

基于阅读理解框架的中文事件论元抽取*

陈敏, 吴凡, 王中卿, 李培峰*, 朱巧明

(苏州大学计算机科学与技术学院, 苏州, 江苏, 215006)

{1131651415, 944721805}@qq.com, {wangzq, pfli, qmzhu}@suda.edu.cn

摘要

传统的事件论元抽取方法把该任务当作句子中实体提及的多分类或序列标注任务, 论元角色的类别在这些方法中只能作为向量表示, 而忽略了论元角色的先验信息。实际上, 论元角色的语义和论元本身有很大关系。对此, 本文提议将其当作机器阅读理解任务, 把论元角色表述为自然语言描述的问题, 通过在上下文中回答这些问题来抽取论元。该方法更好地利用了论元角色类别的先验信息, 在ACE2005中文语料上的实验证明了该方法的有效性。

关键词: 事件论元抽取; 阅读理解; 先验信息; BERT

Chinese Event Argument Extraction using Reading Comprehension Framework

Min Chen, Fan Wu, Zhongqing Wang, Peifeng Li*, Qiaoming Zhu

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

{1131651415, 944721805}@qq.com, {wangzq, pfli, qmzhu}@suda.edu.cn

Abstract

Traditional event argument extraction methods formulated this task as a multi-classification or sequence labeling task mentioned by entities in the sentence. In these methods, the category of argument roles can only be described as vectors, while their prior information are ignored. In fact, the semantics of argument role category is closely related with the argument itself. Therefore, this paper proposes to regard argument extraction as machine reading comprehension, with argument role described as natural language question, and the way to extract arguments is to answer these questions based on the context. this method can make better use of the prior information existed in argument role categories and its effectiveness is shown in the experiments of Chinese corpus of ACE 2005.

Keywords: Event argument extraction, Reading comprehension, Prior information, BERT

基金项目: 国家自然科学基金(61772354, 61836007); 国家自然科学基金青年基金项目(61806137); 江苏高校优势学科建设工程资助项目。

1 引言

作为信息抽取 (Information extraction) 中的重要子任务, 事件 (Event) 抽取是指从描述事件信息的文本中识别并抽取出包含的事件信息, 并以结构化的形式呈现出来。事件抽取任务一般分为2个步骤, 触发词 (Trigger) 抽取和论元 (Argument) 抽取。触发词抽取是根据上下文识别出触发词并判断其事件类型 (Event type); 论元抽取是根据事件类型, 抽取出参与事件的论元, 并分配论元角色 (Argument role)。在ACE2005数据集中, 定义了33种事件子类型 (8种事件大类) 和35种论元角色。例1给出了数据集中包含1个触发词和3个论元角色的事件句。

例1: 法官(A1)随即判(E1)被告(A2)7年预防性监禁(A3)。

触发词抽取部分需要识别出触发词E1, 其对应的事件类型为宣判 (Sentence)。论元抽取部分需要识别出参与宣判事件的论元并分配对应的角色。该事件的论元包括A1、A2和A3, 分别对应角色审裁官 (Adjudicator)、被告 (Defendant) 和判决结果 (Sentence)。

当前中文事件抽取研究更多的是解决触发词抽取问题 (Feng et al., 2018; Lin et al., 2018; Ding et al., 2019; Xiangyu et al., 2019), 而针对中文论元抽取的工作相对较少。Zeng et al. (2016)利用CNN和BiLSTM捕获句子和词汇信息, 然后把论元抽取视为实体提及的多分类任务。He and Duan (2019)利用条件随机场 (CRF) 和多任务学习的框架, 把论元抽取视为序列标注任务。尽管这种多分类或序列标注的方式被认为是事件抽取的一个很好地解决办法, 但是这样的做法仍然存在问题, 论元角色标签本身的语义信息和论元存在着重要关系, 现有的研究工作并不能利用论元角色标签本身的先验信息。如例1中, 判决结果 (Sentence) 这类论元角色出现频率较低, 而这个类别在多分类或序列标注训练中, 仅被视为交叉熵中的一个独热向量, 这种不清楚抽取什么往往导致劣质的性能。

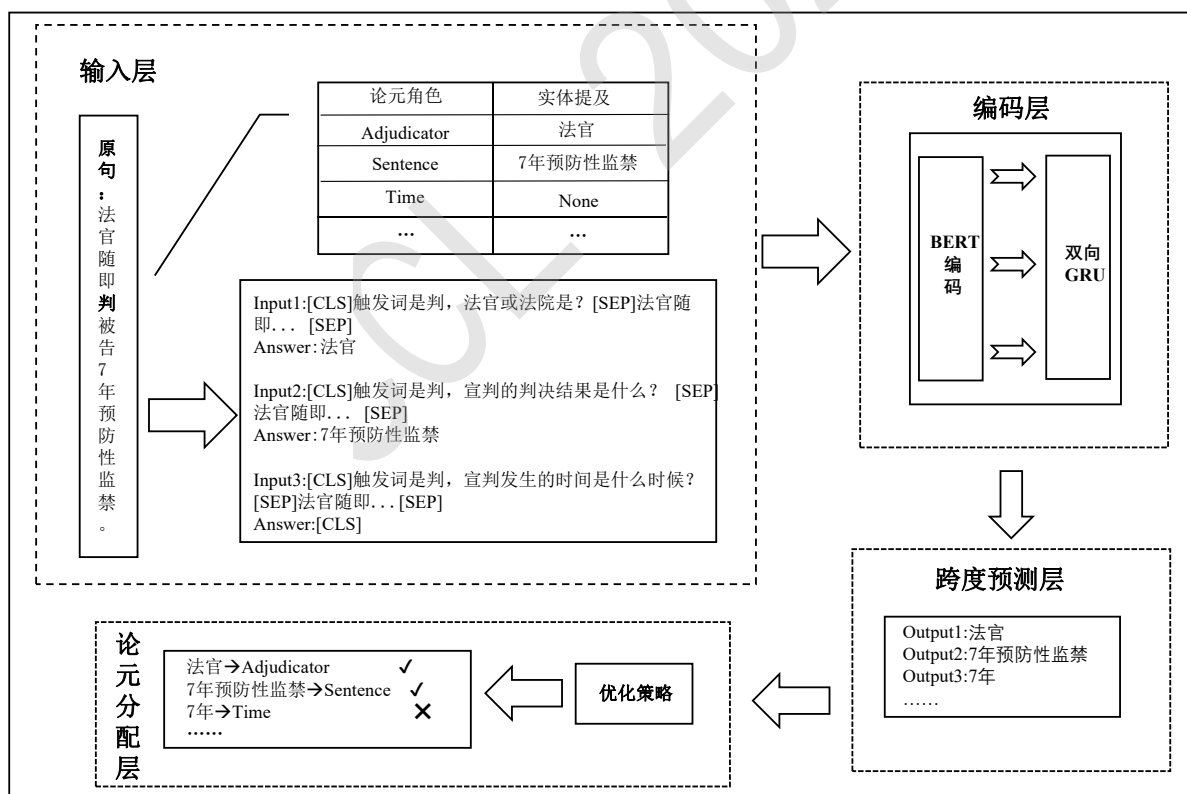


图 1. 论元抽取总体流程

本文的工作主要研究中文事件抽取中的论元抽取。针对论元抽取存在的上述问题, 提出了基于BERT阅读理解框架的论元抽取方法, 将论元抽取视为完型填空式的机器阅读理解

(Machine reading comprehension) 任务。该方法的总体流程如图1所示。如想要抽取的角色为判决结果 (Sentence)，通过回答问题“触发词是判，宣判的判决结果是什么？”来预测该角色对应的论元“7年预防性监禁”，从而实现对论元的识别和角色分配。这样的方式可以编码论元角色的先验信息，能够有效抽取出论元角色类别较少的论元。

总的来说，本文的方法利用已知的事件模式信息，将不同事件类型下的论元特征表述为自然语言问题，论元通过在事件句的上下文中回答这些问题来完成抽取。该方法通过BERT预训练模型学习问题和句子的初始隐向量表示，利用双向GRU更好的学习句子的上下文特征，然后对每个字进行二分类确定论元的跨度，采用合理的规则优化论元跨度，最终利用已知的实体提及完成论元角色识别和分配。在ACE2005中文语料上的实验证明，本文提出的基于阅读理解框架的论元抽取方法，优于传统的多分类或序列标注的方法，验证了阅读理解方式对论元抽取任务的有效性。

2 相关工作

事件抽取一直以来都是自然语言处理研究者们关注的重点领域。大多数工作把事件抽取看成两个阶段的问题，包括事件触发词抽取和论元抽取。触发词抽取工作近年来已经取得了很大的发展，论元抽取成为了事件抽取发展的瓶颈。目前论元抽取相关研究大部分是面向英文文本，中文论元抽取的发展较为缓慢。

在英文上，传统的基于特征表示的方法依靠人工设计的特征和模式。Liao and Grishman (2010)提出跨文档的方法来利用全局信息和其他事件信息。Hong et al. (2011)充分利用了事件句中实体类型的一致性特征，提出一种跨实体推理方法来提高事件抽取性能。Li et al. (2013)提出了一种基于结构预测的联合框架，合并全局特征，显式地捕获多个触发词和论元的依赖关系。随着神经网络的流行，研究者们开始利用预训练好的词向量作为初始化表示 Mikolov et al. (2013a; Mikolov et al. (2013b)，进而建模单词的语义信息和语法信息。Chen et al. (2015)对普通卷积神经网络做出改进，提出一种动态多池化卷积神经网络模型 (DMCNN)，把事件抽取看做两个阶段的多分类任务，先执行触发词分类，再执行论元分类，很好的解决了一个句子中包含多个事件的问题，但没有利用好触发词和论元之间的语义。Nguyen et al. (2016)通过循环神经网络 (RNN) 学习句子表示，联合预测触发词和论元，增加了离散特征。为了捕获触发词和论元之间的依赖关系，引入记忆向量和记忆矩阵来存储在标记过程中的预测信息。Liu et al. (2018)提出了一种新颖的联合多个事件提取框架 (JMEE)，通过引入句法短弧来增强信息流动，以解决句子编码的长距离依赖问题，利用基于注意力的图卷积网络来模型化图信息，从而联合抽取多个事件触发词和论元。Wang et al. (2019b)在DMCNN的基础上，提出了一种分层模块化的论元抽取模型，该模型采用灵活的模块网络 (Modular networks)，利用了论元角色相关的层次概念，作为有效的归纳偏置，不同论元角色共享相同的高层次的单元模块，有助于更好地抽取出特定的事件论元。

随着深度学习的进一步发展，一些先进的技术也被用于英文事件抽取，包括零样本学习 Huang et al. (2018)、远程监督 Chen et al. (2017)、BERT预训练模型 Devlin et al. (2019)等。

相对于英文论元抽取，中文论元抽取的工作发展较缓，中文需要分词、缺少时态等自身特点给该任务带来一定的挑战。尽管如此，近年来也取得了一些进展。传统方法的工作更多的在挖掘语义和语法特征，很大程度上依赖于手工制作的特征和模式。Li and Zhou (2012)引入形态结构来表示隐含在触发词内部的组合语义，提出了一个结合了中文词语的形态结构和义原去推测未知触发词的方法，明显提升了事件抽取的召回率。Chen and Ng (2012)利用局部和全局特征共同抽取触发词和论元。Zhu et al. (2015)利用事件之间的关系来学习实体扮演特定角色的概率，提出了基于马尔可夫逻辑网络的事件论元推理方法。He and Duan (2019)将事件抽取看作序列标注任务，并考虑到数据稀疏问题，对不同事件子类进行互增强，提出基于CRF的多任务学习事件抽取联合模型。神经网络发展起来后，Zeng et al. (2016)提出了一种基于LSTM和CNN的卷积双向LSTM神经网络模型，利用BiLSTM和CNN分别编码句子级别信息和局部词汇特征。

随着预训练语言模型的发展，深度学习提高了许多自然语言处理的性能。很多自然语言理解的任务可以转换为机器阅读理解任务 Mccann et al. (2018)，如文本分类、关系抽取、事件抽取、情感分析、文本蕴含、语言推理、语义角色标注等。机器阅读理解任务是从给定问题的段落中提取答案。将NLP任务转为阅读理解任务成为了新的趋势。Gardner et al. (2019)提出了使

用问答作为特定任务的格式的三种动机，即满足人类信息需求，探查系统对某些上下文的理解以及将学习到的参数从一项任务转移到另一项任务。Li et al. (2019)将实体关系抽取视为一种多轮问答任务，每种实体和关系生成不同的问答模板，这些实体和关系可以通过回答这些模板化的问题来进行抽取。Li et al. (2020)提出使用机器阅读理解框架代替序列标注模型，统一处理嵌套与非嵌套命名实体识别问题，在这种情况下，文本中实体的提取被形式化为回答问题，比如“文本中提到了哪个人？”。

3 基于阅读理解框架的论元抽取

受 Li et al. (2020)工作的启发，本文提出了基于阅读理解框架的论元抽取方法。在标准的机器阅读理解设置中，给定一个问题 $Q = \{Q_1, Q_2, \dots, Q_{N_q}\}$ (N_q 表示问题 Q 中的字数)，上下文 $S = \{S_1, S_2, \dots, S_{N_c}\}$ (N_c 表示句子 S 中的字数)，模型从给出问题的段落中提取答案跨度。该任务可以形式化为两个多分类任务，即预测给定问题的答案跨度的开始位置和结束位置。本文的方法也基于这种设置，该方法的流程和模型如图1和图2所示。

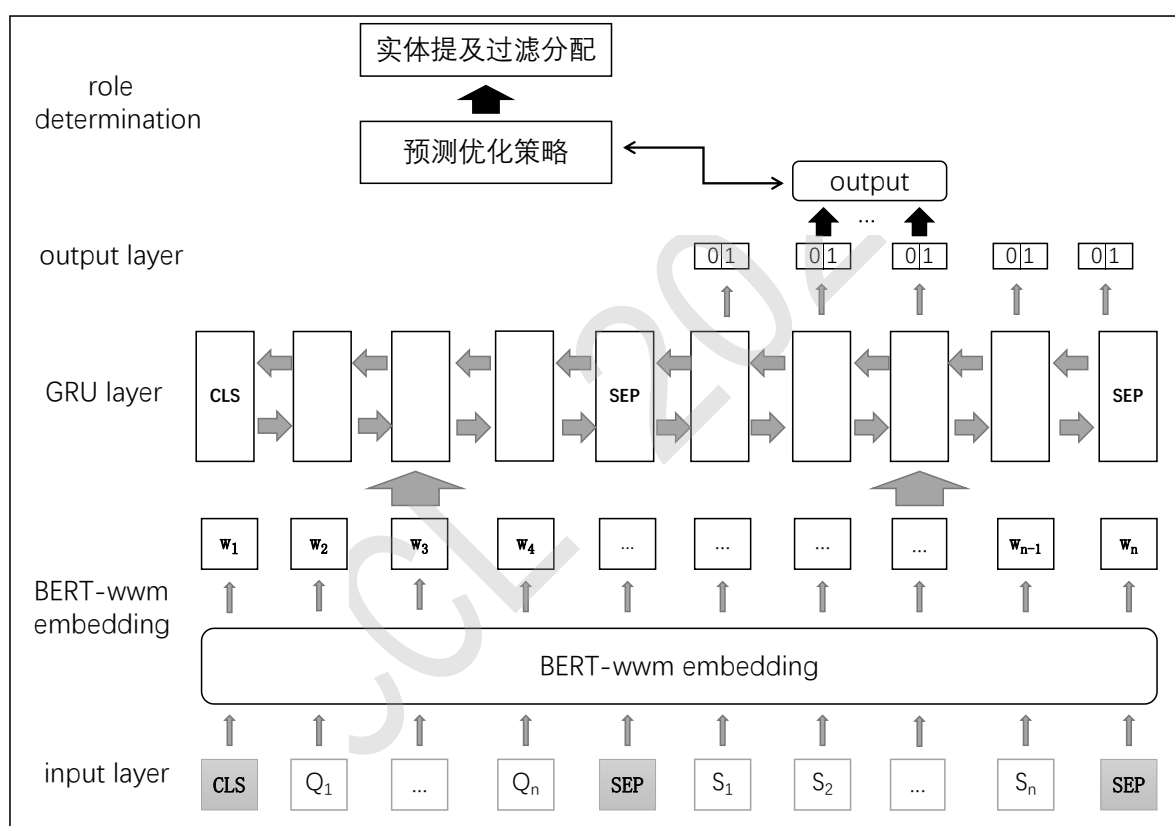


图 2. 基于阅读理解框架的论元抽取模型

主要包含四个部分：1) 输入层，2) 编码层，3) 跨度预测层，4) 论元分配层。其中，输入层是按照机器阅读理解的设置，利用本文采用的语料中的事件模式信息生成具有论元表征的问题和原句子作为初始输入表示；编码层通过BERT预训练模型编码字级别特征，利用双向GRU学习序列特征；跨度预测层根据编码层的输出，对每个字进行二分类来确定答案的跨度；论元分配层利用实体提及过滤抽取结果，最后给实体提及分配论元角色。

3.1 模型输入层

BERT模型的输入序列为句子对所对应的embeddings。句子对包含问题和事件句，并有特殊分隔符“[SEP]”分隔。问题由具有论元表征的论元角色标签构成，事件句是触发词抽取结果中包含事件的文本。同BERT的其他下游任务一样，所有的输入序列的第一个token必须为特殊分类嵌入符“[CLS]”，同时输入序列为字向量、位置向量和句子向量之和。模型的具体输入形式如

下:

[CLS]...Question...[SEP]...Sentence...[SEP]

其中，问题表示的语义信息是很重要的，因为该方法中问题编码了关于论元角色标签的先验知识，并对最终结果有重大影响。本文利用事件模式信息，统计触发词对应事件类型存在的论元角色(这种对应是已知且确定的)，试验了不同问题构成的效果。以3种事件类型为例，事件模式信息如表1所示，不同的问题模板如表2所示，其中Time-*表示与时间相关的论元角色，包括Time-Within、Time-Ending、Time-Starting等。

以受伤事件类型对应的施事者 (Agent) 角色为例，模板1 (伪问题) 使用论元角色作为问题，问题设置为“施事者”；模板2 (触发词+伪问题) 的加入触发词信息，句子中的触发词可以表示触发词信息和触发词位置特征，这也是模型可以学习到的重要特征，问题设置为“触发词是[Trigger]的施事者” (其中[Trigger]表示该事件类型对应的触发词)；模板3 (触发词+自然问题) 利用ACE2005中文语料库中的注释信息，根据事件类型和论元角色生成了更自然的问题 (完整的注释问题参见附录A: 注释问题)，施事者 (Agent) 这一角色在受伤类型下扮演的是该事件下造成伤害的人，问题设置为“触发词是[Trigger]，造成伤害的人是谁？”。本文的实验验证了模板3的问题设置是最合理的。

| 事件类型 | 论元角色 |
|------------|--------------------------------------|
| 受伤(Injure) | Agent Victim Instrument Time-* Place |
| 结婚(Marry) | Person Time-* Place |
| 攻击(Attack) | Agent Target Instrument Time-* Place |
| ... | ... |

表 1. 事件模式表

| 论元角色 | 模板1 (伪问题) | 模板2 (触发词+伪问题) | 模板3 (触发词+自然问题) |
|------------|--------------|-------------------|------------------------------|
| Agent | 施事者 | 触发词是[Trigger]的施事者 | 触发词是[Trigger]，造成伤害的人是谁？ |
| Victim | 受害者 | 触发词是[Trigger]的受害者 | 触发词是[Trigger]，受害者是谁？ |
| Instrument | 工具 | 触发词是[Trigger]的工具 | 触发词是[Trigger]，造成伤害的工具或装置是什么？ |
| ... | ... | ... | ... |

表 2. 不同的问题模板 (以Injure事件为例)

3.2 模型编码层

编码层的主体包括BERT和GRU两部分。

BERT在自然语言处理领域具有里程碑的意义。BERT本质上是通过在大量语料的基础上利用自监督学习的方法为每个字或词学习一个好的特征表示。它使用Transformer捕捉语句中的双向关系，使用遮蔽语言模型 (MLM) 和下一句预测的多任务训练目标。MLM是指在训练时在输入语料上随机遮蔽 (mask) 掉一些单词，然后通过的上下文预测该单词，这样的预训练方式能更好的表示语义特征。在谷歌发布的BERT版本中，中文是以字为粒度进行切分，没有考虑到传统NLP中的中文分词。本文采用哈尔滨工业大学发布的改进版本 (BERT-wwm) (Cui et al., 2019)进行编码，将全词Mask的方法应用在了中文中，即对同属于一个词的汉字mask而不是对单个字的mask。同BERT-base一样，该模型采用12个Transformer Encoder堆叠而成的结构，每一层使用12个独立的注意力机制，包含768个隐层向量。注意力层增加多头注意力机制 (Multi-Head Attention)，扩展了模型专注于不同位置的能力。多头注意力模块的计算如下所示。

$$Attention(Q, K, V) = softmax\left(\frac{QK^t}{\sqrt{d_k}}\right)V \quad (1)$$

$$MultiHead(Q, K, V) = [head_1, \dots, head_h]W^O \quad (2)$$

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (3)$$

多头注意力机制用来学习每个字与其他字的依赖关系和上下文语义，然后通过前馈神经网络对attention计算后的输入进行变换，最终得到序列的全局信息。对于给定的输入序列 $X = \{x_1, x_2, \dots, x_n\}$ ，编码层BERT部分的输出最后一层Transformer的隐层向量，表示为 $W = \{w_1, w_2, \dots, w_n\}$ 。为了更好的学习句子上下文的序列特征，将BERT部分的输出再经过一个双向GRU模型，它可以继承BERT的优点，同时捕获序列语义信息，获取序列的长距离依赖。双向GRU分别从正反两个方向对BERT的输出进行编码，各自得到一个隐层输出，前向GRU层表示为 $\overrightarrow{GRU}(W) = [\vec{h}_1, \vec{h}_2, \dots, \vec{h}_n]$ ，后向GRU层表示为 $\overleftarrow{GRU}(W) = [\vec{h}_1, \vec{h}_2, \dots, \vec{h}_n]$ 。最终，编码层的输出为两个方向的GRU隐层向量的拼接结果，表示为 $h = [\vec{h}_i, \vec{h}_i]$ 。

3.3 跨度预测层

预测层接收编码层的的隐层向量矩阵，答案跨度的预测主要包括开始位置和结束位置的确定，如果答案为空，把BERT输入层的第一个token“[CLS]”作为正确答案。模型在微调期间需要学习的参数是就是每个token作为答案开始位置（start span）和答案结束位置（end span）的向量，隐层向量经过softmax归一化后进行多个二分类，来获得每一个token分别作为开始位置和结束位置的概率，采用概率最高的区间作为预测结果。具体的计算如下。

$$P_{start} = softmax(E \cdot T_{start}) \quad (4)$$

$$P_{end} = softmax(E \cdot T_{end}) \quad (5)$$

$$I_{start} = argmax(P_{start}) \quad (6)$$

$$I_{end} = argmax(P_{end}) \quad (7)$$

其中 E ($E \in \mathbb{R}^{n \times d}$, n 为序列的长度, d 为编码层的输出维度)是编码层输出的隐层向量矩阵; T ($T \in \mathbb{R}^{d \times 2}$)即为需要学习的新参数; P ($P \in \mathbb{R}^{d \times 2}$)为输出概率; I ($I \in [0, n - 1]$)为输出索引。

实验中采用二类交叉熵作为损失函数，在训练过程中，使用Adam优化器优化模型参数，通过最小化交叉熵损失完成训练调优，二类交叉熵具体计算如下。

$$L = -\frac{1}{N} \sum_{i=1}^N y_i \log(\hat{y}_i) \quad (8)$$

$$L_{all} = L_{start} + L_{end} \quad (9)$$

其中， N 表示序列的长度； y_i 表示样本 i 预测为正的的概率； L_{start} 和 L_{end} 分别为开始位置和结束位置的损失。

3.4 论元分配层

传统的论元抽取都是已知实体提及的，本文的方法也一样，利用已知的实体提及完成论元角色分配工作。如果一个问题预测结果与标准答案相同，则该答案的实体被正确分配角色；如果预测出的实体与答案中不同，则该实体认为错误分配了问题中的角色；如果预测的结果没有与实体匹配，判定该句中没有这个角色。

此外，该部分增加了优化策略，用以解决实体不完全匹配的问题。根据标准结果，匹配特定长度的相同开头或结尾的最长实体作为优化后的抽取结果。例如，在标准结果中的实体提及为“26岁”、“人”，预测结果分别为“26岁的时候”、“全家人”。这样的抽取结果也可以判定为正确的抽取。经过预测优化策略后，再根据实体分配不同的论元角色，最终提高论元抽取的性能。

4 实验

4.1 实验数据与评价方法

本文实验基于ACE2005 中文语料库，包含新闻专线、广播、微博等数据。每条数据包含触发词、实体、论元角色标签等标注信息。本文采用文献 (Zhu et al., 2015; He and Duan, 2019) 相同的数据划分方法，从语料库中随机抽取567篇文档作为训练集，66 篇文档作为测试集，并保留训练集中的33 篇文档作为开发集。评判的标准同前人工作一样，一个论元被正确识别当且仅当该论元在文本的位置和类型与标准标注文档中的候选论元的位置和类型完全匹配。采用精确率(P)，召回率(R)，F1值作为本文的评价指标，具体计算如下。

$$P = \frac{TP}{TP + FP} \quad (10)$$

$$R = \frac{TP}{TP + FN} \quad (11)$$

$$F1 = \frac{2 \times P \times R}{P + R} \quad (12)$$

其中，TP为担任角色的实体被正确识别出的个数，FP为角色为None的实体被错误识别的个数，FN为担任角色的实体被错误识别的个数。

4.2 实验参数设置

本文采用哈工大版本的BERT预训练模型 (BERT-wwm) ，其参数字向量维度为768，Transformer层数为12，实验的相关参数设置如表3所示。

| 参数名 | 参数值 |
|--------------|------|
| BERT-wwm基础参数 | 默认值 |
| GRU隐层维度 | 500 |
| 句子最大长度 | 400 |
| 批处理大小 | 8 |
| 训练轮数 | 5 |
| 学习率 | 1e-5 |

表 3. 实验参数设置表

4.3 实验结果

本文工作是针对论元抽取任务，触发词抽取不是重点工作。论元抽取的工作是基于触发词抽取的结果来做，本文的触发词抽取模型利用BERT微调 Xiangyu et al. (2019)的结果，其事件类型分类的精确率(P)为73.9%，召回率(R)为63.8%，F1值为68.5%。

本文主要进行了两组实验对比，一是将本文提出的方法与基准系统进行对比实验，二是设置不同问题策略的对比实验。

4.3.1 与基准系统相比

本文将提出的基于阅读理解框架的方法与现有的论元抽取方法进行了对比。结果如表4所示。

- Rich-C: Chen and Ng (2012)提出的基于特征的模型，该模型针对中文的特殊性开发了一些手工特征，以共同提取事件触发词和论元角色。

- JRNN: Nguyen et al. (2016)提出的一种基于神经网络的模型。它利用双向RNN和手动设计的特征来实现论元抽取。

- C-BiLSTM: Zeng et al. (2016)提出的一种结合LSTM和CNN的卷积双向LSTM神经网络来捕获句级和词汇信息，把论元抽取看成多分类任务的方法。

• MTL-CRF: He and Duan (2019)提出的基于CRF的方法,设计了一个有效挖掘不同事件之间论元相互关系的多任务学习的序列标注模型,联合标注触发词和论元,降低了管道模型带来的级联错误,并没有利用复杂的神经网络,其精确率有明显的提升,但召回率较低。

• DMBERT: Wang et al. (2019a)提出的有效利用预先训练语言模型的方法并使用动态多池化方法来聚合特征。它不同与DMCNN的是利用BERT提取字级别信息和句子信息,获得了较大的性能提升。本文复现了该模型,作为BERT基准。为了公平比较,触发词抽取部分沿用本文的触发词基准结果。

• MRC-EAE: 即本文提出的基于BERT并结合双向GRU的阅读理解模型,本文把传统的论元抽取任务建模成SQUAD风格的机器阅读理解任务,使用了BERT编码问题和句子信息,利用了论元角色的先验信息,同时使用GRU学习句子序列特征。

| 实验系统 | 论元识别(%) | | | 论元分类(%) | | |
|---------------|---------|------|------|-------------|-------------|-------------|
| | P | R | F1 | P | R | F1 |
| Rich-C | 57.3 | 43.6 | 49.5 | 51.6 | 39.2 | 44.6 |
| JRNN | 49.6 | 53.2 | 51.3 | 43.1 | 45.6 | 44.3 |
| C-BiLSTM | 53.0 | 52.2 | 52.6 | 47.3 | 46.6 | 46.9 |
| MTL-CRF | 70.4 | 48.6 | 57.4 | 66.6 | 44.1 | 53.1 |
| DMBERT | 65.2 | 51.0 | 57.3 | 59.1 | 47.8 | 52.8 |
| MRC-EAE(ours) | 59.4 | 55.1 | 57.2 | 56.5 | 52.3 | 54.4 |

表 4. 论元抽取实验结果

从表4中的实验结果可以看出,本文提出的基于阅读理解框架并结合双向GRU的方法优于其他方法。对比多任务学习的序列标注方法MTL-CRF和基于BERT的动态多池化模型DMBERT,本文提出的方法在召回率和F1值上有明显提升,召回率分别提升了8.2%和4.5%,F1值分别提升了1.3%和1.6%。传统的MTL-CRF方法联合抽取触发词和论元,虽然可以降低级联错误,但是这种联合训练的序列标注增加了很多标签,致使类别稀疏,导致召回率较低。同样,在多分类任务DMBERT中,对于有些论元角色较少的类别很难被识别出。而本文提出的方法利用BERT和双向GRU编码,BERT的多头注意力机制和两句输入能充分获取输入文本的语义信息,将问题和句子之间的语义关系充分捕捉,并在句子中获取最终的答案位置。这种阅读理解的方法能够通过问题编码了论元角色的先验信息,这是以往工作中没有利用的重要特征。由于引入论元角色的先验信息,可以有效的识别出角色较少的但是标签有语义区分的类别,如交通工具(Vehicle)、原告(Plaintiff)、卖方(Seller)等,表5给出了5个低频论元在DMBERT和本文方法的结果对比,从结果可以看出,本文提出的方法在这几种少类别的角色标签上有明显的提升效果,更加验证了该方法的有效性。

| 实验系统 | 低频论元(测试集类别数目)F1(%) | | | | | all |
|--------|--------------------|-------------|---------------|--------------|-----------|------|
| | Adjudicator(19) | Vehicle(12) | Plaintiff(12) | Sentence(10) | Seller(2) | - |
| DMBERT | 45.7 | 47.1 | 13.3 | 41.7 | 0.0 | 39.1 |
| ours | 52.6 | 50.0 | 42.1 | 59.2 | 40.0 | 51.2 |

表 5. 低频论元角色对比结果

4.3.2 阅读理解方式不同策略的对比

为了验证编码不同论元角色标签的先验信息对模型的影响,本文设置了不同问题模板进行了消融研究,不同的问题模板设置在第二节给出。实验对比结果如表6所示。

模板1的问题设置仅代表论元角色的语义,在多事件类型的句子中,模型不能正确抽取对应事件类型的论元;模板2的问题设置方式加入了触发词,可以表示句中需要抽取论元具体的触发词语义和触发词的位置信息,但对于论元的描述不够具体;模板3生成了更自然的问题,这种提问策略加入触发词信息的同时融合事件类型信息和论元角色先验信息。从表中实验结果可以发现,性能最好的问题模板3相比模板1和模板2在F1值上分别提升了3.2%和1.7%。当模板3的问

| 策略 | 论元识别(%) | | | 论元分类(%) | | |
|--------------------------|---------|------|------|-------------|-------------|-------------|
| | P | R | F1 | P | R | F1 |
| 模板1 | 57.4 | 51.0 | 54.0 | 53.3 | 47.4 | 50.2 |
| 模板2 | 59.1 | 53.0 | 55.9 | 54.7 | 49.0 | 51.7 |
| 模板3 | 60.2 | 53.2 | 56.5 | 57.0 | 50.3 | 53.4 |
| 模板3 _{触发词} | 58.5 | 51.7 | 54.9 | 54.6 | 48.3 | 51.2 |
| 模板3 _{optimized} | 58.7 | 55.7 | 57.1 | 55.6 | 52.7 | 54.1 |
| Final | 59.4 | 55.1 | 57.2 | 56.5 | 52.3 | 54.4 |

表 6. 不同的策略对比结果

题设置去掉触发词时，性能下降了2.2%，这说明触发词信息的加入可以有效的判断答案的位置和与触发词关系更紧密的论元。此外，在模板3的基础上，对抽取的结果进行优化，在F1值上能提升0.7%；同时利用双向GRU的双向学习序列信息的能力，更好学习输入中问题和句子上下文的关系，在结果优化的基础上F1值提升了0.3%。

4.3.3 错误分析

对实验结果进一步分析发现，本文提出的方法仍然存在不足之处。一方面，本文利用的事件模式信息，存在某些事件句缺失论元角色的情况，即有的问题的答案为空，这种情况下模型往往会被错误预测。例如“法官随即判被告7年预防性监禁。”这一句中并不包含时间相关论元，但是实体提及“7年”会被模型误认为是时间的角色。另一方面，如果一个事件句中某个事件类型存在多个相同的论元角色，受限于本文阅读理解模型的设置，只能识别出其中的一个作为正确答案。例如“而就在吕传升接受记者访问的时候，突然接到了吕秀莲打来的电话，要吕传升暂时封口。”，句中包含2个会面对象 (Entity) ——“吕传升”和“记者”，模型往往只能学习到“吕传升”这个论元而“记者”被忽略。

5 结论

本文采用的基于阅读理解模型的论元抽取方法，将该任务形式化为回答不同的问题来实现不同论元角色的识别和分配，通过优化问题的质量来提升问题回答的性能。通过反复实验证明，这种完型填空式的抽取方式相比基准模型有了明显的提升，也能适用于事件抽取任务上。然而，本文的工作是基于句子级别的论元抽取，缺失了段落信息的句子往往丢失了很多重要的上下文信息。在下一步的研究工作中，还可以考虑基于篇章层面的阅读理解方式来提升论元抽取的效果。

参考文献

- Chen Chen and Vincent Ng. 2012. Joint modeling for Chinese event extraction with rich linguistic features. In *Proceedings of the 24th International Conference on Computational Linguistics*, pages 529–544.
- Yubo Chen, Liheng Xu, Kang Liu, Daojian Zeng, and Jun Zhao. 2015. Event extraction via dynamic multi-pooling convolutional neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 167–176.
- Yubo Chen, Shulin Liu, Xiang Zhang, Kang Liu, and Jun Zhao. 2017. Automatically labeled data generation for large scale event extraction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 409–419.
- Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, Ziqing Yang, Shijin Wang, and Guoping Hu. 2019. Pre-training with whole word masking for Chinese bert. *arXiv preprint arXiv:1906.08101*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.

- Ning Ding, Ziran Li, Zhiyuan Liu, Haitao Zheng, and Zibo Lin. 2019. Event detection with trigger-aware lattice neural network. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 347–356.
- Xiaocheng Feng, Bing Qin, and Ting Liu. 2018. A language-independent neural network for event detection. *Science China Information Sciences*, 61(9):092106.
- Matt Gardner, Jonathan Berant, Hannaneh Hajishirzi, Alon Talmor, and Sewon Min. 2019. Question answering is a format; when is it useful? *arXiv preprint arXiv:1909.11291*.
- Ruifang He and Shaoyang Duan. 2019. Joint Chinese event extraction based multi-task learning. *Journal of Software*, 4:1015–1030.
- Yu Hong, Jianfeng Zhang, Bin Ma, Jianmin Yao, Guodong Zhou, and Qiaoming Zhu. 2011. Using cross-entity inference to improve event extraction. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 1127–1136.
- Lifu Huang, Heng Ji, Kyunghyun Cho, Ido Dagan, Sebastian Riedel, and Clare Voss. 2018. Zero-shot transfer learning for event extraction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2160–2170.
- Peifeng Li and Guodong Zhou. 2012. Employing morphological structures and sememes for Chinese event extraction. In *Proceedings of the 24th International Conference on Computational Linguistics*, pages 1619–1634.
- Qi Li, Heng Ji, and Liang Huang. 2013. Joint event extraction via structured prediction with global features. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 73–82.
- Xiaoya Li, Fan Yin, Zijun Sun, Xiayu Li, Arianna Yuan, Duo Chai, Mingxin Zhou, and Jiwei Li. 2019. Entity-relation extraction as multi-turn question answering. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1340–1350.
- Xiaoya Li, Jingrong Feng, Yuxian Meng, Qinghong Han, Fei Wu, and Jiwei Li. 2020. A unified mrc framework for named entity recognition. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5849–5859.
- Shasha Liao and Ralph Grishman. 2010. Using document level cross-event inference to improve event extraction. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 789–797.
- Hongyu Lin, Yaojie Lu, Xianpei Han, and Le Sun. 2018. Nugget proposal networks for Chinese event detection. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1565–1574.
- Xiao Liu, Zhunchen Luo, and He-Yan Huang. 2018. Jointly multiple events extraction via attention-based graph information aggregation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1247–1256.
- Bryan McCann, Nitish Shirish Keskar, Caiming Xiong, and Richard Socher. 2018. The natural language decathlon: Multitask learning as question answering. *arXiv: Computation and Language*.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 27th Annual Conference on Neural Information Processing Systems*, pages 3111–3119.
- Thien Huu Nguyen, Kyunghyun Cho, and Ralph Grishman. 2016. Joint event extraction via recurrent neural networks. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 300–309.

- Xiaozhi Wang, Xu Han, Zhiyuan Liu, Maosong Sun, and Peng Li. 2019a. Adversarial training for weakly supervised event detection. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 998–1008.
- Xiaozhi Wang, Ziqi Wang, Xu Han, Zhiyuan Liu, Juanzi Li, Peng Li, Maosong Sun, Jie Zhou, and Xiang Ren. 2019b. HMEAE: Hierarchical modular event argument extraction. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5781–5787.
- Xi Xiangyu, Zhang Tong, Ye Wei, Zhang Jinglei, Xie Rui, and Zhang Shikun. 2019. A hybrid character representation for Chinese event detection. In *Proceedings of the 2019 Conference on International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.
- Ying Zeng, Honghui Yang, Yansong Feng, Zheng Wang, and Dongyan Zhao. 2016. A convolution BiLSTM neural network model for Chinese event extraction. In *Proceedings of the Natural Language Understanding and Intelligent Applications*, pages 275–287.
- Shaohua Zhu, Peifeng Li, and Qiaoming Zhu. 2015. A Chinese event argument inference approach based on entity semantics and event relevance. In *Proceedings of Workshop on Chinese Lexical Semantics*, pages 577–586.

附录A.注释问题

| 事件类型 | 论元角色 | 注释问题 | |
|--------------------|-------------|--------------------|-------------|
| declare-bankruptcy | Org | 宣布破产的组织或公司是什么? | |
| | Time | 宣布破产的时间是什么时候? | |
| | Place | 宣布破产的地点在哪里? | |
| end-org | Org | 结束的组织或公司是什么? | |
| | Time | 结束组织或公司的时间是什么时候? | |
| | Place | 结束组织或公司的地点在哪里? | |
| merge-org | Org | 合并的组织或公司是什么? | |
| | Time | 合并组织或公司的时间是什么时候? | |
| | Place | 合并组织或公司发生的地点在哪里? | |
| start-org | Org | 成立的组织或公司是什么? | |
| | Agent | 组织的建立者是谁? | |
| | Time | 成立组织或公司的时间是什么时候? | |
| | Place | 成立组织或公司的地点在哪里? | |
| attack | Target | 攻击目标 (包括意外目标) 是什么? | |
| | Attacker | 攻击者是谁? | |
| | Time | 攻击发生的时间是什么时候? | |
| | Place | 攻击发生的地点在哪里? | |
| Instrument | Instrument | 攻击中使用的工具是什么? | |
| | demonstrate | Entity | 示威游行的人或组织是? |
| | Place | 游行的地点在哪里? | |
| | Time | 游行的时间是什么时候? | |
| meet | Entity | 会面的人或组织是? | |
| | Place | 会面的地点在哪里? | |
| | Time | 会面的时间是什么时候? | |
| phone-write | Entity | 打电话或写信交流的人或组织是? | |
| | Place | 打电话或写信交流的地点在哪里? | |
| | Time | 打电话或写信交流的时间是什么时候? | |

| | | |
|----------------|---|---|
| acquit | Defendant Place Adjudicator | 被无罪释放的是谁? 无罪释放的地点在哪里? 法官或法院? |
| appeal | Plaintiff Adjudicator Time Place | 上诉的原告是谁? 上诉的被告人是谁? 上诉发生的时间是什么时候? 上诉的地点在哪里? |
| arrest-jail | Person Agent Place Time Crime | 被逮捕或被监禁的人是谁? 实施逮捕的人是谁? 逮捕发生的地点在哪里? 逮捕的时间是什么时候? 被逮捕人犯的罪行是什么? |
| charge-indict | Crime Defendant Prosecutor Adjudicator Time | 指控或起诉的罪行是什么? 被指控或起诉的人是谁? 起诉人或法官是谁? 审裁官是谁? 指控的时间是什么时候? |
| convict | Defendant Adjudicator Crime | 被定罪的人是谁? 法官或法院? 被判的罪行是什么? |
| execute | Person Agent Crime | 被处决的人是谁? 负责执行死刑的代理是? 被判的罪行是什么? |
| extradite | Person Origin Destination Agent | 被引渡的人是谁? 被引渡人的原所在地是哪里? 引渡的目的地是哪里? 引渡的代理人是? |
| fine | Crime Entity Adjudicator Money | 被罚款的罪行是什么? 罚款的实体对象是什么? 执行罚款的对象是? 罚款数额是多少? |
| pardon | Defendant Adjudicator Time Crime | 被赦免的人或公司是? 赦免的官员或法官是? 赦免的时间是什么时候? 赦免的罪行是什么? |
| release-parole | Person Time Entity Crime | 被释放的人是谁? 释放的时间是什么时候? 释放之前的劫持者是谁? 被释放的人先前被关押的罪行是什么? |
| sentence | Defendant Sentence Crime Adjudicator Time | 被判刑的人是谁? 宣判的判决结果是什么? 判决的罪行是什么? 法官或法院? 判决发生的时间是什么时候? |
| sue | Plaintiff Defendant Adjudicator | 起诉代理,原告是谁? 被指控或起诉的人是谁? 法官或法院? |

| | | |
|----------------|---|---|
| | Time Prosecutor | 诉讼发生的时间是什么时候? 检察官是谁? |
| trial-hearing | Defendant Adjudicator Time Crime Place | 审判的被告人是谁? 法官或法院? 审判的时间是什么时候? 审判的罪行是什么? 审判的地点在哪里? |
| be-born | Person Place Time | 出生的人是谁? 出生的地点在哪里? 出生的时间是什么时候? |
| die | Victim Agent Time Place Instrument | 受害者是谁? 致人死亡的人是谁? 死亡的时间是什么时候? 死亡的地点在哪里? 致人死亡的装置是什么? |
| divorce | Person Place Time | 离婚的人是谁? 离婚的地点在哪里? 离婚的时间是什么时候? |
| Injure | Agent Victim Instrument Place Time | 进攻者, 造成伤害的人是谁? 受害者是谁? 造成伤害的装置是什么? 受伤的地点在哪里? 受伤的时间是什么时候? |
| marry | Person Place Time | 结婚的人是谁? 结婚的地点在哪里? 结婚的时间是什么时候? |
| transport | Artifact Destination Time Origin Agent Vehicle | 被运送的人或工件是什么? 运送的目的地在哪里? 运送发生的时间是什么时候? 运送的起源地在哪里? 负责移动事件的代理是? 使用的交通工具是什么? |
| elect | Person Time Position Entity | 被选举的人是谁? 选举的时间是什么时候? 选举的职位是什么? 投票代理是? |
| end-position | Person Time Position Entity | 结束任职的人是谁? 卸任的时间是什么时候? 卸任的职位是什么? 雇主是谁? |
| nominate | Person Position Agent Place | 结束任职的人是谁? 卸任的职位是什么? 雇主是谁? 卸任的地点在哪里? |
| start-position | Person Position Time | 被提名的人是谁? 被提名担任的职位是什么? 提名的时间是什么时候? |

| | | |
|--------------------|-----------------|-----------------------|
| | Entity Place | 任职实体是什么？ 提名的地点在哪里？ |
| transfer-money | Time | 钱财转移的时间是什么时候？ |
| | Beneficiary | 在钱财转移中获益的是谁？ |
| | Place | 钱财转移发生的地点在哪里？ |
| | Recipient | 钱财交易的接受者是谁？ |
| | Giver | 捐赠人或公司是什么？ |
| | Money | 给予/捐赠/贷款的金额是多少？ |
| transfer-ownership | Artifact | 进行交易的物品或组织是什么？ |
| | Seller | 销售的代理是？ |
| | Beneficiary | 在交易中获益的人或组织是谁？ |
| | Time | 交易发生的时间是什么时候？ |
| | Buyer | 采购的代理是？ |

JCL 2020