

基于双重注意力机制的事件抽取方法*

朱敏¹, 毛莺池^{1,2}, 程永¹, 陈程军¹, 王龙宝^{1,2}

¹(河海大学 计算机与信息学院, 江苏 南京 211100)

²(水利部水利大数据重点实验室(河海大学), 江苏 南京 211100)

通信作者: 毛莺池, E-mail: yingchimao@hhu.edu.cn



摘要: 针对事件抽取存在未充分利用句法关系、论元角色缺失的情况, 提出了基于双重注意力机制的事件抽取(event extraction based on dual attention mechanism, EEDAM)方法, 有助于提高事件抽取的精确率和召回率。首先, 基于4种嵌入向量进行句子编码, 引入依赖关系, 构建依赖关系图, 使深度神经网络可以充分利用句法关系。然后, 通过图转换注意力网络生成新的依赖弧和聚合节点信息, 捕获长程依赖关系和潜在交互, 加权融合注意力网络, 捕捉句中关键的语义信息, 抽取句子级事件论元, 提升模型预测能力。最后, 利用关键句检测和相似性排序, 进行文档级论元填充。实验结果表明, 采用基于双重注意力机制的事件抽取方法, 在ACE2005数据集上, 较最佳基线联合多中文事件抽取器(joint multiple Chinese event extractor, JMCEE)在精确率、召回率和F1-score分别提高17.82%、4.61%、9.80%; 在大坝安全运行日志数据集上, 较最佳基线JMCEE在精确率、召回率和F1-score上分别提高18.08%、4.41%、9.93%。

关键词: 事件抽取; 双重注意力; 依赖关系; 论元填充; 神经网络

中图法分类号: TP18

中文引用格式: 朱敏, 毛莺池, 程永, 陈程军, 王龙宝. 基于双重注意力机制的事件抽取方法. 软件学报, 2023, 34(7): 3226–3240. <http://www.jos.org.cn/1000-9825/6520.htm>

英文引用格式: Zhu M, Mao YC, Cheng Y, Chen CJ, Wang LB. Event Extraction Method Based on Dual Attention Mechanism. Ruan Jian Xue Bao/Journal of Software, 2023, 34(7): 3226–3240 (in Chinese). <http://www.jos.org.cn/1000-9825/6520.htm>

Event Extraction Method Based on Dual Attention Mechanism

ZHU Min¹, MAO Ying-Chi^{1,2}, CHENG Yong¹, CHEN Cheng-Jun¹, WANG Long-Bao^{1,2}

¹(College of Computer and Information, Hohai University, Nanjing 211100, China)

²(Key Laboratory of Water Big Data Technology of Ministry of Water Resources (Hohai University), Nanjing 211100, China)

Abstract: In view of the fact that the syntactic relationship is not fully utilized and the argument role is missing in event extraction, an event extraction based on dual attention mechanism (EEDAM) method is proposed to improve the accuracy and recall rate of event extraction. Firstly, sentence coding is based on four embedded vectors and dependency relation is introduced to construct dependency relation graph, so that deep neural network can make full use of syntactic relation. Then, through graph transformation attention network, new dependency arcs and aggregate node information are generated to capture long-range dependencies and potential interactions, weighted attention network is integrated to capture key semantic information in sentences, and sentence level event arguments are extracted to improve the prediction ability of the model. Finally, the key sentence detection and similarity ranking are used to fill in the document level arguments. The experimental results show that the event extraction method based on dual attention mechanism can improve the accuracy rate, recall rate, and F1-score by 17.82%, 4.61%, and 9.80% respectively compared with the optimal baseline joint multiple Chinese event extractor (JMCEE) on ACE2005 data set. On the data set of dam safety operation records, the accuracy, recall rate, and F1

* 基金项目: 国家重点研发计划(2018YFC0407105); 江苏省重点研发计划(BE2020729); 江苏省研究生科研创新项目(B200203130); 华能集团总部科技项目(HNKJ19-H12)

收稿时间: 2021-06-05; 修改时间: 2021-08-05; 采用时间: 2021-10-25; jos 在线出版时间: 2022-05-24

CNKI 网络首发时间: 2022-11-15

score are 18.08%, 4.41%, and 9.93% higher than the optimal baseline JMCEE, respectively.

Key words: event extraction; double attention; dependency; argument filling; neural network

互联网发展带来数据爆炸式增长, 如何从互联网大数据中获取想要的信息已成为自然语言处理领域的研究热点. 事件抽取是信息抽取研究中最具挑战性的任务之一, 其目标是根据不同事件类型预定义的事件架构, 从句中抽取相应的事件论元角色, 它依赖于词汇特征和句法关系. 但现有的神经网络无法充分利用句法关系, 且事件信息通常分散在文档的多个句子中, 易出现论元角色缺失的现象, 致使事件抽取的精确率和召回率远低于事件检测. 因此, 提高事件抽取的精确率和召回率具有重要的研究意义.

现有的卷积神经网络和递归神经网络等深度学习网络已被用于生成低维向量, 以自动表示文本语义信息, 并基于这些语义向量进一步提取事件论元. 虽然深度神经网络可以自动学习底层特征, 但没有充分利用句法关系, 引入依赖关系可有效增强单词的信息表示能力和捕获长程依赖关系, 以识别和分类事件的所有参数^[1,2]. 图卷积神经网络 (graph convolutional network, GCN)^[3]常用于表示依赖关系, 但它缺乏对依赖树中的长程依赖或断开连接词的建模, 可通过引入注意力机制来学习具有不同句法距离的单词之间的依赖关系, 以缓解上述问题^[4,5]. 注意力机制还可以提供句子的语义信息^[6]. 虽然上述方法最终解决了句法关系问题, 但是它们的缺点是遗漏了论元角色分类时出现的角色缺失. 可通过检测事件中关键句子, 并用相邻句子中的填充词来补充缺失的事件角色^[7]或者构建事件区域图并利用 GCN 聚合事件信息的方式^[8], 以提升论元角色分类的精确率和召回率.

针对事件抽取存在无法充分利用句法关系和论元角色缺失的问题, 本文提出了一种基于双重注意力机制的事件抽取 (event extraction based on dual attention mechanism, EEDAM) 方法. 将事件抽取转化成序列标注任务, 以抽取事件论元角色和填充缺失部分. 首先, 为了充分利用句法关系, 将构建依赖关系图. 其次, 为了提高事件论元角色抽取的精确率和召回率, 将构建双重注意力网络. 最后, 为了提升论元分类的准确率, 引入关键句检测和相似性排序.

本文的主要贡献包括:

(1) 构建依赖关系图. 借助 4 种嵌入向量进行句子编码, 增强词汇特征和语义信息, 同时引入依赖关系, 辅助模型挖掘句法关系, 使深度神经网络可以充分利用单词的信息表示, 提高事件抽取预测能力.

(2) 构建双重注意力网络. 基于图转换注意力网络, 使用多个候选邻接矩阵来寻找依赖关系图中新的图结构, 弥补现有工具的不足之处, 并加权融合注意力网络, 抽取句子级事件论元, 以提高事件论元角色抽取的精确率和召回率.

(3) 文档级论元填充. 利用关键句检测和相似性排序, 进行文档级论元填充, 提升论元分类的准确性.

本文第 1 节介绍相关工作, 讨论分析目前事件抽取任务的研究进展及其优缺点. 第 2 节概述事件抽取的总体框架. 第 3 节具体描述依赖关系图和双重注意力网络的构建方法以及论元填充方法. 第 4 节介绍本文的实验准备、实验结果与分析, 验证本文方法的有效性. 第 5 节对全文进行总结.

1 相关工作

传统的事件抽取很大程度上依赖词汇和句法特征, 需要大量的人工工程, 并且模型通用性不强. Chen 等人^[9]提出一种动态多池化卷积神经网络 (dynamic multi-pooling convolutional neural networks, DMCNN), 利用动态多池层根据事件触发器和参数使卷积神经网络 (convolutional neural networks, CNN) 保留更重要的信息, 避免了误差传播问题. Zeng 等人^[10]通过卷积双向长短记忆网络 (convolution-BiLSTM, C-BiLSTM), 同时捕获句子和词汇信息. 但是, 绝大多数神经网络模型将每个参数角色独立分类, 且无法充分利用句法关系. Wang 等人^[11]考虑了角色间概念上的相关性, 用分层模块化事件参数提取 (hierarchical modular event argument extraction, HMEAE) 模型将具有逻辑运算的相关单元模块按层次组合成一个面向角色的模块网络, 并对特定的论元角色进行分类. Sha 等人^[1]通过引入依赖关系来辅助模型挖掘句法关系, 增强单词信息表示, 捕捉事件论元之间类型的潜在交互. Balali 等人^[2]通过在依赖图中引入最短依赖路径来同时提取多个事件触发器和参数, 以及消除句中不相关词来捕获长程依赖

关系. 这些方法都采用图卷积网络表示依赖关系, 然而, 它们缺乏对依赖树中的长程依赖或断开连接词的建模. Veysch 等人^[12]认为上述方法未能利用句子的语义结构来诱导有效的事件抽取表征, 可通过图变换网络 (graph transformer networks, GTNs) 学习更有效的句法和语义结构. 目前事件抽取工作基本围绕句子级展开^[8], 事件信息通常分散在文档中的多个句子中, 导致常出现论元缺失问题. Yang 等人^[7]通过检测事件中关键句并利用相邻句子中的填充词来补充缺失的事件角色. 但是, 该方法存在关键句选择错误的问题, 且补充策略是粗粒度的, 不能单独考虑每个候选填充符. Chen 等人^[8]以候选填充符为节点构建图, 将候选填充符之间的边表征为丰富的向量, 从而识别和聚合事件信息. 但是, 该方法会产生许多嘈杂的候选填充符提取器.

通过对传统事件抽取的研究发现, 中文事件抽取是信息抽取领域一项具有挑战性的工作. 卷积双向长短记忆网络是第一个处理中文事件抽取中语言特征问题的模型, 可同时捕获句子和词汇信息^[8]. Yan 等人^[13]首次提出了一种基于图的联合模型, 将中文分词与依存句法分析相联合, 同时结合了神经网络和注意力机制, 并用图卷积学习向量表示, 展示了基于图的模型的优势. 然而, 该方法不能为下游任务提供词性标注信息, 并且基于转换模型的研究证明词性标注任务有助于依存句法分析. 中文字符可提供一些神经网络无法挖掘到的信息. Wu 等人^[14]利用注意力机制学习词与字符联合嵌入的语义特征, 从句中获得更多的单词信息. Liu 等人^[15]通过引入语法快捷弧来增强信息流, 并基于注意力的图卷积网络来建模图信息, 从而联合抽取多个事件触发器和论元. Nguyen 等人^[16]利用事件触发器和论元角色之间的相互依赖性, 提出双向递归神经网络的联合框架 (joint recurrent neural networks, JRNN) 来进行事件抽取. Xu 等人^[17]为解决中文事件抽取中角色重叠问题, 提出中文多事件联合抽取框架, 该框架基于预先训练的语言模型的共享特征表示, 对事件触发器和论元进行联合预测. 以上方法均无法充分利用句法关系, 同时中文分词会导致依存分析受到错误传播的影响.

综上所述, 句法关系在事件抽取中起着关键作用, 引入依赖关系来挖掘句法关系可提高事件抽取的精确率和召回率. 但是, 中文依赖关系生成方面存在一些问题, 同时事件抽取常出现论元缺失问题. 因此, 本文提出基于双重注意力机制的事件抽取 (EEDAM), 借助依赖关系增强信息表示能力, 利用双重注意力网络捕捉句中关键语义信息, 抽取句子级事件论元, 通过关键句子检测并利用相邻句子中的填充词来补充缺失的事件论元角色.

2 总体框架

事件抽取的目标是识别并分类事件论元角色. 目前, 事件抽取存在未充分利用句法关系、论元角色缺失的情况, 致使事件抽取的精确率和召回率较低. 因此, 提高事件抽取的精确率和召回率具有重要的研究意义. 本文提出了基于双重注意力机制的事件抽取 (EEDAM) 框架, 由事件论元角色抽取和文档级论元填充两部分组成, 分别处理上述两种情况. 事件论元角色抽取部分通过双重注意力机制充分利用句法关系, 捕捉句中关键的语义信息, 抽取句子级事件论元, 提高模型精确率和召回率. 论元填充部分则填补缺失的论元角色, 提高论元分类的准确性.

EEDAM 模型框架如图 1 所示. EEDAM 框架由嵌入层、双向长短记忆层、特征提取层、融合门层、条件随机场层、输入层和论元填充网络层 7 部分组成. 其中, 嵌入层中包含 4 种嵌入向量, 可动态学习上下文内容, 获得更丰富的语义信息, 解决传统词嵌入方法无法表示一词多义的难点. 双向长短记忆层用于捕获上下文信息, 捕捉单词中有意义的语义规律. 特征提取层中图转换注意网络使用多个候选邻接矩阵, 寻找依赖关系图中新的依赖弧, 挖掘更丰富的特征信息. 融合门层用于加权融合图转换注意网络和注意力网络, 生成最终的特征向量. 条件随机场层通过添加约束条件, 以保证最后预测的标签是合法的. 输入层和论元填充网络层则用于文档级论元填充, 该部分输入论元标签、实体类型、句子和文档 4 种嵌入向量. 论元标签数量可直接决定句子的重要性程度. 实体类型的数量在判断过程中起辅助作用, 句子可提供语义信息, 文档可提供语境信息, 辅助判断句子的重要性程度. 综合上述信息, 采用文本卷积神经网络 (text convolutional neural networks, TextCNN)^[18]判别关键句, 并利用孪生神经网络将周围句子中最相似的论元角色, 填充到关键事件缺失的部分.

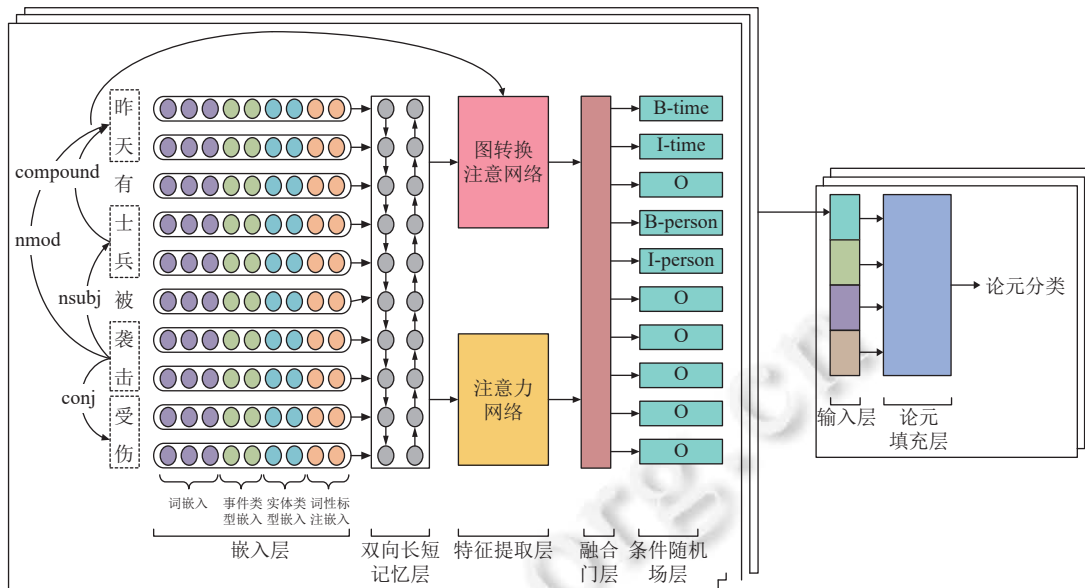


图 1 基于双重注意力机制的事件抽取框架

3 事件抽取框架描述

3.1 构建依赖关系图

本节首先利用 4 种嵌入向量进行句子编码, 增强词汇特征, 然后引入依赖关系, 借助句子结构和语义结构建立依赖关系图。

3.1.1 句子编码

事件抽取可公式化为一个序列标注任务, 通常采用 BIO (B: Begin, I: Inside, O: Outside) 模式标注数据. 将每个元素标记为 B-X、I-X 或 O. 其中, B-X 表示属于 X 类型的关键论元的开始部分, I-X 表示属于 X 类型的关键论元中间部分, O 表示句中除关键论元外其他单词. 例如, 将 X 表示为名词短语 (noun phrase, NP), 则 BIO 模式的 3 个标记为: B-NP 表示名词短语的开始, I-NP 表示名词短语的中间, O 表示不是名词短语. 事件“普京去年在德克萨斯州访问布什。”的 BIO 模式数据标注如图 2 所示.

```

普_B-NP 京_I-NP 去_B-NP 年_I-NP 在_O 德_B-NP 克_I-NP 萨_I-NP 斯_I-NP 州_I-NP 访_O 问_O
布_B-NP 什_I-NP。_O

```

图 2 “普京去年在德克萨斯州访问布什。”的 BIO 模式数据标注

事件抽取任务是识别除 O 以外的关键论元. 输入向量多为句子级, 设为 $W = w_1, w_2, \dots, w_N$, 其中, N 为句子长度. 句子编码主要由词嵌入、事件类型嵌入、实体类型嵌入和词性标注嵌入 4 部分向量构成. 实体类型、词性标注等信息可增强嵌入向量的表示能力^[19], 辅助 EEDAM 提取关键信息. 词嵌入设置每个句子最大长度为 256 维. Devlin 等人^[20]认为双向编码器表示 (bidirectional encoder representation from Transformers, BERT) 采用双向的语言模型, 较其他单向模型, 很好地解决了单向性约束问题. 因此, 本文采用轻量化的双向编码器表示 (a lite bidirectional encoder representation from Transformers, ALBERT)^[21]模型最后一层 312 维向量作为嵌入向量. ALBERT 比 BERT 训练参数更少, 训练速度更快, 可更好表达句子语义信息, 缓解一词多义问题. 事件类型将多事件并存的句子划分成多对数据, 其嵌入向量由可训练查找表生成, 用于标志句中唯一事件类型, 以解决角色重叠问题. 实体类型和词性标注嵌入向量也通过可训练查找表生成, 都有助于模型过滤次要信息. 最后通过双向长短记忆网络 (bi-directional long short-term memory, BiLSTM)^[22]挖掘串联后的嵌入向量, 捕获上下文重要信息, 得到隐藏向量序列

$H = h_1, \dots, h_N$, 并作为下一步表示向量.

3.1.2 依赖关系图建立

EEDAM 根据依存句法分析生成的句子结构和 BiLSTM 生成的语义结构, 构建新的依赖关系图, 采用图转换神经网络做特征提取器, 生成新的表示向量. Sha 等人^[1]证明了依赖树中的依赖关系有助于连接重要的上下文单词, 从而获得有效的事件抽取表示向量. 因此, 本文使用依赖树的邻接矩阵 A^d 作为事件抽取的句法结构之一. 其中, A^d 是一个 $N \times N$ 的二元矩阵, 当单词 w_i 和 w_j 在依赖树中有链接, 则 $A^d(i, j)$ 值设为 1, 否则为 0.

Cui 等人^[23]认为现有方法虽验证了将句法依赖集成到图卷积网络中的有效性. 但是, 其通常忽略依赖标签信息. 因此, 当单词 w_i 和 w_j 之间存在依赖边且依赖标签为 r 时, 本文根据依赖标签初始化矩阵 $A^{dl}(i, j)$, 可以从可训练的嵌入查找表中找到 r 的嵌入向量, 该向量维度为 p . 否则, 用 p 维全零向量初始化 $A^{dl}(i, j)$. 由于依赖关系图本质是分数矩阵, 所以需将依赖标签矩阵 A^{dl} 转化成依赖标签分数矩阵 \tilde{A}^{dl} , 如公式 (1) 所示:

$$\tilde{A}^{dl}(i, j) = \exp(UA^{dl}(i, j)) / \sum_{v=1, \dots, N} \exp(UA^{dl}(i, v)) \quad (1)$$

其中, U 是可训练权重矩阵.

语义结构的目的是利用句子中 w_i 和 w_j 的上下文语义去学习一对词 (w_i, w_j) 的重要性分数, 可为句法结构提供补充信息. 本文通过结合两种结构, 为事件抽取提供有效的表示向量. 在句子编码阶段, 使用 BiLSTM 捕获上下文重要信息, 得到隐藏向量序列 $H = h_1, \dots, h_N$, 以获得本文的语义结构. 通过计算隐藏向量 h_i 和 h_j 之间分数获得语义分数矩阵 A^s , 如公式 (2) 所示:

$$k_i = U_k h_i, q_i = U_q h_i, A^s(i, j) = \exp(k_i q_j) / \sum_{v=1, \dots, N} \exp(k_i q_v) \quad (2)$$

其中, U_k 和 U_q 是可训练权重矩阵.

依赖关系图矩阵 $A = [A^d, \tilde{A}^{dl}, A^s]$ 中的 3 个初始结构, 可解释为 w 中词对之间的 3 种不同类型的关系. 3 种关系共同构成依赖关系图矩阵 A , 以增强单词的信息表示能力, 从而识别和分类事件的所有参数, 提高对事件论元识别的准确率. 每个初始结构中 $cell(i, j)$ 是它们的直接交互或边, 即一跳路径 (w_i, w_j) . 其中, 提供的重要性分数能决定 w_i 和 w_j 之间可能的相互作用. w_i 和 w_j 之间的多跳路径和沿路径的异构边缘类型也有助于事件抽取, 将它们与一跳路径提供的信息相结合, 可提供更丰富的特征向量.

3.2 构建双重注意力网络

本节构建了双重注意力网络. 首先提出图转换注意力网络, 然后生成新的依赖弧和聚合节点信息, 捕获长程依赖关系和潜在交互, 最后引入注意力网络, 将图转换注意力网络层和注意力网络层提取的特征按一定比例融合, 捕捉句中关键的语义信息, 抽取句子级事件论元, 提升模型预测能力.

3.2.1 图转换注意力网络生成

本文采用相对先进的中文依存句法工具 Stanford CoreNLP 生成依赖关系和依赖标签. 但有关中文依存句法研究和工具相对较少, 与英文相比更容易出现依赖弧对象不匹配、依赖弧缺失和依赖标签错误等问题. Yun 等人^[24]提出能够生成新的图结构的图转换网络 GTNs, 与大多数假设图是固定和同构的神经网络不同. GTNs 首先识别原始图中未连接节点之间的有用连接. 然后, 使用多个候选邻接矩阵来寻找新的图结构. 最后, 以端到端的方式学习新图上的有效节点表示. GTNs 在一定程度上可弥补中文依存句法存在的缺陷. 但是, GTNs 仍存在以下 3 个问题.

- (1) 依赖关系中触发器和论元所连接的依赖弧比其他依赖弧的重要性更高.
- (2) GTNs 最后提取特征时采用了 GCN, 该模型无法为每个邻居分配不同的权重, 生成的向量并不是最有效的.
- (3) GTNs 模型的高表征学习能力可能导致记忆仅针对训练数据的信息, 出现过拟合问题.

根据上述问题, 本文提出图转换注意力网络 (graph transformer attention networks, GTANs). 将 GCN 模型替换成图注意力网络 (graph attention network, GAN)^[25], 以解决权重分配问题. 生成的向量会经过 Dropout 层, 防止模型过拟合, 其框架如图 3 所示. GTANs 由图转换层和图注意层两部分组成. 图转换层对矩阵 A 集合采用 1×1 卷积, 软选

择其中两个中间邻接矩阵 Q_1 和 Q_2 , 通过矩阵相乘生成新的元路径图 A^i . 其中, A^i 节点间蓝色线条是初始元路径, 红色线条是新的元路径. 图注意层对元路径图 A^i 每个通道应用 GAN 网络, 并将多个节点表示串联为 Z , 以提高节点分类的性能, 如公式 (3) 所示:

$$Z = \parallel_{i=1}^C \sigma(\bar{D}_i^{-1} \bar{A}_i^{(i)} X V) \quad (3)$$

其中, \parallel 是连接运算符, C 表示通道数量, $\bar{A}_i^{(i)}$ ($\bar{A}_i^{(i)} = A_i^{(i)} + I$) 是 A^i 的第 i 通道的邻接矩阵, \bar{D}_i 是 $\bar{A}_i^{(i)}$ 的度矩阵, V 是一个跨通道共享的可训练权重矩阵, X 是一个特征矩阵. 这里, 依据 GTNs 的实验结果加入单位矩阵 I 助于学习不同长度的多跳路径, 可提高事件抽取的精确率和召回率.

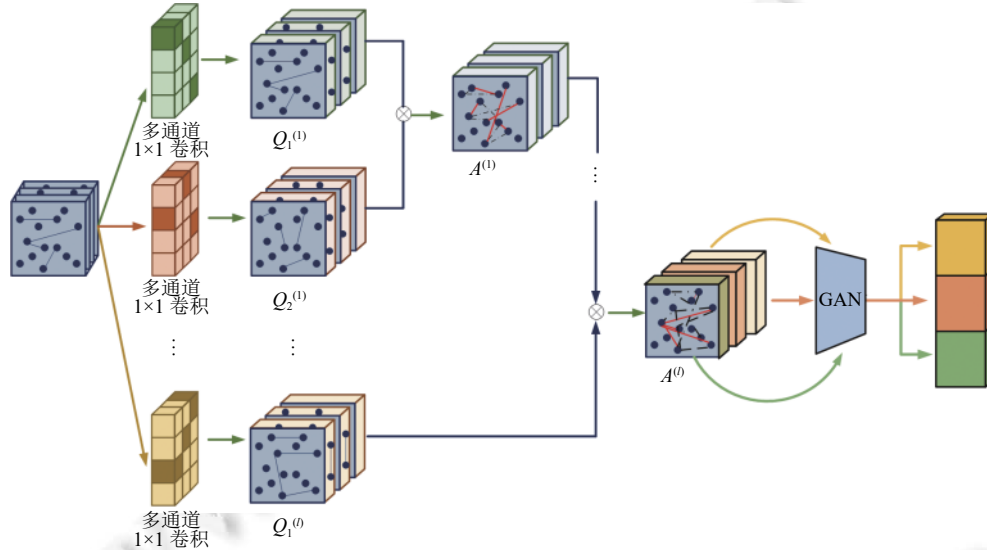


图3 图转换注意网络框架

3.2.2 基于双重注意力的特征融合

图转换注意网络主要从依赖关系图中提取有效特征, 但由于中文依存句法的局限性, 所构建的依赖关系图本身存在一定偏差. 因此, 文本引入双重注意力模式, 将图转换注意网络层和注意力网络层提取的特征按一定比例融合, 相辅相成. 具体方法如下所述.

首先, 将注意力网络层权重矩阵 α_a 点乘隐藏向量 H 生成新的向量 \tilde{H} , 权重矩阵 α_a 的计算公式如公式 (4) 所示. 然后, 用超参数 λ 连接图转换注意网络层生成的 Z 向量和注意力网络层生成的 \tilde{H} 向量, 如公式 (5) 所示. 最后, 通过序列标注的方式进行事件抽取, 利用条件随机场挖掘特征融合向量 \tilde{W} , 以预测每个字符的标签. 其中, 序列标注数据中无用 O 信息过多, 会导致样本不均衡问题. 本文采用 Focal loss 损失函数^[26]和 Adam 优化器解决该问题. 该方法不是本文重点研究内容, 在此不详细叙述.

$$\alpha_a^k = \text{Softmax}(\tanh(W_a^T h_k + b_k)) \quad (4)$$

$$\tilde{W} = \sigma(\lambda \cdot Z + (1 - \lambda) \cdot \tilde{H}) \quad (5)$$

其中, h_k 是 BiLSTM 生成的隐藏向量 H 中的第 k 个向量, W_a 是可训练权重矩阵, b_k 是偏置, σ 是 Sigmoid 函数.

3.3 文档级论元填充

本节先采用 TextCNN 检测文档中具备关键事件的句子, 然后利用孪生神经网络, 将周围句子中相似度最高的论元角色填充到关键事件缺失的部分.

3.3.1 关键事件检测

本文将关键事件检测看作二分类问题, 采用 TextCNN 判断句子是否是关键句子. TextCNN 输入元素包括论元

标签、实体类型、句子和文档 4 种嵌入向量. 一个句子抽取的关键论元个数越多, 说明这个句子是关键句子的概率越大, 即论元标签数量可直接决定句子的重要性程度. 本文采用 one-hot 标签形式. 首先, 设置论元标签初始向量由 1 和 0 组成, 关键论元位置为 1, 其他为 0. 然后, 利用 Word2Vec^[27] 将初始向量训练成 128 维的嵌入向量. 相应的, 实体类型数量越多, 说明这个句子是关键句子的概率越大, 可在判断过程中起辅助作用. 嵌入向量维度设置为 128, 可通过查找随机初始化的嵌入表生成. 句子信息属于关键事件检测的基本数据, 提供语义信息, 可通过 ALBERT 将其转化成 312 维的嵌入向量. 文档信息可辅助判断句子的重要程度, 提供语境信息, 同样可通过 ALBERT 将其转化成 312 维的嵌入向量. 最后, 通过串联上述 4 个嵌入向量, 生成 880 维的新向量.

TextCNN 在预先训练的词向量基础上再训练后, 可较好的分类句子级文本. 同时, 通过微调参数处理特定任务, 可进一步提高性能, 进而适用于本文关键事件检测的应用场景. TextCNN 由嵌入层、卷积层、池化层和全连接层 4 部分组成, 具体框架如图 4 所示. 其中, 嵌入层通过一个隐藏层将输入的 880 维向量投影到维度为 128 的低维空间, 以助于编码语义特征. 卷积层设置 3、4、5 这 3 种卷积核尺寸, 每种卷积核个数都为 128, 其宽度与特征向量维度一致. 通过向下移动卷积核, 提取词与词间的局部相关性. 池化层通过抽取每个特征向量最大值表示该特征, 并拼接各个池化值来生成最终的特征向量. 最后, 通过连接全连接层, 并使用 *Softmax* 激活函数得到每个标签的概率, 判断句子是否包含关键事件.

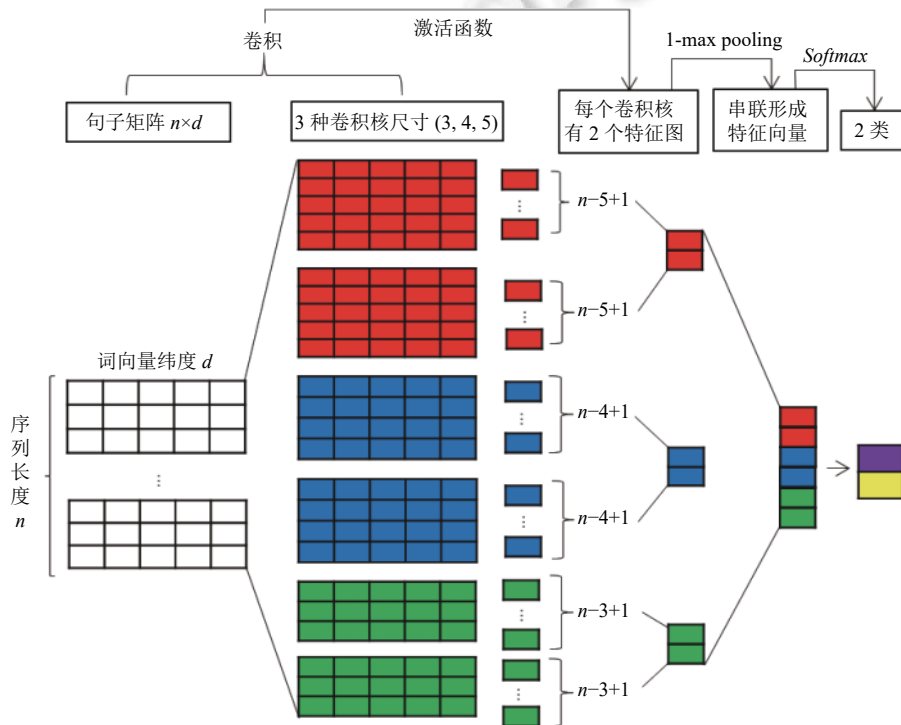


图 4 TextCNN 框架

3.3.2 缺失论元填充

事件信息通常分散在文档中的多个句子中, 导致常出现论元缺失的问题, 影响事件抽取的精确率和召回率. 本文首先检测事件中关键句子, 采用上文提到的 TextCNN 判别关键事件. 然后, 借助相似性排序, 利用相邻句子中的填充词来补充缺失的事件角色. 相邻句筛选的策略是计算它与关键句的相似度. 但是, 与以往文本相似度计算不同的是, 所计算的文本并不是语义和结构相接近的两个句子, 而是同一文档下的两个不同句子. 本文借助孪生神经网络进行相似性排序. 孪生神经网络可从数据中学习相似性度量, 进而匹配未见过类别的新样本. 这里具体采用 Mueller 等人^[28]提出的基于曼哈顿长短记忆网络的孪生神经网络 (Manhattan long short-term memory, MaLSTM),

计算句子间语义相似性.

MaLSTM 由两个长短记忆网络 (long short-term memory, LSTM) 构成, 属于共享权重的孪生神经网络, 具体框架如图 5 所示. 利用 MaLSTM, 具体的缺失论元填充步骤如下.

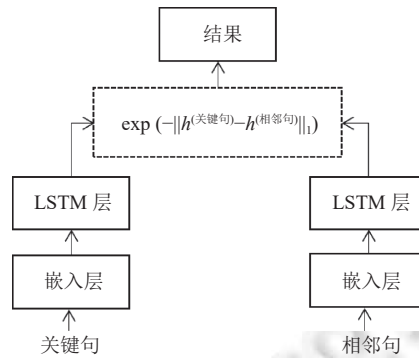


图 5 MaLSTM 框架

- Step 1. 处理<关键句, 相邻句>句子对, 以表征空间中的相似性进而推断句子潜在的语义相似性.
- Step 2. 取其最终隐藏状态作为两个句子的向量表示.
- Step 3. 用曼哈顿距离度量两个句子的相似度.
- Step 4. 计算关键句与其上下文句子的相似度值.
- Step 5. 相似度按从高到低排序.
- Step 6. 寻找对应缺失论元且相似度最高的相邻句中的论元角色进行填充.

4 实验验证

4.1 实验准备

4.1.1 数据集和评价指标

ACE2005 数据集是研究者最为认可的公开数据集, 涉及英语、中文和阿拉伯语 3 种语种, 一共包含 633 篇文章. 该数据集定义了 8 种事件类型、33 种事件子类型和 35 个事件论元^[29], 具体如表 1 所示. 部分事件类型对应的事件论元如表 2 所示.

表 1 ACE2005 数据集中事件类型和子类型

事件类型	事件子类型
生活	出生, 结婚, 离婚, 伤害, 死亡
移动	运输
交流	会议交流, 通过电话、书面交流
冲突	攻击, 示威, 游行等
商业	设立, 合并, 倒闭, 破产
事务	转让所有钱, 转让贷款、捐款等
个人职业	开始职位, 结束职位, 提名职位, 当选职位
司法相关	逮捕入狱, 释放, 审判听证, 指控, 起诉, 被判罪名, 判决有期徒刑, 罚款, 执行、处决, 引渡, 无罪释放, 上诉, 赦免

下面将介绍关于 ACE2005 中文语料库所涉及的相关术语.

(1) 事件触发器: 最能有效表达事件发生的词语, 一般为动词和名词, 是事件的基本要素之一. 例如“开火”“爆炸”等单词就是“攻击”事件的事件触发器.

(2) 事件论元: 事件中涉及的实体, 如施事者、受害者、工具、时间、地点等. 它是事件的重要组成部分, 一个事件可能存在多个论元, 不同事件类型对应不同论元.

(3) 事件角色: 表述论元和事件的具体关系. 如“伤害”事件中“地点”是固定论元, 但该论元事件角色与具体事件有关, 可能是“苏州”, 也可能是“南京”.

表 2 ACE2005 数据集中事件类型和对应的事件论元

事件类型	论元角色
攻击	攻击者, 目标, 工具, 时间, 地点
伤害	施事者, 受害者, 工具, 时间, 地点
逮捕入狱	逮捕者, 被逮捕者, 罪行, 时间, 地点, 待在监狱的时间
开始职位	雇员, 雇主, 职位, 时间, 地点
...	...

ACE2005 数据集将事件抽取任务分为 4 个子任务: 事件触发器识别、事件类型分类、事件论元识别、论元角色分类^[29]. ACE 预定义的事件框架及事件抽取案例如图 6 所示, 左边部分说明了 ACE2005 中的一些预定义的事件模式, 而右边部分用于阐述 4 个子任务工作流程与抽取结果.

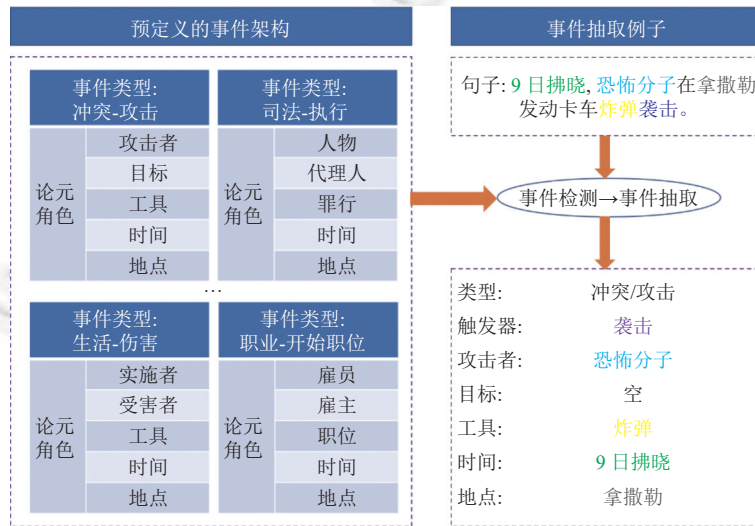


图 6 ACE 预定义的事件框架及事件抽取案例

大坝安全运行日志数据集由历年专项检查报告和日常巡检报告两部分构成, 共计 1000 篇报告, 其案例如表 3 所示. 大坝安全运行日志数据集包含地震、暴雨、泄洪、汛前安全大检查、全面专项检查、日常检修和日常检查 7 种事件类型和 17 个论元角色, 具体如表 4 所示.

表 3 大坝安全运行日志数据集案例

序号	案例
1	2018年8月13日在云南玉溪市通海县发生M5.0级地震, 震源深度7公里, 地震震中距漫湾电站大坝直线距离约为231公里, 漫湾生产区稍有震感, 为掌握地震对漫湾电厂水工建筑物的影响, 电厂及时开展了全面专项检查.
...	...

实验使用的数据集为 ACE2005 数据集和大坝安全运行日志数据集. ACE2005 数据集是包含 33 种事件子类型和 35 个论元角色^[29]. 大坝安全运行日志数据集包含 7 种事件子类型和 17 个论元角色.

表 4 大坝安全运行日志数据集中事件类型和对应的事件论元

事件类型	论元角色
地震	时间, 地点, 震级, 震源深度, 影响范围
暴雨	开始时间, 结束时间, 地点, 降雨量, 预警级别
泄洪	开始时间, 结束时间, 地点, 原因, 监测手段, 监测效果
汛前安全大检查	开始时间, 结束时间, 地点, 原因, 监测手段
全面专项检查	时间, 地点, 原因, 监测手段, 监测效果
日常检修	开始时间, 结束时间, 地点, 检修部位, 级别, 措施, 结果分析
日常检查	时间, 地点, 检查部位, 措施, 结果分析

在所有实验中, 将 80% 的数据作为训练集, 10% 的数据作为验证集和 10% 的数据作为测试集. 本节实验采用精确率 (P)、召回率 (R) 和 $F1$ -score ($F1$) 作为评价指标.

4.1.2 基准模型

实验将 EEDAM 模型与以前最先进的方法进行比较, 具体如下所述.

- (1) DMCNN 模型^[9]: 根据事件触发器和论元使用动态多池层来保留更重要的信息.
- (2) C-BiLSTM 模型^[10]: 采用卷积双向长短期记忆模型从字符级序列标记范式的角度进行中文事件提取.
- (3) JRNN 模型^[16]: 采用基于循环神经网络的事件抽取联合模型, 既避免了管道模型的误差传播问题, 同时也考虑到了事件触发器和事件元素之间的关系.
- (4) HMEAE 模型^[11]: 基于概念层次, 为每个基本单元设计神经模块网络, 再运用逻辑运算组成面向角色的模块网络, 去分类特定的论元角色.
- (5) JMCEE 模型^[17]: 基于预先训练的语言模型的共享特征表示, 联合执行对事件触发器和事件论元的预测, 解决实践中常见的角色重叠问题.

4.1.3 实验设置

EEDAM 模型的各个超参数取值在两个数据集上是一致的, 具体取值如表 5 所示. ALBERT 模型用来生成 312 维单词嵌入向量, 训练查找表用来生成 200 维事件类型嵌入向量、实体类型嵌入向量和词性标注嵌入向量. BiLSTM 隐藏层大小设置为 256, GTANs 层数和通道数分别设置为 2 和 3, 注意力网络隐藏层设置为 128. 最后第 2 层应用 Dropout 层以避免过拟合, 丢弃比例设置为 0.5. 模型训练批次设置为 16, 迭代次数为 100 次, 采用 Adam 优化器以 0.003 的学习率优化模型. λ 用于调节 GTANs 和注意力网络的加权融合比例, 参数值设置为 0.81. 此外, 我们还采用了非 ALBERT 设置, 即使用 Word2Vec 作为词嵌入向量模型 (表示为 EEDAM-W2V).

表 5 超参数设置

参数	参数值
Word embedding size	312
Radical embedding size	200
Entity type embedding size	200
Part of speech tagging embedding size	200
BiLSTM hidden size	256
GTANs layers number	2
GTANs channel number	3
Attention network hidden size	128
Dropout rate	0.5
Batch size	16
Epoch size	100
Learning rate	0.003
λ	0.81
Optimizer	Adam

4.2 实验结果与分析

4.2.1 性能对比分析

为证明 EEDAM 可提高事件抽取的精确率和召回率,将该模型与第 3.1.2 节中提到的 5 种基线方法进行比较,对比实验结果如表 6 所示.从表 6 可知,EEDAM 在两个数据集上的效果均优于其他基线方法,达到了最佳精确率、召回率和 $F1$ -score.从 ACE2005 数据集上的实验结果可知,EEDAM 较最佳基线 JMCEE 在精确率、召回率和 $F1$ -score 分别提高 17.82%、4.61%、9.80%.从大坝安全运行日志数据集上的实验结果可知,EEDAM 较最佳基线 JMCEE 在精确率、召回率和 $F1$ -score 分别提高 18.08%、4.41%、9.93%.从性能对比发现,EEDAM 效果远高于联合式抽取模型 JMCEE.可能是因为 EEDAM 在嵌入向量和双重注意力两个方面做了局部优化,且增加了论元填充环节,以此提升了事件抽取的精确率和召回率.

表 6 不同事件抽取方法对比实验结果 (%)

数据集	模型	论元识别			论元分类		
		P	R	$F1$	P	R	$F1$
ACE2005	DMCNN	51.67	48.36	49.96	45.81	42.55	44.12
	C-BiLSTM	53.04	52.20	52.62	47.35	46.63	46.99
	JRNN	54.75	51.93	53.30	48.27	45.86	47.03
	HMEAE	56.83	50.38	53.41	51.71	44.16	47.64
	JMCEE	66.35	45.28	53.83	53.76	46.74	50.00
	EEDAM-W2V	73.52	53.47	61.91	70.21	48.96	57.69
	EEDAM	75.87	54.96	63.74	71.58	51.35	59.80
大坝安全运行日志	DMCNN	62.48	59.36	60.88	57.64	54.37	55.96
	C-BiLSTM	64.25	63.11	63.67	59.13	57.44	58.27
	JRNN	65.74	62.93	64.30	60.84	56.95	58.83
	HMEAE	73.53	58.07	64.89	62.49	54.81	58.40
	JMCEE	78.63	56.45	65.71	65.37	57.86	61.39
	EEDAM-W2V	84.38	62.78	71.99	80.24	61.06	69.35
	EEDAM	86.59	64.72	74.07	83.45	62.27	71.32

实验结果表明,DMCNN 作为经典的事件抽取模型,效果一般.C-BiLSTM 首次将深度学习应用到中文事件抽取领域,取得了较好的实验结果,原因是卷积双向长记忆神经网络可同时捕获句子和词汇信息,其挖掘语义信息的能力比 DMCNN 强.JRNN 作为同时抽取触发器和事件论元的联合式抽取模型,其整体性能却提升有限, $F1$ -score 只比 C-BiLSTM 高 0.04%,可能因为它虽避免了管道模型的误差传播问题.但是, JRNN 主要围绕英文设计,无法进行局部优化,不能适应中文事件抽取.HMEAE 启发性地将分层模块化方法应用于事件抽取任务,基于灵活的模块化网络,利用模块之间的分层概念相关性做归纳偏差.性能优于模型网路架构几乎相同的 DMCNN 的主要原因是 HMEAE 引入了模块化关注.说明层次模块化方法可以很好地增强事件论元抽取模型的概念层次归纳偏差.但是, HMEAE 缺乏考虑句子中含有多个触发词、多个论元的情况,故性能提升有限. JMCEE 是针对中文提出的联合式事件抽取模型, $F1$ -score 比另外 4 个基准模型都高,其中比 C-BiLSTM 提升 3.01%,性能较佳.原因是 JMCEE 首先将触发器、论元和角色直接关系定义为事件关系三元组.然后,结合预先训练的语言模型的共享特征表示,联合抽取触发器和论元.最后,将论元分类任务转化为关系抽取,以此解决角色重叠问题,提高模型精确率.非 ALBERT 设置的 EEDAM-W2V 模型已经优于最先进的方法,这证实了双重注意力机制的优势.但是, EEDAM-W2V 的性能略低于 EEDAM,原因是 ALBERT 可以弥补 Word2Vec 无法表示一词多义的问题.

EEDAM 整体实验效果最佳,其精确率和召回率均有显著提升的原因如下.

(1) ALBERT 可从文本数据中提取特征,构建词嵌入向量,并结合双向长记忆神经网络捕获上下文重要信息,提升模型的预测能力.

(2) EEDAM 通过图转换注意力网络生成新的依赖弧和聚合节点信息,捕获长程依赖关系和潜在交互,加权结合

注意力网络, 捕捉句中关键的语义信息, 抽取句子级事件论元, 提高事件论元角色抽取的精确率和召回率。

(3) EEDAM 融入事件类型信息, 将含多个事件的句子按事件类型分别进行论元识别, 解决角色重叠的问题。

(4) EEDAM 解决论元缺失问题, 提高论元分类的精确率和召回率。

4.2.2 双重注意力融合比例分析

图转换注意力网络 GTANs 和注意力网络构成的双重注意力是 EEDAM 的核心部分。首先, GTANs 通过引入依赖关系使模型充分挖掘句法关系。然后, 利用多个候选邻接矩阵来寻找新的图结构。最后, 修正依赖弧生成的一些问题, 以提升模型信息表示的能力。但是, GTANs 无法根除已存在的错误依赖弧, 导致这些噪音可能会影响模型的整体性能。因此, 引入注意力网络, 注意力网络可捕获 GTANs 所遗漏的信息, 同时稀释了噪音的影响, 达到互补的功效。由此可见, GTANs 和注意力网络各有优势, 设置一个合适的双重注意力融合比例 λ 尤为重要。

本文通过观察 EEDAM 模型的 $F1$ -score 随 λ 值变化趋势, 选择 $F1$ -score 值最高点作为 λ 的值。EEDAM 模型的 $F1$ -score 随 λ 值变化趋势如图 7 所示。其中, 曲线整体呈先升后降趋势, 峰值是 0.81 (即 λ 最终值), 两个边界点相差约 5.4%。从上述信息可知, GTANs 是提升模型性能的关键, 注意力网络可起到辅助作用。

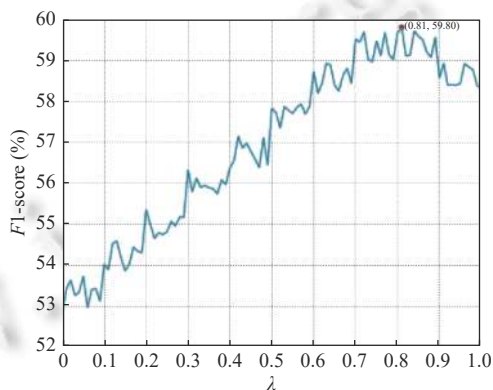


图 7 EEDAM 模型的 $F1$ -score 随 λ 值的变化趋势

4.2.3 消融实验

EEDAM 模型有 4 个重要组件, 分别为词嵌入层、图转换注意力网络层、注意力网络层、论元填充网络层。为确定这 4 个组件对模型的影响程度, 我们分别去掉或替换其中一个组件, 并与原模型进行性能比较。其中, 词嵌入层中实体类型和词性标注的嵌入向量, 在上文已证明其具备过滤功能。然而, 事件类型信息用于辅助解决角色重叠问题。因此, 这 3 项不再进行消融实验来验证它们的有效性。

消融实验结果如表 7 所示。其中, 图转换注意力网络层 GTANs 影响最大, 去掉该层后, $F1$ -score 在两个数据集上分别下降 6.92%、6.94%, 表明该层是 EEDAM 的核心部件。GTANs 可生成新的依赖弧, 弥补现有中文依存句法工具的不足之处, 增强每个单词的信息表示, 捕获长程依赖关系和潜在交互, 以提升模型预测能力。词嵌入层从 ALBERT 替换成 Skip-gram 后, $F1$ -score 在两个数据集上分别下降 2.45%、2.15%, 影响程度仅次于 GTANs。表明 ALBERT 可动态学习上下文信息, 解决 Skip-gram 无法表示一词多义的问题, 同时 ALBERT 表示的词嵌入向量包含更丰富的语义信息, 有助于模型抽取事件论元。去掉注意力网络层后, $F1$ -score 在两个数据集上分别下降 1.48%、1.44%。表明注意力网络虽提升效果不如 GTANs, 但注意力网络能捕获到 GTANs 所遗漏的信息, 两者可互补以达到性能最优。去掉论元填充网络层后, $F1$ -score 在两个数据集上分别下降 1.36%、1.95%。表明解决论元缺失问题确实可提升事件抽取的精确率和召回率, 但提升有限。可能因为前面已经解决论元缺失中角色重叠问题, 致使论元填充效果不明显, 也可能因为论元填充网络层中关键句检测和相似性排序两个方面有待改进。

4.2.4 图转换注意力网络有效性分析

图转换注意力网络 GTANs 将图转化网络 GTNs 中 GCN 模块替换成 GAN 模块, 原因是依赖关系中触发器和关

键论元的弧所占的权重应该更高. 然而, GCN 只能为邻居节点分配相同的权重, 无法充分发挥依赖关系的作用. 为证明 GTANs 效果比 GTNs 好, 本文分别在精确率 (P)、召回率 (R) 和 $F1$ -score ($F1$) 这三方面对比两个模型之间的性能差距. GTNs 与 GTANs 的对比实验如图 8 所示.

表 7 消融实验 (%)

模型	ACE2005数据集			大坝安全运行日志数据集		
	P	R	$F1$	P	R	$F1$
EEDAM	71.58	51.35	59.80	83.45	62.27	71.32
-(Skip-gram)	68.17	49.50	57.35	80.52	60.64	69.17
-(GTANs)	59.36	47.68	52.88	71.69	58.43	64.38
-(Attention)	69.44	50.27	58.32	81.57	61.12	69.88
-(Argument-Fill)	70.62	49.85	58.44	82.26	59.98	69.37

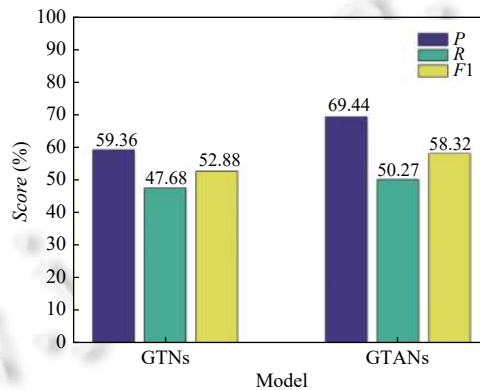


图 8 GTNs 与 GTANs 的对比实验

由图 8 可知, GTANs 在 3 项评价指标上性能均优于 GTNs. 证实了将 GCN 替换 GAN 后, GTANs 拥有为每个邻居分配不同的权重的能力, 充分发挥依赖关系的作用, 同时验证了 GTANs 的有效性.

4.3 实验小结

在 ACE2005 数据集和大坝安全运行日志数据集上的大量实验表明, 基于双重注意力机制的事件抽取方法的性能优于当前所有基线方法. 从 ACE2005 数据集上实验结果可知, EEDAM 较最佳基线 JMCEE 在精确率、召回率和 $F1$ -score 分别提高 17.82%、4.61%、9.80%. 从大坝安全运行日志数据集上实验结果可知, EEDAM 较最佳基线 JMCEE 在精确率、召回率和 $F1$ -score 分别提高 18.08%、4.41%、9.93%. 实验结果验证了 EEDAM 具备充分利用句法关系和解决论元缺失的能力, 从而提高事件抽取的精确率和召回率.

由消融分析可知:

- (1) ALBERT 生成的词嵌入向量具有丰富的语义信息, 可提高模型的准确性.
- (2) GTANs 可弥补依赖弧生成的缺陷, 捕获长程依赖关系和潜在交互, 提高模型精确率和召回率.
- (3) 注意力网络能捕获 GTANs 所遗漏的信息, 与其相辅相成.
- (4) 论元填充的有效性得到验证, 但仍有提升的空间.

最后, 通过 GTANs 有效性分析, 验证了 GCN 替换成 GAN 是有效的, 使模型可按邻居节点重要性程度分配不同的权重, 提高模型的准确性.

5 总结

本文分析了现有自然语言处理研究中的事件抽取方法, 提出基于双重注意力机制的事件抽取方法, 首先进行

句子编码并构建依赖关系图, 然后生成图转换注意网络并融合注意力网络, 最后进行关键事件检测和缺失论元填充. 实验结果表明, 本文提出的事件抽取方法能够充分利用句法关系并拥有解决论元缺失的能力, 同时效果优于当前所有基线方法, 使事件抽取的性能显著提高成为可能.

References:

- [1] Sha L, Qian F, Chang BB, Sui ZF. Jointly extracting event triggers and arguments by dependency-bridge rnn and tensor-based argument interaction. In: Proc. of the 32nd AAAI Conf. on Artificial Intelligence. New Orleans: AAAI, 2018. 5916–5923.
- [2] Balali A, Asadpour M, Campos R, Jatowt A. Joint event extraction along shortest dependency paths using graph convolutional networks. Knowledge-based Systems, 2020, 210: 106492. [doi: [10.1016/j.knosys.2020.106492](https://doi.org/10.1016/j.knosys.2020.106492)]
- [3] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In: Proc. of the 5th Int'l Conf. on Learning Representations. Toulon: OpenReview.net, 2017.
- [4] Ma J, Wang S, Anubhai R, Ballesteros M, Al-Onaizan Y. Resource-enhanced neural model for event argument extraction. In: Proc. of the Findings of the Association for Computational Linguistics: EMNLP 2020. Online: ACL, 2020. 3554–3559. [doi: [10.18653/v1/2020.findings-emnlp.318](https://doi.org/10.18653/v1/2020.findings-emnlp.318)]
- [5] Ahmad WU, Peng NY, Chang KW. GATE: Graph attention transformer encoder for cross-lingual relation and event extraction. In: Proc. of the 35th AAAI Conf. on Artificial Intelligence. Online: AAAI, 2021. 12462–12470.
- [6] Liu J, Chen YB, Liu K, Zhao J. Event detection via gated multilingual attention mechanism. In: Proc. of the AAAI Conf. on Artificial Intelligence. New Orleans: AAAI, 2018. 4865–4872.
- [7] Yang H, Chen YB, Liu K, Xiao Y, Zhao J. DCFEE: A document-level Chinese financial event extraction system based on automatically labeled training data. In: Proc. of the 56th Annual Meeting of the Association for Computational Linguistics-system Demonstration. Melbourne: ACL, 2018. 50–55.
- [8] Chen P, Yang H, Liu K, Huang RH, Chen YB, Wang TF, Zhao J. Reconstructing event regions for event extraction via graph attention networks. In: Proc. of the 1st Conf. of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th Int'l Joint Conf. on Natural Language Processing. Suzhou: AACL, 2020. 811–820.
- [9] Chen YB, Xu LH, Liu K, Zeng DJ, Zhao J. Event extraction via dynamic multi-pooling convolutional neural networks. In: Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int'l Joint Conf. on Natural Language Processing. Beijing: ACL, 2015. 167–176. [doi: [10.3115/v1/p15-1017](https://doi.org/10.3115/v1/p15-1017)]
- [10] Zeng Y, Yang HH, Feng YS, Wang Z, Zhao DY. A convolution BiLSTM neural network model for Chinese event extraction. In: Proc. of the 5th CCF Conf. on Natural Language Processing and Chinese Computing, NLPCC 2016, and 24th Int'l Conf. on Computer Processing of Oriental Languages. Kunming: Springer, 2016. 275–287. [doi: [10.1007/978-3-319-50496-4_23](https://doi.org/10.1007/978-3-319-50496-4_23)]
- [11] Wang XZ, Wang ZQ, Han X, Liu ZY, Li JZ, Li P, Sun MS, Zhou J, Ren X. HMEAE: Hierarchical modular event argument extraction. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing. Hong Kong: ACL, 2019. 5776–5782. [doi: [10.18653/v1/D19-1584](https://doi.org/10.18653/v1/D19-1584)]
- [12] Veyseh APB, Nguyen TN, Nguyen TH. Graph transformer networks with syntactic and semantic structures for event argument extraction. In: Proc. of the Findings of the Association for Computational Linguistics: EMNLP 2020. Online: ACL, 2020. 3651–3661. [doi: [10.18653/v1/2020.findings-emnlp.326](https://doi.org/10.18653/v1/2020.findings-emnlp.326)]
- [13] Yan H, Qiu XP, Huang XJ. A graph-based model for joint Chinese word segmentation and dependency parsing. Trans. of the Association for Computational Linguistics, 2020, 8: 78–92. [doi: [10.1162/tacl_a_00301](https://doi.org/10.1162/tacl_a_00301)]
- [14] Wu Y, Zhang JY. Chinese event extraction based on attention and semantic features: A bidirectional circular neural network. Future Internet, 2018, 10(10): 95. [doi: [10.3390/fi10100095](https://doi.org/10.3390/fi10100095)]
- [15] Liu X, Luo ZC, Huang HY. Jointly multiple events extraction via attention-based graph information aggregation. In: Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing. Brussels: ACL, 2018. 1247–1256. [doi: [10.18653/v1/D18-1156](https://doi.org/10.18653/v1/D18-1156)]
- [16] Nguyen TH, Cho K, Grishman R. Joint event extraction via recurrent neural networks. In: Proc. of the 2016 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego: ACL, 2016. 300–309. [doi: [10.18653/v1/N16-1034](https://doi.org/10.18653/v1/N16-1034)]
- [17] Xu N, Xie HH, Zhao DY. A novel joint framework for multiple Chinese events extraction. In: Proc. of the 19th Chinese National Conf. on Chinese Computational Linguistics. Haikou: Springer, 2020. 174–183.
- [18] Kim Y. Convolutional neural networks for sentence classification. In: Proc. of the 2014 Conf. on Empirical Methods in Natural Language Processing. Doha: ACL, 2014. 1746–1751. [doi: [10.3115/v1/D14-1181](https://doi.org/10.3115/v1/D14-1181)]

- [19] Ji YZ, Lin YF, Gao JW, Wan HY. Exploiting the entity type sequence to benefit event detection. In: Proc. of the 23rd Conf. on Computational Natural Language Learning. Hong Kong: ACL, 2019. 613–623. [doi: [10.18653/v1/K19-1057](https://doi.org/10.18653/v1/K19-1057)]
- [20] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: ACL, 2019. 4171–4186. [doi: [10.18653/v1/N19-1423](https://doi.org/10.18653/v1/N19-1423)]
- [21] Lan ZZ, Chen MD, Goodman S, Gimpel K, Sharma P, Soricut R. ALBERT: A lite BERT for self-supervised learning of language representations. In: Proc. of the 8th Int'l Conf. on Learning Representations. Addis Ababa: ICLR, 2020.
- [22] Feng XC, Qin B, Liu T. A language-independent neural network for event detection. Science China Information Sciences, 2018, 61(9): 092106. [doi: [10.1007/s11432-017-9359-x](https://doi.org/10.1007/s11432-017-9359-x)]
- [23] Cui SY, Yu BW, Liu TW, Zhang ZY, Wang XB, Shi JQ. Edge-enhanced graph convolution networks for event detection with syntactic relation. In: Proc. of the Findings of the Association for Computational Linguistics: EMNLP 2020. Online: ACL, 2020. 2329–2339. [doi: [10.18653/v1/2020.findings-emnlp.211](https://doi.org/10.18653/v1/2020.findings-emnlp.211)]
- [24] Yun S, Jeong M, Kim R, Kang J, Kim HJ. Graph transformer networks. In: Proc. of the 33rd Int'l Conf. on Neural Information Processing Systems. Vancouver: NIPS, 2019. 32.
- [25] Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. Graph attention networks. In: Proc. of the Int'l Conf. on Learning Representations. Vancouver: OpenReview.net, 2018.
- [26] Lin TY, Goyal P, Girshick R, He KM, Dollár P. Focal loss for dense object detection. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318–327. [doi: [10.1109/TPAMI.2018.2858826](https://doi.org/10.1109/TPAMI.2018.2858826)]
- [27] Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. In: Proc. of the 1st Int'l Conf. on Learning Representations. Scottsdale: ICLR, 2013.
- [28] Mueller J, Thyagarajan A. Siamese recurrent architectures for learning sentence similarity. In: Proc. of the 30th AAAI Conf. on Artificial Intelligence. Phoenix Arizona: AAAI, 2016. 2789–2792. [doi: [10.5555/3016100.3016291](https://doi.org/10.5555/3016100.3016291)]
- [29] Xiang W, Wang B. A survey of event extraction from text. IEEE Access, 2019, 7: 173111–173137. [doi: [10.1109/ACCESS.2019.2956831](https://doi.org/10.1109/ACCESS.2019.2956831)]



朱敏(1997—), 女, 硕士生, CCF 学生会会员, 主要研究领域为知识图谱, 事件抽取, 实体链接.



陈程军(1996—), 男, 硕士, 主要研究领域为知识图谱, 事件抽取.



毛莺池(1976—), 女, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为分布式数据处理, 边缘智能计算.



王龙宝(1977—), 男, 博士, 高级工程师, CCF 专业会员, 主要研究领域为行业大数据, 人工智能.



程永(1998—), 男, 硕士生, CCF 学生会会员, 主要研究领域为知识图谱, 事件抽取.