



面向事件时序与因果关系的联合识别方法

张义杰^{1,2}, 李培峰^{1,2}, 朱巧明^{1,2}

(1. 苏州大学 计算机科学与技术学院, 江苏 苏州 215006; 2. 江苏省计算机信息处理技术重点实验室, 江苏 苏州 215006)

摘 要: 从事件时序关系与因果关系的关联性出发, 提出基于神经网络的联合识别方法。将时序关系和因果关系识别分别作为主任务和辅助任务, 设计共享辅助任务中编码层、解码层和编解码层的 3 种联合识别模型, 通过主任务模型和辅助任务模型中的网络层进行信息共享, 学习联合识别模型之间的特征信息。实验结果表明, 联合识别方法能利用事件之间的因果信息有效提升时序关系的识别性能, 且共享辅助任务中编解码层的联合识别模型更适用于事件时序关系与因果关系的联合识别。

关键词: 事件; 时序关系; 因果关系; 神经网络; 联合识别

开放科学(资源服务)标志码(OSID):



中文引用格式: 张义杰, 李培峰, 朱巧明. 面向事件时序与因果关系的联合识别方法[J]. 计算机工程, 2020, 46(7): 65-71.

英文引用格式: ZHANG Yijie, LI Peifeng, ZHU Qiaoming. Joint identification method for temporal and causal relations of events[J]. Computer Engineering, 2020, 46(7): 65-71.

Joint Identification Method for Temporal and Causal Relations of Events

ZHANG Yijie^{1,2}, LI Peifeng^{1,2}, ZHU Qiaoming^{1,2}

(1. School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China;

2. Jiangsu Provincial Key Lab of Computer Information Processing Technology, Suzhou, Jiangsu 215006, China)

[Abstract] Based on the correlation between temporal relations and causal relations of events, this paper proposes a joint identification method using neural network. The method takes the identification of temporal relations as the main task, and that of causal relations as auxiliary task. On this basis, three types of joint identification models of sharing encoding layer, decoding layer and encoding-decoding layer in auxiliary tasks are designed to enable information sharing through the network layer of the main task model and the auxiliary task model. Then feature information of joint identification models is learnt. Experimental results show that the joint identification method can use the causal information between events to significantly improve the identification performance of temporal relations. Also, the joint identification model of sharing encoding-decoding layer in auxiliary tasks is more suitable for the joint identification of temporal and causal relations of events.

[Key words] event; temporal relation; causal relation; neural network; joint identification

DOI: 10.19678/j.issn.1000-3428.0054800

0 概述

事件是一种描述特定人、物、事在特定时间和地点相互作用的客观事实, 文本信息中的事件则是将这一客观事实诉诸文字的独特语言形式。事件研究的核心为识别不同事件之间的关系^[1], 其广泛应用于自动问答、文本摘要和时间线构建等自然语言处理(Natural Language Processing, NLP)任务中。在事件之间可能存在多种关系类型, 早期对于事件

关系的研究多数是针对单一的关系类型, 且许多提出方法是用于提升对应关系的识别性能, 然而 NLP 领域对于事件关系进行联合识别的研究却相对较少。

事件的时序关系表示事件发生的先后顺序, 描述某话题下事件的发生及演变过程, 而因果关系描述原因事件与结果事件之间的因果关联。由于时序关系和因果关系之间存在密切的逻辑关联, 因此在多数情况下其中的一种关系通常受另一种关系的影

基金项目: 国家自然科学基金(61472265, 61772354, 61773276)。

作者简介: 张义杰(1994—), 男, 硕士研究生, 主研方向为自然语言处理; 李培峰(通信作者)、朱巧明, 教授、博士生导师。

收稿日期: 2019-05-05 修回日期: 2019-08-07 E-mail: scuyjzh@qq.com

响,并且一般情况下存在因果关系的事件对在发生顺序上存在先后或重叠关系。本文提出基于神经网络的联合识别方法,并设计共享辅助任务中编码层、解码层和编解码层的联合识别模型。

1 相关工作

在时序关系识别方面,早期的研究工作专注于提取文本中的单一特征。文献[2-3]在标注语料的基础上利用事件属性构造特征向量,包括事件类型、体态、形态、极性、时态等,并使用最大熵分类器进行时序关系识别。文献[4]在此基础上进一步结合词性、句法树结构等语义特征,从 WordNet 中提取词汇和形态学特征,使得特征空间得到极大扩展。文献[5]利用篇章级的特征以及 VerbOcean 中的语义关系,进一步提升事件时序关系的识别性能。文献[6]在英文研究工作的基础上,针对中文语句的特点,提出多种适用于中文语料的新型特征,识别中文事件时序关系。随着深度学习技术的发展,神经网络方法被引入到时序关系识别任务中。文献[7]采取最短依存路径作为输入,构造一种基于 RNN 的神经网络模型,其在不使用任何显性特征和外部知识库的情况下识别效果显著。文献[8]提出基于 LSTM 的网络结构,利用句法依赖关系识别文本中的时序关系。文献[9]通过双通道 LSTM 学习两个事件词上下文的句法和语义表示,用于识别相同句子间的事件时序关系。

在因果关系识别方面,早期的研究工作主要使用模板匹配方法。文献[10]通过从句法分析的结果中提取具有因果关系的词及短语结构来构建因果关系模板,从而识别医疗文本中的因果关系。文献[11-12]在语料库中同时标注时序和因果关系,将人工标注的时序关系类型作为因果关系分类器的补充特征。文献[13]将篇章关系预测和分布相似性方法归纳在一个全局推理过程中,对识别事件因果关系具有重要作用。文献[14]提出一个新的语义标注框架 CaTeRS,对时序关系和因果关系进行联合识别。文献[15]通过使用因果关系的预测结果对时序关系进行分析,研究两种事件关系之间的相互作用。文献[16]利用时序和因果关系之间存在的约束条件和语言学规则,将联合识别任务转化为一个整数线性规划问题。

2 基于神经网络的联合识别方法

2.1 最短依存路径

基于依存路径的神经网络模型在时序关系识别任务中的性能表现较好^[7,9]。本文基于该思想,以事件句的依存解析树为基础,生成两个最短依存路径分支序列,其中左分支形成从源事件词结点到最近

公共祖先结点的依存路径,而右分支形成从目标事件词结点到最近公共祖先结点的依存路径。为简化网络结构,本文将两个最短依存路径分支进行拼接,形成一个完整的依存路径序列作为模型输入。对于句内事件对,直接将依存解析树中源事件词结点到目标事件词结点的依存路径作为序列输入;对于句间事件对,本文假设对应的两颗依存解析树共享一个“公共根结点”,并以该结点为中心,将左右两条依存路径拼接为完整的依存路径序列。

对于依存路径序列中的每一个单词,首先将词(W)、词性(P)和该词结点在依存解析树中与其父结点的依存关系(D)映射到低维实值向量空间中,其中词通过使用 Word2Vec^[17]预训练词向量的方式进行查表映射,词性与依存关系通过随机初始化的方法进行映射,然后分别获得词向量 x_w 、词性向量 x_p 和依存向量 x_d ,并将其进行拼接形成每个单词的向量表示 x_i :

$$x_i = x_w \oplus x_p \oplus x_d \quad (1)$$

该序列的矩阵化表示为 $X = \{x_1, x_2, \dots, x_s\}$,其中 s 表示序列长度。

2.2 神经网络编解码层

目前,主流的事件关系识别方法主要为基于神经网络的识别方法。此类方法通常包含两个组成部分,即编码(Encoding)层和解码(Decoding)层。

2.2.1 编码层

在形式上,使用 X 表示输入序列,该序列通过编码层进行传播得到中间表示 h ,即:

$$h = \text{Encoding}(X) \quad (2)$$

近年来,多种神经网络被作为事件关系识别模型的编码层。本文采用两个最具代表性的编码层,具体为:

1) Bi-LSTM 编码层

文献[7]使用双通道 LSTM(Bi-LSTM)对依存路径输入进行编码,在时序关系识别任务中效果显著。Bi-LSTM 用于解决传统 RNN 中长期依赖的学习问题,是事件关系识别任务中最常见的神经网络编码层。由于其具有较好的文本建模能力,在自然语言处理领域得到广泛应用,因此本文使用 Bi-LSTM 作为编码层。

2) Self-Att 编码层

Self-Att 编码层采用一种融合自注意力(Self-Attention)机制的神经网络模型。Google 于 2017 年提出基于 Self-Attention 机制的 Transformer 模型^[18],在机器翻译任务中效果显著。Self-Attention 机制直接捕获输入序列的全局依赖关系,从而加强语义信息的编码能力。与传统 RNN 相比,该机制的主要优势在于其能够在句子中任意两个词例之间建立直接的语义关联。因此,句子中相离的元素之间能够在更短的路径上产生相互作用,使信息在网络

中实现畅通无阻的传输。并且, Self-Attention 机制提供了更加灵活的方式对序列中的信息进行选择、表示以及合成,并且可作为非线性网络模型的补充形式。由于 Self-Attention 机制使用加权求和的方法计算输出向量,对特征的表示能力相对有限,因此为进一步提高其表征能力,本文使用双通道 GRU 网络子层^[19]对序列输入进行初步编码。

2.2.2 解码层

解码层将经由编码层得到的中间表示 h 转换成一个隐层表示 r , 即:

$$r = \text{Decoding}(h) \quad (3)$$

在基于序列到序列学习的自然语言处理任务中,如神经机器翻译,解码层通常将源序列经由编码层进行非线性转换得到的作为中间表示的上下文向量再转化为相应的目标序列。与这些成分较为复杂的解码层不同,事件关系识别中的解码层结构相对简单,因为模型的最终输出只是基于所有标签的概率得分。

本文将一层全连接层 (Fully Connected Layer) 作为解码层,其中使用 ReLU 函数^[20]作为激活函数,此解码层的目的是将从编码层学习到的分布式特征表示映射到样本标签空间,计算公式为:

$$r = \text{ReLU}(Wh + b) \quad (4)$$

其中, W 表示权重参数矩阵, b 为偏置量。

2.3 联合识别模型

联合识别模型将时序关系识别作为主任务 (Main Task), 因果关系识别作为辅助任务 (Auxiliary Task)。

2.3.1 主任务

联合识别模型通过从辅助任务中学习特征表示来促进并提升主任务的识别性能。

1) 共享辅助任务中编码层的联合识别模型。图 1 描述了通过共享辅助任务中的编码层来实现时序和因果关系联合识别的模型架构。

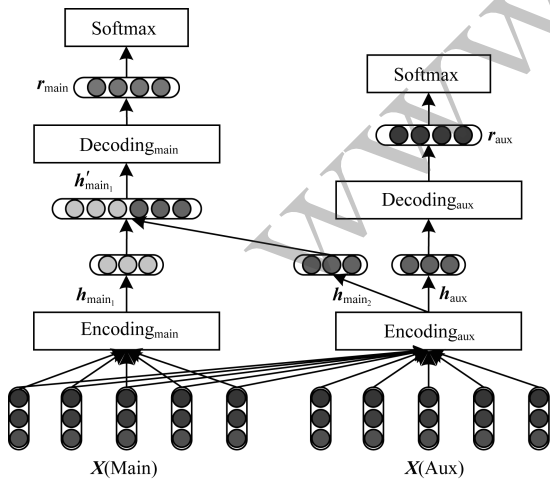


图 1 共享辅助任务中编码层的联合识别模型

Fig. 1 Joint recognition model of sharing encoding layer in auxiliary task

在一般情况下,主任务的编码层和辅助任务的编码层对生成主任务中间表示的贡献相等,即:

$$h_{\text{main}_1} = \text{Encoding}_{\text{main}}(X) \quad (5)$$

$$h_{\text{main}_2} = \text{Encoding}_{\text{aux}}(X) \quad (6)$$

通过将两个任务的中间表示向量 h_{main_1} 和 h_{main_2} 进行拼接来更新主任务的中间表示 h'_{main} , 并得到解码后的隐层表示, 即:

$$h'_{\text{main}} = h_{\text{main}_1} \oplus h_{\text{main}_2} \quad (7)$$

$$r_{\text{main}} = \text{Decoding}_{\text{main}}(h'_{\text{main}}) \quad (8)$$

2) 共享辅助任务中解码层的联合识别模型。图 2 描述了通过共享辅助任务中的解码层来实现时序和因果关系联合识别的模型架构。

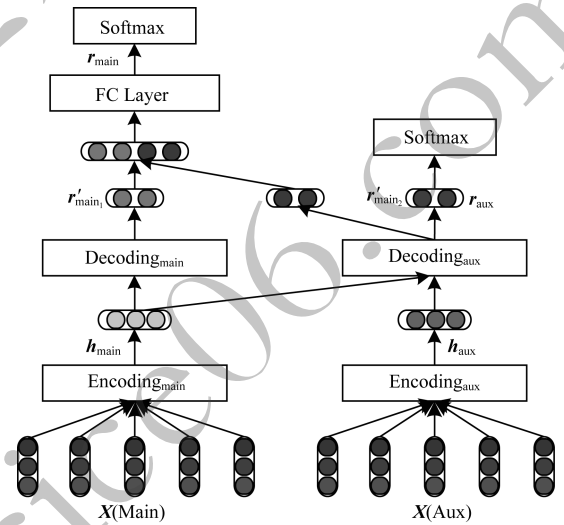


图 2 共享辅助任务中解码层的联合识别模型

Fig. 2 Joint recognition model of sharing decoding layer in auxiliary task

同样地,主任务的解码层和辅助任务的解码层对生成主任务隐层表示的贡献相等,即:

$$r'_{\text{main}_1} = \text{Decoding}_{\text{main}}(h_{\text{main}}) \quad (9)$$

$$r'_{\text{main}_2} = \text{Decoding}_{\text{aux}}(h_{\text{main}}) \quad (10)$$

将主任务的两个输出 r'_{main_1} 和 r'_{main_2} 进行拼接,并通过一个非线性全连接层重新学习到一个新的结果 r_{main} , 即:

$$r_{\text{main}} = \text{ReLU}(W'[r'_{\text{main}_1}, r'_{\text{main}_2}] + b') \quad (11)$$

其中, W' 表示权重参数矩阵, b' 表示偏置量。

3) 共享辅助任务中编解码层的联合识别模型。图 3 描述了通过同时共享辅助任务中的编码层和解码层来实现时序和因果关系联合识别的模型架构。

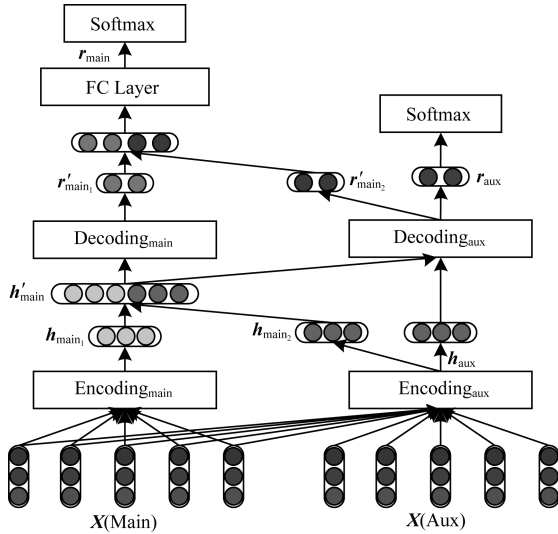


图3 共享辅助任务中编解码层的联合识别模型

Fig.3 Joint recognition model of sharing encoding-decoding layer in auxiliary task

模型3是模型1和模型2的融合。在形式上,与模型2不同的是,辅助任务的输出学习自中间表示的拼接结果,即:

$$r'_{main_2} = \text{Decoding}_{aux}(h'_{main_1}) \quad (12)$$

其中, h'_{main_1} 表示模型1的中间表示向量的拼接结果。

2.3.2 辅助任务

辅助任务模型采用标准的神经网络方法进行因果关系识别,其包含一个编码层和一个解码层,即:

$$h_{aux} = \text{Encoding}_{aux}(X) \quad (13)$$

$$r_{aux} = \text{Decoding}_{aux}(h_{aux}) \quad (14)$$

其中, h_{aux} 表示辅助任务的中间表示向量, r_{aux} 表示解码层输出的隐层结果。主任务可共享辅助任务的编码层和解码层。

2.4 关系分类

主任务和辅助任务分别采用一个 Softmax 层作为模型的输出层来获得其关系标签的预测结果:

$$o_{main} = \text{Softmax}(W_{o_{main}} r_{main} + b_{o_{main}}) \quad (15)$$

$$o_{aux} = \text{Softmax}(W_{o_{aux}} r_{aux} + b_{o_{aux}}) \quad (16)$$

其中, r_{main} 和 r_{aux} 表示主任务和辅助任务中解码层的输出向量, $W_{o_{main}}$ 和 $W_{o_{aux}}$ 表示对应的权重参数矩阵, $b_{o_{main}}$ 和 $b_{o_{aux}}$ 表示偏置量参数。

3 联合识别模型训练

联合识别模型通过最小化负对数似然函数进行训练。为学习模型参数,本文将两个关系识别任务的损失函数进行线性组合得到一个新的损失函数,具体为:

$$\begin{aligned} \tau_{\text{joint}} = & -\lambda \cdot \sum_{i=1}^T \log_a p(y_{\text{main}}^i | x_{\text{main}}^i) - \\ & (1-\lambda) \cdot \sum_{i=1}^T \log_a p(y_{\text{aux}}^i | x_{\text{aux}}^i) + \frac{l}{2} \|\theta\|_2^2 \end{aligned} \quad (17)$$

其中, (x^i, y^i) 表示训练样本中第 i 个样本对应的关系类型标签, T 表示训练样本个数, θ 表示模型训练参数集合, λ 表示保持主任务和辅助任务之间损失平衡的权重参数,在本文实验中设置为 0.6。

在训练过程中,本文采用 L2 正则化方法对网络参数进行约束,并使用 Adam 方法对模型进行优化。所有参数均以均匀分布的方式进行初始化。

4 实验结果与分析

4.1 数据集设置

RED 数据集由 95 篇涵盖新闻专线、论坛日志和叙述性文本等体裁的文档组成。RED 标注模式不仅需要理解不同事件的相对顺序,而且需要标注一个事件导致 (CAUSE) 或预设 (PRECONDITION) 另一个事件的关系。标注人员通过对之前 (BEFORE) 事件链和重叠 (OVERLAP) 事件链进行子类型标注来构造结果 (EFFECT) 事件链。CAUSE 和 PRECONDITION 本质上是存在时序链的: 先决条件必须发生在它所作用的事件之前。CAUSE 或 PRECONDITION 将先于 EFFECT 发生,但它们与结果最终可能重叠,比如“跑步”和“流汗”事件。然而,CAUSE 关系表明原因先于结果发生。因此,对于 CAUSE 和 PRECONDITION 总存在时序关系。根据两个事件之间的时序和因果关系,对因果关联事件使用 4 种标注类型: BEFORE/CAUSES, OVERLAP/CAUSES, BEFORE/PRECONDITIONS 和 OVERLAP/PRECONDITIONS。

由于 CAUSE 和 PRECONDITION 都属于因果关系范畴,本文将其融合统称为 CAUSE。表 1 展示了 RED 数据集的事件关系类型。

表1 RED 数据集事件关系类型

Table 1 Types of event relation in RED dataset		
关系	关系类型	样例数
时序	BEFORE	459
	CONTAINS	191
	OVERLAP	354
	BEGINS-ON	282
	ENDS-ON	135
	SIMULTANEOUS	34
因果	BEFORE/CAUSE	596
	OVERLAP/CAUSE	432

本文实验采用句子级的五折交叉验证方式对实验结果进行评估,并且以 15% 的比例从训练集中随机抽样得到验证集。

4.2 超参数设置

将模型输入中事件词、词性、依存关系的词嵌入向量维度分别设为 300、50、50。模型中隐藏单元数量设为 200。在输入层和解码层后采用 Dropout 策略防止模型过拟合,其保留率设为 0.5;并在训练时使用早停法(Early Stopping)保存验证集上的最优模型,置信度设为 10,具体超参数设置如表 2 所示。

表 2 超参数设置
Table 2 Setting of hyperparameters

参数	参数取值
LSTM 单元大小	100
GRU 单元大小	200
Adam 学习率	10^{-3}
Batch 大小	32
Epoch 数	50
L2 惩罚权重系数	10^{-5}
Dropout 保留率	0.5
早停法置信度	10

表 3 在 RED 数据集上的时序关系单任务识别 F1 值对比

Table 3 Comparison of F1 value of single task recognition of temporal relation on RED dataset

单任务识别方法	BEFORE	CONTAINS	OVERLAP	BEGINS-ON	ENDS-ON	SIMULTANEOUS	Overall
基于 LR 的单任务识别方法	56.5	42.5	30.6	18.1	12.3	—	48.9
基于 Bi-LSTM 的单任务识别方法	59.9	45.9	35.5	24.0	17.6	—	52.7
基于 Self-Att 的单任务识别方法	61.8	47.2	36.8	24.2	19.2	—	53.8

4.3.2 时序关系和因果关系联合识别实验结果分析

本文通过共享 Bi-LSTM 编码层和 Self-Att 编码层实现联合识别,其识别方法具体如下:

1) 基于 Encoding Sharing 的联合识别方法:通过共享辅助任务(因果关系识别)的编码层来实现时序关系和因果关系的联合识别,Encoding Sharing 的具体模型架构如图 1 所示。

2) 基于 Decoding Sharing 的联合识别方法:通过共享辅助任务(因果关系识别)的解码层来实现时序关系和因果关系的联合识别,Decoding Sharing 的具体模型架构如图 2 所示。

3) 基于 BOTH Sharing 的联合识别方法:通过对辅助任务(因果关系识别)的编码层与解码层同时进行共享来实现联合识别,BOTH Sharing 的具体模型架构如图 3 所示。

另外,本文构建一个 Baseline 基准方法,表示在时序关系和因果关系上的单任务识别方法。表 4 和表 5 分别给出了使用 Bi-LSTM 和 Self-Att

4.3 结果分析

4.3.1 时序关系单任务识别实验结果分析

本文通过 3 种方法进行时序关系单任务识别实验:

1) 基于 LR 的时序关系单任务识别方法。LR 是一种基于特征工程的统计机器学习方法。本文使用带 L2 正则化的 Logistic 回归模型构建分类器,其中使用的特征来自文献[21]。

2) 基于 Bi-LSTM 的时序关系单任务识别方法。Bi-LSTM 是一种对依存路径进行编码的方法,其采用 Bi-LSTM 网络层作为编码层,非线性全连接层作为解码层。

3) 基于 Self-Att 的时序关系单任务识别方法。Self-Att 是一种融合 Self-Attention 机制的神经网络方法,其编码层由 GRU 网络子层和自注意力网络层两部分组成,采用非线性全连接层作为解码层。

表 3 给出了上述 3 种方法在 RED 数据集上的时序关系单任务识别性能对比结果,其中,加粗数据表示最优结果,“—”表示 SIMULTANEOUS 类型的样例数过于稀少,因此导致上述 3 种方法都无法对其进行识别。可以看出,基于 Bi-LSTM 和 Self-Att 的时序关系单任务识别方法通常优于基于 LR 的时序关系单任务识别方法。因此,下文实验中使用 Bi-LSTM 编码层和 Self-Att 编码层实现联合识别。

作为模型编码层的事件时序关系和因果关系联合识别结果。可以看出,无论使用 Bi-LSTM 还是 Self-Att 作为编码层,识别效果都优于 Baseline 方法,该结果验证了本文提出的联合识别方法的有效性,即对于事件时序关系识别任务,可以利用因果关系识别模型学习到有效的特征,从而采用其因果信息促进时序关系识别性能的提升。特别地,在 3 种联合识别方法中,基于 BOTH Sharing 的联合识别方法在具体关系类别及整体上几乎都取得了最优的实验结果,这表明对辅助任务的编码层和解码层同时进行共享的方式最适用于事件时序关系和因果关系联合识别任务。其主要原因为当联合学习模型对辅助任务的编码层和解码层同时进行信息共享时,既充分利用了事件之间的因果信息来促进时序关系的识别,且扩充了主任务模型中编码层输出的特征向量空间,又在解码阶段使模型在进行特征压缩时最大限度地保留了网络中的有用信息。

表 4 在 RED 数据集上的联合识别结果 (Bi-LSTM 编码层)
Table 4 Joint recognition results on RED dataset (Bi-LSTM encoding layer)

联合识别方法	时序关系							因果关系
	BEFORE	CONTAINS	OVERLAP	BEGINS-ON	ENDS-ON	SIMULTANEOUS	Overall	CAUSE
基于 Baseline 的联合识别方法	59.9	45.9	35.5	24.0	17.6	—	52.7	64.2
基于 Encoding Sharing 的联合方法	61.4	46.7	36.8	24.8	19.6	—	54.3	65.1
基于 Decoding Sharing 的联合方法	60.8	46.2	35.7	24.6	18.2	—	52.9	64.4
基于 BOTH Sharing 的联合方法	62.5	48.3	36.2	25.2	21.4	—	55.2	65.7

表 5 在 RED 数据集上的联合识别结果 (Self-Att 编码层)
Table 5 Joint recognition results on RED dataset (Self-Att encoding layer)

联合识别方法	时序关系							因果关系
	BEFORE	CONTAINS	OVERLAP	BEGINS-ON	ENDS-ON	SIMULTANEOUS	Overall	CAUSE
基于 Baseline 的联合识别方法	61.8	47.2	36.8	24.2	19.2	—	53.8	66.4
基于 Encoding Sharing 的联合方法	62.9	48.3	35.5	25.0	20.6	—	55.7	67.2
基于 Decoding Sharing 的联合方法	62.1	48.0	36.9	24.4	19.5	—	54.2	66.8
基于 BOTH Sharing 的联合方法	63.4	48.8	37.2	25.9	21.6	—	56.8	68.0

联合识别方法与时序关系单任务识别方法相比,其因果关系识别性能的提升幅度相对较小,主要因为本文提出的联合识别方法主要是通过共享辅助任务(因果关系识别)网络中的信息来更新主任务(时序关系识别)网络中的特征向量表示,而没有将主任务中的信息向辅助任务进行反馈。可见,该结果与本文利用事件之间的因果信息来促进时序关系识别的动机相一致,验证了本文联合识别方法的正确性与有效性。

5 结束语

本文考虑到事件时序关系与因果关系之间的联系,提出 3 种联合识别模型,通过共享因果关系识别任务中的特征信息提升时序关系的识别性能。实验结果表明,联合识别方法的性能要优于时序关系单任务识别方法,并且共享辅助任务中编解码层的联合识别效果最佳。下一步将研究联合识别方法在事件时序关系和因果关系识别中的应用,并尝试通过其他编码层信息共享方式对主任务中的信息向辅助任务进行有效反馈,提升因果关系的识别性能。

参考文献

- [1] ZHAO Yanyan, QIN Bing, CHE Wanxiang, et al. Research on Chinese event extraction [J]. Journal of Chinese Information Processing, 2008, 22 (1): 3-8. (in Chinese)
赵妍妍, 秦兵, 车万翔, 等. 中文事件抽取技术研究[J]. 中文信息学报, 2008, 22 (1): 3-8.
- [2] MANI I, VERHAGEN M, WELLNER B, et al. Machine learning of temporal relations [C]//Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the ACL. Philadelphia, USA: Association for Computational Linguistics, 2006: 753-760.
- [3] MANI I, WELLNER B, VERHAGEN M, et al. Three approaches to learning TLINKs in TimeML; CS-07-268 [R]. Boston, USA: Brandeis University, 2007.
- [4] CHAMBERS N, WANG S, JURAFSKY D. Classifying temporal relations between events [C]//Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions. New York, USA: ACM Press, 2007: 173-176.
- [5] D'SOUZA J, NG V. Classifying temporal relations with rich linguistic knowledge [C]//Proceedings of NAACL-HLT'13. Philadelphia, USA: Association for Computational Linguistics, 2013: 918-927.
- [6] ZHENG Xin, LI Peifeng, ZHU Qiaoming. Global inference for temporal relations between Chinese events [J]. Journal of Chinese Information Processing, 2016, 30 (5): 129-135. (in Chinese)
郑新, 李培峰, 朱巧明. 基于全局优化的中文事件时序关系推理方法 [J]. 中文信息学报, 2016, 30 (5): 129-135.
- [7] CHENG F, MIYAO Y. Classifying temporal relations by bidirectional LSTM over dependency paths [C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Philadelphia, USA: Association for Computational Linguistics, 2017: 1-6.
- [8] MENG Y, RUMSHISKY A, ROMANOV A. Temporal information extraction for question answering using syntactic dependencies in an LSTM-based architecture [C]//Proceedings of 2017 Conference on Empirical Methods in Natural Language Processing. Philadelphia, USA: Association for Computational Linguistics, 2017: 887-896.

- [9] CHOUBEY P K, HUANG R. A sequential model for classifying temporal relations between intra-sentence events [C]//Proceedings of 2017 Conference on Empirical Methods in Natural Language Processing. Philadelphia, USA: Association for Computational Linguistics, 2017:1796-1802.
- [10] KHOO C S G, CHAN S, NIU Y. Extracting causal knowledge from a medical database using graphical patterns [C]//Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics. Philadelphia, USA: Association for Computational Linguistics, 2002:336-343.
- [11] BETHARD S, CORVEY W, KLINGENSTEIN S, et al. Building a corpus of temporal-causal structure [C]//Proceedings of the 6th International Conference on Language Resources and Evaluation. Marrakech, Morocco: European Language Resources Association, 2008:908-915.
- [12] BETHARD S, MARTIN J H. Learning semantic links from a corpus of parallel temporal and causal relations [C]//Proceedings of ACL '08. Philadelphia, USA: Association for Computational Linguistics, 2008:177-180.
- [13] DO Q X, CHAN Y S, ROTH D. Minimally supervised event causality identification [C]//Proceedings of 2011 Conference on Empirical Methods in Natural Language Processing. Philadelphia, USA: Association for Computational Linguistics, 2011:294-303.
- [14] MOSTAFAZADEH N, GREALISH A, CHAMBERS N, et al. CaTeRS: causal and temporal relation scheme for semantic annotation of event structures [C]//Proceedings of the 4th Workshop on Events: Definition, Detection, Coreference, and Representation. Philadelphia, USA: Association for Computational Linguistics, 2016:51-61.
- [15] MIRZA P, TONELLI S. CATENA: causal and temporal relation extraction from natural language texts [C]//Proceedings of the 26th International Conference on Computational Linguistics. Philadelphia, USA: Association for Computational Linguistics, 2016:64-75.
- [16] NING Qiang, FENG Zhili, WU Hao, et al. Joint reasoning for temporal and causal relations [C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Philadelphia, USA: Association for Computational Linguistics, 2018:2278-2288.
- [17] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space [EB/OL]. [2019-04-23]. <https://arxiv.org/abs/1301.3781>.
- [18] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need? [EB/OL]. [2019-04-23]. <https://arxiv.org/abs/1706.03762>.
- [19] DEY R, SALEM F M. Gate-variants of Gated Recurrent Unit (GRU) neural networks [EB/OL]. [2019-04-23]. <https://arxiv.org/abs/1701.05923>.
- [20] NAIR V, HINTON G E. Rectified linear units improve restricted Boltzmann machines [C]//Proceedings of the 27th International Conference on Machine Learning. New York, USA: ACM Press, 2010:807-814.
- [21] MIRZA P, TONELLI S. On the contribution of word embedding to temporal relation classification [C]//Proceedings of the 26th International Conference on Computational Linguistics. Philadelphia, USA: Association for Computational Linguistics, 2016:2818-2828.

编辑 陆燕菲