doi:10.3969/j.issn.1673-9833.2024.01.010

基于序列增强的事件主体抽取方法

沈加锐[1]，朱艳辉[1]，金书川[2]，张志轩[1]，满芳滕[2]

（1. 湖南工业大学 轨道交通学院，湖南 株洲 412007；2. 湖南工业大学 计算机学院，湖南 株洲 412007）

摘 要：为了解决在事件抽取中使用固定文本长度造成短句子填充过多，从而引发语义偏移的问题，提出一种基于序列增强的事件主体抽取方法。具体而言，首先，将固定长度文本通过预训练模型映射到 1 个稠密向量中；然后，将文本对应的稠密向量与自定义 Mask 层和 SpatialDropout 层按位相乘，得到编码输出；最后，将该输出连接 BiGRU 层以及 Mask 层，得到解码输出，并将其映射在 MLP 层中，得到最后的结果。该模型既能够避免预训练模型对文本表征过拟合的问题，也能够限制填充文本在语义上的过度表达。使用 CCKS 2022 所供金融领域事件主体作为数据集进行不同模型读取对比实验，所得实验数据表明，对填充文本加以负影响的增强序列比传统序列在事件主体识别的正确性、F_1 值上皆显著提升。

关键词：序列增强；事件主体；抽取；掩码模型

中图分类号：TP389.1 **文献标志码**：A **文章编号**：1673-9833(2024)01-0070-08

引文格式：沈加锐，朱艳辉，金书川，等 . 基于序列增强的事件主体抽取方法 [J]. 湖南工业大学学报，2024，38(1)：70-77.

Event Subject Extraction Method Based on Sequence Enhancement

SHEN Jiarui[1]，ZHU Yanhui[1]，JIN Shuchuan[2]，ZHANG Zhixuan[1]，MAN Fangteng[2]

（1. College of Railway Transportation，Hunan University of Technology，Zhuzhou Hunan 412007，China；
2. College of Computer Science，Hunan University of Technology，Zhuzhou Hunan 412007，China）

Abstract：In view of a solution of semantic deviation brought about by overfilling short sentences with fixed text length in event extraction, a sequence enhancement based event subject extraction method has thus been proposed. Specifically, an initial mapping of the fixed-length text is given to a dense vector through a pre-trained model. Subsequently, the dense vector corresponding to the text is bitwise multiplied by the custom Mask layer and SpatialDropout layer, thus obtaining the encoded output. Finally, the output is connected with BiGRU and Mask layers to get the decoded output, which is then mapped to an MLP layer to obtain the final result. This model can not only avoid the problem of overfitting the text representation in the pre-trained model, but also limit the semantic overexpression of the filled text. By using the financial field event subjects provided by CCKS 2022 as a dataset for different model reading comparative experiments, the experimental data obtained shows that the enhanced sequence with negative impact on filled text significantly improves the accuracy and F_1 value of event subject recognition compared to traditional sequences.

Keywords：sequence enhancement；event subject；extraction；masked model

收稿日期：2023-02-26

基金项目：国家自然科学基金资助项目（62106074）；湖南省教育厅基金资助重点项目（22A0408，21A0350）；湖南省自然科学基金资助项目（2022JJ50051）

作者简介：沈加锐，男，湖南工业大学硕士生，主要研究方向为自然语言处理与事件抽取，E-mail：1046188545@163.com

通信作者：朱艳辉，女，湖南工业大学教授，硕士生导师，主要研究方向为自然语言处理与知识工程，E-mail：swayhzhu@163.com

1　研究背景

事件是指发生在某个特定的时间点或者特定的时间段 [1]，在特定区域内由一个或者多个角色做出一组或者多组动作而造成的状态改变的行为 [2]。事件抽取，旨在将包含有事件信息的非结构化文本和半结构化文本以结构化的形式展示出来 [3−4]。近年来，事件抽取吸引了较多的研究机构和科研工作者们的注意，其中，MUC（Message Understanding Conference）会议、ACE（Automatic Content Extraction）会议等，就是典型的事件抽取评测会议。在各有关的事件抽取实验之中，ACE 2005 数据集 [5] 更是被作为绝大部分实验的评测语料，出现在各个事件抽取任务中。此外，常用于事件抽取任务的测评语料还有 TAC-KBP（Text Analysis Conference Knowledge Base Population）语料库、TDT（Topic Detection and Tracking）语料库和其他特定领域的语料库，如 BioNLP 语料库、TimeBANK 语料库、CEC（Chinese Electronic Corpora）语料库、MUC 语料库等 [6−7]。

在 ACE 中定义的事件包括事件元素与事件触发词两个部分。其中，事件触发词是事件语句的核心部分，ACE 将事件抽取任务分为触发词检测、触发器 / 事件类型检测、参数检测、参数角色识别 4 个阶段。其中触发词检测是检测事件是否存在的依据，可用其判断语句是否具有后续抽取价值。触发器 / 事件类型检测是通过不同触发词归类事件类型，并由研究人事先定义好的模式组成。参数检测类似于命名实体识别，抽取事件语句中的各种论元实体。参数角色识别将论元实体分类到相应角色，如时间、地点、涉事公司等。

根据抽取对象领域的不同，可以将事件抽取分为开放领域和专业 (封闭) 领域 [8−9]；根据文本粒度的不同，可以将事件抽取分为句子级和篇章级；根据模型结构的不同，可以将事件抽取分为 pipeline 式和联合模型式。在开放领域中，有更多的数据集可选。不同的文本粒度对事件抽取也有影响，篇章级包含更多事件，模型进行分辨时难度较大，而细粒度一般只包含一个扁平或嵌套事件，抽取准确性较高。根据抽取方法的不同，可以将事件抽取分为模式匹配和机器学习，例如，文献 [10−12] 提出了一种由人工构建的事件抽取模式，其通过对文本进行匹配，从而提取出文本中的事件信息，这种方法可以获得较高的精确率，但是人工参与程度较高，且受制于模式搭建，不利于新事件的抽取。文献 [13−14] 使用机器学习方法，将信息处理重点放在特征项上，通过词汇特征、句法特征、语义特征等获得文本信息，从而提取事件信息，但其对数据标注的质量有极高的依赖，并且无法抽取较复杂的事件。

近年来，随着深度学习不断地发展，抽取方法的中心也从模式匹配方向逐渐转移到机器学习，再到深度学习方法上。文献 [15] 提出使用 HMM（hidden markov model）与句法分析相结合的方法，对事件进行抽取。首先，使用句法分析对中文文本进行分析，随后将得到的句法结构交给 HMM 学习，得到一个抽取模型，其模型在 200 篇网络地震文本中的 F_1 值达 86.184。但 HMM 与语法分析对语言的依赖性较强，增加了开发与移植不同语言的难度，其对于复杂语法结构也表现较差。文献 [16] 抽取了微博交通信息，经过去噪、句子分割、词性标注及命名实体识别后，使用 CRF（conditional random field）与基于规则的正则表达方法抽取文本中的事件信息，结果 F_1 值为 62.5，在标准化的语料库上 F_1 值为 66.5。但这种方法会依赖正则表达的设计好坏，极大限制了在不同任务中使用同一套模型的效率。

2019 年，BERT（bidirectional encoder representations from transformers）模型横空出世，其有效推动了 NLP（natural language processing）领域各任务的发展。文献 [17] 用 BERT 微调稠密词向量作为中文字词表示，使用 BiLSTM+CRF（bidirectional long short-term memory + conditional random field）方法，在突发公共卫生事件上，以 pipeline 方式进行事件抽取，建立的模型在该数据集上的 F_1 值得分为 86.32，相较于只使用 BiLSTM+CRF 的 F_1 值 78.3，有了较大的提升。但是此种方法使用文本截断时的大小对最终的结果有较大的影响，同时在多语言文本中难以实现事件抽取；文献 [18] 提出了一个基于 GAT（graph attention networks）的模型，其利用 Sentence Community 缓解多事件和角色重叠问题，提高了事件的抽取效率，其平均 F_1 值为 78.9。但是其对低频出现事件抽取表现较差，需要每类事件有较多数据可供训练；文献 [19] 使用 BiLSTM+Attention 机制整合信息的基础上，利用 top-k 注意力机制，构建语法依赖图，学习隐藏的语义上下文表示，抽取事件时间关系，其在 Micro 数据集上的 F_1 值为 73.2，相较 CAEVO（CAscading EVent Ordering architecture）有较大提升。但是这种方法在整合信息层面有较高要求，在这个过程中需要作出大量预处理工作，以防误差传播到语法依赖图上 [20]。

上述研究均获得了一定的成果，但是很多工作

对数据的依赖性较强或者需要前期的大量准备工作，而使用 BERT 预训练模型也未注意到填充文本对语义表达的影响。为了解决预训练模型对文本填充部分过分表达的问题，本文拟提出一种序列增强的事件主体抽取方法，用于解决在事件抽取中使用固定文本长度造成短句子填充过多，从而引发语义偏移的问题。本研究进行了如下创新：

1）融合预训练模型对字符级特征表达能力和序列模型对语义的表达能力，构建编码器－解码器架构实现事件主体抽取任务；

2）加入 Mask 层，增强序列抽取能力，可有效抑制文本填充部分对语义空间的表达能力，从而提高最终的识别精确率；

3）将损失函数与文本填充部分通过 Mask 层相关联，减少了模型过拟合现象的发生。

2 基于序列增强事件主体抽取模型构建

本文提出一种序列增强的事件主体抽取方法，该方法能有效解决抽取过程中填充文本在训练中错误传播的问题，模型整体结构见图 1。

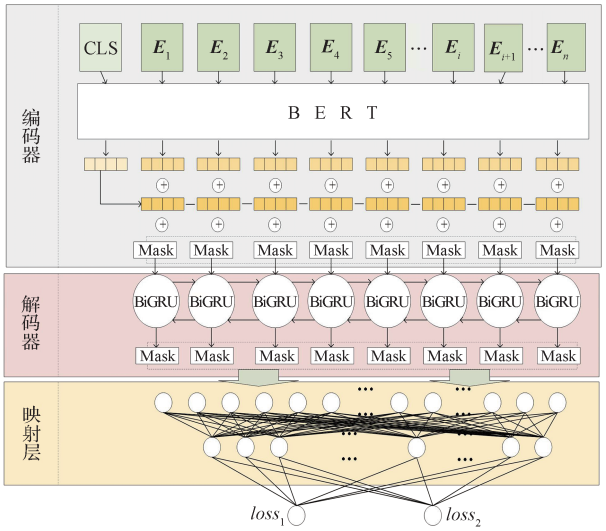

图 1 模型整体结构示意图

Fig. 1 Model overall structure diagram

为提高事件的主体抽取精确率，模型构建编码器－解码器－MLP（multi-layer perceptrons）映射层用于解析文本语义。其中：编码器以 BERT 层为主体，将输入文本以字符级切分为 token，经过 BERT 层和 Dropout 层训练后，得到每个 token 的稠密词向量表达；解码器以增强序列层为主体；MLP 映射层由使用神经元个数递减的全连接层组成。这样的组织结构有效将各种长度的文本序列映射到相同语义空间之中，学习一种共同的语义表达。

2.1 编码器

本文构建的编码器主要由 BERT 层、Mask 层、Dropout 层和 SpatialDropout 层组成，如图 2 所示。

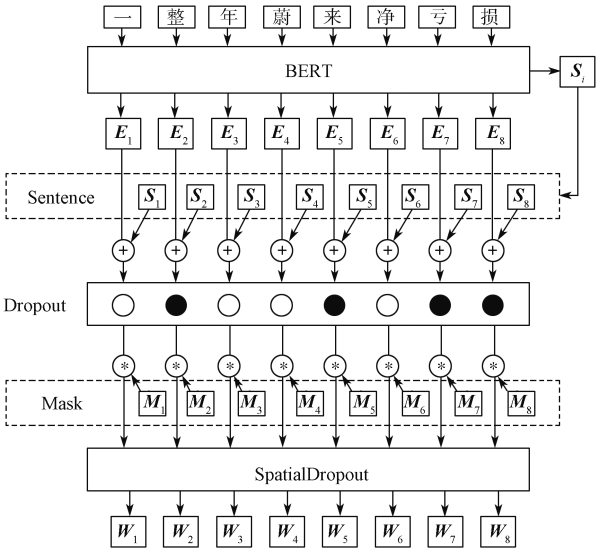

图 2 编码器结构示意图

Fig. 2 Encoder structure diagram

为便于对文本整体进行语义编码，本文将文本截断、填充后，以字符为单位划分 token，让模型学习中文文本更细粒度的语义表达，token 输入 BERT 模型中得到字符级的向量表达，表达公式如下：

$$\{[\boldsymbol{E}_1, \boldsymbol{E}_2, \boldsymbol{E}_3, \cdots, \boldsymbol{E}_n],\ \boldsymbol{S}\} = BERT\left(Tokenizer\left(sentence\_text\right)\right), \tag{1}$$

式中：$\boldsymbol{E}_i$（i=1, 2, ··, n）为词向量；

$\boldsymbol{S}$ 为句向量。

Dropout 层以字向量 $\boldsymbol{E}_i$ 与句向量 $\boldsymbol{S}$ 的和作为输入，随机舍弃更新，其中更新公式如下：

$$\boldsymbol{r}_j^{(l)} \sim Bernoulli(p), \tag{2}$$

$$\boldsymbol{z}_i^{(l+1)} = \boldsymbol{w}_i^{(l+1)}\left(\boldsymbol{r}^{(l)} * \boldsymbol{y}^{(l)}\right) + \boldsymbol{b}_i^{(l+1)}, \tag{3}$$

$$\boldsymbol{y}_i^{(l+1)} = f\left(\boldsymbol{z}_i^{(l+1)}\right)。 \tag{4}$$

式（2）～（4）中：*Bernoulli* 函数生成概率向量 $\boldsymbol{r}$；

p 为控制随机舍去的概率；

$f(\cdot)$ 为使用的相应激活函数；

$\boldsymbol{z}$ 为中间结果值；

$\boldsymbol{w}$ 为可训练权重；

$\boldsymbol{y}$ 为第 l 层的结果；

$\boldsymbol{b}$ 为 l 层偏置项；

l 为神经层数。

本文在编码器结构中加入了一个 Mask 层，它将原文本对应的位置用 1 表示，填充部分用一个极小的数值表示，以此形成一个掩码序列向量，并将其与 BERT 模型微调后得到的稠密词向量相加，以此增加

文本部分的权重，降低填充部分的重要性。

本文编码器结构中的最后一层采用了SpatialDropout，它最早是在图像领域提出的，与Dropout相比，它不仅能够将部分元素置零，还可以随机对某一维度向量全部置零[21]。此外，该方法可以防止模型对特定特征项过度依赖，并通过强制学习所有特征来获取其语义表达。

2.2　解码器

本研究构建的解码器由两个双向序列模型、两个Mask层组成，如图3所示。

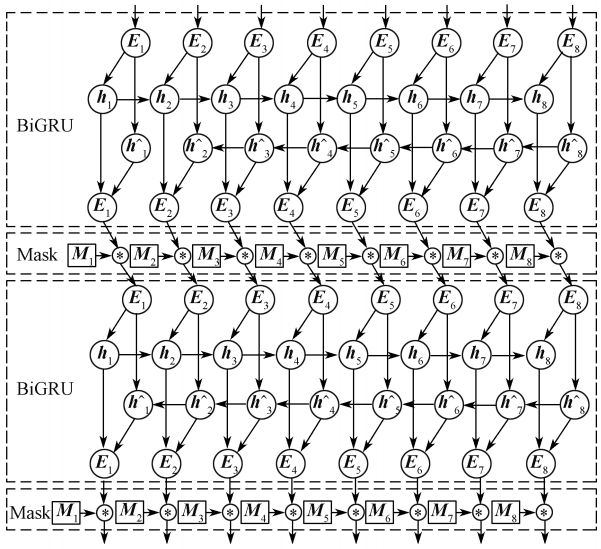

图3　解码器结构示意图

Fig. 3　Decoder structure diagram

从解码器角度来看，Bidirectional-RNN模型、Bidirectional-GRU（BiGRU）模型、BiLSTM模型或者是GRU模型、LSTM模型，都是可行的序列模型，本文拟通过实验证明，采用BiGRU可以在最终训练结果上达到最好的效果。

在经过编码层后得到的词向量表达 $\boldsymbol{E}_i$ 被用作解码器的输入，并通过第一层BiGRU处理，以形成新的 $[\boldsymbol{E}_1, \boldsymbol{E}_2, \boldsymbol{E}_3, \cdots, \boldsymbol{E}_i, \cdots, \boldsymbol{E}_n]$，其计算公式如下：

$$\boldsymbol{z}_t=\sigma\left(\boldsymbol{W}_{\mathrm{z}}\cdot\left[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t\right]\right), \tag{5}$$

$$\boldsymbol{r}_t=\sigma\left(\boldsymbol{W}_{\mathrm{r}}\cdot\left[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t\right]\right), \tag{6}$$

$$\tilde{\boldsymbol{h}}_t=\tanh\left(\boldsymbol{W}\cdot\left[\boldsymbol{r}_t\times\boldsymbol{h}_{t-1}, \boldsymbol{x}_t\right]\right), \tag{7}$$

$$\boldsymbol{h}_t=\left(1-\boldsymbol{z}_t\right)\times\boldsymbol{h}_{t-1}+\boldsymbol{z}_t\times\tilde{\boldsymbol{h}}_t。 \tag{8}$$

式（5）~（8）中：

$\sigma(\cdot)$ 为sigmoid函数；

$\boldsymbol{W}_{\mathrm{z}}$、$\boldsymbol{W}_{\mathrm{r}}$ 和 $\boldsymbol{W}$ 分别为更新门、重置门以及候选隐藏状态的权重矩阵。

首先，$\boldsymbol{E}_i$ 与Mask层进行点乘操作，用以降低第一层BiGRU的前半部分中填充部分文本的权重，并且将其输入至后半部分BiGRU中，得到更新后的 $[\boldsymbol{E}_1, \boldsymbol{E}_2, \boldsymbol{E}_3, \cdots, \boldsymbol{E}_i, \cdots, \boldsymbol{E}_n]$；然后通过Mask层对其进行处理，以减少填充文本的权重，从而得出解码器的输出结果。

2.3　MLP 映射层整合特征项

在经过编码器和解码器处理后，模型已学习到许多字符信息特征，并通过设计MLP映射层以抽取出事件主体，其模型结构如图4所示。

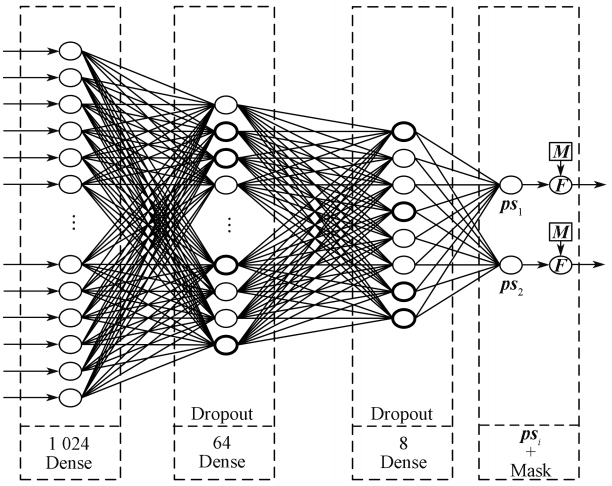

图4　MLP层结构示意图

Fig. 4　MLP layer structure diagram

如图4所示，MLP层前半部分由3层神经元以及2层Dropout组成。其中，3层神经元的数量呈指数递减，将高维信息映射至低维以有利于最后的表达输出，而2层Dropout则可以防止过拟合。MLP层后半部分对应2个输出，均由1个神经元以及1个InvMask层构成。在InvMask层中，该模型将会将所有的填充部分设定为微小数字，并进行相减处理；此外，该函数也会将神经元的输出内容减去InvMask中的反转函数，以得到如下的最后表达式：

$$\boldsymbol{ps}_i=\boldsymbol{Wve}_n[:,0]-(1-\mathrm{Mask}[:,0]*\alpha),\ i\in(1,2)。 \tag{9}$$

式中：α 为掩码部分掩盖权重超参数；

$\boldsymbol{Wve}_n$ 为经过上一层神经元后的特征表达；

$\boldsymbol{ps}_i$ 为经过InvMask层后的输出。

2.4　损失函数

为了提高事件主体抽取的精确率，本文设计了一个损失函数，其由两部分组成，分别对应于MLP层后的 $\boldsymbol{ps}_1$ 和 $\boldsymbol{ps}_2$，用于进行综合损失值计算，并进行反向传播。具体的计算公式如下。

$$\boldsymbol{cc}_i loss=-\sum_{t}^{outsize}\sum_{i} sen_i*\log\widehat{\boldsymbol{ps}_1}, \tag{10}$$

$$loss_1=\theta\left(\exp\left(\boldsymbol{cc}_i^1\right)\Big/\sum_{i=1}^{n}\exp\left(\boldsymbol{cc}_i^1\right)\right)。 \tag{11}$$

式（10）（11）中：

sen_i 为第 i 个输入句子，将输入 sen_i 与对应第 i 个 MLP 层输出做交叉熵，当求得第一部分损失函数时 i=1，求得第二部分损失函数时 i=2；

$\theta(\cdot)$ 为张量均值。

与第一部分损失函数类似，第二部分损失函数为了减少 Mask 层对抽取精确率的影响，需在计算损失函数时减去 InvMask 的改变量，公式如下：

$$\widetilde{\boldsymbol{ps}_2} = \boldsymbol{ps}_2 - \sigma(\boldsymbol{s}_1, axis)*\alpha, \tag{12}$$

式中$\sigma(\cdot, axis)$ 为输入值在 $axis$ 轴上的累积张量。

第二部分损失值通过式（10）得到 $\boldsymbol{cc}_2$，则第二部分损失函数表达式为

$$loss_2 = \theta\left(\exp\left(\boldsymbol{cc}_i^2\right) \Big/ \sum_{i=1}^{n} \exp\left(\boldsymbol{cc}_i^2\right)\right), \tag{13}$$

最终损失函数为

$$loss = loss_1 + loss_2。 \tag{14}$$

3 实验设计及数据分析

3.1 实验数据集分析

每年，CCKS 都发布各种高质量数据集，供研究者进行实验。这些数据集包括大量可用于事件抽取的数据，对于使用不同框架进行事件抽取具有显著的帮助[22–24]。为了证明本文提出方法的有效性，采用 CCKS 2022 评测任务九金融领域事件抽取数据集进行实验，其中训练数据集共 59 143 条，测试数据共 15 265 条。数据集文本长度分布如图 5 所示。

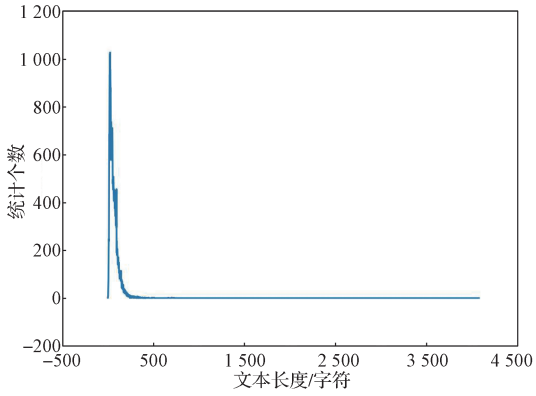

图 5 数据集文本长度分布图

Fig. 5 Dataset text length distribution

由图 5 可以得知，本研究选用的数据集文本长度集中分布在 20~140 字符之间。训练数据集事件类型分布如表 1 所示。

数据集中有 173 种不同的事件类型，但是这些类型的数量分布并不均匀。这是由金融领域事件发展所决定的，例如高层变动、与其他机构合作、资产重组等事件的概率较高，因此其对应的信息也会更加密集。

表 1 训练数据集类型统计结果

Table 1 Training dataset type statistics

事件类型	总数 / 个	事件类型	总数 / 个
高层变更	5 735	盈利能力下降	1 788
与其他机构合作	4 876	破产清算	1 192
业务 / 资产重组	2 400	公司退市	1 179
债务违约	2 311	股票转让 – 股权受让	1 178
股票转让 – 股权转让	1 957		

3.2 实验环境

本实验基于 TensorFlow 框架搭建模型，并使用 GPU 加速训练过程，以便快速收敛。为此，本模型封装原有优化器以实现梯度累积功能，将计算得到的梯度方向保存在内存中，在获取一个 batch 数据后，当进行一定的累加步数后，根据之前保存的梯度方向来对神经网络进行参数调整，接下来将所有保存的参数全部重新归零，并进行迭代训练。具体的实验硬件配置如下：操作系统为 Windows 10 × 64 位；CPU 为 i7-10510U@ 2.30 GHz，GPU 为 NVIDIA GeForce MX250；内存为 16 GB；Python 版本为 3.7.13；TensorFlow 版本为 1.14.0。

3.3 实验设计

为了验证本模型在金融领域文本中的事件主体抽取能力，本文设计如下对比实验方案：

1）BiLSTM+CRF 模型。这是一类典型的序列抽取模型，通过训练语料生成一个 200 维的词向量，将待抽取文本通过 BiLSTM 表达出来，再由 CRF 对其进行约束，最后得到序列预测结果。

2）BERT+CRF 事件抽取模型。事件先经过 BERT 编码后，将文本语义投影到 768 维的特征向量中，然后由 CRF 层对其进行约束，最后得到序列预测结果。

3）DMCNN 模型。它是一种 pipeline 式抽取方案，其具有自动学习特征的能力。该模型通过无监督的方式学习词嵌入以及字典级别特征表达能力，并且具备句子级特征抽取能力。同时，通过使用动态卷积神经网络来实现事件抽取。

4）序列增强事件抽取模型，即本文所提模型。首先构建了一个中文词表，供 BERT 模型查询使用。接着，使用 tokenizer 对文本进行细粒度切分，并将其输入 BERT 模型中。然后，将 BERT 模型输出的 token 向量作为双向 LSTM 模型的输入，以学习文本的隐藏语义特征。最后，运用 MLP 层对这些特征进行预测，得出最终结果。

除此之外，本文还开展了模型对比实验，比较了

基线模型和本研究所提出的模型在信息抽取方面的表现。接着，进行了一系列重要的超参数选取实验，以期望能够进一步改良本文的模型，并且达到更优秀的试验效果。

4　实验结果

4.1　超参数选取实验

为了寻求更高的抽取精确率和效率，分别对序列模型层数、α 值、MLP 层内神经元个数以及文本截取长度进行取值实验。本实验采用精确率 P 作为评价指标，实验结果如表 2 所示。

表 2　MLP 层、GRU 层数、α 值超参数结果对比

Table 2　Result comparison between MLP layer, layers of GRU and α values exceeding parameters

MLP 层	P/%	GRU	P/%	α 值	P/%
64-4-1	46.12	100	76.58	1e9	75.47
512-32-8	73.51	200	81.14	5e9	78.06
1 024-256-64	78.26	300	74.37	1e10	81.14
1 024-64-8	81.14	500	76.37	5e10	74.75

在经过 BERT 模型微调后，每个 token 都会具备高维度的语义表达能力，这就要求使用更多层数的神经元来承载信息。从表 2 所示实验结果可以看出，当 MLP 层神经元个数为 64-4-1 时，由于神经元较少，对特征空间的表达能力不足，因此精确率仅为 46.12%。在提高 MLP 每层神经元个数到 1024-64-8 后，精确率提高到 81.14%。

在序列模型层数实验中发现，随着层数的增加，精确率随之增加，且当层数达到最大值 200 后，精确率开始减小。实验结果表明，模型层数 GRU 为 200 时 P 值最大，因此本模型最终将序列模型的层数定为 200。

虽然 α 值会影响模型对非文本序列的权重，但从结果中发现 α 值也并非越大越好，当其值超过一定范围时，其精度将大幅度下降，根据实验结果显示，当 α 值为 1e10 时，结果最优。

同时，本文设计采用不同截取长度的文本，探究填充长度对模型提取事件主体的精确率和运行时间的影响，所得实验结果如表 3 所示。

表 3　文本截取填充长度结果对比

Table 3　Comparison of text intercept filling length results

文本长度	P/%	每个 Epoch 运行时间 /s
maxlen-32	72.62	11 843
maxlen-64	75.90	21 575
maxlen-140	78.86	53 816
maxlen-200	75.87	116 081

根据前 3 个实验结果可以得出，随着截取文本长度的增加，训练精度也相应增加，但是训练时间也会急剧上升。当文本长度从 140 字增加到 200 字时，训练时间大幅度上升，而训练精度却开始下降。这是由于当将截取文本填充到 200 字以后，会引入大量不相关的信息对语义造成干扰。这一实验结果进一步证明了使用 Mask 层的必要性，其可以有效降低语义干扰。权衡训练效果与运行时间两个因素，本文选择 maxlen 值为 140 进行后续实验。

除此之外，模型所采用的 BERT 模型是中文的 L-12_H-768_A-12 预训练权重，其中包含 12 层 Encoder，768 个隐藏神经单元以及 12 个 attention heads。考虑到每一轮的学习率应该有所不同，本研究使用了 warm up 方式来在不同轮数时改变学习率：在第一轮，以 0.005 的学习率开始，随后以 step 依次减少，直到 0.000 1。Dropout 层都使用 20% 的舍弃率进行随机舍弃；SpatialDropout 层则将舍弃率降低到 10%。

4.2　序列对比实验

为了评估 BiLSTM、BiGRU、LSTM 和 GRU 这 4 种序列模型的表达能力，在其他参数不变的情况下，采用 10 折交叉验证的方式进行实验，最终将表示能力取平均值作为最佳结果。所得各序列模型的实验结果如表 4 所示。

表 4　序列模型结果对比

Table 4　Comparison of sequence model results

序列模型	P/%	R/%	F_1
GRU	77.90	62.17	69.15
LSTM	76.17	62.11	68.42
BiLSTM	78.39	69.89	73.89
BiGRU	81.14	68.11	74.05

由表 4 可知，双向模型比单向模型具有更好的表达能力，本文采用的 BiGRU 模型比 BiLSTM 在理解句子语义逻辑上具有更强的优势。

4.3　模型对比实验

BiLSTM+CRF、BERT+CRF、DMCNN 和本文所给出模型的对比实验结果如表 5 所示。

表 5　4 种模型实验结果对比

Table 5　Comparison of experimental results of four models

模型	P/%	R/%	F_1
BiLSTM+CRF	71.66	69.31	70.46
BERT+CRF	72.80	70.44	71.60
DMCNN	80.40	67.70	73.50
本文模型	82.59	68.79	75.06

通过分析表 5 所示实验结果可以得出，DMCNN

模型经过动态卷积层后可以很好地表达出句子级的文本理解能力，抽取出事件主体。而简单使用 BiLSTM 较难对句子进行编码操作，很难理解句子内主体逻辑能力，精确度相较别的模型更低。且结果显示，本文模型具有最高的识别精确率，并且在 4 种模型中，F_1 得分最高。

5 结语

为了解决填充文本对语义表达带来的偏差，本文提出了一种增强序列模型，它对输入文本中的填充部分与非填充部分进行不同的处理，从而使模型能够更加高效地专注于原始文本部分，进而提升文本的表达能力。该方法结合了 BERT 字符级的语义映射能力与序列技术对文本特征的抽取能力，取得了优异的效果。

从实验数据集中的事件类型数量分布来看，大量不平衡样本类型对召回率造成影响，因此后续工作可以从以下方面进行优化：

1）通过采用样本均衡的方法来缓解样本类型数量间的不平衡，从而改善效果较差类别的召回率。

2）增加注意力机制，将文本语义空间与类型相关联，以便更好地提升抽取效果。

3）寻找更细粒度字符表达或增加字符表达能力，提高 BERT 模型的训练效果，进一步增加文本语义表达能力。

参考文献：

[1] BOGURAEV B，MUÑOZ R，PUSTEJOVSKY J. Proceedings of the Workshop on Annotating and Reasoning About Time and Events[C/OL]//Workshop on Annotating & Reasoning About Time and Events. Sydney：Association for Computational Linguistics 2006. [2023-02-10]. https://aclanthology.org/w06-0900.pdf.

[2] 项 威，王 邦 . 中文事件抽取研究综述 [J]. 计算机技术与发展，2020，30(2)：1-6.
XIANG Wei，WANG Bang. A Survey of Chinese Event Extraction[J]. Computer Technology and Development，2020，30(2)：1-6.

[3] CHEN Chen，VINCENT N G. Joint Modeling for Chinese Event Extraction with Rich Linguistic Features[C]// Proceedings of COLING 2012. Mumbai：The COLING 2012 Organizing Committee，2012：529-544.

[4] 李旭晖，程 威，唐小雅，等 . 基于多层卷积神经网络的金融事件联合抽取方法 [J]. 图书情报工作，2021，65(24)：89-99.
LI Xuhui，CHENG Wei，TANG Xiaoya，et al. A Joint Extraction Method of Financial Events Based on Multi-Layer Convolutional Neural Networks[J]. Library and Information Service，2021，65(24)：89-99.

[5] AGUILAR J，BELLER C，MCNAMEE P，et al. A Comparison of the Events and Relations Across ACE，ERE，TAC-KBP，and FrameNet Annotation Standards[C]// In Proceedings of the Second Workshop on EVENTS. Definition，Detection，Coreference，and Representation. Baltimore：Association for Computational Linguistics，2014：45-53.

[6] SUNDHEIM B M，CHINCHOR N A. Survey of the Message Understanding Conferences[C]// Proceedings of the Workshop on Human Language Technology (HLT'93). Princeton，New Jersey：Association for Computational Linguistics，1993：56-60.

[7] HIRSCHMAN L. The Evolution of Evaluation：Lessons from the Message Understanding Conferences[J]. Computer Speech and Language，1998，12(4)：281-305.

[8] MARCO A，GUS H，MIHAI S，et al. A Domain-Independent Rule-based Framework for Event Extraction. [C]//In Proceedings of ACL-IJCNLP 2015 System Demonstrations. Beijing：Association for Computational Linguistics and the Asian Federation of Natural Language Processing，2015：127-132.

[9] LIU J W，MIN L Y，HUANG X H. An Overview of Event Extraction and Its Applications[EB/OL]. [2023-02-10]. 2021: arXiv: 2111.03212. https://arxiv.org/abs/2111.03212.

[10] RILOFF E. Automatically Constructing a Dictionary for Information Extraction Tasks[C]//Proceedings of the Eleventh National Conference on Artificial Intelligence (AAAI'93). Washington，D.C.：AAAI Press/MIT Press，1993：811-816.

[11] BORSJE J，HOGENBOOM F，FRASINCAR F. Semi-Automatic Financial Events Discovery Based on Lexico-Semantic Patterns[J]. International Journal of Web Engineering and Technology，2010，6(2)：115-140.

[12] LI P F，ZHOU G D，ZHU Q M. Minimally Supervised Chinese Event Extraction from Multiple Views[J]. ACM Transactions on Asian Low-Resource Language Information Processsing（TALLIP），2016，16(2)：1-16.

[13] 王东波，吴 毅，叶文豪，等 . 多特征知识下的食品安全事件实体抽取研究 [J]. 数据分析与知识发现，2017(3)：54-61.
WANG Dongbo，WU Yi，YE Wenhao，et al. Research on Entity Extraction of Food Safety Events Based on Multi-Feature Knowledge[J]. Data Analysis and

Knowledge Discovery，2017(3)：54−61.

[14] 赵圆圆 . 基于机器学习的抑郁症电子病历时间事件信息抽取研究 [D]. 北京：北京工业大学，2020：68.
ZHAO Yuanyuan. Research on Time Event Information Extraction from Electronic Medical Records of Depression Based on Machine Learning[D]. Beijing：Beijing University of Technology，2020：68.

[15] 吴家皋，周凡坤，张雪英 . HMM 模型和句法分析相结合的事件属性信息抽取 [J]. 南京师大学报（自然科学版），2014，37(1)：30−34.
WU Jiagao，ZHOU Fankun，ZHANG Xueying. Research of the Extraction Method of Event Properties Based on the Combining of HMM and Syntactic Analysis [J]. Journal of Nanjing Normal University (Natural Science Edition)，2014，37(1)：30−34.

[16] 熊佳茜 . 基于 CRF 的中文微博交通信息事件抽取 [D]. 上海：上海交通大学，2014：67.
XIONG Jiaqian. CRF-Based Traffic Information Event Extraction in Chinese Weibo[D]. Shanghai：Shanghai Jiao Tong University，2014：67.

[17] 石　磊，李敬明，朱家明 . 基于 BERT-BiLSTM-CRF 的突发公共卫生事件抽取研究 [J]. 哈尔滨师范大学自然科学学报，2022，38(2)：37−42.
SHI Lei，LI Jingming，ZHU Jiaming. Event Extraction of Public Health Emergencies Based on BERT-BiLSTM-CRF[J]. Natural Science Journal of Harbin Normal University，2022，38(2)：37−42.

[18] HUANG Yusheng，JIA Weijia. Exploring Sentence Community for Document-Level Event Extraction[C]// Findings of the Association for Computational Linguistics. EMNLP 2021.Punta Cana：Association for Computational Linguistics，2021：340−351.

[19] XU X L，GAO T，WANG Y X，et al. Event Temporal Relation Extraction with Attention Mechanism and Graph Neural Network[J]. Tsinghua Science and Technology，2022，27(1)：79−90.

[20] 万齐智，万常选，胡　蓉，等 . 基于句法语义依存分析的中文金融事件抽取 [J]. 计算机学报，2021，44(3)：508−530.
WAN Qizhi，WAN Changxuan，HU Rong，et al. Chinese Financial Event Extraction Base on Syntactic and Semantic Dependency Parsing[J]. Chinese Journal of Computers，2021，44(3)：508−530.

[21] TOMPSON J，GOROSHIN R，JAIN A，et al . Efficient Object Localization Using Convolutional Networks[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston：IEEE，2015：648−656.

[22] SHENG J，LI Q，HEI Y M，et al. A Joint Learning Framework for the CCKS-2020 Financial Event Extraction Task[J]. Data Intelligence，2021，3(3)：444−459.

[23] YU W T，YI M Z，HUANG X H，et al. Make It Directly: Event Extraction Based on Tree-LSTM and Bi-GRU[J]. IEEE Access，2020，8：14344−14354.

[24] WANG Z Q，WANG X Z，HAN X，et al. CLEVE：Contrastive Pre-Training for Event Extraction[EB/OL]. [2023−02−22]. 2021：arXiv：2105.14485. https://arxiv.org/abs/2105.14485.

（责任编辑：廖友媛）