

# 基于信息交互增强的事件时序关系分类方法



周新宇 李培峰

苏州大学计算机科学与技术学院 江苏 苏州 215006

江苏省计算机信息处理技术重点实验室 江苏 苏州 215006

(20185227075@stu.suda.edu.cn)

**摘要** 事件时序关系分类任务是信息抽取领域的一个分支,由于其对多个自然语言处理任务具有很好的辅助作用,近年来得到了越来越多的关注。目前,已有的神经网络方法对事件间信息交互的考虑相对缺乏。针对这个问题,提出了一种通过参数共享来增强事件间信息交互的方法。该方法首先通过门控卷积神经网络(Gated Convolutional Neural Network,GCNN)学习句子的语义信息和上下文信息,并将其融入最短依存路径序列作为输入;然后使用双向长短期记忆网络(Bidirectional Long Short-Term Memory network,Bi-LSTM)对输入进行编码以获取其语义表示,并通过参数共享来增强事件之间的信息交互;最后将获得语义表示输入全连接层,使用 Softmax 函数进行分类预测。TimeBank-Dense 语料库上的实验结果表明,所提方法在分类精度上优于现有的大多数神经网络方法。

**关键词:** 时序关系分类; 句子表示; 信息交互; 最短依存路径; 双向长短期记忆网络

**中图法分类号** TP391.1

## Event Temporal Relation Classification Method Based on Information Interaction Enhancement

ZHOU Xin-yu and LI Pei-feng

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

Provincial Key Laboratory for Computer Information Processing Technology, Suzhou, Jiangsu 215006, China

**Abstract** As a branch of information extraction, event temporal relation classification has attracted more and more attention in recent years due to its good auxiliary effect on many natural language processing tasks. At present, the existing neural network approaches lack of consideration for the information interaction between events. To address this issue, this paper proposes a method of event temporal relation classification based on parameter sharing to enhance information interaction between events. This method firstly learns the semantic information and context information of sentences through gated convolutional neural networks (GCNN), and incorporates them into the shortest dependency path sequence as input. Then, it uses Bidirectional long short-term memory network (Bi-LSTM) to encode the input and capture its semantic representation. In addition, it enhances the information interaction between event pairs by parameter sharing. Finally, the obtained semantic representation is input into the fully connected layer, and the softmax function is used for classification prediction. Experimental results on TimeBank-Dense show that the proposed method outperforms most of the existing neural network methods in classification accuracy.

**Keywords** Temporal relation classification, Sentence representation, Information interaction, Shortest dependency path, Bidirectional long and short-term memory network

### 1 引言

事件时序关系描述了不同事件发生的先后顺序,揭示了某一话题下事件从开始到结束的演变过程。事件时序关系的识别是自然语言处理任务中的重要课题,与其相关的研究具有一定的现实应用价值。比如,在医疗领域中,医生可以通过相关技术构建时间表,自动为病人制订就诊及服药计划。在自然语言处理领域,由于事件时序关系在文本摘要、自动问答

等任务中展现出的良好辅助功能,以及 TempEval 评测任务的推动,时序关系识别任务吸引了越来越多研究者的关注。

在 TempEval 评测任务中,事件时序关系识别被分为 3 个子任务:事件与时间表达式之间的时序关系识别(Event-Timex)、事件与文档创建时间之间的时序关系识别(Event-DCT),以及事件对之间的时序关系识别(Event-Event)。本文主要对事件对之间的时序关系识别展开研究。目前的主流观点将时序关系识别看作分类问题,事件时序关系分类旨在

收稿日期:2019-09-09 返修日期:2019-11-25 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61836007,61772354,61773276)

This work was supported by the National Natural Science Foundation of China (61836007,61772354,61773276).

通信作者:李培峰(pfli@suda.edu.cn)

对事件对之间发生时间的先后关系(如“**AFTER**”“**BEFORE**”等)进行正确分类。在 Timebank-Dense 语料库中,通常将一个单独的词标注为事件。例如,在句子  $S_1$  中,事件“**advanced**”与“**begin**”之间的时序关系是“**BEFORE**”,即“**advanced**”发生于“**begin**”之前。

$S_1$ : One of the scenarios widely **advanced** before the visit is that through the Pope, Cuba, still led by Castro, can **begin** a more productive relationship with the world.

早期的时序关系研究工作<sup>[1-5]</sup>倾向于使用基于特征工程的统计机器学习方法,并在此基础上结合全局优化推理方法来进行关系分类。但是,这些方法对人工标注的特征具有较强的依赖性,而这些特征的标注工作往往需要耗费大量的人力与财力资源,在实际应用中往往很难实现。

近年来,随着深度学习技术的发展,各种不同的神经网络方法被引入到自然语言处理领域中。神经网络方法<sup>[6-12]</sup>由于对外部特征的依赖性较小且实验效果较好,受到了大多数研究者的青睐。之前的研究工作通常将事件对中两个事件所在的序列分别输入两个独立的神经网络模型进行训练,对事件词所在序列之间信息交互的重视程度较弱,使得模型不容易捕获到事件之间的关系。

针对这类问题,本文提出了一种基于信息交互增强的时序关系分类方法。鉴于基于最短依存路径的神经网络模型在多个自然语言处理任务中具有良好的表现,本文采用最短依存路径作为模型的输入。针对最短依存路径会丢失部分句子信息的缺点,本文在输入信息中融合了完整的句子表示,然后使用 Bi-LSTM 对输入序列进行编码以生成其语义表示。为了让神经网络能够更好地捕获到要判别的事件对之间的语义联系,本文采用了参数共享机制来增强事件对所在序列之间的信息交互。最后,综合利用上述信息识别时序关系。在 TimeBank-Dense 语料库上的实验结果表明,相比已有方法,本文的方法在微平均 F1 值、宏平均 F1 值以及除“SIMULTANEOUS”以外的时序关系类型上的识别精度均有提升。

## 2 相关工作

事件时序关系分类任务的传统方法专注于提取文本中的各种有效特征,并使用基于特征的分类器进行关系分类。Mani 等<sup>[1]</sup>在 TimeBank 语料库上提出了实体属性特性,如事件类别、时态、体态、极性,并通过最大熵分类器进行时序关系分类。Chambers 等<sup>[2]</sup>在 Mani 等<sup>[1]</sup>工作的基础上拓宽了特征空间,加入了词性、句法树等语义特征,并提取 WordNet 中派生的词汇和形态学特征,进一步提升了分类器性能。

基于统计机器学习的方法只是针对单一的事件对进行识别,但是多个不同事件对之间的关联对于时序关系的判别有着明显的辅助作用。因此,一些全局优化推理方法被结合到分类器中用于时序关系分类。Chambers 等<sup>[3]</sup>在之前工作的基础上,使用整数线性规划方法对分类结果进行传递性约束,提升了实验性能。Do 等<sup>[4]</sup>将文章中出现的所有事件按发生的时间顺序构建事件链,以获取完整的事件对联系,并通过整数线性规划方法对事件链进行优化。D'Souza 等<sup>[5]</sup>在 Do 等<sup>[4]</sup>工作的基础上,将篇章级别的特征加入特征空间,利用篇章结构的特点获取到了更有用的语义信息,并将 VerbOcean

中动词之间的语义关系合并后加入分类器,进一步提升了系统的识别性能。

近年来,基于深度学习框架的神经网络方法在很多自然语言处理任务中展现出优异的性能,一些研究人员尝试将其引入到时序关系分类任务中。Cheng 等<sup>[6]</sup>借鉴 Xu 等<sup>[7]</sup>在实体关系分类任务上基于依存路径的研究工作,将事件词之间的最短依存路径作为模型输入,构造了一种基于双向长短期记忆网络的神经网络模型。在没有使用任何显式特征(类、时态、极性)和外部资源的情况下,该模型取得了与当时最好的基准系统相当的效果。Choubey 等<sup>[8]</sup>通过采用顺序学习能力良好的双通道长短期记忆网络来学习两个事件词上下文的句法和语义表示,提出了一种使用序列模型进行细粒度时序分类的方法。Yao 等<sup>[9]</sup>根据叙事类文本在事件时间关系上的有序性,从大量新闻、小说、博客中提取出叙事类文本,再从叙事文本中抽取事件时间知识,并将事件时间知识加入神经网络的训练中,提升了时序关系的分类效率。Tourille 等<sup>[10]</sup>提出了一种基于长短期记忆网络(Long Short-Term Memory, LSTM)的神经网络架构,用于识别医学事件或时间表达式之间的包含关系。Meng 等<sup>[11]</sup>以神经图灵机为蓝本,通过模仿人类的阅读习惯构建了一个记忆池,以叙述顺序存储处理过的时序关系,使用可更新外部存储器的神经网络模型来整合全局上下文信息,从而提升了时序关系的分类效率。Zhang 等<sup>[12]</sup>提出了一种深度 Bi-LSTM 网络模型,将多层 Bi-LSTM 网络中所有前置层的输出拼接后作为后置层的输入来构造深度连接的网络结构,从而捕捉到深层的语义信息,使得实验性能明显提升。

## 3 基于信息交互增强的事件时序关系分类模型

本文提出了一种基于信息交互增强的事件时序关系分类模型,模型架构如图 1 所示。其中,通过 GCNN 训练出的句子表示分别被加入到对应的最短依存路径序列中。

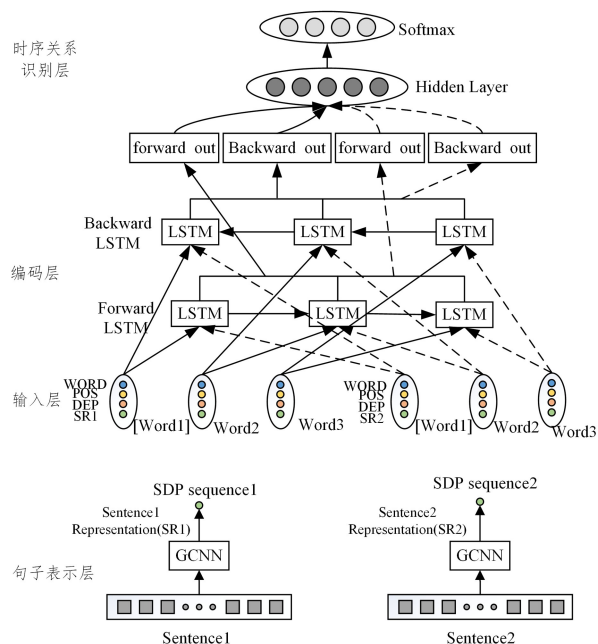


图 1 模型架构图

Fig. 1 Model architecture diagram

(1) 句子表示层: 将事件词所在的句子序列输入 GCNN 中, 学习句子的语义信息和上下文信息。

(2) 输入层: 该模型的输入为融合了句子表示的最短依存路径序列, 序列中的每个单词包含 4 种特征信息, 即单词本身、词性、依存关系表示和完整的句子表示, 并将输入的特征转化为低维实值向量。

(3) 编码层: 使用双向长短期记忆网络学习输入序列的抽象语义表示, 并加入参数共享机制以增强事件对之间的信息交互。

(4) 时序关系识别层: 将模型训练得到的结果输入一层全连接层, 然后使用 Softmax 函数对事件对之间的时序关系进行分类预测。

### 3.1 最短依存路径

依存解析树是一种表示词与词之间依存关系的树状图, 最短依存路径就是在依存解析树上从一个词结点到另一个词结点的最短路径。Xu 等<sup>[7]</sup>率先将最短依存路径引入实体关系分类任务中, 并取得了优异的效果, 证明了最短依存路径可以为关系分类任务提供有效信息。因此, 本文引入最短依存路径作为模型的输入。

根据两个事件词相对位置的不同, 本文在抽取最短依存

路径时将事件对分为同句和跨句两种类型。对于同句中的事件对, 以两个事件词的最近公共祖先结点  $A$  为中心, 将最短依存路径分为两个子路径: 从源事件词到  $A$  的路径, 以及从目标事件词到  $A$  的路径。对于跨句的事件对, 本文沿用了 Cheng 等<sup>[6]</sup>的做法, 假设两个句子的依存解析树之间存在一个公共根结点, 同样将最短依存路径分为两个子路径, 每个子路径从一个事件词指向虚构的公共根结点。以句子  $S_2$  和  $S_3$  中的事件对 (search, bombings) 为例, 从  $S_2$  的依存解析树中提取事件词 search 到根结点的路径 (search, assigned) 作为左分支, 再从  $S_3$  的依存解析树中提取事件词 bombings 到根结点的路径 (bombings, link, suspect) 作为右分支, 最后将它们连接到假设的公共根结点, 组成最短依存路径。

$S_2$ : And this week, FBI director Louie Freeh assigned more agents to the search.

$S_3$ : Freeh also ordered the investigation consolidate information under the Atlanta task force, another indication officials suspect a link between all the bombings.

图 2 中展示了  $S_2$  和  $S_3$  对应的依存解析树, 并用红色实线标出了事件对 (search, bombings) 之间的最短依存路径。

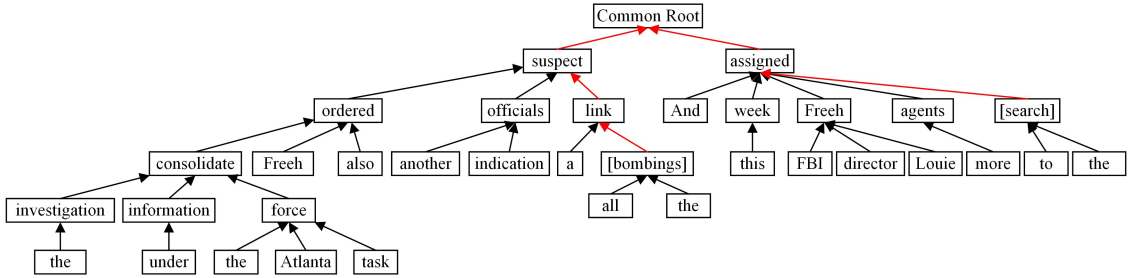


图 2 依存解析树及最短依存路径

Fig. 2 Dependency parse tree and the shortest dependency path

### 3.2 句子表示层

最短依存路径可以获取句子中的关键信息, 但被舍弃掉的部分有时可能对关系识别有重要作用。为了解决这个问题, 本文将句子的语义表示融合到输入中。

之前的工作通常采用 Bi-LSTM 和 CNN 等神经网络对序列进行编码, 但 Bi-LSTM 训练较为费时, CNN 则难以捕获全局信息。为此, 本文引入了 Dauphin 等<sup>[13]</sup>提出的门控卷积神经网络抽取文本的语义特征, 具体做法如下: 把含有  $m$  个单词的句子序列记作  $X = \{x_1, x_2, x_3, \dots, x_m\}$ ,  $x_i$  表示句子中第  $i$  个词在单词字典中的编号; 然后把句子序列中的每一个单词进行查表映射, 将其转化为低维实值向量; 再将句子序列输入门控卷积神经网络中, 学习句子序列的特征表示, 原理如式 (1) 所示。

$$Y = (X * W + b) \otimes \sigma(X * V + c) \quad (1)$$

其中,  $W$  和  $V$  代表两个不同的卷积核;  $b$  和  $c$  为偏置项;  $\otimes$  表示逐元素相乘;  $\sigma$  表示 sigmoid 函数, 带有该激活函数的卷积即为门控单元; GCNN 的输出记为  $Y$ 。

本文将  $Y$  作为句子表示加入到输入层, 以期为时序关系分类提供帮助。

### 3.3 输入层

在 Cheng 等<sup>[6]</sup>工作的基础上, 本文将融合了句子表示的

最短依存路径序列作为输入, 要判别时序关系的两个事件词分别被置于左右两个子路径序列中。对于左右两个子路径序列, 使用以下特征作为输入。

(1) 词法特征: 最短依存路径中的单词 (WORD)、单词词性 (POS)。

(2) 句法特征: 单词结点在依存解析树中与其父结点之间的依存关系 (DEP)、通过门控卷积神经网络训练得到的句子表示 (SR)。

输入的特征需要通过映射转化为低维实值向量。Cheng 等<sup>[6]</sup>使用 Word2Vec 来训练词向量, 但经过对比实验发现, 使用 Glove 预训练的词向量表现出了更好的性能。因此, 本文对单词本身通过使用 Glove 预训练词向量进行查表映射, 对词性和依存关系通过随机初始化的方法进行映射, 分别获得词向量  $V_w$ 、词性向量  $V_p$ 、依存关系向量  $V_d$ , 再将已训练好的句子表示向量  $V_s$  与这 3 个向量拼接起来, 形成每个单词的向量表示  $V_i$ :

$$V_i = V_w \oplus V_p \oplus V_d \oplus V_s \quad (2)$$

最终, 分别将左右两个分支序列中每个单词的向量表示堆叠起来, 组成左分支序列的矩阵表示  $R_1 = \{V_{11}, V_{12}, \dots, V_{1m}\}$  和右分支序列的矩阵表示  $R_2 = \{V_{21}, V_{22}, \dots, V_{2n}\}$ , 其中  $m$  和  $n$  分别表示左右分支序列的长度。

### 3.4 编码层

循环神经网络(Recurrent Neural Network, RNN)是一种十分适合处理序列化数据的神经网络模型,在自然语言处理领域得到了广泛的应用。原始 RNN 的隐藏层仅有一个状态向量  $h$ ,虽然对于短期输入非常敏感,但无法处理长距离依赖问题。而在时序关系分类任务中,对于一些跨句的事件对,长距离依赖信息显得尤为重要。改进之后的长短期记忆网络在传统 RNN 的基础上增加了一个新的单元状态  $c$  来保存长期信息,并加入遗忘门、输入门和输出门来控制长期状态  $c$ ,从而较好地捕获长距离依赖信息。

由于从目标词到根结点的路径与从根结点到目标词的路径可能包含不同的有效信息,本文使用双向长短期记忆网络在前向和逆向两个方向对输入序列进行建模,并将前向和逆向 LSTM 的输出结果拼接起来。

之前的研究工作通常使用两个独立的模型对左右两个子路径序列分别进行训练,这种做法限制了两个输入序列在训练过程中的信息交互。而在时序关系分类任务中,事件词所在句子之间的联系对于关系判别具有重要作用。针对这个问题,本文提出了一种参数共享机制,即使用同一个双向长短期记忆网络先后对左右两个子路径序列进行训练。在批训练(mini-batch)过程中,先将一个 batch 的左分支序列输入模型中进行训练,再将一个 batch 的右分支序列输入模型中进行训练,然后进行反向传播和权值更新。由于是双向神经网络模型,左右分支序列在训练时可以获取彼此的信息,并且共同对权值更新产生影响。

同时,输入时在每个单词的向量表示中都加入了句子的语义信息,一些在抽取最短依存路径时被舍弃的信息也被模型捕获到,从而对关系识别起到辅助作用。

### 3.5 时序关系识别层

时序关系识别主要包括两部分:全连接层和 Softmax 层。模型训练完毕之后,将两个子路径序列的输出拼接起来形成最后的输出  $h$ ,其通过全连接层后又作为输入送入 Softmax 层,最终生成模型预测的时序关系结果  $P$ 。公式表示为:

$$P = \text{Softmax}(W_0 * h + b_0) \quad (3)$$

其中,  $W_0$  和  $b_0$  是 Softmax 函数的权重矩阵和偏置量参数。

## 4 实验

### 4.1 语料库和实验设置

TimeBank-Dense 语料库是在 TimeBank 语料库的基础上构建而成的,旨在解决 TimeBank 语料库中存在的时序关系标注过于稀疏的问题。在 TimeBank-Dense 语料库中,对 36 篇文档进行了标注,共标注出 12715 条时序关系,比 TimeBank 语料库中的标注内容稠密得多,涵盖了“BEFORE”“AFTER”“SIMULTANEO-US”“INCLUDES”“IS\_INCLUDED”以及“VAGUE”6 种关系类型。其中,“VAG-UE”类型用来表示一些无法判别或模棱两可的时序关系。由于时序关系的判断相对比较困难,这一类别在 TimeBank-Dense 语料库中占据了较大比重。

由于 TimeBank-Dense 语料库中的文档数较少,且样本数量分布不均衡,本文将所有文档中的文本抽取出来,以实体为单位划分训练集和测试集。本文主要研究事件与事件之间

(E-E)的时序关系分类,因此从 TimeBank-Dense 语料库中将该类型的数据抽取出来,共计 6 088 个事件对。经统计,各关系类别的样本数如表 1 所列。

表 1 TimeBank-Dense 语料库分类别统计数据

Table 1 Statistics of TimeBank-Dense corpus by category

Relation	Count
AFTER	1 120
BEFORE	1 348
SIMULTANEOUS	93
INCLUDES	276
IS_INCLUDED	347
VAGUE	2 904

为了更加精准地评估模型的性能,缓解数据集较小带来的分布敏感性问题,本文采用了句子级五折交叉验证的方法。五折交叉验证是指将实验数据随机划分为 5 份相互独立且样本数量相当的子集,每次验证选择其中的 1 份作为测试集,剩下的数据作为训练集,最后取 5 次实验的平均性能作为该实验的性能指标。

为了与 Cheng 等<sup>[6]</sup>的实验性能进行对比,本文模型尽可能遵循其研究中的实验设置。在输入方面,对于单词本身,使用通过 Glove 方法预训练的词向量文件进行查表映射,预训练词向量的维度为 300。词性向量的维度设置为 50,依存表示(单词结点在依存解析树中与其父结点之间的依存关系)向量的维度设置为 50,句子表示向量的维度设置为 100。编码层中的双向长短期记忆网络单元的大小设置为 128,全连接层的维度设置为 200。训练过程中,本文采取了批训练和随机梯度下降优化算法(Stochastic Gradient Descent, SGD)训练模型,样本批次大小设置为 64,SGD 优化器中的动量因子设置为 0.9,学习率设置为 0.01。此外,为防止过拟合,本文在嵌入层、编码层和全连接层均引入了 Dropout 机制,Dropout 比率设置为 0.5。

为了与较优的基准系统进行比较,本文使用微平均 F1 值(Micro-F1)和宏平均 F1 值(Macro-F1)对模型性能进行评估。F1 值可通过准确率(P)及召回率(R)求得,具体计算方法如式(4)~式(6)所示。

$$P = \frac{TP}{TP + FP} \quad (4)$$

$$R = \frac{TP}{TP + FN} \quad (5)$$

$$F1 = \frac{2 * P * R}{P + R} \quad (6)$$

其中,  $TP$  表示正例(真实值为真的样本)中判断正确的样本数,  $FP$  表示负例(真实值为假的样本)中判断错误的样本数,  $FN$  表示正例中判断错误的样本数。

微平均 F1 值就是计算出所有类别总的准确率和召回率,然后计算出 F1 值。宏平均 F1 值则是分别计算出每一个类别的 F1 值,然后将它们取平均。在时序关系分类任务中,将微平均 F1 值作为主要评判指标,而将宏平均 F1 值作为辅助评判指标。

### 4.2 实验结果

为了验证本文提出的模型在事件时序关系分类上的性能,设置了 3 个性能优异的基准系统进行比较。

(1)MIRZA:由 Mirza 等<sup>[14]</sup>于 2016 年提出,将稀疏的传

特征与低维词嵌入结合起来作为分类器的输入,并在分类器中加入实体属性、时间标志词以及 WordNet 语义信息等特征;在传统方法中取得了较好的效果。

(2)CHENG:由 Cheng 等<sup>[6]</sup>于 2017 年提出,是最早将神经网络方法引入时序关系识别任务的方法之一;使用在实体关系分类任务中表现良好的最短依存路径作为输入,采用了单层 Bi-LSTM 网络进行编码;在没有使用任何显式特征和外部资源的情况下取得了优于传统方法的结果。

(3)ZHANG:由 Zhang 等<sup>[12]</sup>于 2018 年提出,堆叠了多层 Bi-LSTM 构建出一种深度双向长短期记忆网络(Deep Bidirectional Long Short-Term Memory network, DBiLSTMs),旨

在获取数据的深层语义表示,并使用平均池化层对深度 Bi-LSTM 生成的语义表示进行特征压缩,取得了良好的效果。

在 Cheng 等<sup>[6]</sup>工作的基础上,本文使用通过 Glove 方法预训练的词向量,实验性能得到了一定的优化。在编码部分,本文加入了参数共享机制,增强了两个子路径序列之间的信息交互,使模型捕获到它们之间的语义联系。最后,本文使用门控卷积神经网络训练出句子的向量表示,并将其融入到最短依存路径输入序列中,在微平均 F1 值上取得了最好的效果,具体的实验结果如表 2 所列。从实验结果中可以发现,无论是从整体性能还是单独时序关系类型上的性能来看,本文提出的方法都达到了目前最好的性能。

表 2 TimeBank-Dense 数据集上的实验性能

Table 2 Experimental performance on TimeBank-Dense dataset

	AFT.	BEF.	SIM.	INC.	IS_I.	VAG.	Macro-F1	Micro-F1
MIRZA(2016)	0.430	0.471	—	0.049	0.250	0.613	0.302	0.519
CHENG(2017)	0.440	0.460	—	0.025	0.170	0.624	0.287	0.529
ZHANG(2018)	0.526	0.503	—	0.106	0.325	0.626	0.348	0.548
Proposed	<b>0.601</b>	<b>0.573</b>	—	<b>0.205</b>	<b>0.345</b>	<b>0.676</b>	<b>0.400</b>	<b>0.613</b>

### 4.3 实验分析

为了验证本文所提方法的有效性,下面对影响实验结果的几个主要因素进行分析。

#### (1)信息交互增强机制

在神经网络方法中,已有研究工作通常使用两个独立的模型分别对左右两个子路径序列进行训练,这种做法限制了两个输入序列之间的信息交互。针对这个问题,本文提出了一种通过参数共享来增强两个子路径序列之间信息交互的方法。

本文将使用 Glove 预训练词向量的系统作为初始系统,并在其基础上增加信息交互增强机制,进行对比实验。表 3 列出了使用信息交互增强机制前后各时序关系类型上的 F1 值,以及 6 种时序关系类型上的 Macro-F1 值和 Micro-F1 值。

表 3 加入信息交互增强机制前后实验效果的对比

Table 3 Comparison of experimental results before and after adding information interaction enhancement mechanism

Relation	Without	Added
AFTER	0.538	0.590
BEFORE	0.507	0.544
SIMULTANEOUS	—	—
INCLUDES	0.146	0.210
IS_INCLUDED	0.379	0.394
VAGUE	0.652	0.669
Macro-F1	0.362	0.401
Micro-F1	0.574	0.604

实验效果表明,相对于未使用信息交互增强机制的初始系统,在相同的实验设置下,使用信息交互增强机制的系统在 Macro-F1 值和 Micro-F1 值上分别取得了 3.9% 和 3% 的提升,证明了该方法是有效果的。

以句子  $S_4$  为例:事件词“said”与“forecast”之间的时序关系为“AFTER”,最短依存路径的左分支序列为⟨said⟩,右分支序列为⟨forecast, far, added, said⟩。使用信息交互增强机制的模型捕获到了事件词“said”即为事件词“forecast”所在分支路径的根结点这一信息,轻易地将事件对间的时序关系预测为“AFTER”,而未使用信息交互增强机制的模型则错误地将其预测为“VA-GUE”。

$S_4$ : The Bureau of Labor Statistics **said** the economy ad-

ded 358000 jobs last month, far above the 235000 forecast by economists.

通过分析各关系类型的样本数量发现,“V-AGUE”类型样本占了总样本数的一半,使得模型会倾向于将样本预测为“VAGUE”类型,从而对关系预测的精度带来影响,而上面所举的例子证明了信息交互增强机制可以较好地解决这个问题。

#### (2)句子表示

由于仅通过事件词本身很难准确地判别出时序关系,句法信息对于时序关系分类十分重要。但是,构建最短依存路径时需要丢弃句中的大部分单词,而这些被丢弃的单词可能对时序关系分类起到关键作用。以  $S_2$  和  $S_3$  中的事件对为例,最短依存路径的左分支序列为⟨search, assigned⟩,右分支序列为⟨bombings, link, suspect⟩,根据上述信息很难辨别出事件对之间的时序关系。将句子表示融入到输入后,可以获取两个句子中的行为主体都是“Freeh”;再通过“also”一词,就可以轻易捕获到  $S_2$  和  $S_3$  中的几个动词(“assigned”“ordered”“suspect”)是在同一时间段发生的,“assigned”与“search”也在同一时间段,而“bombings”则显然发生在“suspect”之前,从而帮助模型推断出该事件对之间的时序关系为“AFTER”。因此,本文在加入信息交互增强机制的系统上进行了改进,在输入的最短依存路径序列中融入了句子表示信息。

为了证明 GCNN 的有效性,本文使用多种神经网络模型进行了对比实验,如表 4 所列。

表 4 加入句子表示信息前后实验效果的对比

Table 4 Comparison of experimental results before and after adding sentence representation information

Relation	Without	With Bilstm	With Bigru	With GCNN
AFTER	0.590	0.595	0.602	0.601
BEFORE	0.544	0.565	0.572	0.573
SIMULTANEOUS	—	—	—	—
INCLUDES	0.210	0.191	0.169	0.205
IS_INCLUDED	0.394	0.395	0.401	0.345
VAGUE	0.669	0.661	0.662	0.676
Macro-F1	0.401	0.401	0.402	0.400
Micro-F1	0.604	0.602	0.606	0.613

可以看出,使用 GCNN 训练出的句向量在微平均 F1 值上取得了最好的效果。另外,该方法在样本数据量较大的时序关系类型以及微平均 F1 值上取得了一定的提升,证明了额外提供的句法信息对于提升模型预测的预测精度是有帮助的。而在一些数据样本较少的数据类型上,该方法则没有表现出很好的效果,这是由于测试集中这些类型的样本较少,几个样本的误差就会对实验结果造成很大的影响。

另外,表 5 列出了本文所提方法在实验中的混淆矩阵,矩阵中的行代表真实值,列代表预测值。

表 5 在 TimeBank-Dense 语料库上的混淆矩阵

Table 5 Confusion matrix on TimeBank-Dense corpus

	AFT.	BEF.	SIM.	INC.	IS_I.	VAG.
AFT.	613	45	1	6	9	400
BEF.	50	694	0	4	14	546
SIM.	18	12	0	1	0	56
INC.	9	23	0	36	1	206
IS_I.	13	25	0	0	93	202
VAG.	247	310	1	22	58	2170

从表中数据可以看出,虽然信息交互增强机制能在一定程度上降低数据分布不均衡带来的影响,但是“VAGUE”类型的数据仍然是影响模型判断的主要因素,这种影响在数据样本较少的关系类型上尤为明显。

**结束语** 本文提出了一种基于信息交互增强的事件时序关系分类方法。首先,使用门控卷积神经网络提取每个句子的语义信息以生成句子表示。然后,抽取事件词之间的最短依存路径,将其划分成两个子路径序列作为模型的输入,并将句子表示融合到最短依存路径序列内每个单词的向量表示中。接着,使用双向长短期记忆网络对输入信息进行编码以获取其语义表示,并加入参数共享机制来增强输入序列间的信息交互。最终,将编码层获得的语义表示输入全连接层,使用 Softmax 函数对事件对间的时序关系进行分类预测。在 TimeBank-Dense 语料库上的实验结果表明,本文方法取得的效果达到了先进水平。

时序关系分类对理解与推理能力有一定的要求,很多样本即使是人类也很难辨别,目前事件时序关系分类任务的总体性能仍处于较低水平。在今后的工作中,主要考虑对篇章级的时序关系分类任务进行研究。

## 参 考 文 献

- [1] MANI I, VERHAGEN M, WELLNER B, et al. Machine learning of temporal relations[C]// Proceedings of the Association for Computational Linguistics. Association for Computational Linguistics, 2006: 753-760.
- [2] CHAMBERS N, WANG S, JURAFSKY D. Classifying temporal relations between events[C]// Proceeding of the ACL on Interactive Poster and Demonstration Sessions. Association for Computational Linguistics, 2007: 173-176.
- [3] CHAMBERS N, JURAFSKY D. Jointly combining implicit constraints improves temporal ordering[C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2008: 698-706.
- [4] DO Q, LU W, ROTH D. Joint inference for event timeline construction[C]// Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computa-

tional Natural Language Learning. Association for Computational Linguistics, 2012: 677-687.

- [5] D'SOUZA J, NG V. Classifying temporal relations with rich linguistic knowledge[C]// Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics. Human Language Technologies, 2013: 918-927.
- [6] CHENG F, MIYAO Y. Classifying temporal relations by bidirectional LSTM over dependency paths[C]// Proceedings of the Association for Computational Linguistics. Association for Computational Linguistics, 2017: 1-6.
- [7] YAN X, MOU L, LI G, et al. Classifying relations via long short term memory networks along shortest dependency path[J]. Computer Science, 2015, 42(1): 56-61.
- [8] CHOUBEY P K, HUANG R H. A sequential model for classifying temporal relations between intra-sentence events[C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2017: 1796-1802.
- [9] YAO W, HUANG R. Temporal eventknowledge acquisition via identifying narratives[C]// Proceedings of the Association for Computational Linguistics. Association for Computational Linguistics, 2018: 537-547.
- [10] TOURILLE J, FERRET O, TANNIER X, et al. Neural architecture for temporal relation extraction: A Bi-LSTM approach for detecting narrative containers[C]// Proceedings of the Association for Computational Linguistics. Association for Computational Linguistics, 2017: 224-230.
- [11] MENG Y L, RUMSHISKY A. Context-Aware neural model for temporal information extraction[C]// Proceedings of the Association for Computational Linguistics. Association for Computational Linguistics, 2018: 527-536.
- [12] ZHANG Y J, LI P F, ZHOU G D. Classifying temporal relations between events by deep BiLSTM[C]// Proceedings of International Conference on Asian Language Processing. Institute of Electrical and Electronics Engineers, 2018: 267-272.
- [13] DAUPHIN Y N, FAN A, AULI M, et al. Language modeling with gated convolutional networks[C]// Proceedings of the International Conference on Machine Learning. International Machine Learning Society, 2017: 933-941.
- [14] MIRZA P, TONELLI S. On the contribution of word embeddings to temporal relation classification[C]// Proceedings of the International Conference on Computational Linguistics. 2016: 2818-2828.



**ZHOU Xin-yu**, born in 1996, postgraduate, is a member of China Computer Federation. His main research interests include natural language processing.



**LI Pei-feng**, born in 1971, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include natural language processing and machine learning.