浙江科技学院学报，第 34 卷第 1 期，2022 年 2 月
Journal of Zhejiang University of Science and Technology
Vol. 34 No. 1, Feb. 2022

doi：10.3969/j.issn.1671-8798.2022.01.006

基于深度学习的交通事故文本因果关系抽取

周龚雪，马伟锋，龚一飞，王柳迪
（浙江科技学院 信息与电子工程学院，杭州 310023）

摘　要：针对交通事故文本因果关系抽取过程中因果事件边界难以识别及连锁因果关系难以抽取的问题，将抽取问题转化为序列标注问题，提出了相对逗号位置特征及基于该特征与字词向量混合的多头注意力卷积双向长短时记忆网络的因果关系抽取方法。首先将字词分别编码后与相对逗号位置特征拼接，其次通过卷积神经网络（convolutional neural network，CNN）、双向长短时记忆网络（bidirectional long and short-term memory networks，Bi-LSTM）及多头注意力机制（multihead self-attention，MHSA）提取深层次的语义信息及长距离特征信息，最后采用条件随机场（conditional random field，CRF）分类器进行分类，得到最终的输出结果。在我们创建的交通事故文本数据集上将本模型与主流模型进行比较，结果表明：本模型抽取结果的召回率与 F_1 值分别提高了 5.75% 和 2.54%，可以更有效地抽取交通事故文本中的因果关系。较完整地抽取因果关系有利于人们分析交通事故的成因，从而为如何有效地预防和避免交通事故的再次发生提供参考。

关键词：因果关系抽取；序列标注；双向长短时记忆网络；多头注意力机制
中图分类号：TP391.1　　**文献标志码**：A　　**文章编号**：1671-8798(2022)01-0042-10

Causality extraction from traffic accidents texts based on deep learning

ZHOU Gongxue, MA Weifeng, GONG Yifei, WANG Liudi
(School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, Zhejiang, China)

Abstract: In response to the problem that the event boundary and chain causalities are difficult to identify in the process of causality extraction from traffic accident texts, the causality extraction was transformed into a sequence labeling task, in which a method was proposed to extract the relative comma position feature. This method encoded the words and chars, then combined them with the relative comma position features, introduced convolutional neural networks (CNN), bidirectional long and short-term memory networks (Bi-LSTM) and multihead self-attention (MHSA) to extract deep features and long-distance features, and finally used a conditional random field (CRF) classifier for classification to obtain the ultimate outputs. By comparing the model with mainstream models on the traffic accident text dataset,

收稿日期：2021-03-26

通信作者：马伟锋（1979—　），男，浙江省绍兴人，副教授，硕士，主要从事大数据与人工智能应用研究。E-mail：mawf@zust.edu.cn。

the results show that the model has a marked increase in the recall rate and F_1 value, by 5.57% and 2.54% respectively, capable of more effectively extracting the causal relationship in traffic accident texts. And it is more conducive to the analysis into the causes of traffic accidents, so it can provide reference for effectively preventing and avoiding traffic accidents.

Keywords: causality extraction; sequence labeling; bidirectional long short-term memory networks; multihead self-attention

随着汽车技术的发展，车辆给人们带来了便捷的生活。与此同时，频发的交通事故也严重危害了人民群众的生命及财产安全。交通事故的发生具有随机性，但也具有相当的统计规律性和必然性[1]，故交通事故产生的原因备受人们的关注。对交通事故中原因的分析是预防与避免交通事故再次发生的有效手段。孙轶轩等[2]构建了事故数据的特征变量集，利用支持向量机探究影响交通事故严重程度的核心因素；贾熹滨等[3]利用关联规则，结合新闻报道具有的真实性和时效性特点来进行造成交通事故的原因及责任的分析。然而，目前针对交通事故的分析一般从人、车、路、环境等方面展开，为了更好地探究事故发生的原因，需要对事故中事件因果关系的演变过程进行挖掘与分析。

因果关系是“原因”与“结果”之间的关系，是引起和被引起的关系[4]，在事件检测与预测、情景生成、问答等任务中起着十分重要的作用。许晶航[5]将一个实体既是一个实体的原因，同时又是另一个实体的结果的特殊因果关系定义为“连锁因果”。故文本中除了原因事件与结果事件外，还存在连锁因果事件，即一个事件既是上一事件的结果，又是导致下一事件发生的原因。对文本中的因果关系抽取经历了长时间的发展。早期常构造规则用于抽取因果关系，Khoo 等[6-7]根据因果关系句子的结构特点，利用关键词通过模式匹配的方式进行因果关系抽取。而后随着机器学习与深度学习的发展，基于机器学习的方法与基于深度学习的方法都被提出来用于文本中因果关系的抽取。Zhao 等[8]提出一种新的贝叶斯网络，利用上下文特征、句法特征、位置特征与新构建的因果连接词类别特征，有效提升了因果关系抽取的准确率；Zeng 等[9-10]利用卷积神经网络(convolutional neural network，CNN)来进行文本因果关系抽取，以捕获句子中的深层语义信息，从而改善因果关系抽取的效果；田生伟[11]等提取 10 项事件内部结构信息特征，同时用双向长短时记忆网络来抽取用维吾尔语表述的事件因果关系，该方法改善了事件因果关系抽取的效果。

上述研究在文本因果关系抽取中均取得了较好的效果，但目前针对特定交通事故领域的相关研究还不多见。交通事故产生的原因常常是复杂多变的，因此对句子中多因多果与连锁因果的挖掘十分重要，然而目前针对交通事故文本中因果关系的抽取还存在以下问题：一是现有中文因果关系抽取语料库不足，难以完成交通事故因果关系抽取任务；二是交通事故文本中的因果事件边界难以寻找，不能直接利用分词工具得出；三是交通事故的产生常由多种因素造成，文本中含有多个因果事件，因此句子中常包含多种因果关系，对其中的连锁因果关系难以抽取。针对上述问题，我们创建了交通事故文本因果关系抽取数据集，采用序列标注的方法来进行抽取，并且在原因标签与结果标签的基础上，增加连锁因果标签，在抽取原因与结果的同时抽取连锁因果来表示因果事件的演变过程。为正确识别交通事件的边界，我们利用交通事故文本中交通事件的位置特性，提出相对逗号的位置特征，将该特征与字向量、词向量、位置向量拼接后一起输入卷积神经网络中进行编码，并在双向长短时记忆网络(bidirectional long and short-term memory networks，Bi-LSTM)与条件随机场(conditional random field，CRF)结合的序列标注模型上添加多头注意力机制，提出基于字词向量与相对位置向量相融合的多头注意力卷积双向长短时记忆网络(multihead self-attention convolution Bi-LSTM network with mixed char-word and relative position embeddings)，以下简称 MACL。

1　MACL 模型的构建

序列标注是自然语言处理中的任务之一，它可以用于解决分词、词性标注、命名实体识别、关系抽取等问题。它给一维线性序列中的每个元素都打上标签集合中的某个标签，若输入为中文句子，则给句子中的每个汉

字都标记指定标签中的某个标签。因此,将因果关系抽取任务转化为序列标注任务是解决因果关系抽取问题的一种有效途径,它可以通过对句子中的每个汉字进行标注从而确定句子中所包含的因果事件及其关系。

交通事故文本因果关系的抽取流程如图 1 所示,对交通事故文本进行预处理后,将语料库中的因果关系对标签转化为序列标注任务所需的标注方式,从而给每个字都打上对应的标签。把数据集分为训练集、验证集、测试集三部分,将训练集与验证集放入搭建的 MACL 模型中进行训练,得到训练好的模型。利用模型对测试集进行预测,输出句子中每个字对应的标签,将此标签进行还原,得到句子中的因果关系对。

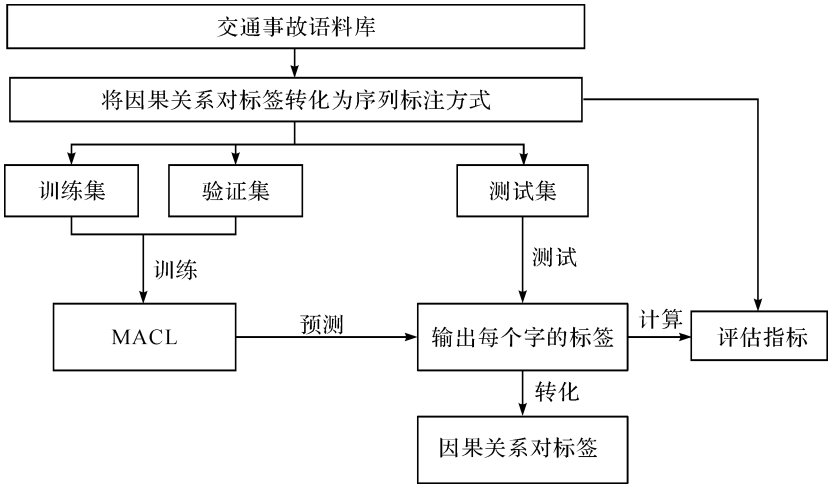

图 1 交通事故文本因果关系的抽取流程图

Fig. 1 Flow chart of causality extraction of traffic accident text

利用序列标注方法完成因果关系的抽取,MACL 模型的结构如图 2 所示。模型包括输入层、嵌入层、卷积层、双向长短时记忆网络、注意力层、条件随机场分类器、输出层。MACL模型工作流程如下:

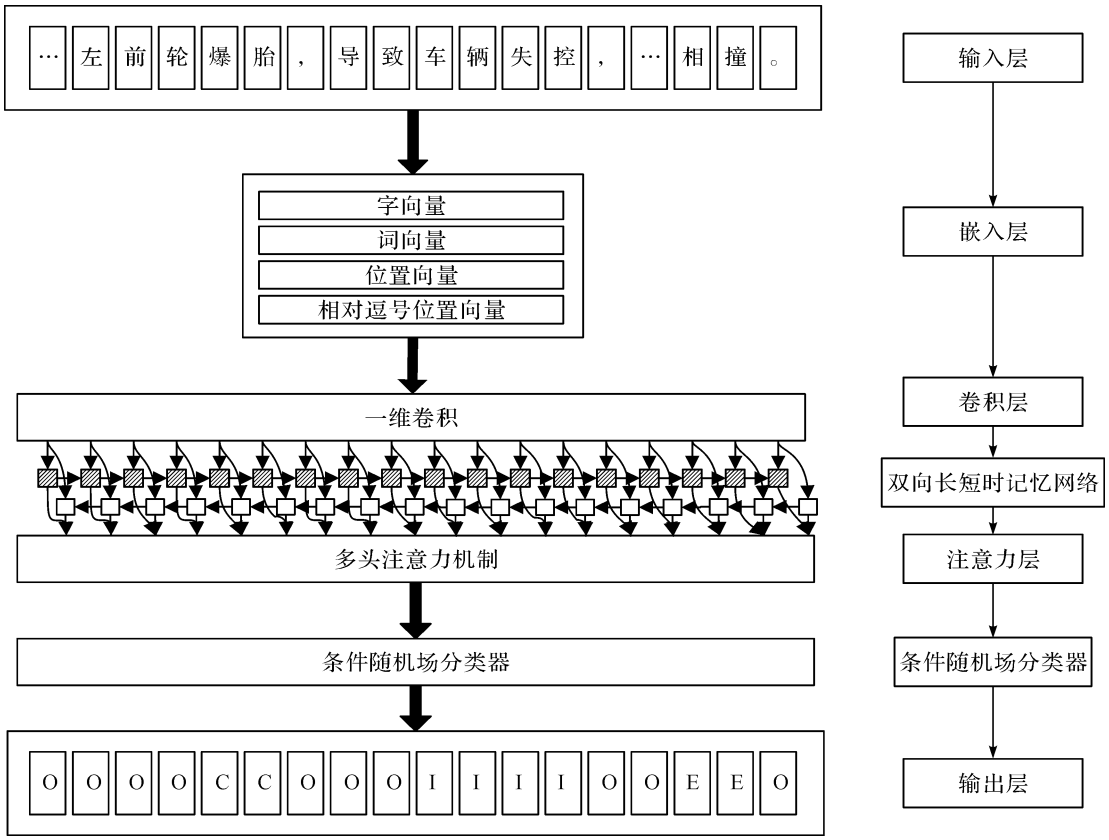

图 2 MACL 模型的结构

Fig. 2 Model structure of MACL

1）将句子从输入层输入后，通过嵌入层向量化；2）通过卷积层提取特征，将卷积层提取的特征放入 Bi-LSTM 中挖掘潜在语义信息，利用上下文信息提取远距离语义特征；3）将 Bi-LSTM 得到的结果通过注意力机制后放入 CRF 层，CRF 利用相邻词的信息，通过极大似然函数训练后得到每个字的因果标签；4）输出层输出最终的分类结果，其中“C”代表原因事件词，“I”代表连锁因果事件词，“E”代表结果事件词，“O”代表该词没有因果关系。

1.1　嵌入层

输入层将含有因果关系的句子输入后，为增加文本表示能力，将字向量与训练好的词向量、位置向量拼接。为了使词向量更符合交通领域的特征，本研究采集了与交通相关的 22 683 条微博信息，对数据进行预处理后，使用 Word2Vec 模型中的 CBOW（continuous bag-of-words）模式训练得到词向量，设定共现窗口大小为 10，词向量维度为 256。

为了更准确找到因果事件的边界，提高因果事件的抽取准确率，本研究强化了因果事件的位置特征。交通事故文本中的因果事件与一般的实体不同，在描述交通因果事件时常使用具有不同词性的词相结合的形式，因此并不能简单地使用分词工具而得到因果事件。Chen 等[12]为在整个句子中分割出因果事件，将因果连接词与逗号共同作为分割的依据。因此借鉴 Chen 等的方法，本研究利用交通事故因果事件边界与逗号位置的相关性，抽取得到一个相对逗号位置的二维向量 $\boldsymbol{l}$，$\boldsymbol{l}\in\mathbb{R}^{a\times b}$，$a$ 表示与前一个逗号的相对距离，b 表示与后一个逗号的相对距离。将此二维向量与字、词、位置向量进行如下拼接：

$$\boldsymbol{E}=\boldsymbol{C}+\boldsymbol{W}+\boldsymbol{P}+\boldsymbol{l}。\tag{1}$$

式(1)中：$\boldsymbol{C}$ 为字向量；$\boldsymbol{W}$ 为词向量；$\boldsymbol{P}$ 为位置向量；$\boldsymbol{l}$ 为相对逗号位置向量。

1.2　卷积层

将嵌入层得到的向量放入一维卷积中，卷积层可以有效提取深层特征。本研究采用单层卷积网络结构，为减少信息的丢失未设池化层。在进行卷积操作时，选取过滤器数量为 30 个，卷积核大小为 3，卷积核维度为 $3\times v$（v 为字向量、词向量、位置向量、相对逗号位置向量维度之和），卷积过程如图 3 所示。

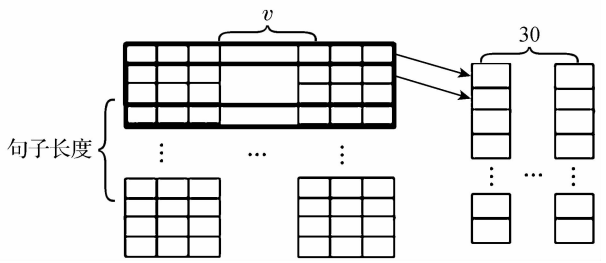

图 3　卷积过程

Fig. 3　Convolution procedure

1.3　双向长短时记忆网络

在传统的网络中，输入与输出都是假设相对独立的，但是这与很多实际情况不符。想预测下一个词是什么的时候，我们最好可以知道它前面有哪些词。循环神经网络（recurrent neural network，RNN）利用记忆单元解决了这个问题，然而对距离较远的文本信息仍无法有效利用。为更好利用远距离文本信息长短时记忆网络（long short-term memory，LSTM）被提出，作为 RNN 的变体，它是一种链式结构，如图 4 所示。它拥有 3 个特殊的“门”结构——遗忘门、输入门、输出门。遗忘门可以决定要抛弃这个单元中的某些信息，可以“忘记”之前没用的信息；信息被遗忘后，输入门从当前的输入中选择哪些部分可以加入；输出门则根据当前的状态计算产生当前时刻的输出。

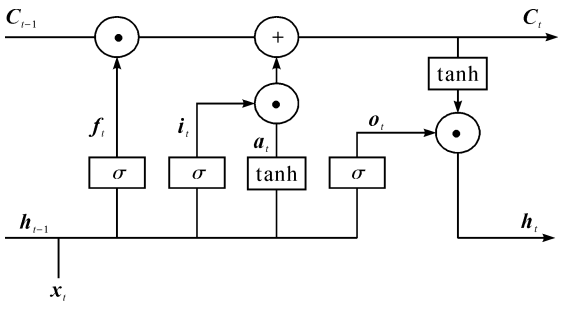

图 4　LSTM 结构

Fig. 4　Structure of LSTM

图 4 中：$\boldsymbol{h}_{t-1}$ 为上一序列隐藏状态；$\boldsymbol{C}_{t-1}$ 为旧细胞状态；$\boldsymbol{x}_t$ 为本序列数据；$\boldsymbol{C}_t$ 为新细胞状态；$\boldsymbol{h}_t$ 为当前时刻隐藏状态；$\boldsymbol{f}_t$ 为遗忘门的输出；$\boldsymbol{i}_t$、$\boldsymbol{a}_t$ 为输入门的两部分；$\boldsymbol{o}_t$ 为输出信息。

上一序列隐藏状态 $\boldsymbol{h}_{t-1}$ 和本序列数据 $\boldsymbol{x}_t$ 通过 sigmoid 激活函数得到遗忘门的输出 $\boldsymbol{f}_t$。$\boldsymbol{f}_t$ 计算方式如下：

$$\boldsymbol{f}_t = \sigma(\boldsymbol{W}_f \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_f)。 \tag{2}$$

式(2)中:σ 为 sigmoid 激活函数;$\boldsymbol{W}_f$ 为系数矩阵;$\boldsymbol{b}_f$ 为偏置。

输入门由两部分组成,一部分由 sigmoid 激活函数得到,设为 $\boldsymbol{i}_t$,另一部分由 tanh 激活函数得到,设为 $\boldsymbol{a}_t$,计算方式如下:

$$\begin{cases} \boldsymbol{i}_t = \sigma(\boldsymbol{W}_i \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_i); \\ \boldsymbol{a}_t = \tanh(\boldsymbol{W}_a \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_a)。 \end{cases} \tag{3}$$

式(3)中:$\boldsymbol{W}_i$、$\boldsymbol{W}_a$ 为系数矩阵;$\boldsymbol{b}_i$、$\boldsymbol{b}_a$ 为偏置项。

遗忘门和输入门的结果都会被用于更新细胞状态 $\boldsymbol{C}_t$:

$$\boldsymbol{C}_t = \boldsymbol{f}_t \cdot \boldsymbol{C}_{t-1} + \boldsymbol{i}_t \cdot \boldsymbol{a}_t。$$

最后,本序列数据 $\boldsymbol{x}_t$ 和上一序列隐藏状态 $\boldsymbol{h}_{t-1}$ 及 sigmoid 激活函数一起得到输出信息 $\boldsymbol{o}_t$,将之前得到的新细胞状态 $\boldsymbol{C}_t$ 通过 tanh 层后与输出信息 $\boldsymbol{o}_t$ 相乘后得到当前时刻隐藏状态 $\boldsymbol{h}_t$:

$$\begin{cases} \boldsymbol{o}_t = \sigma(\boldsymbol{W}_o \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_o); \\ \boldsymbol{h}_t = \boldsymbol{o}_t \cdot \tanh(\boldsymbol{C}_t)。 \end{cases}$$

在单向 LSTM 中,模型实际上只使用到了“过去”的数据,而没有考虑“未来”的数据,因此,将前向的 LSTM 与后向的 LSTM 相结合来组成 Bi-LSTM,在获取“上文”的同时可以获取“下文”的信息,其结构如图 5 所示。

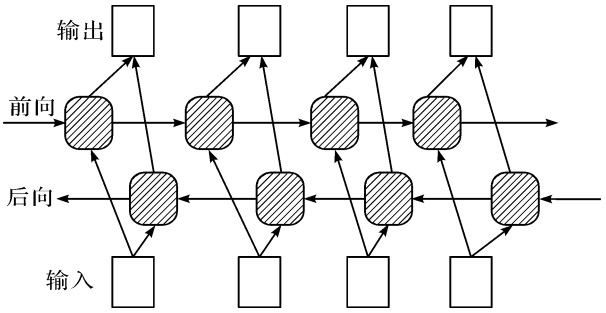

图 5 Bi-LSTM 结构

Fig. 5 Structure of Bi-LSTM

我们将得到的第 i 个字符的编码信息放入 Bi-LSTM 中,前向 LSTM 与后向 LSTM 分别得到隐藏特征:$\overrightarrow{\boldsymbol{l}_i}$与$\overleftarrow{\boldsymbol{l}_i}$,2 个隐藏特征拼接后得到:$\boldsymbol{l}_i = [\overrightarrow{\boldsymbol{l}_i}, \overleftarrow{\boldsymbol{l}_i}]$,最终得到 Bi-LSTM 的输出 $\boldsymbol{L} = [\boldsymbol{l}_1, \boldsymbol{l}_2, \cdots, \boldsymbol{l}_n]$。

1.4 多头注意力机制

自注意力机制自提出就受到广泛的关注,它被用于自然语言处理的各种任务中,如语义角色标注[13]、医学关系抽取[14]等。注意力机制源自人们视觉感知物体时的特性,即人们常观察特定需要的部分。Vaswani 等[15]提出了多头注意力机制,其结构如图 6 所示,查询(Query,Q)、键(Key,K)、值(Value,V)首先经过一个线性变换,然后做 h 次的放缩点积,每次算一个头,即 h 个头,再拼接在一起后进行线性变换得到结果 $\boldsymbol{M}$。本研究将经过 Bi-LSTM 层后的结果矩阵 $\boldsymbol{L}$ 放入注意力层,使用多头注意力机制重新分配权重,设定 h 值为 3,计算过程如下:

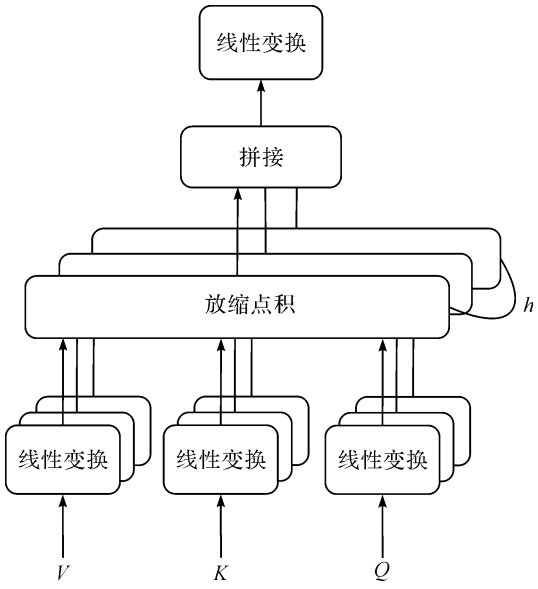

图 6 多头注意力机制结构

Fig. 6 Structure of multihead self-attention

$$\begin{cases} f_{\mathrm{a}}(\boldsymbol{LW}_i^q, \boldsymbol{LW}_i^k, \boldsymbol{LW}_i^v) = f_{\mathrm{s}}\left(\dfrac{(\boldsymbol{LW}_i^q)(\boldsymbol{LW}_i^k)}{\sqrt{d_k}}\right)\boldsymbol{LW}_i^v; \\ \boldsymbol{H}_i = f_{\mathrm{a}}(\boldsymbol{LW}_i^q, \boldsymbol{LW}_i^k, \boldsymbol{LW}_i^v); \\ \boldsymbol{M} = f_{\mathrm{c}}(\boldsymbol{H}_1, \cdots, \boldsymbol{H}_h)\boldsymbol{W}^o。 \end{cases} \tag{4}$$

式(4)中:$\boldsymbol{W}_i^q$、$\boldsymbol{W}_i^k$、$\boldsymbol{W}_i^v$、$\boldsymbol{W}^o$ 为参数矩阵;d_k 为 $\boldsymbol{L}$ 的维度;f_{s} 为 Softmax 函数;f_{a} 为放缩点积函数;f_{c} 为拼接函数。

1.5 条件随机场

条件随机场[16]是一种判别式的无向图模型,可以使用句子级别的标注信息,且不需要严格的独立性假设,可以有效克服标注偏置等问题[17],因此在词性标注和命名实体识别等序列标注任务中可得到较好

的效果。它是在给定一组输入随机变量条件下另一组输出随机变量的条件概率分布模型，其输出变量构成马尔可夫随机场。最简单且最常用的是一阶链式结构，即线性链结构。令 $X=\{x_1,x_2,\cdots,x_n\}$ 表示观察序列（因果关系抽取任务中的句子），$Y=\{y_1,y_2,\cdots,y_n\}$ 是有限状态的集合（句子中每个汉字所对应的语义角色），设 $P(Y|X)$ 为线性链条件随机场，则得到的 Y 条件概率为

$$\begin{cases} Z(x)=\sum\limits_{y\in Y}\exp\left(\sum\limits_{n=1}^{n}\sum\limits_{k}\lambda_k f_k(y_{n-1},y_n,X,n)\right); \\ P(Y|X)=\dfrac{1}{Z(x)}\exp\left(\sum\limits_{n=1}^{n}\sum\limits_{k}\lambda_k f_k(y_{n-1},y_n,X,n)\right)。 \end{cases} \tag{5}$$

式(5)中：$f_k(y_{n-1},y_n,X,n)$ 为特征函数；λ_k 为对应的权重；$Z(x)$ 为规范化因子。特征函数的权重可以通过极大似然法训练得到，再利用得到的条件概率模型进行预测。

2　数据与试验

2.1　试验数据

本研究使用的数据集为自定义数据集，来源于政府部门网站发布的交通事故调查报告与安全管理网中的车辆损坏案例。选择交通事故报告与案例文档中描述事故的直接原因语句，并对它进行标注，构建为交通事故数据集。因果事件的判断具有很强的主观性，为保证标注结果的客观性，选择 4 人同时进行标注，选择标注一致的结果作为最终结果。因果关系对的标签形式与 SCITE 数据集[18]一致，具体标注方式如下：

＜item id＝"1" label＝"Cause-Effect((e1,e2),(e2,e3))"＞

＜sentence＞左前轮＜e1＞爆胎＜/e1＞，导致＜e2＞车辆失控＜/e2＞，与对向正常行驶的大货车＜e3＞相撞＜/e3＞。＜/sentence＞

使用标签对＜e1＞＜/e1＞来表示原因事件、连锁因果事件或结果事件，label＝Cause-Effect((e1,e2),(e2,e3))用来表示各事件之间的关系，(e1,e2)表示事件 e1 是事件 e2 的原因，(e2,e3)表示事件 e3 是 e2 所导致的，由此可知，e1 为原因事件，e2 为连锁因果事件，e3 为结果事件。

由于本研究采用序列标注任务的方式来完成抽取，因此需要将标签转化为更简洁的形式并给每个字都进行具体标注。标注样例如图 7 所示。

图 7　序列标注任务的标注样例

Fig. 7　Labeling sample of a sequence labeling task

将原因事件用 C 表示，将结果事件用 E 表示，将连锁因果事件用 I 表示，其他无关文本用 O 表示。

经过整理标注结果之后，本研究的数据集中含有因果关系的句子 1 084 条，其中训练集包含 975 条，测试集包含 109 条数据；训练集中共含有因果事件 3 003 个，测试集中含有 300 个。数据集构成见表 1。

表 1　数据集构成

Table 1　Dataset composition

标签类型	训练集	测试集
C	14 128	1 943
E	3 844	836
I	4 992	758
O	38 236	7 423
合计	61 200	10 960

2.2　评估指标的选择

选用精确率 P、召回率 R、F_1 值作为评估指标，计算公式如下：

$$\begin{cases} P=\dfrac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{P}}}; \\ R=\dfrac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{N}}}; \\ F_1=\dfrac{2PR}{P+R}。 \end{cases} \tag{6}$$

式(6)中：T_P 为实际为正抽取为正的标签数；F_P 为实际为负抽取为正的标签数；F_N 为实际为正但抽取为负的标签数。

试验的目的是抽取原因 C、结果 E 和连锁因果 I，因此无关事件词 O 不是本试验关注的重点。由于文本中含有大量的无关文本，若将其放入计算指标则会对结果产生影响。因此，在计算时，先将标注为 O 的数据排除，再计算 P、R、F_1 的值，用于比较 C、E、I 标签的抽取结果。由表 1 的数据集数量可知，由于交通事故文本的特殊性，数据存在不均衡现象，不同标签类型的数据量差距较大，P、R 的值无法全面衡量，因此本研究重点比较 F_1 值的大小。

2.3 对比模型的选择

为验证本模型的有效性，本文选取多种模型在同一数据集下进行比较，包含基准模型 LSTM、LSTM+CRF、Bi-LSTM，除此之外还包括如下模型：

1) Bi-LSTM+CRF[19]，经典的序列标注模型，将 Bi-LSTM 与 CRF 分类器相结合，当下的许多序列标注任务都是基于此种模型而展开的。

2) Bi-LSTM+self-ATT[13]，将 Bi-LSTM 与自注意力机制相结合，将原模型中的语义角色标注修改为本文所需标注后进行比较。

3) BERT+Bi-GRU+CRF[10]，用于识别突发事件，将原模型中的标注修改为本文所需标注后，将该模型用于交通事故文本因果关系抽取后与本模型的结果进行比较。

4) Attention-based Bi-LSTM[20]，基于注意力机制的双向 LSTM 网络，该模型被提出用于进行突发事件演化关系的抽取，将该模型用于交通事故文本因果关系抽取后与本模型的结果进行比较。

5) BERT+Bi-LSTM+MHSA+CRF，将 BERT 预训练模型与 Bi-LSTM、多头注意力机制、CRF 分类器相结合后，将该模型用于交通事故文本因果关系抽取后与本模型的结果进行比较。

2.4 试验结果与分析

试验训练 100 个轮次后，将本文模型与其他模型就准确率、召回率、F_1 值在测试集上做对比，结果见表 2。

表 2 本文模型与其他模型性能的对比情况

Table 2 Comparison between MACL model and other models on performance

模型	P	R	F_1
LSTM	0.618 6	0.593 8	0.606 0
LSTM+CRF	0.597 3	0.547 8	0.571 5
Bi-LSTM	0.670 6	0.700 5	0.685 2
Bi-LSTM+CRF	0.692 2	0.716 7	0.704 3
Bi-LSTM+self-ATT	**0.697 0**	0.685 6	0.691 3
BERT+Bi-GRU+CRF	0.461 7	0.415 8	0.437 6
Attention-based Bi-LSTM	0.678 4	0.703 3	0.690 6
BERT+Bi-LSTM+MHSA+CRF	0.581 0	0.686 6	0.629 4
MACL	0.692 1	**0.743 1**	**0.716 7**

由表 2 可知，在 F_1 值的比较上本文模型优于其他模型。通过模型性能比较可知，一些经典的基准模型在序列标注任务中仍有十分出色的表现，特别是 Bi-LSTM 模型，有较强的捕捉上下文信息的能力。Bi-LSTM+self-ATT 模型虽然在准确率上比本文模型提高了 0.49%，但本文模型在召回率与 F_1 值上均比它效果更好，分别提高了 5.75%和 2.54%。对比注意力机制的效果，本文模型利用多头注意力机制分配权重，使得注意力更关注于描述因果事件的词上，而不是其他无关词，可以更完整地抽取因果关系。同时与 BERT+Bi-LSTM+MHSA+CRF 模型的对比中可以看出，本研究提出的字词相对位置融合向量的编码方式，可以增强对语义特征提取与边界识别的能力，从而提高因果关系的抽取准确率。

在测试集上计算各标签的准确率、召回率、F_1 值的平均值，本文模型与其他模型在测试集上的对比

情况见表 3，本文模型在标签 C、E 上的 F_1 值皆为最高。

表 3　本文模型与其他模型在各标签上对比情况

Table 3　Comparison between MACL model and other models on every label

模型	C-P	C-R	C-F_1	E-P	E-R	E-F_1	I-P	I-R	I-F_1
LSTM	0.434 8	0.375 4	0.395 5	0.769 9	0.193 9	0.308 6	0.479 1	0.148 4	0.220 0
LSTM＋CRF	0.505 7	0.352 2	0.410 1	0.640 5	0.206 6	0.309 5	0.462 8	0.126 6	0.186 0
Bi-LSTM	0.336 8	0.437 1	0.373 7	0.641 7	0.2855	0.384 5	0.063 5	0.204 4	0.087 2
Bi-LSTM＋CRF	0.632 9	0.434 8	0.511 7	0.774 1	0.296 5	0.428 0	0.636 3	0.256 1	0.350 2
Bi-LSTM＋self-ATT	0.572 2	0.461 7	0.507 1	0.220 3	0.270 8	0.241 8	**0.681 8**	0.200 4	0.298 7
BERT＋Bi-GRU＋CRF	0.420 6	0.189 7	0.257 0	0.651 8	0.187 5	0.287 4	0.199 3	0.105 8	0.135 3
Attention-based Bi-LSTM	0.650 4	0.485 1	0.555 6	0.734 3	0.313 8	0.439 6	0.657 7	0.339 3	**0.447 4**
BERT＋Bi-LSTM＋MHSA＋CRF	0.534 6	0.425 5	0.471 1	0.737 3	0.297 5	0.420 9	0.584 7	0.256 8	0.347 6
MACL	**0.663 3**	**0.499 6**	**0.569 6**	**0.793 4**	**0.320 1**	**0.456 0**	0.632 9	**0.339 7**	0.440 6

本研究选取了 5 种模型（Bi-LSTM、Bi-LSTM＋CRF、Bi-LSTM＋self-ATT、BERT＋Bi-LSTM＋MHSA＋CRF、MACL）对比在训练过程中验证集的 F_1 值随轮次变化的情况。为尽可能保持数据分布的一致性，将训练集随机分成 10 组后，随机选取一组作为验证集。将 5 种模型在验证集上进行比较，验证集 F_1 值随迭代次数变化的情况如图 8 所示。

由图 8 可知，本文模型 MACL 在开始训练时的 F_1 值已高于其他模型，随着轮次的增加，模型的 F_1 值保持稳定且在迭代 80 次后达到最高值；其他模型的 F_1 值随着迭代次数的增加也不断增加，但在迭代 50 次后逐渐趋于稳定并缓慢上升。可见 MACL 模型与其他模型相比稳定性更好。

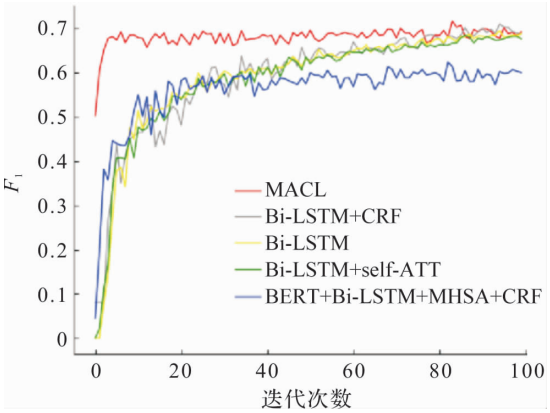

图 8　验证集 F_1 值随迭代次数变化的情况

Fig. 8　Variation of validation set F_1 value with number of iterations

2.5　预测结果错误分析

模型预测错误结果列举见表 4，从表中可知，在因果类型是一因一果的情况下，预测结果与真实值相近，一因一果的抽取成功率较高。但在多因多果或存在连锁因果的情况下，模型预测准确率有所降低，对此做如下分析：

1）标签错误。为提高标注的准确率采用人工标注的方式，但人工标注存在一定的主观性，且存在具有争议而无法确定的因果事件。对于人工尚难以识别的因果事件，模型则更难以提取准确的特征信息，因此容易导致预测错误。如表 4 中，在多因一果的句子中，“安全法律意识淡薄”与“超载”“逆向行驶”同作为原因事件，但是句子中还包含着由于“安全法律意识淡薄”从而导致“超载”“逆向行驶”的因果关系，“超载”“逆向行驶”可作为连锁因果词，因此对标签的标注存在歧义。

2）因果事件数量较多且分布不均衡。交通事故文本数据总量较少，但每条数据中因果事件数量较多。交通事故文本中常存在多因一果的情况，每个句子中的因果事件常大于 3 个。由表 1 标注情况可知，标签类型为 C 的数量远大于标签类型为 E 和 I 的数量，因果标签分布不均衡，原因事件数量远大于连锁事件与结果事件。

3）模型结构。因模型自身结构的不完善，导致未能更好地提取句子更深层的语义特征，从而使得部分因果关系抽取失败，从表 4 中可以看出，在对连锁因果关系进行抽取时，未将事件“换道行驶”抽取成功。

表 4 模型预测错误结果列举

Table 4 Sample of model prediction error results

因果类型	真实结果	预测结果
一因一果	label="Cause-Effect((e1,e2))" ……车辆行驶<e1>跟前车距离太近</e1>,因此发生<e2>追尾事故</e2>。	label="Cause-Effect((e1,e2))" ……车辆行驶<e1>跟前车距离太近</e1>,因此发生<e2>追尾事故</e2>。
多因一果	label="Cause-Effect((e1,e4),(e2,e4),(e3,e4))" 驾驶人……交通<e1>安全法律意识淡薄</e1>,驾驶<e2>超载</e2>的重型货车<e3>逆向行驶</e3>,是造成此<e4>事故</e4>的直接原因。	label="Cause-Effect((e1,e3),(e2,e3))" 驾驶人……交通安全法律意识淡薄,驾驶<e1>超载</e1>的重型货车<e2>逆向行驶</e2>,是造成此<e3>事故</e3>的直接原因。
连锁因果	label="Cause-Effect((e1,e3),(e2,e3),(e3,e4))" ……驾驶机动车<e1>换道行驶</e1>,<e2>占道停车</e2>,<e3>影响后方车辆通行秩序</e3>,也是导致<e4>事故</e4>发生的原因之一。	label="Cause-Effect((e1,e2),(e2,e3))" ……驾驶机动车换道行驶,<e1>占道停车</e1>,<e2>影响后方车辆通行秩序</e2>,也是导致<e3>事故</e3>发生的原因之一。

3 结 语

本研究针对交通事故文本因果关系难以抽取的问题提出了 MACL 模型,先将逗号相对位置特征与字词向量、位置向量相融合后采用 CNN 进行编码,然后放入 Bi-LSTM 中挖掘长距离语义特征,接着利用多头注意力机制进行处理,最后采用 CRF 分类器进行分类判断。使用我们构造的交通事故语料集,将 MACL 模型与其他几种模型进行多方面试验对比,结果表明本文模型能更有效地抽取交通事故文本中所包含的因果关系。本研究仍存在一些不足之处,如数据集标注较复杂,数据量较少,连锁因果没有再进行细粒度分析以致未抽取出更深层嵌套因果关系等,这些都有待进一步研究。

参考文献:

[1] 王洪明. 我国公路交通事故的现状及特征分析[J]. 中国安全科学学报,2009,19(10):122.

[2] 孙轶轩,邵春福,岳昊,等. 基于 SVM 灵敏度的城市交通事故严重程度影响因素分析[J]. 吉林大学学报(工学版),2014,44(5):1315.

[3] 贾熹滨,叶颖婕,陈军成. 基于关联规则的交通事故影响因素的挖掘[J]. 计算机科学,2018,45(增刊 1):447.

[4] 冯冲,康丽琪,石戈,等. 融合对抗学习的因果关系抽取[J]. 自动化学报,2018,44(5):811.

[5] 许晶航. 基于深度学习与图注意力网络的因果关系抽取研究[D]. 长春:吉林大学,2020.

[6] KHOO C, KORNFILT J, ODDY R, et al. Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing[J]. Literary and Linguistic Computing,1998,13(4):177.

[7] ZHAO S D, WANG Q, MASSUNG S, et al. Constructing and embedding abstract event causality networks from text snippets[C]//Proceedings of the 10th ACM International Conference on Web Search and Data Mining. New York: Association for Computing Machinery,2017:335.

[8] ZHAO S D, LIU T, ZHAO S C, et al. Event causality extraction based on connectives analysis[J]. Neurocomputing, 2016,173(3):1943.

[9] ZENG D J, LIU K, LAI SW, et al. Relation classification via convolutional deep neural network[C]//Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. Dublin: Dublin City University and Association for Computational Linguistics,2014:2335.

[10] 郑巧夺. 基于深度学习的突发事件关系识别研究[D]. 成都:四川师范大学,2020.

[11] 田生伟,周兴发,禹龙,等. 基于双向 LSTM 的维吾尔语事件因果关系抽取[J]. 电子与信息学报,2018,40(1):200.

[12] CHEN D, CAO Y, LUO P. Pairwise causality structure: towards nested causality mining on financial statements [C]//CCF International Conference on Natural Language Processing and Chinese Computing. Cham: Springer, 2020:725.

[13] TAN Z X, WANG M X, XIE J, et al. Deep semantic role labeling with self-attention[C]//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans: AAAI,2018:4929.

[14] VERGA P, STRUBELL E, MCCALLUM A. Simultaneously self-attending to all mentions for full-abstract biological relation extraction[C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. New Orleans: Association for Computational Linguistics, 2018: 872.

[15] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: Curran Associates Inc, 2017: 6000.

[16] LAFFERTY J D, MCCALLUM A, PEREIRA F C N. Conditional random fields: probabilistic models for segmenting and labeling sequence data[C]//Proceedings of the Eighteenth International Conference on Machine Learning. San Francisco: Morgan Kaufmann Publishers Inc, 2001: 282.

[17] 付剑锋，刘宗田，刘炜，等. 基于层叠条件随机场的事件因果关系抽取[J]. 模式识别与人工智能，2011，24(4)：569.

[18] LI Z N, LI Q, ZOU X T, et al. Causality extraction based on self-attentive BiLSTM-CRF with transferred embeddings[J]. Neurocomputing, 2021, 33(5): 207.

[19] HUANG Z H, XU W, YU K. Bidirectional LSTM-CRF models for sequence tagging[EB/OL]. (2015-08-01)[2021-02-26]. https://arxiv.org/pdf/1508.01991.pdf.

[20] 闻畅，刘宇，顾进广. 基于注意力机制的双向长短时记忆网络模型突发事件演化关系抽取[J]. 计算机应用，2019，39(6)：1646.

[21] 刘苏文，邵一帆，钱龙华. 基于联合学习的生物医学因果关系抽取[J]. 中文信息学报，2020，34(4)：60.

[22] 仇培元，张恒才，余丽，等. 微博客蕴含交通事件信息抽取的自动标注方法[J]. 中文信息学报，2017，31(2)：107.

[23] CAO Y, CHEN D, XU Z, et al. Nested relation extraction with iterative neural network[J]. Frontiers of Computer Science, 2021, 15(3): 1.

(上接第 6 页)

[8] 李媛，夏锦. 复对称 Toeplitz 算子与向量值函数空间上的 Toeplitz 算子[J]. 四川轻化工大学学报(自然科学版)，2021，34(2)：79.

[9] 夏锦，王晓峰，曹广福. Dirichlet 空间上的 Toeplitz 算子的一些性质[J]. 数学物理学报，2012，32(2)：395.

[10] 胡云重，陈泳. 小 Hankel 算子和 Toeplitz 算子的乘积的一些代数性质[J]. 湖州师范学院学报，2014，36(4)：1.

[11] 卢玉峰，张波. Bergman 空间上可交换的 Hankel 算子和 Toeplitz 算子[J]. 数学年刊 A 辑，2011，32(5)：519.

[12] 黄辉斥. Bergman 空间上小 Hankel 算子的代数性质[J]. 复旦学报(自然科学版)，2005，44(3)：370.

[13] 邓燕，徐宪民. Bergman 空间上拟齐次 Toeplitz 算子的乘积[J]. 应用泛函分析学报，2009，11(3)：211.

[14] KANG D O, KIM H J. Products of truncated Hankel operators[J]. Journal of Mathematical Analysis and Applications, 2016, 435(2): 1804.

[15] CUCKOVIC Z, RAO N V. Mellin transform, monomial symbols, and commuting Toeplitz operators[J]. Journal of Function Analysis, 1998, 154(1): 195.

[16] 郑涛涛，来越富. 非齐次空间上的双线性广义分数次积分算子[J]. 浙江科技学院学报，2018，30(3)，181.