

# 基于信息交互增强的时序关系联合识别

戴倩颖  
苏州大学  
计算机科学与技术学院  
20185227029@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

时序关系识别是信息抽取领域的一个重要分支，对文本理解发挥着关键作用。按照关联对象的不同，时序关系分为三大类：事件对(E-E)间的时序关系，事件与时间表达式间(E-T)的时序关系，事件与文档建立时间(E-D)间的时序关系。不同关系类型孤立识别的方法忽视了其间隐含的关联信息，针对这一问题构建了基于信息交互增强的时序关系联合识别模型。通过在不同神经网络层之间共享参数实现E-E与E-T时序关系的语义交流，利用两者的潜在联系提高识别精度。在TimeBank-Dense语料上的一系列实验表明，该方法优于现有的大多数神经网络方法。

**关键词：** 时序关系；参数共享；信息交互；联合识别

## Joint Recognition of Temporal Relation Based on Information Interaction Enhancement

Qianying Dai  
School of Computer  
Science and Technology  
Soochow University  
20185227029@stu.suda.edu.cn

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

## Abstract

Temporal relation recognition is an important branch of information extraction and plays a key role in text understanding. According to the different associated objects, temporal relation can be divided into three categories: event-event(E-E), event-time expression (E-T), and event-document creation time (E-D) temporal relation. The isolation identification method of different relation types ignores the implicit correlation information among them. To solve this problem, this paper proposes a joint identification model of temporal relation based on information interaction enhancement. By sharing parameters between different neural network layers, the semantic communication of E-E and E-T temporal relations is realized, and the potential connection between them is used to improve the recognition accuracy. A series of experiments on TimeBank-Dense corpus show that this method is superior to most existing neural network methods.

**Keywords:** Temporal relation, Parameter sharing, Information interaction, Joint recognition

基金项目:

国家自然科学基金项目，面向领域大数据的知识图谱构建（61836007）；  
国家自然科学基金项目，面向篇章信息性的汉语篇章结构多层次联合分析研究（61876118）；

## 1 引言

文本中的事件与时间表达式往往不是独立存在的，在时间角度上存在着一定的先后顺序关系。时序关系揭示了事物从开始到结束的发展过程。作为自然语言处理的基本任务之一，时序关系识别对自动问答、文本摘要、信息抽取等任务具有重要意义。近年来，随着时序关系在自然语言处理各项下游应用中作用的显现以及TempEval系列共享任务的推动，时序关系识别吸引了众多学者的关注，逐渐成为自然语言处理领域的研究热点。

时序关系识别是对具有时序关联的事件与时间表达式进行识别，并对它们之间的时序关系进行分类的一项任务(Cheng and Miyao, 2017)。在TempEval评测任务中，根据所关联对象的不同，时序关系被分为三大类：事件对间的时序关系 (E-E)、事件与时间表达式之间的时序关系 (E-T)、事件与文档建立时间之间的时序关系 (E-D)。

例1中展示了TimeBank-Dense语料中的一个句子，句中的事件与时间表达式都以粗斜体标出，其中包含两个E-T时序关系与一个E-E时序关系。

例1: In Washington *today*, the Federal Aviation Administration *released* air traffic control tapes from the night the TWA Flight eight hundred *went* down.

relation1(E-T): *today*与*released*之间的时序关系为“INCLUDES”，表示在*today*这个时间跨度内*released*事件发生。

relation2(E-T): *today*在*went*之间的时序关系为“AFTER”，表示*today*时间段是在*went*事件之后。

relation3(E-E): *released*事件在*went*事件之间的时序关系为“AFTER”，说明*released*在*went*之后发生。

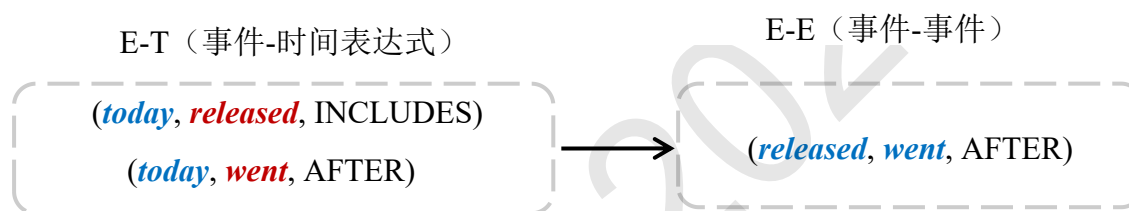


图 1 时序关系的传递性

本文将事件与时间表达式统称为时序对象。从例句中可以看出一个时序对象可与其他多个时序对象相关联，并且时序关系可以时序对象为载体进行传递。图1中，可由前两个E-T时序关系直接推导出*released*事件与*went*事件间的时序关系，由此可看出不同类型时序关系之间存在着密切的联系，但已有的研究多对不同种类的时序关系进行孤立识别，如何挖掘它们之间的联系成了关键。为此，本文提出了一种增强信息交互的时序关系联合识别方法。为了让神经网络能够更好的捕获E-E与E-T时序关系之间的语义联系，本文在编码与解码阶段同时采用参数共享机制以增强两者之间的信息交互。在TimeBank-Dense语料上的实验表明，这一模型可以捕获到更多的有效信息，时序关系识别的性能得到了显著提升。

## 2 相关工作

自Allen等人(1983; 2005)通过“区间代数法”编写出规则，由此提出的13种时序关系问世后，后期的一些研究工作主要在语言学假设或人工制定限制的基础上构建更为全面的规则集。如Dowty等人(2005)运用了叙事惯例，假设当前句子的事件在前一个句子的事件之后发生，Passonneau等人(2005)提出的时序系统中增加了对时态的分析。Chklovski等人(2005)利用事件词构建关系图，根据事件节点间的多条间接路径对时序关系进行调整，并将其应用于外部知识库verboclean上，实现了该知识库上时序关系的精炼。

随着标注体系的规范化，诞生了一些高质量的语料。由此研究者们开始尝试使用机器学习方法。Mani等人(2007)结合timebank语料库中手工标记的体态、极性、时态等特征，在时序关系抽取中应用最大熵算法完成了模型的训练。Chambers等人(2018)拓展了特征集，加

入WordNet与VerbOcean等外部知识库中的词汇和形态学特征,进行时序关系的抽取实验并取得优异结果。Bethard等(2007)使用句法特征与语义信息结合的方法,在此基础上应用了支持向量机抽取出时序对象间的关系。D'souza等人(2013)通过融合篇章级特征与语义关系特征实现了特征集的多层次覆盖,使得时序关系抽取任务的性能得到优化。Mirza等(2014)专注于为模型选取合适的特征,对特征集进行简化,在减少特征提取工作的基础上取得了相当的效果。郑新等(2015)基于中文语料的特性,在借鉴英文特征的基础上增加了相关特征,并在中文语料上对所做改进的正确性进行测试。之后一批研究人员尝试结合整数线性规划方法实现时序关系在全局层面的优化,代表性工作有(Chambers and Jurafsky, 2008)、(Li et al., 2016)、(Do et al., 2012)等。

近年来,深度学习方法被引入到时序关系抽取任务上来。Xu等人(2015)使用最短依存路径的做法在实体关系抽取任务中表现出色,Cheng等(2017)受此启发,将该方法移植到时序关系抽取任务,在此基础上使用双向LSTM实现特征的自动抽取,相较于传统方法,识别效果与效率都取得了很大提升。Choubey等人(2017)通过引入事件上下文的构成、句法和语义表示区分细粒度的时序关系,构建了基于双通道LSTM的序列模型,将其应用于句内事件对的时序关系识别。Zhang(2019)将自注意力机制引入时序关系识别,弥补了普通的非线性网络层对长序列信息捕获不足的问题,在多个语料上证实了该方法的有效性。Dai(2020)提出将句子层面的依存特征与篇章级别特征进行融合,构建了一个能捕获多层次信息的时序关系抽取模型。Zhou(2020)将事件对的两个序列放入同一个空间中进行训练增强事件之间的上下文信息挖掘,并且结合完整的句子表示弥补最短依存路径的信息缺失问题,进一步增强了事件时序关系识别性能。

已有研究多对不同类型时序关系独立识别,而时序关系的传递性使得不同类型的时序关系之间并不是孤立的,单独识别的方式割裂了他们之间的联系。针对这一问题,本文从E-E时序关系和E-T时序关系的关联性出发,通过编码层与解码层的信息交互挖掘两种类型时序关系间的语义联系,提取出时序对象间更多的潜在关联信息,从而提升时序关系识别的整体性能。

### 3 增强信息交互的时序关系联合识别模型

本文所提出的增强信息交互的时序关系联合识别模型的网络构架如图1所示,该模型左侧为E-E时序关系识别模块,右侧为E-T时序关系识别模块,两个模块具有相同的架构,底端使用BERT来获得事件或时间表达式上下文相关的词向量表征,顶端是用于关系分类的全连接层,值得注意的是,在这里本文除了识别时序关系,同时还增加了对VAGUE类别的识别,详细情况后续会进行说明。模型中两个模块的编码与解码层是共用的,最终通过损失函数进行结合。接下来对模型进行详细介绍。

#### 3.1 编码层

单词所表示的含义依赖于其所在上下文,比如“今年苹果大丰收”与“乔布斯创建了苹果”这两个句子中的“苹果”传达了不同的含义。而word2vec、Glove等传统词向量在转换过程中忽视上下文语境,固定表征,缺乏对单词多义性的建模能力,上述两句中的“苹果”在向量空间中会被映射成相同表示,缺乏合理性。BERT(Devlin et al., 2019)作为一种新型的预训练语言模型,其在预训练词向量时能够挖掘词序列中丰富的句法与语义信息,很好的理解文本中的一词多义。简单的说,BERT可获得上下文相关的动态预训练表示。鉴于BERT优秀的多义词建模与序列表示能力,本节所提出的多任务系统将其作为词向量模型。

不同于以往多数研究将两个时序对象所在句子放入两个独立的神经表示学习器进行训练,为了能够捕获到要判别的时序对象之间的语义联系,本文将两个句子拼接而成的长序列作为输入,使得两个时序对象在一个空间内共同编码,每个句子以[SEP]标签为结尾,拼接后的序列头部添加[CLS]。BERT模型中[CLS]位置的输出向量是整个序列的语义表示,包含了完整的上层特征。为了避免丢失时序对象在序列中的位置信息,本节又分别提取出两个时序对象A与B经过BERT训练后的最高隐层输出,拼接三者,融合整体与局部信息,获得了更完整的表征:

$$v = e_{CLS} \oplus e_A \oplus e_B \quad (1)$$

其中,  $e_{CLS}$  代表[CLS]输出的向量,  $e_A$ 和 $e_B$  为两个时序对象的输出向量。

之前的研究对于不同类型的时序关系独立进行识别,这种做法限制了两者在训练过程中的信息交互,而不同类型时序关系之间联系的挖掘对于关系判别是具有促进作用的。针对这个问

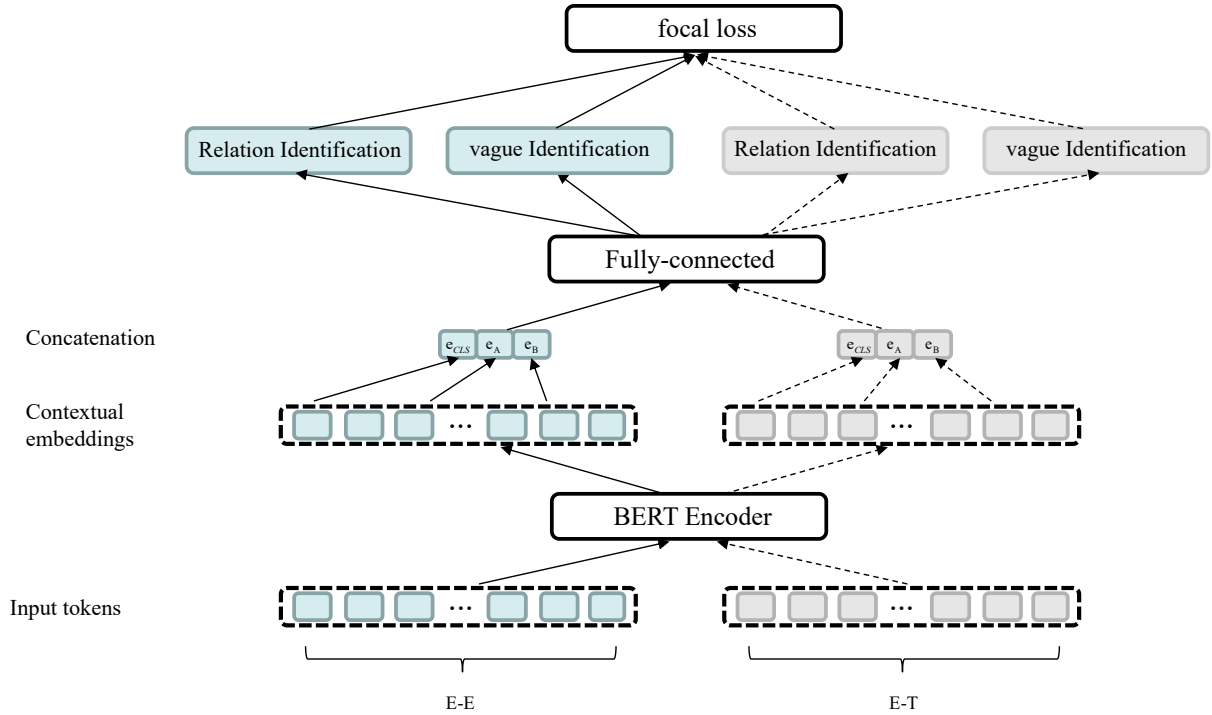


图 2 增强信息交互的时序关系联合识别模型

题，本文采用了一种参数共享机制，即使用一个BERT编码层先后对E-E与E-T时序关系中时序对象所在序列进行训练。在批训练（mini-batch）过程中，先将一个batch的E-E时序关系序列输入到BERT模型中进行词向量的预训练，再将一个batch的E-T时序关系序列输入到BERT模型中进行词向量的预训练，两种类型的时序关系在训练过程中可以获得彼此的信息，完成信息的交互过程，对彼此的关系识别产生辅助作用。

### 3.2 解码层

为了把BERT编码层得到的分布式特征表示映射到样本标签空间，我们在BERT后面接入一个全连接层，对之前得到的特征表示进行压缩：

$$h = \text{relu}(vW_h + b_h) \quad (2)$$

之后是时序关系识别模块。在TB-D语料中，不同时序关系类型分布数量差异过大导致的数据不均衡问题使得低频的时序关系类别经常被误分到高频的类别中，尤其是语料中占比过半的VAGUE类型，很大程度上限制了识别的性能。本文通过在解码层引入一个辅助任务来提升模型对不平衡数据的处理能力。简单的说，以时序关系识别任务为主完成时序关系的六分类，二元关系识别任务为辅，用于判断两个时序对象之间的时序关系是否为VAGUE。一方面，模型底部的隐藏层为这两个相关任务所共享，训练过程中利用两者之间的关联，互相传播监督信号实现彼此性能的提升。VAGUE识别任务中的信号作为归纳偏置传播给时序关系识别任务，能够有效避免时序类别的错分。另一方面，任务中不相关的部分对彼此而言相当于噪声，噪声的引入能够实现任务间的约束，避免单个模型的过度拟合，训练过程中通过平衡两个任务的噪声提升模型的泛化能力。

模型采用softmax层作为输出层来获得主任务与辅助任务的关系标签的预测结果：

$$o_{main} = \text{softmax}(hW_{o_{main}} + b_{o_{main}}) \quad (3)$$

$$o_{aux} = \text{softmax}(hW_{o_{aux}} + b_{o_{aux}}) \quad (4)$$

其中， $h$ 对应完整句子表征经过全连接层的输出， $W_o$ 和 $b_o$ 是softmax函数的权重矩阵和偏置量。



在这里，本文为了进一步增强E-E时序关系识别与E-T时序关系识别的信息交互，在解码层同样采用了参数共享机制，使用同一个全连接层与softmax层对两者进行解码。与编码阶段类似，先将一个batch的E-E时序关系的隐层序列表示输入解码层进行关系标签的预测，再将一个batch的E-T时序关系的隐层序列表示输入解码层进行关系标签的预测。

### 3.3 focal loss

数据集中正负样本、难易样本分布的不平衡会使得模型的优化脱离设想。比如，在TB-D语料的E-E时序关系中，“SIMULTANEOUS”表示两个事件在同一时间发生或发生时间接近，足以忽略时间差。这种关系类型仅占样本总数的1.5%，而“AFTER”类别占比高达18.1%。后者相比前者容易识别的多。训练过程中，大量的简单“AFTER”样本会将难识别的“SIMULTANEOUS”样本淹没，不利于稀疏关系类别的预测。聚焦损失（focal loss）(Lin et al., 2020)的引入能够减小训练过程中大量简单负样本所占的权重，更好的实现对困难样本信息的挖掘，提升模型的识别效果。

假设 $t_i$ 是第 $i$ 个样本的真实概率值， $p_i$ 表示条件概率值，则focal loss可以定义为：

$$J(\theta) = -\sum_{i=1}^n (\alpha_i (1 - p_i)^\gamma t_i \log(p_i)) \quad (5)$$

focal loss 是在交叉熵损失的基础上改进的。相比交叉熵损失，focal loss增加了一个权重因子 $\alpha$ 以及一个调制因子 $(1 - p_i)^\gamma$ ，其中 $\gamma$ 是一个难易样本调节参数。 $\alpha$ 用来缓解样本数据不平衡问题，根据类别出现的频率决定，对于数量较少的类别，使得其 $\alpha$ 取值相对更大。这样，模型会更加专注于优化样本少的那些损失。 $(1 - p_i)^\gamma$ 调制因子用来调整易正确判别的样本损失和不易正确判别的样本损失之间的比例，能够减少易分类样本对于损失函数的贡献，从而使得模型在训练时更专注于难分类的样本，本次实验中 $\alpha$ 取值为0.25， $\gamma$ 取值为2。

将时序关系识别与“VAGUE”类别识别两个任务的损失函数进行结合，得到的E-E时序关系模块与E-T时序关系模块的损失函数计算公式如下：

$$J_1 = -\lambda \sum_{i=1}^m (\alpha_{main_i} (1 - p_{main_i})^\gamma t_{main_i} \log(p_{main_i})) \quad (6)$$

$$J_2 = -(1 - \lambda) \sum_{i=1}^m (\alpha_{aux_i} (1 - p_{aux_i})^\gamma t_{aux_i} \log(p_{aux_i})) \quad (7)$$

$$J_{E-E}(\theta) = J_1 + J_2 + \frac{1}{2} \|\theta\|^2 \quad (8)$$

$$J_3 = -\lambda \sum_{i=1}^n (\alpha_{main_i} (1 - p_{main_i})^\gamma t_{main_i} \log(p_{main_i})) \quad (9)$$

$$J_4 = -(1 - \lambda) \sum_{i=1}^n (\alpha_{aux_i} (1 - p_{aux_i})^\gamma t_{aux_i} \log(p_{aux_i})) \quad (10)$$

$$J_{E-T}(\theta) = J_3 + J_4 + \frac{1}{2} \|\theta\|^2 \quad (11)$$

其中 $m$ 表示E-E时序关系训练样本个数， $n$ 表示E-T时序关系训练样本个数， $\theta$ 为模型训练参数集合， $\lambda$ 为权重参数，用来分配主任务与辅助任务训练过程中的损失函数的重要性，使得二者保持平衡，本实验中取值为0.8。

结合E-E是时序关系模块与E-T时序关系模块的损失函数作为最终的目标函数，这个目标函数用于反向传播时，可以使得两个模块的信息共同对模型参数更新产生影响。

$$J_{joint}(\theta) = \omega J_{E-E}(\theta) + (1 - \omega) J_{E-T}(\theta) \quad (12)$$

其中 $\omega$ 是保持E-E是时序关系模块与E-T时序关系模块损失平衡的权重参数，本实验中取值为0.8。

## 4 实验与分析

### 4.1 数据集及评价指标

本文在TimeBank-Dense(TB-D)语料上对所提方法进行验证。

TB-D语料中根据所关联对象的不同，将时序关系分为三大类：事件与事件(Event-Event, E-E)，事件与时间表达式(Event-TimeX, E-T)，事件与文档建立时间(Event-Document Creation Time, E-D)，总共包含了六种不同的时序关系：BEFORE, AFTER, SIMULTANEOUS, IS-INCLUDED, INCLUDES, VAGUE。本节关注于E-E时序关系与E-T时序关系。语料中的具体分布情况如表1所示。

Relation	E-E	E-T
AFTER	1120	439
BEFORE	1348	418
SIMULTANEOUS	93	34
INCLUDE	278	196
IS-INCLUDED	347	243
VAGUE	2904	852
Overall	6088	2182

表 1 TB-D不同时序关系类型分布情况

为了评价本文方法的有效性，我们采用准确率，召回率以及F1值作为评价指标。

### 4.2 实验设置

由于TB-D语料中文档数过少，本文采用文档+句子id级别的五折交叉验证方法进行训练集与测试集的划分，从训练集中随机抽取15%的样本用作验证集，由于实验结果使用多分类微平均作为评价标准，所以总体P、R、F1值相同，因此本文的实验结果部分中仅将F1值列出。模型采用了小批次训练(mini-batch)的策略，总共训练15轮，E-E模块的批次大小为24，E-T模块的批次大小为8。模型参数的训练更新通过Adam优化算法实现，初始学习率大小为 $2e-5$ 。训练过程中使用L2正则化防止过拟合。本节使用“bert-base-uncased”进行词向量的预训练，全连接层的输出特征宽度为512。在各个层间引入dropout机制，丢弃几率设置为0.5。最终保存在验证集上效果最好的模型对测试集进行时序关系抽取。

### 4.3 实验结果与分析

#### 4.3.1 与已有研究的比较

为了验证本文提出的模型在时序关系分类上的性能，我们选择了目前传统方法以及神经网络架构下性能表现优异的以下系统做对比：

CAEVO:2014年，Chambers等人(2014)将多个基于规则与机器学习的分类算法进行结合，提出了一个用于时序关系识别的混合模型。

MIRZA: 2016年，Mirza等人(2016)结合低维词向量与传统特征向量，应用逻辑斯蒂回归算法完成了时序关系分类器的训练。

Cheng:Cheng 等人(2017)于2017年提取事件词的最短依存路径作为输入，结合路径中的单词、词性以及依存关系，使用BiLSTM识别时序关系。

Dai:Dai(2020)等人于2019年提出将句子所在的依存路径信息与事件所在基本篇章单元间的修辞关系信息相结合，给出了一种能捕获更多有效上下文信息的表征策略。

Zhou: Zhou(2020)等人于2020年针对最短依存路径信息缺失的问题融合了完整的句子表示，并且将事件对所在序列放入同一个LSTM中训练，提升了事件对时序关系识别的效果。

表2与表3中列出了具体的比较结果，ours表示本文提出的联合学习模型，从表中可以看到，与其他start-of-the-art的方法相比，ours模型显示出在时序关系识别任务上的优越性。

E-E关系识别方面，相较于Dai与Zhou，ours模型的总体F1值分别提升了9.2%、4.8%。单独类别上的表现也更为出色。性能的提升一方面得益于编解码层的信息交互机制对不同类型时序关系潜在联系的挖掘作用，另一方面，解码层时序关系识别任务与VAGUE类别识别任务的

Relation	Cheng	Dai	zhou	ours
AFTER	44.0	56.7	60.1	72.3
BEFORE	46.0	55.9	57.3	67.8
SIMULTANEOUS	-	-	-	2.1
INCLUDE	2.5	22.9	20.5	33.7
IS-INCLUDED	17.0	30.4	34.5	58.5
VAGUE	62.4	63.4	67.6	68.4
Overall	52.9	56.9	61.3	66.1

表 2 E-E时序关系识别实验结果对比

结合也可以有效避免关系类别的错分。值得关注的是, 本文提出的系统在SIMULTANEOUS类别的识别方面突破了以往研究的困境, F1值达到了2.1%。

Relation	CAEVO	MIRZA	Cheng	ours
AFTER	-	-	-	69.5
BEFORE	-	-	-	71.4
SIMULTANEOUS	-	-	-	5.4
INCLUDE	-	-	-	55.5
IS-INCLUDED	-	-	-	61.0
VAGUE	-	-	-	60.1
Overall	49.4	46.8	47.1	63.6

表 3 E-T时序关系识别实验结果对比

E-T关系识别方面, 与使用传统方法的CAEVO以及MIRZA系统相比, 本文的方法分别提升了14.2与16.8个百分点, 与同样使用神经网络的Cheng相比, 本文方法提升了16.5个百分点, 性能提升显著, 同样证明了信息交互机制的有效性。

#### 4.3.2 实验分析

本文所提的方法在时序关系识别任务上表现优异, 下面对影响E-E时序关系识别实验结果的几个主要因素进行分析。

##### (1) 联合识别

在验证时序关系联合识别效果的同时, 为了分析不同网络层信息交互对联合识别的贡献度, 本文设置了以下4个模型进行对比。

Baseline: 单独的E-E时序关系识别。

Encoder sharing(ES): E-E与E-T时序关系联合识别, 只在编码层参数共享进行信息交互。

Decoder sharing(DS): E-E与E-T时序关系联合识别, 只在解码层参数共享进行信息交互。

ours: E-E与E-T时序关系联合识别, 编码与解码层同时参数共享进行信息交互。

Relation	Bs	ES	DS	ours
AFTER	69.7	72.5	70.6	72.3
BEFORE	66.3	67.2	66.9	67.8
SIMULTANEOUS	6.4	3.5	5.4	2.1
INCLUDE	32.5	33.6	33.2	33.7
IS-INCLUDED	46.3	54.4	50.7	58.5
VAGUE	65.2	64.6	64.8	68.4
Overall	63.5	64.0	63.8	66.1

表 4 不同网络层信息交互联合识别的实验结果

从表4中给出的结果可以看出，无论在哪个网络层进行信息交互，联合识别的方法都优于baseline方法中单独的E-E时序关系识别，ours模型与之相比总体F1值提升了2.6%，从各个类别上来看除了“SIMULTANEOUS”性能都有了较大的提升。原因在于，单独的E-E时序关系识别仅仅专注于挖掘句中事件对之间的直接关系，与E-T时序关系之间信息的交互使得事件对之间的内在联系被提取出来，有效提升了时序关系识别的实验效果。SIMULTANEOUS类别由于数量稀少，所以实验性能不太稳定。

以句子“The gain wages 3.8 percent higher than a year earlier, extending a trend that has given back to workers some of the earning power they lost to inflation in the last decade.”为例，事件extending与事件lost的时序关系为after，BS模型将其误分为“VAGUE”，与E-T时序关系进行联合学习后，增加了与时间表达式a year earlier的关联，判别出a year earlier与事件extending关系为“BEFORE”，与事件lost时序关系为“AFTER”，E-T时序关系的加入使得extending与lost更多的上下文相关信息被挖掘，学习到更多的依赖信息，帮助该事件对的时序关系被成功预测。

并且，相较于ES与DS，ours模型在单个时序关系类型与整体上几乎都取得了最优的结果，证明编码层与解码层的双重参数共享机制可以实现更深层次的语义信息的交流，加强对不同种类时序关系联系的挖掘，最大幅度的提升了联合学习的实验性能。

### (2) “VAGUE”类别识别

为了验证解码层中增加的辅助任务，即“VAGUE”类别识别的效果，我们设置了一个基准模型without，与ours模型相比，该基准模型去除了E-E与E-T时序关系模块中的辅助任务部分。

Relation	without	ours
AFTER	71.4	72.3
BEFORE	67.4	67.8
SIMULTANEOUS	1.9	2.1
INCLUDE	32.8	33.7
IS-INCLUDED	55.3	58.5
VAGUE	67.5	68.4
Overall	65.4	66.1

表 5 加入辅助任务前后的实验结果对比

从表5中列出的实验结果可以看出，辅助任务的加入对于时序关系识别任务是具有帮助作用的，与去除辅助任务的模型without相比，ours模型的整体F1值上升了0.7%，单个关系类别上也都有不同程度的提升，说明辅助任务的加入能帮助系统更好地区分“VAGUE”和其他类别，在一定程度上缓解了语料库本身存在的数据不平衡问题，并最终提升其他类别的时序关系的识别性能。

### (3) focal loss

为了验证引入focal loss的效果，我们设置了一个基准模型cross，与ours模型相比，该基准模型使用了交叉熵损失作为目标函数。

Relation	cross	ours
AFTER	72.1	72.3
BEFORE	66.9	67.8
SIMULTANEOUS	2.1	2.1
INCLUDE	33.0	33.7
IS-INCLUDED	51.5	58.5
VAGUE	67.8	68.4
Overall	65.6	66.1

表 6 使用不同损失函数的实验结果对比



表6中列出了使用不同损失函数的两个模型的实验结果，与使用普通的交叉熵损失相比，使用聚焦损失后总体F1值提升了0.5%，说明损失函数的改进也有助于时序关系的正确分类。在数据集中占比较小的INCLUDES时序关系类别上F1值提升了0.7%，对IS-INCLUDED关系类别的促进作用则更加明显，F1值提升了7个百分点。聚焦损失的引入对于识别难分类样本的有效性得到充分验证。

## 5 结束语

本文针对TB-D语料不同种类时序关系联合研究有限的现状，提出了一种增强信息交互的方式实现E-E与E-T时序关系的联合识别，一方面，引入参数共享机制，通过编码与解码层的信息交互挖掘两种类型时序关系的语义联系，使得时序对象间更多的潜在关联信息被提取出来。另一方面，在解码层设计了一个辅助任务，通过利用二元分类的感应信号，提升模型对不平衡数据的处理能力。一系列实验表明，经过改进后的模型在时序关系识别任务上能够取得更好的表现。

## 致谢

感谢各位匿名评审专家对本文提出的宝贵修改意见。本次工作受到国自然重点项目(No.61836007)与国自然面向项目(No.61876118)支持。

## 参考文献

- James F. Allen. 1983. Maintaining knowledge about temporal intervals. *Commun. ACM*, 26(11):832–843.
- James F. Allen. 2005. Towards a general theory of action and time. In Inderjeet Mani, James Pustejovsky, and Robert J. Gaizauskas, editors, *The Language of Time - A Reader*, pages 251–276. Oxford University Press.
- Steven Bethard and James H. Martin. 2007. CU-TMP: temporal relation classification using syntactic and semantic features. In Eneko Agirre, Lluís Màrquez i Villodre, and Richard Wicentowski, editors, *Proceedings of the 4th International Workshop on Semantic Evaluations, SemEval@ACL 2007, Prague, Czech Republic, June 23-24, 2007*, pages 129–132. The Association for Computer Linguistics.
- Nathanael Chambers and Daniel Jurafsky. 2008. Jointly combining implicit constraints improves temporal ordering. In *2008 Conference on Empirical Methods in Natural Language Processing, EMNLP 2008, Proceedings of the Conference, 25-27 October 2008, Honolulu, Hawaii, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 698–706. ACL.
- Nathanael Chambers, Taylor Cassidy, Bill McDowell, and Steven Bethard. 2014. Dense event ordering with a multi-pass architecture. *Trans. Assoc. Comput. Linguistics*, 2:273–284.
- Fei Cheng and Yusuke Miyao. 2017. Classifying temporal relations by bidirectional LSTM over dependency paths. In Regina Barzilay and Min-Yen Kan, editors, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 2: Short Papers*, pages 1–6. Association for Computational Linguistics.
- Timothy Chklovski and Patrick Pantel. 2005. Global path-based refinement of noisy graphs applied to verb semantics. In Robert Dale, Kam-Fai Wong, Jian Su, and Oi Yee Kwong, editors, *Natural Language Processing - IJCNLP 2005, Second International Joint Conference, Jeju Island, Korea, October 11-13, 2005, Proceedings*, volume 3651 of *Lecture Notes in Computer Science*, pages 792–803. Springer.
- Prafulla Kumar Choubey and Ruihong Huang. 2017. A sequential model for classifying temporal relations between intra-sentence events. In Martha Palmer, Rebecca Hwa, and Sebastian Riedel, editors, *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, pages 1796–1802. Association for Computational Linguistics.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. 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, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.
- Quang Do, Wei Lu, and Dan Roth. 2012. Joint inference for event timeline construction. In Jun'ichi Tsujii, James Henderson, and Marius Pasca, editors, *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, EMNLP-CoNLL 2012, July 12-14, 2012, Jeju Island, Korea*, pages 677–687. ACL.
- David R. Dowty. 2005. The effects of aspectual class on the temporal structure of discourse: Semantics or pragmatics? In Inderjeet Mani, James Pustejovsky, and Robert J. Gaizauskas, editors, *The Language of Time - A Reader*, pages 333–352. Oxford University Press.
- Jennifer D'Souza and Vincent Ng. 2013. Classifying temporal relations with rich linguistic knowledge. In Lucy Vanderwende, Hal Daumé III, and Katrin Kirchhoff, editors, *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 9-14, 2013, Westin Peachtree Plaza Hotel, Atlanta, Georgia, USA*, pages 918–927. The Association for Computational Linguistics.
- Peifeng Li, Qiaoming Zhu, Guodong Zhou, and Hongling Wang. 2016. Global inference to chinese temporal relation extraction. In Nicoletta Calzolari, Yuji Matsumoto, and Rashmi Prasad, editors, *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*, pages 1451–1460. ACL.
- Tsung-Yi Lin, Priya Goyal, Ross B. Girshick, Kaiming He, and Piotr Dollár. 2020. Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 42(2):318–327.
- Inderjeet Mani, Ben Wellner, Marc Verhagen, and James Pustejovsky. 2007. Three approaches to learning tlinks in timeml. *Technical Report CS-07-268, Computer Science Department*.
- Paramita Mirza and Sara Tonelli. 2014. Classifying temporal relations with simple features. In Gosse Bouma and Yannick Parmentier, editors, *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2014, April 26-30, 2014, Gothenburg, Sweden*, pages 308–317. The Association for Computer Linguistics.
- Paramita Mirza and Sara Tonelli. 2016. On the contribution of word embeddings to temporal relation classification. In Nicoletta Calzolari, Yuji Matsumoto, and Rashmi Prasad, editors, *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*, pages 2818–2828. ACL.
- Rebecca J. Passonneau. 2005. A computational model of the semantics of tense and aspect. In Inderjeet Mani, James Pustejovsky, and Robert J. Gaizauskas, editors, *The Language of Time - A Reader*, pages 129–172. Oxford University Press.
- Yan Xu, Lili Mou, Ge Li, Yunchuan Chen, Hao Peng, and Zhi Jin. 2015. Classifying relations via long short term memory networks along shortest dependency paths. In Lluís Màrquez, Chris Callison-Burch, Jian Su, Daniele Pighin, and Yuval Marton, editors, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015*, pages 1785–1794. The Association for Computational Linguistics.
- Yijie Zhang, Peifeng Li, and Guodong Zhou. 2018. Classifying temporal relations between events by deep bilstm. In Minghui Dong, Moch Arif Bijaksana, Herry Sujaini, Ade Romadhony, Fariska Z. Ruskanda, Elvira Nurfadhilah, and Lyla Ruslana Aini, editors, *2018 International Conference on Asian Language Processing, IALP 2018, Bandung, Indonesia, November 15-17, 2018*, pages 267–272. IEEE.
- 周新宇 and 李培峰. 2020. 基于信息交互增强的事件时序关系分类方法. *计算机科学*, 47(11):244–249.
- 张义杰, 李培峰, and 朱巧明. 2019. 基于自注意力机制的事件时序关系分类方法. *计算机科学*, 46(8):244–248.
- 王俊, 史存会, 张瑾, 俞晓明, 刘悦, and 程学旗. 2020. 融合上下文信息的篇章级事件时序关系抽取方法. *计算机研究与发展*, page 1.
- 郑新, 李培峰, 朱巧明, et al. 2015. 中文事件时序关系的标注和分类方法. *计算机科学*, 42(7):276–279.