

# 面向对话文本的实体关系抽取

陆亮  
苏州大学  
计算机科学与技术学院  
20185227045@stu.suda.edu.cn

孔芳\*  
苏州大学  
计算机科学与技术学院  
kongfang@suda.edu.cn

## 摘要

实体关系抽取旨在从文本中抽取出实体之间的语义关系，是自然语言处理的一项基本任务。在新闻报道、维基百科等规范文本上该任务的研究相对丰富，已经取得了一定的效果，但面向对话文本的相关研究还处于起始阶段。相较于规范文本，用于实体关系抽取的对话语料规模较小，对话文本的有效特征难以捕获，这使得面向对话文本的实体关系抽取更具挑战。该文针对这一任务提出了一个基于Star-Transformer的实体关系抽取模型，通过融入高速网络进行信息桥接，并在此基础上融入交互信息和知识，最后使用多任务学习机制进一步提升模型的性能。在DialogRE公开数据集上实验得到 $F_1$ 值为55.7%， $F_{1c}$ 值为52.3%，证明了提出方法的有效性。

**关键词：** 对话；实体关系抽取；注意力机制；高速网络；多任务

## Entity Relation Extraction for Dialogue Text

Liang Lu  
School of Computer  
Science and Technology  
Soochow University  
20185227045@stu.suda.edu.cn

Fang Kong\*  
School of Computer  
Science and Technology  
Soochow University  
kongfang@suda.edu.cn

## Abstract

Entity relation extraction, a fundamental task of natural language processing, aims to extract semantic relations between entities from texts. Up to now, related works on entity relation extraction mainly focus on written texts, such as news and Wikipedia texts, and have achieved considerable success. However, the research for dialogue texts is still in the initial stage. Compared with written texts, the dialogue corpus used for entity relation extraction is currently small in scale, so it is difficult to capture the effective features of dialogue texts, which makes entity relation extraction more challenging. Aiming at this task, this paper proposes an entity relation extraction model based on Star-Transformer, which bridges information by integrating into Highway Networks, integrates interactive information and knowledge on this basis, and finally uses multi-task learning mechanism to further improve the performance of the model. The experiment results on DialogRE dataset show that  $F_1$  value is 55.7% and  $F_{1c}$  value is 52.3%, which proves the effectiveness of proposed method.

**Keywords:** Dialogue, Entity relation extraction, Attention mechanism, Highway networks, Multi-Task

## 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(Golshan et al., 2018)的核心技术之一，实体关系抽取在知识图谱(Ji et al., 2020)、自动问答(Asai et al., 2020)、聊天机器人(Gopalakrishnan et al., 2019)等下游任务中起着关键作用。

目前，受限于语料资源，实体关系抽取相关研究多集中在新闻报道、维基百科等规范文本上。而这类文本通常由专业人士编辑，行文规范，文本内容的信息密度大。随着对话系统需求及应用的日趋广泛，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、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(Yu et al., 2020)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，对话语料的特点主要有：(1)对话语料目前规模较小，大型深度学习模型在训练时容易陷入过拟合的困境。(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)对话内容通常围绕某些特定主题在交互过程中向前推进，因此对话的交互过程蕴含了大量重要信息。(4)人们在对话时会进行知识联想，很多信息不会在对话中体现，理解对话需要知识的融入。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。

针对上述问题，本文提出了面向对话文本的实体关系抽取模型，具体贡献如下：(1)在Star-Transformer(Q et al., 2019)中融入高速网络提高模型提取特征的能力。(2)使用交叉注意力机制挖掘对话中的交互信息，丰富模型的输入特征。(3)通过融入知识帮助模型关注更重要的关系来指导模型进行实体关系抽取。(4)使用多任务学习机制缓解数据分布不均衡的问题。

## 2 相关研究

近年来深度学习逐渐崛起，推动了各类人工智能任务的研究，学者们开始将深度学习应用到实体关系抽取任务中，CNN(Krizhevsky et al., 2012)、RNN(Zaremba et al., 2014)和LSTM(Staudemeyer and Morris, 2019)等经典模型的应用都取得了一定成果。而后更多基于基础方法改进的模型被提出，Katiyar(2017)等人首次在递归神经网络Bi-LSTM(Huang et al., 2015)中融入注意力机制。Wang(2016)等人提出的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(2018)等人将一种新的修剪策略和定制的图卷积网络(Chen et al., 2020)结合。李青青(2019)等人提出了一种基于Attention机制的主辅多任务模型，通过多个关联任务共享信息提升性能。Guo(2019)等人提出了注意引导图卷积网络，直接以全依赖作为输入，该模型可以更好的利用全依赖树的结构信息。陈佳沣(2019)等人提出一种基于强化学习(Feng et al., 2018)的实体关系抽取模型。Wang(2019a)等人使用预训练语言模型BERT(Devlin et al., 2019)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(2019)等人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(Liu et al., 2020)模型应用到该任务中。随着深度学习模型共同发展的还有相关实验数据集，学者们不再满足于句子级的实体关系抽取，规模更大、难度更大的篇章级实体关系抽取数据集也如雨后春笋般出现。Quirk(2017)等人发布了生物领域的跨句子实体关系抽取数据集。Yao(2019)等人发布了DocRED数据集，该数据集覆盖丰富的领域，并且提供了人工标注和远程监督(Mintz et al., 2009)两个版本。

然而目前主流的用于实体关系抽取的数据都是针对新闻报道和维基百科等这类规范文本的，DialogRE数据集的出现将学者们的目光吸引到针对人类对话的实体关系抽取任务上。Yu(2020)等人在发布DialogRE数据集的同时将一些主流的神经网络模型如CNN、LSTM、Bi-LSTM和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。本文针对对话文本的特点，提出使用融合了高速网络的Star-Transformer捕获特征，融入交互信息和知识来丰富特征，最后使用多任务学习缓解数据分布不均衡问题，提高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

本节介绍对话关系抽取任务的定义以及针对该任务提出的评价方法。

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

\*通信作者：kongfang@suda.edu.cn

### 3.1 任务定义

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说的话， $m$  代表总的对话次数。模型需要提取出出现在  $D$  中的  $a_1$  和  $a_2$  的关系，图1给出了一个具体的例子，在给定的对话文本中，根据下划线部分可以判断出“Speaker 2”和“Frank”之间的关系是“per:siblings”，即兄弟姐妹关系，而“brother”就是指出这一关系的触发词(Trigger)。将对话拼接成一段形如： $s_1 : t_1 s_2 : t_2 \dots s_m : t_m [E_1] a_1 T_1 [E_1] [E_2] a_2 T_2 [E_2]$  的长文作为模型的输入，其中  $[E_1]$  和  $[E_2]$  是表示其中内容是实体的特殊字符， $T_1$  和  $T_2$  分别代表实体  $a_1$  和  $a_2$  的实体类型。需要特别说明的是，模型中并没有加入位置编码。因为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，具体的实验结果见表1。关系标签的设置如下：数据集中包含37种关系，对每个训练样例设置一个长度为36的列表，如果是unanswerable关系，则列表中的数值全是0，如果包含其他关系，则把对应位置的0变成1。

Speaker 1: Hey Pheebs.  
 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.  
  
 <Speaker2, per:siblings, Frank>

图 1 对话语料及其提取的关系三元组

### 3.2 评价方法

本文采用Yu等人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  $d$ ，模型的输入长文拼接的是对话全文  $d$ ，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  $R$  和  $F_1$  值对识别结果进行评价。

#### (2) 对话评价方法

在对话评价方法中，将第  $i$  ( $i \leq m$ ) 轮对话视为  $d$ ，采用一种新的衡量标准：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。下面介绍如何定义  $F_{1c}$  值。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  $f(x)$ ，若  $x$  没有出现在对话  $D$  中返回  $m$ ，否则返回  $x$  第一次出现时的对话轮次。定义辅助函数  $\Phi(r)$ ：(1) 对于每个关系类型  $r \in L$ ，如果存在一个  $r$  的人工标注的触发词，则  $\Phi(r) = f(\lambda_r)$ ， $\lambda_r$  表示触发词，否则  $\Phi(r) = m$ 。(2) 对于每个  $r \in R \setminus L$ ， $\Phi(r) = 1$ 。第  $i$  轮对话的关系类型集合可以用  $E_i$  来评价， $E_i$  的定义如公式(1)所示。

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

公式(1)的意思是，给出第  $i$  轮对话中包含的  $d$ ，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。该定义是基于以下假设：根据实体对和触发词的位置，可以大致估计需要多少轮对话来预测两个参数之间的关系。

对话准确率  $P_c$  的定义如公式(2)所示，对话召回率  $R_c$  的定义如公式(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，以获得最终的值，计算过程如公式(4)~(6)所示。

$$P_c = \frac{\sum_{D', a'_1, a'_2} P_c(D', a'_1, a'_2)}{\sum_{D', a'_1, a'_2} 1} \quad (4)$$

$$R_c = \frac{\sum_{D', a'_1, a'_2} R_c(D', a'_1, a'_2)}{\sum_{D', a'_1, a'_2} 1} \quad (5)$$

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

在标准评价方法中模型的输入可以利用整个对话全文，而对话评价方法中模型的输入只可以利用根据上述方法得到部分对话。通俗来讲， $F_{1c}$ 值就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 面向对话文本的实体关系抽取模型

图2给出了面向对话文本的实体关系抽取模型的具体结构，主要包含4个模块：

- (1)融入高速网络的Star-Transformer：通过高速网络进行信息桥接来增强表征。
- (2)获取交互信息：通过交叉注意力机制挖掘交互信息。
- (3)融入知识：使用关键词共现的方式构建关系集合并通过attention机制融入知识。
- (4)多任务学习：使用全共享多任务层缓解数据分布不均衡的问题。

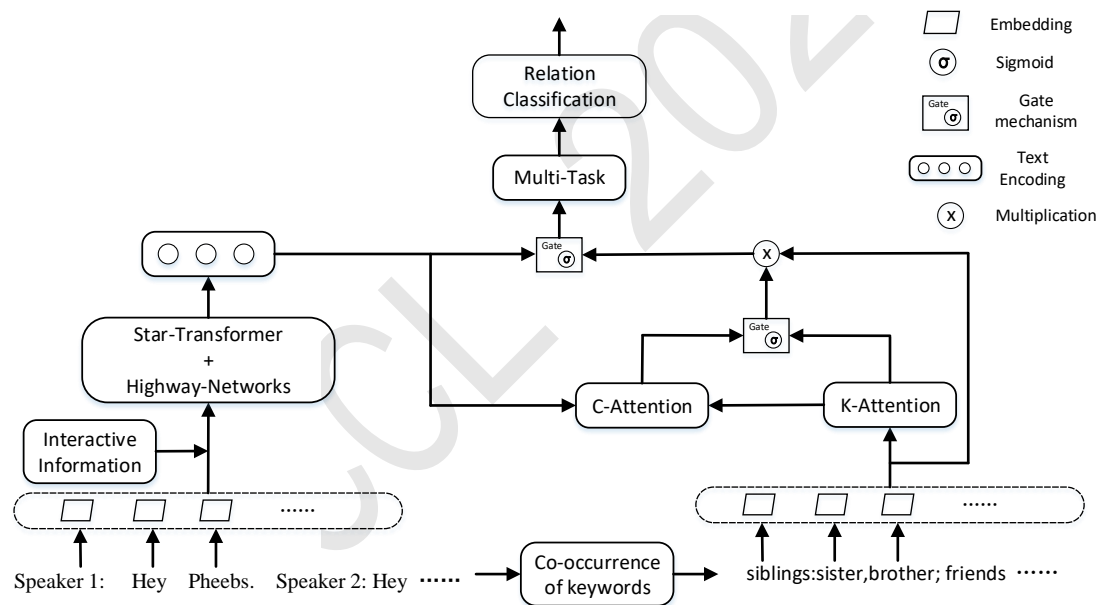


图 2 模型结构

下面对模型各个模块进行详细介绍。

##### 4.1 融入高速网络的Star-Transformer

高速网络(Highway Networks)是一种新型网络结构，用于优化深度网络的训练。随着神经网络深度增加，梯度信息回流受阻，训练难度大幅上升。高速网络在深度网络中架起信息高速公路，允许信息在其中畅通无阻，有效缓解了过深的网络中无法有效传播梯度的问题。

Chai(2020)等人的实验证明了高速网络结构可以有效提升Transformer模型组件的训练效果。Star-Transformer核心思想是通过星型结构来简化体系结构，削减参数量，同时减轻对样本量的依赖，但是这样也造成了信息一定程度的损失。因此，我们考虑在Star-Transformer框架中对每个卫星结点 $h_i$ 进行参数更新时加入高速网络，以增强表征。高速网络的结构如图3所示。

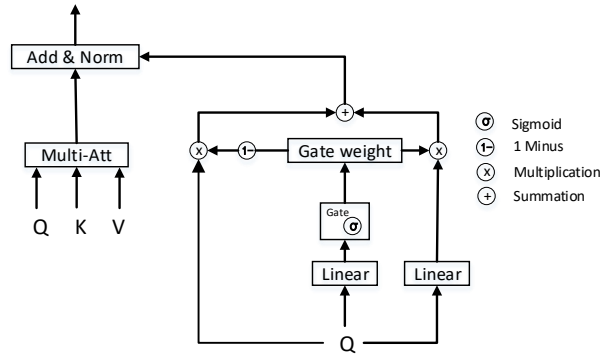


图 3 在Star-Transformer中融入高速网络

在Star-Transformer计算完多头注意力后，加入高速网络的输入 $HW(h_i)$ ，对卫星结点 $h_i$ 进行自我桥接，如公式(7)~(9)所示：

$$gated = \sigma(w_1 h_i + b_1) \tag{7}$$

$$f(h_i) = w_2 h_i + b_2 \tag{8}$$

$$HW(h_i) = (1 - gated) \cdot h_i + gated \cdot f(h_i) \tag{9}$$

其中， $w_1, w_2$ 表示门控机制的权重参数， $b_1, b_2$ 表示门控机制的偏差参数， $\sigma$ 表示激活函数 $Sigmoid$ 。最后将高速网络的结果和多头注意力的结果相结合来增强表征，经过层归一化后得到新的卫星结点 $H_i$ ，如公式(10)所示：

$$H_i = LayerNorm(HW(h_i) + MultiAtt(h_i, C_i, C_i)) \tag{10}$$

#### 4.2 获取交互信息

本文将对话者的一问一答看作一轮对话 $u_i$ ，每个训练样本中平均包含7.5轮对话，依次从中提取出两轮对话来模拟交互过程，使用交叉注意力机制来捕获这两轮对话的关联信息，最后把所有得到的关联信息进行整合得到我们需要的交互信息 $C$ ，具体过程如图4所示。

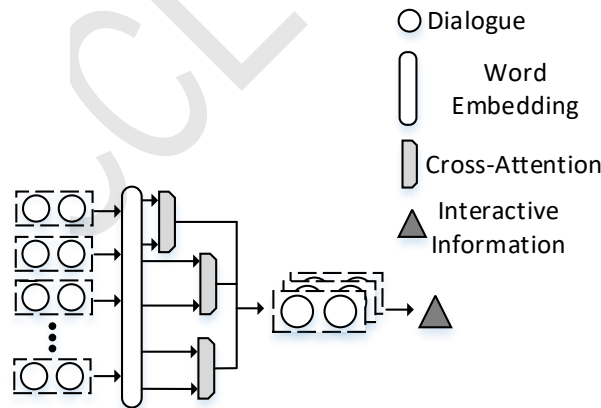


图 4 获取对话交互信息

使用交叉注意力挖掘交互信息的具体计算过程如公式(11)~(12)所示：

$$u_{i+1}^c = CrossAttention(u_i, u_{i+1}) \tag{11}$$

$$C = mean([u_1^c; u_2^c; \dots; u_{L-1}^c]) \tag{12}$$

其中 $[u_1^c; u_2^c; \dots; u_{L-1}^c]$ 表示将所有向量在0维上进行堆叠， $L$ 表示总的对话轮数。交叉注意力机制工作原理如图5所示。

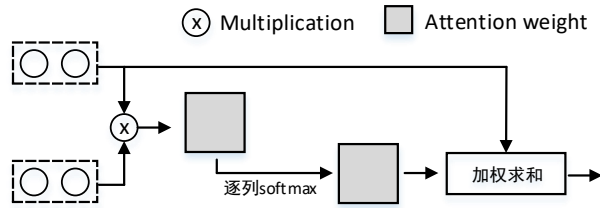


图 5 交叉注意力机制

### 4.3 融入知识

Yang(2019)等人在BERT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，最终提高了模型在阅读理解任务上的性能。Wang(2019b)等人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。上述的研究都证明了通过融入知识帮助提升模型性能这一方法的可能性，人们在进行对话时会进行知识联想，通过特定词语联想到对应的关系，可以融入关系的相关知识特征帮助模型关注对话中更重要的关系进而提升模型的性能，下面介绍如何在本文提出的模型中融入知识。

#### 4.3.1 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语义特征，因此我们把所有的关系形成一个关系集合 $C$ ，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，我们采用了两种方式：

(1)数据集中为了每种关系标注了触发词，我们把训练集的触发词收集后，在进行去重、去除停用词等处理后，为每种关系选取出现频率最高的前五个触发词作为该关系的语义关键词。

(2)针对每种关系，我们在Wikipedia上搜索关系名词，选取内容相关的文章，使用TF-IDF技术提取关键词，选取的TF-IDF值最高的前五个作为该关系的语义关键词。

在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，遍历其中词语，如果出现了和语义关键词相同的词语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的后面，用于丰富该关系的语义，提升其重要性。

#### 4.3.2 相关性计算层

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，需要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，本章提出了两种注意力机制（K-Attention，C-Attention）来实现。

#### 4.3.3 K-Attention

Knowledge Attention(K-Attention)用于计算关系集合中每个关系自身的重要性，其计算过程如公式(13)所示。

$$\alpha_i = \text{softmax}(w_1^T f(W_1 c_i) + b_1) \quad (13)$$

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 $\tanh$ ， $W_1$ 是权重矩阵， $w_1$ 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，而不重要的关系的权重就会较小。

#### 4.3.4 C-Attention

Correlation Attention(C-Attention)用于计算对话文本和关系集合的语义相关性，减少由于关键词相同或者关系含义相近而引入噪声带来的负面影响，其计算过程如公式(14)所示。

$$\beta_i = \text{softmax}(w_2^T f(W_2 [c_i; e] + b_2)) \quad (14)$$

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文本内容越相关。 $W_2$ 是权重矩阵， $w_2$ 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，我们使用门控机制将二者进行结合，其计算过程如公式(15)所示。此处使用门控机制可以让模型自己学习到合适的权重分配给 $\alpha_i$ 和 $\beta_i$ ，在重点关注关系自身还是重点关注关系与对话文本特征的相关性之间取得一个最优解。

$$\gamma_i = Gate_1 \cdot \alpha_i + Gate_2 \cdot \beta_i \tag{15}$$

其中， $Gate_1$ 和 $Gate_2$ 分别代表 $\alpha_i$ 和 $\beta_i$ 所占的权重向量， $Gate_1 + Gate_2 = 1$ 。最后，通过计算关系集合向量的加权和得到关系向量 $P$ ，并将该向量同样使用门控机制与模型的输出向量结合得到最终的输出 $O$ ，最后送入多任务学习层。其计算过程如公式(16)~(17)所示。

$$P = \sum_{i=1}^M \gamma_i c_i \tag{16}$$

$$O = Gate_3 \cdot P + Gate_4 \cdot e \tag{17}$$

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， $Gate_3 + Gate_4 = 1$ 。

#### 4.4 多任务学习

多任务学习 (Multi-Task Learning) 的过程是利用多个不同但具有一定相关性的任务共同进行学习，不同任务中的有用信息可以相互联系，提高各个子任务的性能，从而提高模型整体的泛化能力。Hard参数共享是多任务学习中最常用的方法，其结构如图6所示。在硬共享机制下，模型的底层参数一致，顶层各个子任务的模型是独立的，这样共享参数的好处是可以降低模型在训练时过拟合的概率，以此缓解数据分布不均衡带来的负面影响。

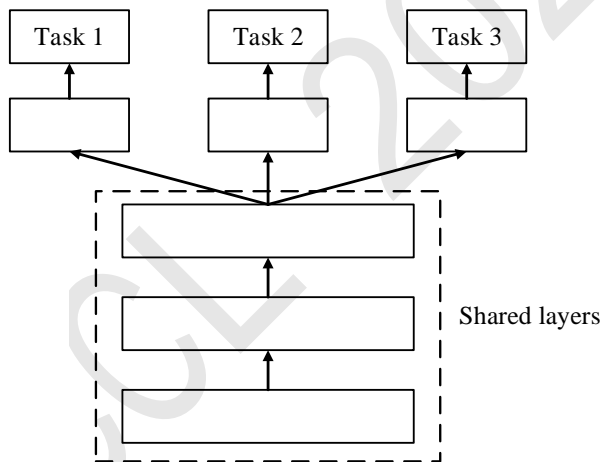


图 6 Hard参数共享

我们通过添加两个辅助任务来解决关系类型不平衡带来的问题。具体来说，实体关系识别为主任务，二元关系识别为辅助任务，第一个辅助任务判断两个实体之间关系类型是否为unanswerable，第二个辅助任务判断两个实体在对话中是否含有触发词。三个任务分类器共享相同的上下文表征嵌入，因此它们可以通过互相传播监督信号来彼此帮助，最终的Loss由三个任务各自的loss按照一定权重相加得到，其计算公式如公式(18)所示，其中 $\alpha_1$ 和 $\alpha_2$ 是两个辅助任务各自的权重。

$$Loss = loss_{main} + \alpha_1 \cdot loss_{aux_1} + \alpha_2 \cdot loss_{aux_2} \tag{18}$$

## 5 实验设置与结果分析

### 5.1 实验数据集

本文采用了Yu等人发布的已经划分好的DialogRE语料，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的基础上对错误的标注进行了修正，所以本文采用了V2版本进行实验。V2版本数据集的详细结构如表1所示，可以看出该语料整体规模不大，大型深度学习模型在训练时容易出现过拟合现象。

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

表 1 DialogRE 数据集结构

DialogRE语料的具体分析如表2所示，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，所以有大量信息隐藏在对话的交互过程中，这些交互信息在面向对话的实体关系抽取任务中起着重要作用。此外，我们分析语料发现在37中关系中无明确关系（unanswerable）占到了21.3%，远大于其他类型的关系，关系分布不均衡。语料中38%的实例中包含触发词（Trigger）。

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

表 2 DialogRE数据集对话分析

## 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。和Yu等人的基准模型一样，本文使用GloVe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是30，学习率为3e-5，损失函数使用的是BCEWithLogitsLoss，它将Sigmoid层和BCELoss层融合在了一起，用Adam (Adaptive Moment Estimation)算法优化模型参数，Dropout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。多任务层包含两个辅助任务，第一个辅助任务的loss值权重为0.0003，第二个辅助任务的loss值权重为0.0003。

## 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，表3给出了各个模型的实验结果，其中前三个模型是Yu等人给出的基准模型。从表中可以看出，我们提出的模型的综合性能 $F_1$ 值达到了55.7%， $F_{1c}$ 值达到了52.3%，显著优于基准模型。

模型	$F_1$ (%)	$F_{1c}$ (%)
CNN(Yu et al., 2020)	47.1	44.7
LSTM(Yu et al., 2020)	47.2	44.1
Bi-LSTM(Yu et al., 2020)	47.9	44.4
Star-Transformer	52.3	48.3
Star-Transformer(no Positional Encoding)	49.1	45.7
Our Work	55.7	52.3

表 3 实验结果



## 5.4 实验分析

我们在Star-Transformer模型的基础上，依次融入高速网络、交互信息、多任务和知识，每一种模块的融入都使得前一种模型的性能得到了提升，表4给出每一种模型的具体结果。

模型	$P(\%)$	$R(\%)$	$P_c(\%)$	$R_c(\%)$	$F_1(\%)$	$F_{1c}(\%)$
Star-Transformer	49.2	55.9	51.2	45.7	52.3	48.3
+高速网络	50.1	58.5	51.5	48.5	54	49.9
+交互信息	50.1	60.2	52.7	49.4	54.7	51
+多任务	53.3	57.1	57.6	46.7	55.1	51.6
+融入知识	52.7	59.2	56.3	48.8	55.7	52.3

表 4 消融实验结果

Star-Transformer模型取得了一定的效果，但它是为小规模数据集设计的，不可避免地牺牲了部分提取特征的能力。此外分析语料发现，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，如图7所示的例子。

```

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, Speaker 2>
<Speaker 2, per:alternate_names, Paisan>
<Speaker 1, per:negative_impression, Frankie>

```

图 7 对话中局部信息密度高的情况

例子中虽然对话者只进行了一轮对话，文本中却包含了五对实体关系对，Star-Transformer模型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。此外，Transformer模型中引入的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。

在引入高速网络后，召回率 $R$ 值上升了2.6%，对话召回率 $R_c$ 值上升了2.8%，说明引入高速网络，通过信息的自我桥接来增强表征，一定程度上弥补了损失掉的信息，同时缓解了深度网络梯度信息回流受阻的问题，优化了模型的训练，使得整体性能上升。

在融入对话交互信息后， $P_c$ 提升了1.2%，说明模型在得到较少轮数的对话特征时，交互信息显得尤为重要，它可以帮助模型快速辨别实体对之间的关系。例如图1的例子中，对话“Speaker1: Any sign of your brother?”中的“your brother”和对话“Speaker2: No, but he’s always late.”中的“he”有关联语义，而其中的“he’s always late”又与对话“Speaker2: ... Frank’s always late.”中的“Frank’s always late”有关联语义，交叉注意力机制可以挖掘上下文中的这种有关联性的词，帮助模型把“he”和“Frank”联系在一起，然后通过触发词“brother”识别出“Speaker2”和“Frank”之间的关系是“per:siblings”。

在使用多任务学习后，召回率 $R$ 值和对话召回率 $R_c$ 值都出现了一定程度的下降，说明原先存在无关系的实体对被错分到了别的关系，模型在结合两个辅助任务的学习后将原本的错误分类进行了纠正。准确率 $P$ 值上升了3.2%，对话准确率 $P_c$ 值上升了4.9%，二者的巨大提升使得模型的整体性能提升。同时我们分析结果发现仅使用Star-Transformer时模型对unanswerable关系的识别准确率是64.1%，而在使用了多任务学习后模型对该关系的识别准确率达到86.1%，说明多任务学习确实帮助模型缓解了语料库数据分布不均衡的负面影响，只是目前语料库的规模较小，所以多任务学习只发挥了部分作用。我们认为在扩大语料库规模后，多任务学习可以更好地帮助模型提升性能。此外，因为我们添加的两个辅助任务都是帮助模型更好地识别unanswerable关系，相关性较大，所以使用一个辅助任务的效果略差于使用两个辅助任务，但是差别不大。

在融入了知识后，可以看到准确率 $P$ 值下降了0.6%，对话准确率 $P_c$ 值下降了1.3%，说明虽然使用了门控机制进行筛选，但是在知识融入过程中还是引入了部分噪声，对模型的性能造成了一些负面影响。但是召回率 $R$ 值和对话召回率 $R_c$ 值都提升了2.1%，说明知识特征的融入弥补了匮乏的对话文本特征，因而提取出了更多原先被遗漏的实体关系。比如关系“per:negative\_impression”是表示人类情感的，在对话中通常只是由一些带负面情感的词汇来表达，不会长篇大论的去表达，模型很容易因为损失掉该部分的特征而判断错误。而通过对话中出现的该关系的语义关键词，如：“bad”，“mad”等这类带负面情感的词语，模型可以得到对该关系进行语义增强后的知识特征，进而提取出正确的关系。

#### 5.4.1 在BERT上的进一步实验

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。表5给出了不同的模型在测试集上的耗时，可以看出本文提出的模型在响应速度上优于BERT模型，因此我们并未直接比较两者的性能，而是把相关模块添加到BERT模型中，实验结果如表6所示，其中BERT模型的结果来自Yu等人的实验。

模型	总耗时(ms)	平均耗时(ms)
Star-Transformer	603.7	0.3
Our work	1120.8	0.6
BERT	28059.2	15

表 5 模型耗时情况

在添加了交互信息和多任务模块后， $F_1$ 值和 $F_{1c}$ 值都提升了0.4%，进一步提升了BERT的性能，证明了方法的有效性。在此基础上进一步融入知识后综合性能 $F_1$ 值达到了58.4%， $F_{1c}$ 值上达到了53.7%，但是 $F_1$ 值只提升了0.1%， $F_{1c}$ 值提升了0.2%，提升并不是很明显，说明目前融入知识的方式还不能很好地与BERT模型结合，需要进一步优化二者的结合方式。

模型	$F_1$ (%)	$F_{1c}$ (%)
BERT	57.9	53.1
+交互信息&多任务	58.3	53.5
+融入知识	58.4	53.7

表 6 在BERT上的实验结果

## 6 结论

本文提出面向对话文本进行实体关系抽取的模型，在Star-Transformer中融入高速网络增强了模型提取对话文本的能力，使用交叉注意力机制挖掘对话中交互信息进一步丰富特征，通过关键词共现的方式融入知识指导模型关注更重要的关系，最后利用多任务学习的方式缓解数据分布不均衡的问题，从而提升了模型的性能。

我们在分析语料时发现对话中存在大量的指代情况，指代情况十分复杂。在未来的工作中可以考虑引入指代消解模型，把存在指代情况的句子重写成语义完整的句子，帮助关系抽取模型理解语义，提升模型整体的性能。

## 参考文献

- Akari Asai, Kazuma Hashimoto, Hannaneh Hajishirzi, R. Socher, and Caiming Xiong. 2020. Learning to retrieve reasoning paths over wikipedia graph for question answering. In *8th International Conference on Learning Representations*. Addis Ababa: ICLR.
- Yekun Chai, Jin Shuo, and X. Hou. 2020. Highway transformer: Self-gating enhanced self-attentive networks. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6887–6900. ACL.

- M. Chen, Zhewei Wei, Zengfeng Huang, B. Ding, and Y. Li. 2020. Simple and deep graph convolutional networks. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pages 1725–1735. Vienna: Proceedings of Machine Learning Research.
- J. Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186. Minneapolis: ACL.
- J. Feng, Minlie Huang, L. Zhao, Y. Yang, and Xiaoyan Zhu. 2018. Reinforcement learning for relation classification from noisy data. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18)*, pages 5779–5786. New Orleans: AAAI.
- Parisa Naderi Golshan, HosseinAli Rahmani Dashti, S. Azizi, and L. Safari. 2018. A study of recent contributions on information extraction. *CoRR*, abs/1803.05667.
- Karthik Gopalakrishnan, Behnam Hedayatnia, Q. Chen, Anna Gottardi, Sanjeev Kwatra, Anu Venkatesh, R. Gabriel, and D. Hakkani-Tur. 2019. Topical-chat: Towards knowledge-grounded open-domain conversations. In *Interspeech 2019, 20th Annual Conference of the International Speech Communication Association*, pages 1891–1895. Graz: ISCA.
- Zhijiang Guo, Yan Zhang, and Wei Lu. 2019. Attention guided graph convolutional networks for relation extraction. In *Proceedings of the 57th Conference of the Association for Computational Linguistics*, pages 241–251. Florence: ACL.
- Zhiheng Huang, W. Xu, and Kai Yu. 2015. Bidirectional lstm-crf models for sequence tagging. *CoRR*, abs/1508.01991.
- Shaoxiong Ji, Shirui Pan, E. Cambria, P. Marttinen, and Philip S. Yu. 2020. A survey on knowledge graphs: Representation, acquisition and applications. *CoRR*, abs/2002.00388.
- Arzoo Katiyar and Claire Cardie. 2017. Going out on a limb: Joint extraction of entity mentions and relations without dependency trees. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pages 917–928. Vancouver: ACL.
- A. Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84 – 90.
- Xiaoya Li, Fan Yin, Zijun Sun, Xiayu Li, Arianna Yuan, Duo Chai, Mingxin Zhou, and J. Li. 2019. Entity-relation extraction as multi-turn question answering. In *Proceedings of the 57th Conference of the Association for Computational Linguistics*, pages 1340–1350. Florence: ACL.
- X. Liu, K. Liu, X. Li, Jinsong Su, Y. Ge, Bin Wang, and Jiebo Luo. 2020. An iterative multi-source mutual knowledge transfer framework for machine reading comprehension. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, pages 3794–3800. IJCAI.
- M. Mintz, Steven Bills, R. Snow, and Dan Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 1003–1011. Singapore: ACL.
- Guo Q, Qiu X, Liu P, and et al. 2019. Star-transformer. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1315–1325. Minneapolis: ACL.
- Chris Quirk and Hoifung Poon. 2017. Distant supervision for relation extraction beyond the sentence boundary. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pages 1171–1182. Valencia: ACL.
- R. Staudemeyer and Eric Rothstein Morris. 2019. Understanding lstm - a tutorial into long short-term memory recurrent neural networks. *CoRR*, abs/1909.09586.
- L. Wang, Z. Cao, Gerard de Melo, and Zhiyuan Liu. 2016. Relation classification via multi-level attention cnns. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Berlin: ACL.

- Haoyu Wang, M. Tan, Mo Yu, S. Chang, Dakuo Wang, Kun Xu, Xiaoxiao Guo, and Saloni Potdar. 2019a. Extracting multiple-relations in one-pass with pre-trained transformers. In *Proceedings of the 57th Conference of the Association for Computational Linguistics*, pages 1371–1377. Florence: ACL.
- Xiang Wang, Dingxian Wang, Canran Xu, X. He, Yixin Cao, and Tat-Seng Chua. 2019b. Explainable reasoning over knowledge graphs for recommendation. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019*, pages 5329–5336. Honolulu: AAAI.
- An Yang, Q. Wang, Jing Liu, Kai Liu, Yajuan Lyu, H. Wu, Qiaoqiao She, and Sujian Li. 2019. Enhancing pre-trained language representations with rich knowledge for machine reading comprehension. In *Proceedings of the 57th Conference of the Association for Computational Linguistics*, pages 2346–2357. Florence: ACL.
- Yuan Yao, D. Ye, Peng Li, Xu Han, Yankai Lin, Zhenghao Liu, Z. Liu, Lixin Huang, Jie Zhou, and M. Sun. 2019. Docred: A large-scale document-level relation extraction dataset. In *Proceedings of the 57th Conference of the Association for Computational Linguistics*, pages 764–777. Florence: ACL.
- Dian Yu, Kai Sun, Claire Cardie, and Dong Yu. 2020. Dialogue-based relation extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4927–4940. ACL.
- W. Zaremba, Ilya Sutskever, and Oriol Vinyals. 2014. Recurrent neural network regularization. *CoRR*, abs/1409.2329.
- Yuhao Zhang, Peng Qi, and Christopher D. Manning. 2018. Graph convolution over pruned dependency trees improves relation extraction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2205–2215. Brussels: ACL.
- 李青青, 杨志豪, 罗凌, 林鸿飞, 王健. 2019. 基于多任务学习的生物医学实体关系抽取. *中文信息学报*, 33(8):84–92.
- 陈佳泮, 滕冲. 2019. 基于强化学习的实体关系联合抽取模型. *计算机应用*, 39(07):1918–1924.