims to extract semantic relations between entities from text. Up to now, related work on entity relation extraction mainly focuses on written texts, such as news and Wikipedia text, and has achieved considerable success. However, the research for dialogue texts is still in initial stage. At present, the dialogue corpus used for entity relation extraction is small in scale and low in information density, so it is difficult to capture effective features. The deep learning model does not associate knowledge like human beings, so it is difficult to understand the dialogue content in detail and depth simply by increasing the amount of annotation data and enhancing the computing power. In response to the above problems, this paper proposes a knowledge-integrated entity relation extraction model, which uses Star-Transformer to effectively capture features from dialogue texts, and constructs a relation set containing relations and their semantic keywords through the co-occurrence of keywords. The important relation features obtained by calculating the correlation between the set and dialogue text are integrated into the model as knowledge. Experiment results on the DialogRE dataset show that the  $F_1$  value is 53.6% and the  $F_{1c}$  value is 49.5%, which proves the effectiveness of proposed method.

**Keywords** Dialogue context, Entity relation extraction, Transformer, Attention mechanism, Integrate knowledge

## 1 引言

实体关系抽取旨在从非结构化文本中获取实体间所具有的语义关系,抽取出来的信息以三元组  $\langle X, R, Y \rangle$  的形式存储,其中  $X$  和  $Y$  表示实体,  $R$  表示实体间的语义关系。作为信息抽取<sup>[1]</sup>的核心技术之一,实体关系抽取不仅是自动构建知识库的核心部件之一,还在知识图谱<sup>[2]</sup>、自动问答<sup>[3]</sup>和聊天机器人<sup>[4]</sup>等下游任务中起着关键作用。

目前,受限于语料资源,实体关系抽取的相关研究多集中

在新闻报道、维基百科等规范文本上。这类文本通常由专业人士编辑,行文规范且文本内容的信息密度大。在规范文本上,句子级的实体关系抽取日趋成熟,大量研究转向了篇章级的实体关系抽取。相比句子级,篇章级蕴含的信息更丰富,涉及的实体及关系也更复杂,而对话文本中存在大量需要跨句子推理才能抽取出来的实体关系,适合展开篇章级实体关系抽取研究。

此外,随着对话系统需求的增加及应用的日趋广泛,特别是多轮对话相关研究的推进,基于对话文本的实体关系抽取

到稿日期:2021-03-19 返修日期:2021-07-16

基金项目:国家自然科学基金面上项目(61876118);国家自然科学基金重点项目(61836007)

This work was supported by the General Program of National Natural Science Foundation of China(61876118) and Key Program of National Natural Foundation of China(61836007).

通信作者:孔芳(kongfang@suda.edu.cn)

成为了研究焦点之一。研究者们试图借助抽取到的实体关系进行对话文本的场景焦点识别和对话场景的知识增强,以期更好地理解对话文本。近期,随着 DialogRE<sup>[5]</sup>数据集的发布,针对人类对话的实体关系抽取研究得以展开。相较于规范文本,对话语料的主要特点有:1)对话语料目前规模较小,大型深度学习模型在小规模语料上难以发挥优势;2)对话中通常包含日常闲聊且对话轮数较多,整体信息密度低,模型难以捕获有效特征;3)人们在对话时会进行知识联想,而模型无法捕获这种支持对话推进的“背景知识”,因此单纯依靠扩大数据量和增强硬件的算力无法对对话文本做到精细而有深度的语义理解。对话文本的这些特殊性,使得面向对话文本的实体关系抽取更具挑战性。

Yang等<sup>[6]</sup>在BERT<sup>[7]</sup>模型的基础上融入了知识图谱中的语义信息,这种外部知识增强了阅读理解模型感知上下文和知识的能力,最终提升了模型在阅读理解任务上的性能。Wang等<sup>[8]</sup>将知识图谱中用户到商品的路径信息融入推荐系统,为用户与商品的交互行为提供丰富的补充信息,从而增强了模型的推理能力。Hou等<sup>[9]</sup>改进实体链指模型时,收集Wikipedia上与实体相关的文章,从中提取出频繁出现的名词以及相关的词组成实体的语义词典,通过这些语义词对实体进行语义实体嵌入后提升了模型的性能。上述研究都证明了通过引入知识来提升模型性能这一方法的可能性。在进行实体关系抽取时,我们发现关系名词本身可以作为相关知识,在此基础上还可以对每种关系进行语义增强,即在关系名词后添加语料中的触发词以及对该关系在Wikipedia中的相关文章使用TF-IDF提取出来的关键词。使用Attention机制将这些知识与对话进行相关性计算后,将得到的重要关系特征融入模型中。

本文选取了轻量级的Star-Transformer<sup>[10]</sup>模型作为总体框架,利用其星型拓扑结构捕获对话中的有效特征。此方法有效缓解了语料库规模小导致模型训练容易出现过拟合的问题。然后在模型中融入上述知识进行实体关系抽取,在DialogRE语料上融入知识的实体关系抽取模型的 $F_1$ 值达到53.6%, $F_{1c}$ 值达到49.5%。

## 2 相关研究

近年来,随着深度学习逐渐崛起,学者们将深度学习应用到实体关系抽取任务中。Zeng等<sup>[11]</sup>于2014年首次使用CNN进行关系抽取;Katiyar等<sup>[12]</sup>于2017年首次将注意力机制Attention与递归神经网络Bi-LSTM结合使用来提取实体和分类关系;Wang等<sup>[13]</sup>提出的CNN架构使用了一种新颖的结合了指定实体的注意力和指定关系的池化注意力的多层次注意力机制;Zhang等<sup>[14]</sup>于2018年将一种新的修剪策略和定制的图卷积网络相结合;Zhang等<sup>[15]</sup>于2019年提出了注意力引导图卷积网络,该模型直接以全依赖作为输入,可以更好地利用全依赖树的结构信息;Wang等<sup>[16]</sup>使用预训练语言模型BERT,通过一次段落编码对不同的实体对进行不同的关系抽取;Li等<sup>[17]</sup>将关系抽取任务转换为多轮问答的问题,将完善的机器阅读理解<sup>[18]</sup>模型应用到该任务中。随着越来越多的改进深度学习模型在实体关系抽取任务中取得巨大进步,学者们不再满足于句子级的实体关系抽取,因此,难度更大的

篇章级实体关系抽取任务渐渐成为研究重点。Quirk等<sup>[19]</sup>于2017年借助远程监督生成了生物领域的跨句子实体关系抽取数据集;Yao等<sup>[20]</sup>于2019年发布了DocRED数据集,该数据集覆盖领域广阔,并且提供了人工标注和远程监督两个版本。

然而,目前主流的用于实体关系抽取的数据集都是针对新闻报道和维基百科等规范文本的,DialogRE数据集的出现将学者们的目光吸引到了针对人类对话的实体关系抽取任务上。Yu等<sup>[5]</sup>在发布DialogRE数据集的同时,将一些主流的神经网络模型,如CNN,LSTM,Bi-LSTM以及BERT等应用到了该任务上,但是并没有依据对话文本的特点对模型做出相应的改进。

本文针对对话语料库规模小、对话文本难以捕获有效特征以及深度学习模型自身局限性的问题,提出使用Star-Transformer来捕获特征,同时通过融入知识的方式帮助模型理解对话语义,从而提升了模型在对话文本中进行实体关系抽取的性能。

## 3 任务定义和评价方法

### 3.1 面向对话文本的实体关系抽取

给定一段对话 $D = s_1 : t_1, s_2 : t_2, \dots, s_m : t_m$ 和一个实体对 $(a_1, a_2)$ ,其中 $s_i$ 和 $t_i$ 分别代表第 $i$ 轮对话的对话者和他所说的话, $m$ 代表总的对话轮数。模型需要提取出现在 $D$ 中的 $a_1$ 和 $a_2$ 之间的关系,如表1所列。根据对话中的下划线部分可以判断出Speaker 2和Frank之间的关系是per:siblings,而触发词就是brother。本文实验中的输入是形如 $s_1 : t_1 s_2 : t_2 \dots s_m : t_m [E_1] a_1 T_1 [E_1] [E_2] a_2 T_2 [E_2]$ 的一段长文,其中 $[E_1]$ 和 $[E_2]$ 是用于分隔实体的特殊字符, $T_1$ 和 $T_2$ 分别代表实体 $a_1$ 和 $a_2$ 的实体类型。

表1 对话语料以及提取的关系三元组

Table 1 Dialogue and relational triples

| 对话者       | 对话  |
|-----------|---|
| Speaker 1 | Hey Pheebes.  |
| Speaker 2 | Hey!  |
| Speaker 1 | Any sign of your brother?   |
| Speaker 2 | No, but he's always late.   |
| Speaker 1 | I thought you only met him once?  |
| Speaker 2 | Yeah, I did. I think it sounds y'know big sistery, y'know, 'Frank's always late'. |
| Speaker 1 | Well relax, he'll be here.  |

(Speaker 2, per, siblings, Frank)

### 3.2 评价方法

本文采用Yu等<sup>[5]</sup>提出的两种方法来评估模型性能,下面分别介绍两种评价方法。

#### (1) 标准评价方法

在标准评价方法中,将对话 $D$ 视为文档 $d$ ,模型输入的长文拼接的是所有对话,输出是基于 $d$ 的 $a_1$ 和 $a_2$ 之间的关系。实验采用准确率 $P$ 、召回率 $R$ 以及 $F_1$ 值对识别结果进行评价。

#### (2) 对话评价方法

在对话评价标准中,将第 $i$  ( $i \leq m$ )轮对话视为 $d$ ,采用一种新的衡量标准,即将对话准确率 $P_c$ 和对话召回率 $R_c$ 的调和平均值 $F_{1c}$ 值作为 $F_1$ 值的补充。下面介绍 $F_{1c}$ 值的定义。当输入为 $a_1, a_2$ 和第 $i$ 轮对话时, $O_i$ 表示输出的预测关系集合,

对于实体对 $(a_1, a_2)$ ,  $L$  表示它对应的人工标注的关系类型集合。 $R$  表示 37 种关系类型的集合, 且  $O_i, L \subseteq R$ 。定义一个辅助函数  $\iota(x)$ , 若  $x$  没有出现在对话  $D$  中, 则返回  $m$ , 否则返回  $x$  第一次出现时的对话轮次。定义辅助函数  $\tau(r)$ : 1) 对于每个关系类型  $r \in L$ , 若存在一个  $r$  的人工标注的触发词, 则  $\tau(r) = \iota(\lambda_r)$ ,  $\lambda_r$  表示触发词, 否则  $\tau(r) = m$ ; 2) 对于每个  $r \in R \setminus L$ ,  $\tau(r) = 1$ 。

第  $i$  轮对话的关系类型集合可以用  $E_i$  来评价,  $E_i$  的定义如式(1)所示, 即给出第  $i$  轮对话  $d$ , 如果  $a_1, a_2$  和  $r$  的触发词都在  $d$  中, 那么  $a_1$  和  $a_2$  之间的关系  $r$  是可评估的。

$$E_i = \{r | i \geq \max\{\iota(a_1), \iota(a_2), \tau(r)\}\} \quad (1)$$

对话准确率的定义如式(2)所示, 对话召回率的定义如式(3)所示。

$$P_c(D, a_1, a_2) = \frac{\sum_{i=1}^m |O_i \cap L \cap E_i|}{\sum_{i=1}^m |O_i \cap E_i|} \quad (2)$$

$$R_c(D, a_1, a_2) = \frac{\sum_{i=1}^m |O_i \cap L \cap E_i|}{\sum_{i=1}^m |L \cap E_i|} \quad (3)$$

对所有实例的对话准确率和对话召回率进行平均, 以获得最终值, 计算过程如式(4)一式(6)所示。

$$P_c = \frac{\sum_{D', a_1', a_2'} P_c(D', a_1', a_2')}{\sum_{D', a_1', a_2'} 1} \quad (4)$$

$$R_c = \frac{\sum_{D', a_1', a_2'} R_c(D', a_1', a_2')}{\sum_{D', a_1', a_2'} 1} \quad (5)$$

$$F_{1c} = \frac{2 \cdot P_c \cdot R_c}{P_c + R_c} \quad (6)$$

在标准评价方法中, 模型的输入可以利用整个对话全文, 而在对话评价方法中, 模型的输入只可以利用上述方法得到部分对话。通俗来讲,  $F_{1c}$  值是用来评价模型在尽可能少的对话轮数中快速识别出实体间关系的性能。

## 4 融入知识的实体关系抽取模型

### 4.1 模型概览

图 1 给出了融入知识模型的具体结构。可以看到, 该模型的核心思想是使用 Star-Transformer 提取特征, 在此基础上, 通过融入知识来提升模型的性能。需要特别说明的是, 模型中并没有加入位置编码, 因为在对话文本中, 大量特征的分布是零散的, 位置编码的加入一定程度上会使模型关注语句的前后顺序而分散了注意力, 导致模型性能下降。

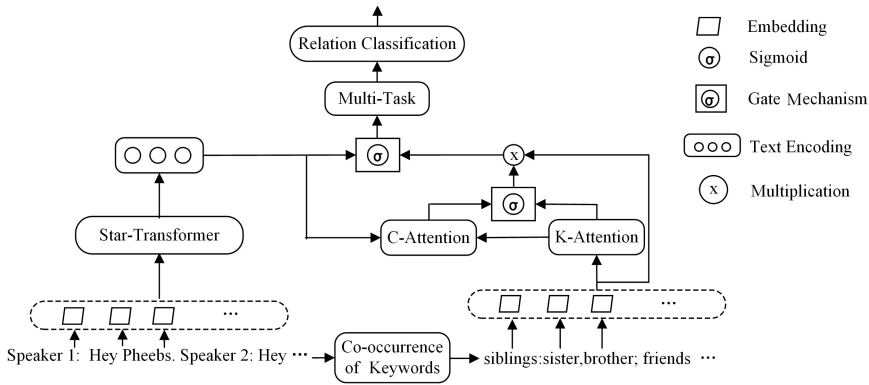


图 1 模型结构

Fig. 1 Model structure

下面分别介绍模型中的两个核心部件: 融入知识模块和 Star-Transformer 结构。

### 4.2 融入知识模块

要想做到精细而有深度的自然语言理解, 一味地依靠增大数据量、提升硬件的算力和优化模型是无法实现的, 需要教会模型在进行语义理解时像人一样进行知识联想。

#### 4.2.1 构建关系集合

DialogRE 数据集中共涉及 37 种关系, 而关系名词本身就包含一定的语义特征, 因此我们把所有的关系组合成一个关系集合  $C$ , 每种关系用特殊符号隔开。接下来需要为关系集合中的每种关系都提取出一些可以用来增强其语义的关键词, 本文采用了如下两种方式:

(1) 数据集为每种关系标注了触发词, 收集训练集的触发词后, 对其进行去重、去除停用词等处理, 并为每种关系选取出现频率最高的前 5 个触发词作为该关系的语义关键词。

(2) 针对每种关系, 在 Wikipedia 上收集相关主题的文章, 使用 TF-IDF 技术来提取关键词, 选取 TF-IDF 值最高

的前 5 个作为该关系的语义关键词。TF-IDF 算法的公式如式(7)一式(9)所示。

$$TF_w = \frac{N_w}{N} \quad (7)$$

$$IDF_w = \log\left(\frac{Y}{Y_w + 1}\right) \quad (8)$$

$$TF-IDF_w = TF_w \times IDF_w \quad (9)$$

其中,  $N_w$  是某一文章中词  $w$  出现的次数,  $N$  是该文章总的词语个数;  $Y$  是与某关系相关的所有文章数,  $Y_w$  是包含词语  $w$  的文章数,  $Y_w + 1$  是为了避免因  $w$  未出现在所有文章中致使分母为 0 的情况;  $TF-IDF_w$  的值由  $TF_w$  值和  $IDF_w$  值相乘得到, 该值越大说明这个词语越重要, 其越能代表关系集合中的关系。

得到关系集合和每种关系对应的语义关键词后, 使用关键词共现的方式对重要的关系进行语义增强。得到对话文本时, 遍历其中的词语, 如果出现了和语义关键词相同的词, 则考虑该关系可能存在于对话信息中, 并把相应的关键词补充到关系集合中该关系之后, 用于增强该关系的语义并提升其重要性。

#### 4.2.2 相关性计算

通过上述关键词共现的方式,得到了大小为  $M$  的关系集合  $C$ ,表示为  $(c_1, c_2, \dots, c_M)$ ,其中  $c_i$  是第  $i$  个关系或该关系的关键词,计算每个关系与对话文本的相关性,最后得到相关性较高的关系向量。本节提出了两种注意力机制(K-Attention和C-Attention)来实现相关性计算。

Knowledge Attention(K-Attention)用于计算关系集合中每个关系自身的重要性,其计算过程如式(10)所示:

$$\alpha_i = \text{softmax}(\mathbf{w}_1^\top f(\mathbf{W}_1 c_i) + b_1) \quad (10)$$

其中,  $\alpha_i$  代表第  $i$  个关系或该关系的关键词  $c_i$  在整个关系集合中的注意力权重,  $f(\cdot)$  是非线性激活函数  $\tanh$ ,  $\mathbf{W}_1$  是权重矩阵,  $\mathbf{w}_1$  是权重向量,  $b_1$  是偏差参数。最后使用  $\text{softmax}$  对每种关系的注意力权重进行归一化。

K-Attention 的作用是进行特征选择,重要的关系会分配较大的权重,而不重要的关系分配的权重会较小。

Correlation Attention(C-Attention)用于计算对话文本和关系集合的语义相关性,减小关键词相同或者关系含义相近引入的一些噪声而带来的负面影响,其计算过程如式(11)所示:

$$\beta_i = \text{softmax}(\mathbf{w}_2^\top f(\mathbf{W}_2 [c_i; e] + b_2)) \quad (11)$$

其中,  $\beta_i$  代表第  $i$  个关系或该关系的关键词相对于对话文本编码  $e$  的注意力权重,  $\beta_i$  越大,说明该关系与对话内容越相关;  $\mathbf{W}_2$  是权重矩阵,  $\mathbf{w}_2$  是权重向量,  $b_2$  是偏差参数。

在得到注意力权重  $\alpha_i$  和  $\beta_i$  后,使用门控机制将二者进行结合,其计算过程如式(12)所示:

$$\gamma_i = \text{Gate}_1 \cdot \alpha_i + \text{Gate}_2 \cdot \beta_i \quad (12)$$

其中,  $\text{Gate}_1$  和  $\text{Gate}_2$  分别代表  $\alpha_i$  和  $\beta_i$  所占的权重向量,  $\text{Gate}_1 + \text{Gate}_2 = 1$ 。

门控机制可以让模型自己学习到合适的权重并分配给  $\alpha_i$  和  $\beta_i$ ,在重点关注关系自身还是重点关注关系与对话的相关性之间取得一个最优解。

最后,通过计算关系集合中向量的加权和得到关系向量  $\mathbf{P}$ ,并对该向量同样使用门控机制与模型的输出向量相结合的方式得到最终的输出。其计算过程如式(13)、式(14)所示。

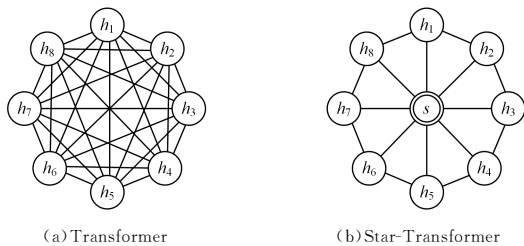
$$\mathbf{P} = \sum_{i=1}^M \gamma_i c_i \quad (13)$$

$$\mathbf{O} = \text{Gate}_3 \cdot \mathbf{P} + \text{Gate}_4 \cdot e \quad (14)$$

其中,  $c_i$  代表第  $i$  个关系或该关系的关键词,  $M$  是关系集合的长度,  $e$  是模型对对话文本的编码,  $\text{Gate}_3 + \text{Gate}_4 = 1$ 。

#### 4.3 Star-Transformer

虽然 Transformer<sup>[21]</sup> 在很多 NLP 任务上都展示了优秀的性能,但是它的结构比较复杂,注意力是完全连接的,这导致了它对大量训练数据的依赖。为了降低模型的复杂性,Guo 等<sup>[10]</sup> 提出利用星型拓扑结构代替全连通结构来简化架构,其中每两个相邻结点通过一个共享中继结点进行连接,因此,模型复杂性从二次降低到线性,同时保留捕获局部成分和长期依赖关系的能力。Star-Transformer 的星型结构如图 2 所示,由一个中继结点  $s$  和  $n$  个卫星结点组成。第  $i$  个卫星结点的状态表示文本序列中第  $i$  个 Token 的特征;中继结点  $s$  充当虚拟中心,在所有卫星结点之间收集和散布信息。



(a) Transformer

(b) Star-Transformer

图 2 Star Transformer 星型结构图

Fig. 2 Star diagram of Star-Transformer

Star-Transformer 提出了基于 time step 的循环更新方式:每个 Token 由 embedding 进行初始化,中继结点初始为所有 Token 的平均值,每个 Token 依次通过 Attention 机制更新。在更新卫星结点时,每个卫星结点  $h_i$  的状态根据其相邻的结点更新,包括上一轮的上一个结点的隐态、上一轮该结点的隐态、上一轮下一个结点的隐态、本结点的 embedding 以及上一轮的中继结点状态,如式(15)、式(16)所示:

$$C_i = [h_{i-1}^{-1}; h_i^{-1}; h_{i+1}^{-1}; e^i; s^{-1}] \quad (15)$$

$$h_i^i = \text{MultiAtt}(h_i^{i-1}, C_i) \quad (16)$$

其中,  $C_i$  表示第  $i$  个卫星结点的上下文信息。

在更新信息后,使用层归一化操作,如式(17)所示:

$$h_i^i = \text{LayerNorm}(\text{ReLU}(h_i^i)), i \in [1, n] \quad (17)$$

在更新中继结点时,中继结点  $s$  汇总所有卫星结点的信息以及之前的状态,如式(18)、式(19)所示:

$$s^i = \text{MultiAtt}(s^{i-1}, [s^{i-1}; H^i]) \quad (18)$$

$$s^i = \text{LayerNorm}(\text{ReLU}(s^i)) \quad (19)$$

通过交替更新卫星结点和中继结点,Star-Transformer 可以捕获局部特征和长期依赖关系。

自注意力机制的  $\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V})$  函数的本质可以描述为:把一个查询(query)和(key-value)键值对的集合映射成一个输出,其中 **query**, **key**, **value**, **output** 都是向量。在处理文本时,自注意力机制会将较高的权重分配给文本中重要的信息,其他无关紧要的信息则分配到较低的权重。

自注意力机制中的核心模块缩放点乘积注意力(Scaled Dot-Product Attention)本质上是使用点积进行相似度计算。Scaled Dot-Product Attention 的计算公式如式(20)所示。

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_K}}\right)\mathbf{V} \quad (20)$$

其中,  $\mathbf{Q}, \mathbf{K}, \mathbf{V}$  均为为向量形式,  $\sqrt{d_K}$  起到了调节作用,用以控制  $\mathbf{Q}, \mathbf{K}$  的内积不会太大,确保该 Attention 为软分布注意力(Soft Attention)<sup>[22]</sup>。

## 5 实验设置与结果分析

### 5.1 实验数据集

本文实验在 DialogRE 数据集上进行,该数据集收集了美剧《老友记》中的人物对话,通过人工注释的方式在 1788 段对话中标注了 10168 个实体关系对,涵盖了 37 种关系。目前,该数据集的英文语料分为 V1 和 V2 两个版本,其中 V2 版本在 V1 的基础上修改了错误标注,因此本文采用了 V2 版本进行实验。表 2 列出了该语料的结构,可以看出,该语料整体

规模不大,深度学习模型在训练时容易出现过拟合现象。

表 2 DialogRE 数据集

|     | 实例数  | 涉及关系数 |
|-----|------|-------|
| 训练集 | 5997 | 35    |
| 验证集 | 1914 | 37    |
| 测试集 | 1862 | 35    |

DialogRE 语料的具体分析如表 3 所列,每段对话平均包含 3.6 个对话者和 15.1 次对话,65.9% 的关系三元组需要通过跨句子抽取,导致模型难以捕获有效特征。

表 3 DialogRE 语料分析

Table 3 Analysis of DialogRE corpus

| 平均对话长度 | 平均对话次数 | 平均对话者数 | 平均涉及关系数 |
|--------|--------|--------|---------|
| 190.6  | 15.1   | 3.6    | 3.0     |

## 5.2 实验设置

实验采用了 Pytorch 1.4.0 框架,并用 NVIDIA 的 1080GPU 进行了加速。与 Yu 等<sup>[5]</sup>的基准模型一样,本文使用 GloVe<sup>[23]</sup> 预训练词向量进行初始化,Embedding\_dim 为 100。模型的输入是一个形状为  $[Batch-size, Max-seq-length, Embedding-dim]$  的张量,其中  $Batch-size$  为 24,  $Max-seq-length$  表示最大句子长度,设置为 512。在训练模型时,Epochs 为 30,学习率为  $3 \times 10^{-5}$ ,损失函数使用的是 BCEWithLogitsLoss,它将 Sigmoid 层和 BCELoss 层相融合,用 Adam (Adaptive Moment Estimation)<sup>[24]</sup> 算法优化模型参数,Drop-out 设置为 0.1。Star-Transformer 中多头注意力机制的头数设置为 4,layer 设置为 2 层。

## 5.3 实验结果

实验采用  $F_1$  值和  $F_{1c}$  值对识别结果进行评价。我们在新数据集上进行了实验,结果如表 4 所列。其中前 3 个模型是 Yu 等<sup>[5]</sup> 给出的基准模型,可以看出,本文所提模型的  $F_1$  值达到了 53.6%, $F_{1c}$  值达到了 49.5%,显著优于基准模型。

## 5.4 实验分析

### (1) Star-Transformer & Transformer

大量实验证明 Transformer 模型在提取全局特征方面具有巨大优势。本文首先采用 Transformer 进行实验,结果发现模型效果不佳。考虑到 DialogRE 数据集的规模不大,Transformer 模型无法发挥全部性能,于是改用轻量型的 Star-Transformer 模型进行实验。Transformer 模型与 Star-Transformer 模型的实验结果如表 4 所列,对比两者可以看到,Star-Transformer 模型在数据规模较小情况下的表现明显优于 Transformer 模型。

表 4 实验结果

Table 4 Experimental results

|                   | Model            | $F_1$ | $F_{1c}$ |
|-------------------|------------------|-------|----------|
| Yu <sup>[5]</sup> | CNN              | 47.1  | 44.7     |
|                   | LSTM             | 47.2  | 44.1     |
|                   | Bi-LSTM          | 47.9  | 44.4     |
|                   | Transformer      | 27.8  | 22.6     |
| Our work          | Star-Transformer | 52.3  | 48.3     |
|                   | ST+Knowledge     | 53.6  | 49.5     |

### (2) 融入知识模块的贡献

表 5 列出了两个模型在准确率和召回率上的详细结果。在融入知识后,模型的  $P$  值下降了 0.2%, $P_c$  值下降了 1%,说明知识的融入带来了一部分噪声,对模型的查准率造成了负面影响。但是  $R$  值上升了 3.3%, $R_c$  值上升了 3.2%,说明融入的知识特征弥补了原先匮乏的对话文本特征,因此模型正确判断出了更多原先被遗漏的实体关系,两者的显著提升提高了模型整体的性能。

表 5 实验结果对比

Table 5 Comparison of experimental results

|  | Model            | $P$  | $R$  | $P_c$ | $R_c$ |
|--|------------------|------|------|-------|-------|
|  | Star-Transformer | 49.2 | 55.9 | 51.2  | 45.7  |
|  | ST+Knowledge     | 49.0 | 59.2 | 50.2  | 48.9  |

表 6 列出了一个示例,对话文本虽然整体信息的密度较低,但存在局部信息密度较高的情况。虽然对话者只进行了一轮对话,但文本中却包含了 5 对实体关系对。

表 6 对话中局部信息密度高的情况

Table 6 High local information density in dialogue

| 对话者       | 对话   |
|-----------|--|
| Speaker 1 | Yo,paisan! Can I talk to you for sec? Your tailor is a very bad man! |
| Speaker 2 | Frankie? What're you talking about?                                  |
|           | <Frankie,per:title,tailor>   |
|           | <Frankie,per:client,Speaker 1>                                       |
|           | <Frankie,per:client,Spieaker 2>                                      |
|           | <Speaker 2,per:alternate_names,Paisan>                               |
|           | <Speaker 1,per:negative_impression,Frankie>                          |

Star-Transformer 模型中的 Self-Attention 机制在处理文本时会给不同的信息分配不同的权重,但在这种局部高密度信息的场景下,Self-Attention 机制可能会漏掉部分重要信息,在损失有效特征的情况下,很容易遗漏掉存在关系的实体对,比如 per:negative\_impression 关系,利用关键词 bad 对该关系进行语义增强后将其融入到模型中,模型更加准确地判断出了该关系。

### (3) 在 BERT 上的实验

面向对话的实体关系抽取用于在对话过程中进行用户画像,因此对模型的响应速度有很高的要求。Star-Transformer 这一轻量级框架在响应速度上优于 BERT 模型,因此我们并未直接比较两者的性能,而是把相关知识融入 BERT 模型中,实验结果如表 7 所列。

表 7 在 BERT 上的实验

Table 7 Experiments on BERT

|  | Model          | $F_1$ | $F_{1c}$ |
|--|----------------|-------|----------|
|  | BERT           | 57.9  | 53.1     |
|  | BERT+Knowledge | 58.1  | 53.2     |

可以看出,融入知识后,BERT 模型的性能有了一定的提升,证明了本文方法的有效性,但是提升效果并不明显,说明目前融入的知识特征的方式还不能很好地与 BERT 模型相结合,因此需要进一步优化两者的结合方式。

**结束语** 本文提出了面向对话的融入知识的实体关系

抽取模型,通过 Star-Transformer 的星型拓扑结构,在有效捕获对话文本特征的同时缓解了深度学习模型在小规模数据集上训练容易出现过拟合的问题,融入的知识特征帮助模型关注对话中更重要的关系,进而提升了模型的性能。

未来我们将针对对话文本的特点,从篇章的视角跟踪对话者与其对应的话语信息,尝试引入指代消解模型,把信息不完整的句子重写为信息完整的句子,帮助模型更好地提取对话文本的信息特征,从而提升模型的性能。

## 参考文献

- [1] GOLSHAN P N, DASHTI H A R, AZIZI S, et al. A study of recent contributions on information extraction[J]. arXiv:1803.05667,2018.
- [2] JI S, PAN S, CAMBRIA E, et al. A survey on knowledge graphs: Representation, acquisition and applications[J]. arXiv:2002.00388,2020.
- [3] ASAI A, HASHIMOTO K, HAJISHIRZI H, et al. Learning to retrieve reasoning paths over wikipedia graph for question answering[J]. arXiv:1911.10470,2019.
- [4] ZHOU X, LI L, DONG D, et al. Multi-turn response selection for chatbots with deep attention matching network[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. 2018;1118-1127.
- [5] YU D, SUN K, CARDIE C, et al. Dialogue-based relation extraction[J]. arXiv:2004.08056,2020.
- [6] YANG A, WANG Q, LIU J, et al. Enhancing pre-trained language representations with rich knowledge for machine reading comprehension[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019;2346-2357.
- [7] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv:1810.04805,2018.
- [8] WANG X, WANG D, XU C, et al. Explainable reasoning over knowledge graphs for recommendation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 5329-5336.
- [9] HOU F, WANG R, HE J, et al. Improving Entity Linking through Semantic Reinforced Entity Embeddings[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020;6843-6848.
- [10] GUO Q, QIU X, LIU P, et al. Star-transformer[J]. arXiv:1902.09113,2019.
- [11] ZENG D, LIU K, LAI S, et al. Relation classification via convolutional deep neural network[C]// 25th International Conference on Computational Linguistics; Technical Papers(COLING 2014). 2014;2335-2344.
- [12] KATIYAR A, CARDIE C. Going out on a limb: Joint extraction of entity mentions and relations without dependency trees[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017;917-928.
- [13] WANG L, CAO Z, DE MELO G, et al. Relation classification via multi-level attention cnns[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. 2016; 1298-1307.
- [14] ZHANG Y, QI P, MANNING C D. Graph convolution over pruned dependency trees improves relation extraction[J]. arXiv:1809.10185,2018.
- [15] ZHANG Y, GUO Z, LU W. Attention guided graph convolutional networks for relation extraction[J]. arXiv:1906.07510,2019.
- [16] WANG H, TAN M, YU M, et al. Extracting multiple-relations in one-pass with pre-trained transformers[J]. arXiv:1902.01030,2019.
- [17] LI X, YIN F, SUN Z, et al. Entity-relation extraction as multi-turn question answering[J]. arXiv:1905.05529,2019.
- [18] LIU X, LIU K, LI X, et al. An iterative multi-source mutual knowledge transfer framework for machine reading comprehension[C]// Twenty-Ninth International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence. 2020;1-7.
- [19] QUIRK C, POON H. Distant supervision for relation extraction beyond the sentence boundary[J]. arXiv:1609.04873,2016.
- [20] YAO Y, YE D, LI P, et al. DocRED: A large-scale document-level relation extraction dataset[J]. arXiv:1906.06127,2019.
- [21] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. arXiv:1706.03762,2017.
- [22] YAO L, TORABI A, CHO K, et al. Video description generation incorporating spatio-temporal features and a soft-attention mechanism[J]. arXiv:1502.08029,2015.
- [23] PENNINGTON J, SOCHER R, MANNING C D. Glove: Global vectors for word representation[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014;1532-1543.
- [24] KINGMA D P, BA J. Adam: A method for stochastic optimization[J]. arXiv:1412.6980,2014.



**LU Liang**, born in 1995, postgraduate, is a member of China Computer Federation. His main research interests include natural language processing and so on.



**KONG Fang**, born in 1977, Ph. D, professor, Ph.D supervisor, is a member of China Computer Federation. Her main research interests include natural language processing and discourse analysis.

(责任编辑:李亚辉)