计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
[doi: 10.15888/j.cnki.csa.009842] [CSTR: 32024.14.csa.009842]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于 BiLSTM-Attention 的议论文篇章要素识别①

刘佳旭[1], 白再冉[1], 张艳菊[2]

[1](辽宁工程技术大学 软件学院, 葫芦岛 125105)
[2](辽宁工程技术大学 工商管理学院, 葫芦岛 125105)
通信作者: 刘佳旭, E-mail: liujiaxu@buaa.edu.cn

摘　要: 篇章要素识别 (discourse element identification) 的主要任务是识别篇章要素单元并进行分类. 针对篇章要素识别对上下文依赖性理解不足的问题, 提出一种基于 BiLSTM-Attention 的识别篇章要素模型, 提高议论文篇章要素识别的准确率. 该模型利用句子结构和位置编码来识别句子的成分关系, 通过双向长短期记忆网络 (bidirectional long short-term memory, BiLSTM) 进一步获得深层次上下文相关联的信息; 引入注意力机制 (attention mechanism) 优化模型特征向量, 提高文本分类的准确度; 最终用句间多头自注意力 (multi-head self-attention) 获取句子在内容和结构上的关系, 弥补距离较远的句子依赖问题. 相比于 HBiLSTM、BERT 等基线模型, 在相同参数、相同实验条件下, 在中文数据集和英文数据集上准确率分别提升 1.3%、3.6%, 验证了该模型在篇章要素识别任务中的有效性.

关键词: 双向长短期记忆网络; 注意力机制; 位置编码; 篇章要素识别; 多头注意力

引用格式: 刘佳旭,白再冉,张艳菊.基于 BiLSTM-Attention 的议论文篇章要素识别.计算机系统应用. http://www.c-s-a.org.cn/1003-3254/9842.html

Discourse Elements Identification in Argumentative Essays Based on BiLSTM-Attention

LIU Jia-Xu[1], BAI Zai-Ran[1], ZHANG Yan-Ju[2]

[1](Software College, Liaoning Technical University, Huludao 125105, China)
[2](College of Business Management, Liaoning Technical University, Huludao 125105, China)

Abstract: The main task of discourse element identification is to identify discourse element units and classify them. Aiming at the lack of understanding of context dependence in discourse element identification, this study proposes a discourse element identification model based on BiLSTM-Attention to improve the accuracy of discourse element identification in argumentative essays. The model uses sentence structure and positional encoding to identify sentence component relationships and further acquires deep context-related information through bidirectional long short-term memory (BiLSTM). Attention mechanism is introduced to optimize the model feature vectors and improve the accuracy of text classification. Finally, inter-sentence multi-head self-attention is used to obtain the relationships between the content and structure of sentences, so as to make up for the distant sentence dependence. Compared with baseline models such as HBiLSTM and BERT, the accuracy on Chinese and English datasets is improved by 1.3% and 3.6% respectively under the same parameters and the same environmental conditions, which verifies the effectiveness of the model in the discourse element identification task.

Key words: bidirectional long short-term memory (BiLSTM); attention mechanism; positional encoding; discourse element identification; multi-head attention

① 基金项目: 辽宁省社会科学规划基金 (L22BJY034)
收稿时间: 2024-10-17; 修改时间: 2024-11-19; 采用时间: 2024-12-04; csa 在线出版时间: 2025-02-28

议论文是教育领域锻炼学生写作能力的一种常见文体. 而作文自动评分 (automatic essay scoring, AES) 是利用语言学、自然语言处理 (natural language processing, NLP)[1]等技术对作文进行自动评分, 对教育行业和自动评分领域具有重要意义. 传统评分方法依靠人工评分, 容易忽略一些重要评估指标, 存在不公平性、主观性强、效率低等问题. 而自动评分从连贯性、完整性等多个维度评估文本, 全面准确地评估文本质量, 做到作文评分的客观化, 提高效率, 同时又节约人力、物力和财力.

篇章结构[2]是自然语言处理领域的重要研究方向, 其目的在于深入理解篇章的结构和语义. 篇章不仅是文本序列本身, 更是句子或段落构成的整体[3]. 它不是孤立存在的, 而是各自承担着部分结构从而表达完整的语义. 早期篇章结构研究针对概念提出了简单的模型. Li 等人[4]提出了推理机制, 通过触发词内部的组成语义和中文触发词之间的一致性探索中文的特殊性. Chen 等人[5]在 Li 等人的基础上研究了用于中文事件提取各种丰富的知识源. 其结果表明, 该方法明显优于 Li 等人的方法. Li 等人[6]提出了使用循环神经网络学习句子的句法语义表示. 该方法在一致性评估任务中取得了最优效果. Song 等人[7]提出篇章解析是篇章分析中的一个重要研究课题, 旨在推断篇章结构, 用篇章结构学习特定或一般目的的句子和文档的表征. Song 等人[8]对记叙文中陈述句、说明句、描述句、议论句和情感句进行手动识别和自动识别的研究, 注释语料库用来研究篇章模式的特征, 并用于识别神经序列标记模型, 结果表明篇章模式可以自动识别, 篇章模式可以用作改进自动论文评分的功能.

运用篇章要素表示篇章结构, 使其更具有可解释性.篇章要素在多个方面辅助作文自动评分, 例如作文结构建模、主题和观点识别, 可以利用篇章要素获取信息、表达观点, 篇章要素在理解和表达过程中具有重要作用, 广泛应用于自然语言处理中的其他任务, 包括问答系统[9]、信息抽取[10]、文本摘要[11]以及分类任务[12]等. 早期研究大多基于手工设计的规则和启发式方法, 通过识别文本中的标题、标点符号等对文本分类. 随着深度学习的迅速发展, 运用神经网络识别篇章要素的方法受到了广泛关注. 利用模型的学习能力, 对文本进行分类, 实现了较高的篇章要素识别准确率. 刘海顺等人[13]以 Layer-attentive 进行特征融合的语言模型为解码器, 基于 LSTM 的序列生成模型作为解码器. 结果表明, 该方法对分类任务有效果. Mim 等人[14]提出了一种无监督的预训练方法获得论文篇章结构连贯性和衔接性, 不需要任何篇章注释.该方法在论文组织评分任务上取得了最先进的结果. Song 等人[15]使用衔接性提高学生论文中句子的篇章要素识别的方法. Fu 等人[16]将注意力机制运用到语义分割上均取得了良好的效果. 张周彬等人[17]建立了相互循环注意力模型用于情感分析任务中. 程艳等人[18]基于注意力机制提出了多通道 CNN (convolutional neural network) 和双向门控循环单元 (bidirectional gated recurrent unit, BiGRU) 的文本情感分析模型, 提取丰富的文本特征. Su 等人[19]提出旋转位置编码, 将位置编码融入到 Self-Attention 的计算中, 用绝对位置编码表达词语相对位置信息. 但相对位置编码和旋转位置编码都需要学习额外参数, 增加了模型训练成本. Daxenberger 等人[20]利用 CNN 和 LSTM 对句子进行分类, 识别不同领域的文章. 肖琳等人[21]利用标签语义注意力进行多标签文本分类, 效果优于传统的多标签分类模型. Song 等人[22]将相对位置编码和句间注意力运用到篇章要素识别的工作中, 准确率为 0.681. 在识别篇章要素[22]基础上结合组织评估[23], 将组织评估表示为网格, 模拟论文的视觉布局并整合多个语言层面的篇章要素, 该模型取得显著改进. Sun 等人[24]提出基于图卷积网络 (graph convolutional network, GCN) 的篇章论证模型, 将篇章关系识别转化为节点分类任务, 该模型能有效识别篇章关系. Wang 等人[25]提出了一种基于图的篇章要素识别模型, 该模型可以有效地捕捉句子级篇章要素之间的交互关系.

目前, 篇章要素识别还存在着许多问题. 句子的多义性[26,27]是模型很难确定的, 不同语境中句子代表的含义会有所不同; 其次, 某些句子级篇章要素对上下文的依赖性理解不够充分, 只看单个句子不能确定属于哪一类别.

因此, 为了解决上下文依赖性理解不足的问题, 结合议论文句子结构的特殊性, 提出了一种基于 BiLSTM-Attention 的议论文篇章要素识别模型. 本文主要针对句子级篇章要素进行研究, 结合 BiLSTM 与 Attention 机制的优点, 建立 BiLSTM-Attention 篇章要素识别模型. 因为议论文使用特定的结构进行写作, 所以位置信息与文本内容是不相关的. 运用句间多头自注意力识别数量少且与其他句子具有不同关系模式的句子级篇

章要素, 可以捕捉不同的关注点和特征表示, 从而获取更全面和多样化的表达能力, 来获取文本中的关键部分, 进一步提高篇章要素识别的准确性.

1 模型设计与框架

基于 BiLSTM-Attention 的议论文篇章要素识别模型整体结构图如图 1 所示.

模型将篇章要素标签$y=(y_1,\cdots,y_n)$分配给文本中的句子$(x_1,\cdots,x_n)$, 其中x_i $(1\leqslant i\leqslant n)$是句子的词嵌入表示, $y_i\in Y(1\leqslant i\leqslant n)$是一组预定义的篇章要素, 包括引言、中心论点、分论点、事实论据、理论论据、结论和其他. 模型运用 BiLSTM-Attention 将词嵌入转换为句向量, 作为模型输入. 该模型将句子全局、段落、局部相对位置与句子表示内容相结合, 经过 BiLSTM-Attention 层获得更全面的表示. 该模型有一个句间多头自注意力模块, 获取句子间要素和内容注意力向量. 将注意力向量和要素表示进行拼接得到最终表示, 送到 Linear 层和 *Softmax* 层进行下一步工作.

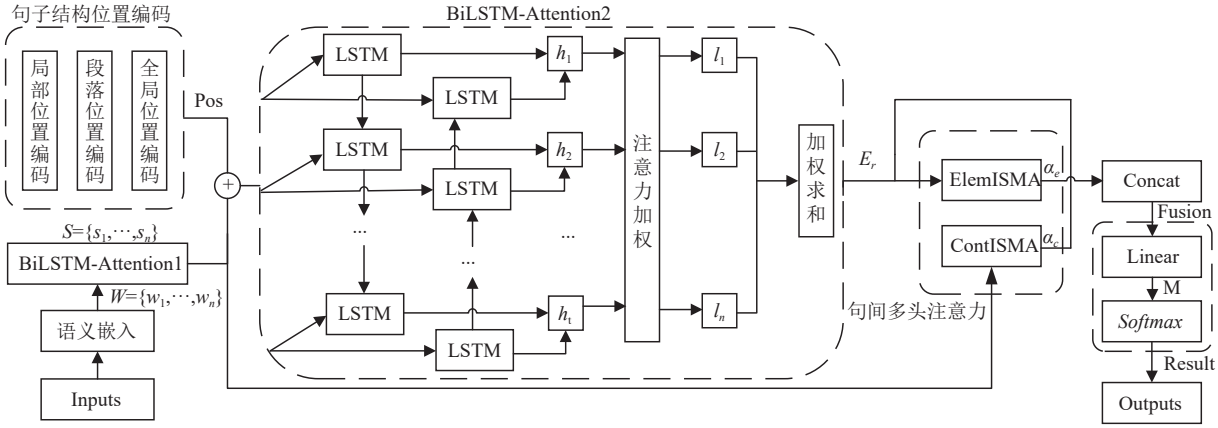


图 1 模型结构

1.1 语义嵌入过程

中文数据集采用腾讯预训练词向量. 英文采用 BERT-Case-Uncased 模型提取英文议论文单词和句子级别的嵌入表示, 其具体过程如图 2 所示.

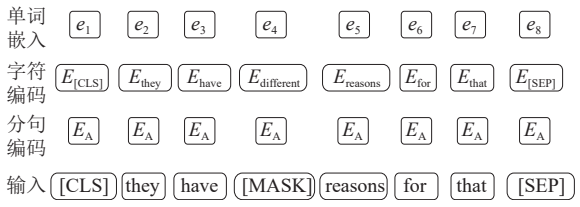


图 2 单词嵌入过程

经 BERT-Case-Uncased 模型训练后得到词嵌入矩阵, 针对单词嵌入情况需去除首位[CLS]编码向量, 得到词嵌入表示$W=\{w_1,\cdots,w_n\}$.

1.2 BiLSTM-Attention

将得到的词嵌入$W=\{w_1,\cdots,w_n\}$输入到 BiLSTM-Attention1, 并进行计算, 其具体过程如下:

第 1 步: 计算正向 LSTM.

$$h_t^{\text{forward}}=LSTM^{\text{forward}}(h_{t-1},x_t,c_{t-1}) \tag{1}$$

第 2 步: 计算反向 LSTM.

$$h_t^{\text{backward}}=LSTM^{\text{backward}}(h_{t-1},x_t,c_{t-1}) \tag{2}$$

第 3 步: 将正向和反向拼接.

$$h_t=\left(h_t^{\text{forward}},h_t^{\text{backward}}\right) \tag{3}$$

其中, h_{t-1}表示$t-1$时刻输出值, c_{t-1}表示$t-1$时刻单元状态, h_t表示t时刻输出值.

运用 BiLSTM 学习输入参数的特征, 将各个时刻输出的信号状态作为输入传递到 Attention 层, 运用注意力机制去除无效信息, 并得到句子表示$S=\{s_1,\cdots,s_n\}$.

1.3 句子结构位置编码

全局位置表示句子全文中的位置. 将文章句子当成一个序列, 描述该句子在文章的全局位置. 段落位置表示该句子段落全文中的位置. 局部位置表示句子在段落内的位置. 计算上述 3 种位置类型的相对位置. 例如, 文章E中第i $(i\geqslant 1)$个句子的相对全局位置表示为:

$$pos_{\text{global}}(i)=\frac{i}{|E|} \tag{4}$$

其中, $|E|$是文章中句子的数量. $pos_{\text{para}}(i)$表示段落的相对位置和$pos_{\text{local}}(i)$表示局部相对位置, 以相同的方式计算全局的相对位置. 最终位置$pos(i)$表示 3 个相对位置表示的线性组合, 表示为:

$$pos(i)=\sum_{t\in\{\text{global,local,para}\}}\beta_t pos_t(i) \tag{5}$$

其中, β_t表示训练中要学习的参数.

最后, 将句子位置编码pos和句子语义表示$S=\{s_1,\cdots,s_n\}$输入到 BiLSTM-Attention2 模块中, 通过 Attention 层对其进行权重分配, 并使用非线性层将语义表示映射到篇章要素表示E_r.

1.4 句间多头注意力

句间多头注意力 (ISMA) 是将多头自注意力应用于句子表示S和要素表示E_r来计算句子与其他句子的相关性.

要素多头自注意力 (ElemISMA): ElemISMA 主要是篇章要素之间的关系进行建模. 用要素表示E_r计算 Elem 向量α_e.

用要素表示E_r得到Q, K和V, 其中, $Q\in R^{n\times dk}$, $K\in R^{m\times dk}$, $V\in R^{m\times dv}$, m, n, dk, dv分别代表着矩阵的维度. 其具体计算过程表示为:

$$Attention(Q,K,V)=Softmax\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right)V \tag{6}$$

多头注意力可以同时计算多次缩放点积注意力, 并将多次的计算结果进行拼接, 最终能够得到权重求和结果. 其中, 第i头注意力计算过程表示为:

$$O_i=Attention(QW_i^Q,KW_i^K,VW_i^V) \tag{7}$$

其中, $W_i^Q\in R^{dk\times dk}$, $W_i^K\in R^{dk\times dk}$, $W_i^V\in R^{dv\times dv}$.

$$D=Concatenate(O_1,O_2,\cdots,O_N)W^0 \tag{8}$$

其中, $Concatenate$为拼接函数, W^0为训练参数.

内容多头自注意力 (ContISMA): ContISMA 主要是对内容相关性进行建模. 与 ElemISMA 类似, 使用句子表示S来计算 Cont 向量α_c.

1.5 特征融合层

将 Elem 向量α_e、Cont 向量α_c、要素表示E_r拼接, 得到融合后的向量$Fusion$, 具体计算过程表示为:

$$Fusion=Concat(\alpha_e,E_r,\alpha_c) \tag{9}$$

其中, $Concat$表示向量拼接函数.

1.6 线性层和 *Softmax* 层

将融合后的向量$Fusion$先通过线性层降低维度得到M, 再输入到 *Softmax* 层进行归一化操作得到概率, 取数值最大值作为最终结果.

$$M=Softmax(linear(Fusion)) \tag{10}$$

$$Result=Max(M) \tag{11}$$

1.7 对数似然损失函数

为训练该模型, 采用前向和反向传播数据更新迭代算法. 在反向传播过程的每次迭代中, 通过损失值计算梯度值从而执行模型参数更新. 采用对数似然损失函数来计算, 其具体计算过程表示为:

$$L(T,P(T\mid X))=-\log P(T\mid X)=-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}t_{ij}\log\left(p_{ij}\right) \tag{12}$$

其中, T是输出变量, X是输入变量, L是损失函数. N是输入样本数量, M表示类别数, t_{ij}表示输入实例x_i真实类别是否是类别j. p_{ij}表示属于类别j的模型预测的输入实例x_i的概率.

2 实验

2.1 数据集

本文采用 Burstein 等人[28]提出的篇章要素的定义和分类, Song 等人[22]提出的数据集, 共 1 230 篇文章, 按 9:1 的比例划分训练集和测试集, 验证集占训练集的 10%, 将测试集实验结果作为实验结果. 主要考虑以下篇章要素: 引言、中心论点、分论点、事实论据、理论论据、结论和其他. 表 1 显示了数据集的基本统计数据.

表 1 中文数据集概况

要素	训练	测试	总计	百分比 (%)
引言	2 859	285	3 144	9.5
中心论点	881	151	1 032	3.1
分论点	4 443	578	5 021	15.2
事实论据	5 972	679	6 651	20.1
理论论据	12 405	1 127	13 532	41.0
结论	3 086	333	3 419	10.3
其他	170	20	190	0.6
总计	29 816	3 173	32 989	100

引言表示提出论点之前介绍背景, 中心论点表达了作者关于主题的最重要的论点, 分论点要素是与论文相关的基本思想, 建立和支持这些思想的论据, 事实论据是支持主要观点和论点的示例, 理论论据进一步解释主要观点或提供理由, 不包含示例或其他证据, 结论是分论点的延伸, 概括全文, 呼应文章论点, 其他指与上述类别不匹配的要素.

本文采用了 Stab 等人[29]提出的英语议论文数据集, 共 402 篇文章[22], 按 4:1 的比例划分训练集和测试集, 验证集占训练集的 10%, 将测试集的实验数据作为

实验结果. 主要考虑以下篇章要素: 主旨、论点、前提和其他. 如表 2 所示.

表 2 英文数据集概况

要素	训练	测试	总计	百分比 (%)
主旨	598	153	751	10.3
论点	1 202	304	1 506	20.6
前提	3 023	809	3 832	52.3
其他	999	232	1 231	16.8
总计	5 822	1 498	7 320	100

主旨要素代表作者对该主题的立场, 每个论点要素都包含一个主旨和至少一个前提, 前提要素是论点的理由, 其他要素是指与上述类别不匹配的要素.

本文使用准确率 (Acc) 和 Macro-$F1$ 分数作为实验评价指标来评估系统的精度.

2.2 实验环境

实验基于 PyTorch 1.3.1+CPU, 实验环境配置如表 3 所示. 超参数设置如表 4 所示.

表 3 实验环境配置

项目	环境
操作系统	Windows 11 64位
CPU	Intel(R) Core(TM) i5-1135G7
内存	40 GB
集成开发环境	PyCharm
深度学习框架	PyTorch 1.3.1+CPU
脚本语言	Python 3.7.0

表 4 超参数设置

参数	中文	英文
优化器	SGD	SGD
epoch	700	1 500
BiLSTM	256	128
Batch_size	50	30
Learning rate	0.2	0.1
多头注意力机制头数	8	8

使用 PyTorch 作为深度学习框架, 每个句子最大长度为 40, 词嵌入维度 200, 采用 SGD 优化器, epoch 分别设置为 700 和 1 500, 中文数据集的 BiLSTM 隐藏层维数为 256, 英文数据集的 BiLSTM 隐藏层维数为 128, batch_size 分别设置为 50、30, learning rate 分别设置为 0.2、0.1, 多头注意力机制头数都为 8.

2.3 消融实验

本实验的所有参数和环境配置一致, 在中、英文数据集上进行实验, Acc 和 Macro-$F1$ 分数作为实验指标. 对改进部分进行消融实验, 结果如表 5 所示.

Base 表示 DiSA 模型; BiLSTM-Attention1 表示改进了 base 使用 BiLSTM-Attention1 模块; +MHA 表示在 base 的基础上改进多头自注意力模块; +BiLSTM-Attention1+MHA 表示集成使用 BiLSTM-Attention1 和 MHA; +BiLSTM-Attention 表示集成使用 BiLSTM-Attention1 模块和 BiLSTM-Attention2 模块; +BiLSTM-Attention+MHA 表示是将 3 个模块全部集成使用.

表 5 消融实验结果

模型	中文		英文	
	Acc	Macro-$F1$	Acc	Macro-$F1$
Base	0.679	0.657	0.831	0.811
+BiLSTM-Attention1	0.684	0.646	0.844	0.841
+MHA	0.685	0.653	0.845	0.842
+BiLSTM-Attention1+MHA	0.684	0.657	0.851	0.850
+BiLSTM-Attention	0.688	**0.660**	0.858	0.847
+BiLSTM-Attention+MHA	**0.692**	0.659	**0.867**	**0.864**

从表 5 中可以看出, 在中文数据集上单独改进两个模块后, Acc 分别提升 0.5%、0.6%; 在英文数据集上单独改进两个模块后, Acc 分别提升 0.3%、0.1%, 表明各个模块均有助于中英文篇章要素识别任务; +BiLSTM-Attention1+MHA 在中文数据集上, Acc 提升了 0.5%. 在英文数据集上, Acc 提升了 2%. 由实验结果可以看出, BiLSTM-Attention1 模块和多头注意力模块结合使用对篇章要素识别任务的准确精度有所提升; 当两个 BiLSTM-Attention 模块集成使用时, 与+BiLSTM-Attention1+MHA 相比, 在中文数据集上, Acc 提升了 0.4%, 与原模型相比提升了 0.9%; 在英文数据集上, Acc 提升了 0.7%, 与原模型相比提升了 2.7%; 说明改进还是有效果的. 最后, 将改进模块全部集成后, 在中文数据集和英文数据集上, Acc 分别达到 0.692、0.867.

中文数据集实验效果没有英文数据集实验效果好, 是由于中文语料与英文语料存在巨大差异, 中文语料存在大量的成语、俗语、惯用词和固定搭配等. 同一个词在不同语境中所表达的含义也会有所不同, 语义需要根据语境来确定. 所提出模型与 base 模型相比 Macro-$F1$ 提高 0.2%, 是由于 Macro-$F1$ 指标平等的对待所有的篇章要素标签, 不考虑不同篇章要素对文章的重要性, 而对于议论文来说, 中心论点要素最重要, 但在 Macro-$F1$ 的计算过程中, 将中心论点要素与其他要素平等对待, 从而导致 Macro-$F1$ 有微弱的提高. 综上所述, 表明模块改进在中英文篇章要素识别任务上都有效果.

2.4 对比实验

为了进一步证明本文改进方法的有效性, 与

Featured-based[15]、HBiLSTM、BERT[30]、DiSA[8]等方法进行比较.

Featured-based: 该方法用基于特征的方法来构建基于特征的 CRF 模型.

HBiLSTM: 该方法用两个 BiLSTM 编码单词序列和句子.

BERT: 该方法是对 BERT 进行微调.

DiSA: 该方法用 BiLSTM 编码句子, 用相对位置编码、句间注意力构建 DiSA 模型.

中文实验结果如表 6 所示. BERT 的性能最差. HBiLSTM 的性能略高于 BERT. HBiLSTM 的 Macro-*F*1 分数较低, 表明它不擅长于识别篇章要素. DiSA 的性能比 BERT 好. 而 BiLSTM-Attention 效果最佳.

表 6 模型比较

模型	Acc	Macro-*F*1
Featured-based	0.633	0.589
BERT	0.575	0.513
HBiLSTM	0.582	0.522
DiSA	0.679	0.642
BiLSTM-Attention	**0.692**	**0.658**

BiLSTM-Attention 模型在识别篇章要素方面具有最高准确率和 *F*1 分数. 相比之下, HBiLSTM 无法准确识别部分篇章要素. Featured-based 在识别篇章要素方面比 HBiLSTM 表现得好. DiSA 的性能与前几个模型相比都有所提升. BiLSTM-Attention 实验结果显示其性能较之前模型都有提高.

图 3 说明了基线模型识别篇章要素的性能. HBiLSTM 无法准确识别中心论点和分论点. Featured-based 的方法在识别中心论点和分论点方面比 HBiLSTM 表现更好, 但由于它们依赖手工特征, 因此识别事实论据要素方面表现较差. DiSA 的性能优于 HBiLSTM 和 Featured-based 方法, 与 HBiLSTM 和 Featured-based 方法相比, 分论点的 *F*1 分数提高了 29.2% 和 38.5%, 这表明 DiSA 模型中的位置编码在评分任务有重要作用. BiLSTM-Attention 模型的 *F*1 分数在引言、分论点、事实论据和结论等方面优于 DiSA 模型, BiLSTM-Attention 能捕获深层次的上下文关联信息, 句间多头自注意力更有助于获取句子在内容和结构上的关系.

DiSA-SPE: 该方法用 BiLSTM 编码句子, 用句间注意力构建模型.

DiSA+Feature[29]: 该方法用 BiLSTM 编码句子, 融合指标特征和位置特征, 再用句间注意力构建模型.

Joint-Best[29]: 将识别论证关系作为辅助任务.

英文实验结果如表 7 所示. ISMA 有助于识别部分篇章要素. DiSA 效果优于 DiSA-SPE. Joint-Best 效果优于 DiSA-SPE. BiLSTM-Attention 优于 Joint-Best. 基线方法使用了位置编码等手工特征. 因此, 通过合并特征和位置编码构建特征向量, 在句间使用多头自注意力来获取上下文信息, 这种组合效果优于 Single-Best 的结果, 在 Acc 和 Macro-*F*1 方面分别提升 2.4%、4.7%.

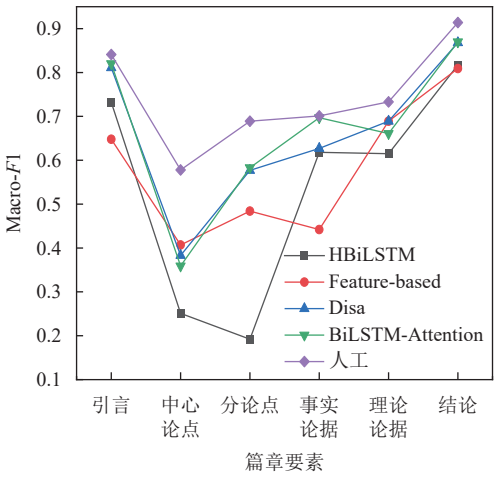


图 3 识别篇章要素的 *F*1 分数

表 7 模型比较

模型	Acc	Macro-*F*1
DiSA	0.806	0.742
DiSA-SPE	0.710	0.534
DiSA+Feature	0.832	0.798
Joint-Best	0.844	0.817
BiLSTM-Attention	**0.868**	**0.864**

2.5 句间注意力分析

如图 4 所示, 中文数据集上 ISMA 对部分要素有一定的影响, 对识别事实论据影响最大. 事实论据常与其他句子相关联, 提供事实或例子. ISMA 有助于捕获上下文信息, 能够更好地识别论据要素. ISMA 对其他的篇章要素识别也有一定帮助.

BiLSTM-Attention 在英文数据集上篇章要素识别的 Macro-*F*1 分数, 如图 5 所示. 添加了 ISMA 模块, 识别主旨要素和论点要素的 Macro-*F*1 分数分别提升了 6.8%、4.7%. 进一步说明 ISMA 有助于该分类任务.

2.6 注意力头个数分析

注意力头对输入数据进行线性变换, 结果是离散的, 头数越多累积误差越大, 导致模型表现变差. 针对此问题, 在中、英文数据集分别进行注意力头个数为 2、4、8、16、32 和 64 这 6 次实验.

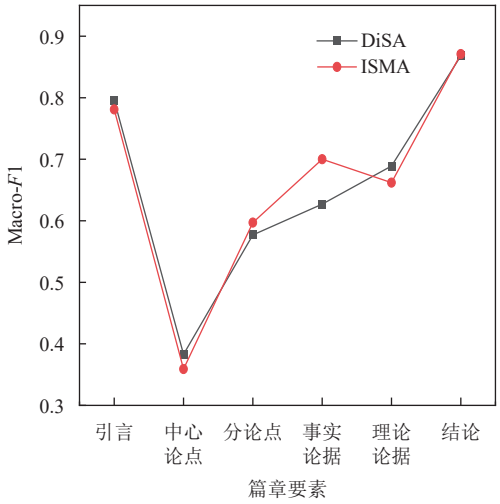


图 4 识别篇章要素句间注意力 Macro-*F*1 分数分析

如图 6 所示, 结果表明注意力头个数为 8 时, 中英文 Acc、*F*1 分数最高. 注意力头的个数为 2、4, 还没有达到最好的效果. 注意力头的个数为 16、32、64, Acc、*F*1 分数都有所下降, 表示该模型过拟合, 导致模型的表现变差.

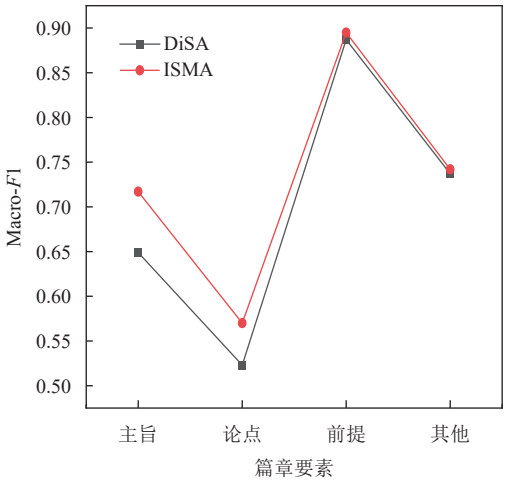
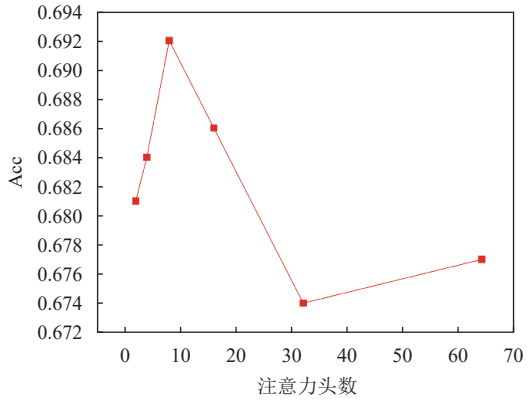
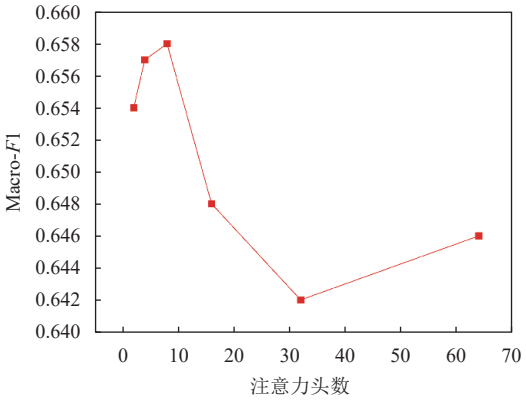


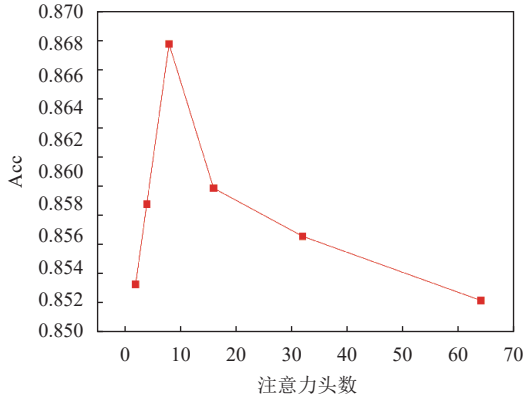
图 5 识别篇章要素的句间注意力 Macro-*F*1 分数分析



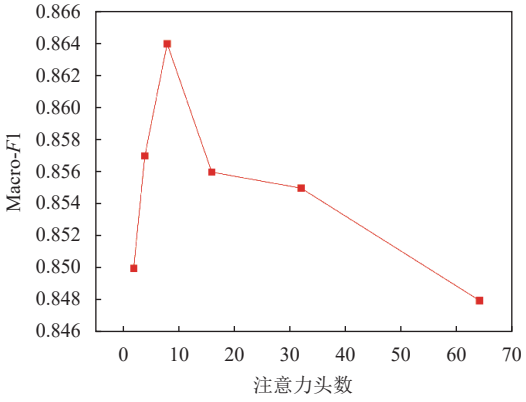
(a) 中文 Acc



(b) 中文 Macro-*F*1



(c) 英文 Acc



(d) 英文 Macro-*F*1

图 6 注意力头的数量实验结果

图 7 所示是中文数据集各要素的对比图. 由图 7 中实验结果可知, 当注意力头 *n* 为 8 时, 总体效果是最好的. *n* 为 2 时, 引言准确率最高, 分论点和结论要素准确率最低; 当 *n* 为 4 时, 分论点、结论和事实论据要素的准确率有明显的提升, 但引言、中心论点和理论论据要素有明显的下降; 当 *n* 为 16 时, 分论点和理论

论据要素准确率最差, 论点的准确率最高; 当 n 为 32 时, 在识别引言、中心论点和事实论据要素的效果最不佳; 当 n 为 64 时, 各篇章要素的准确率都有明显的下降.

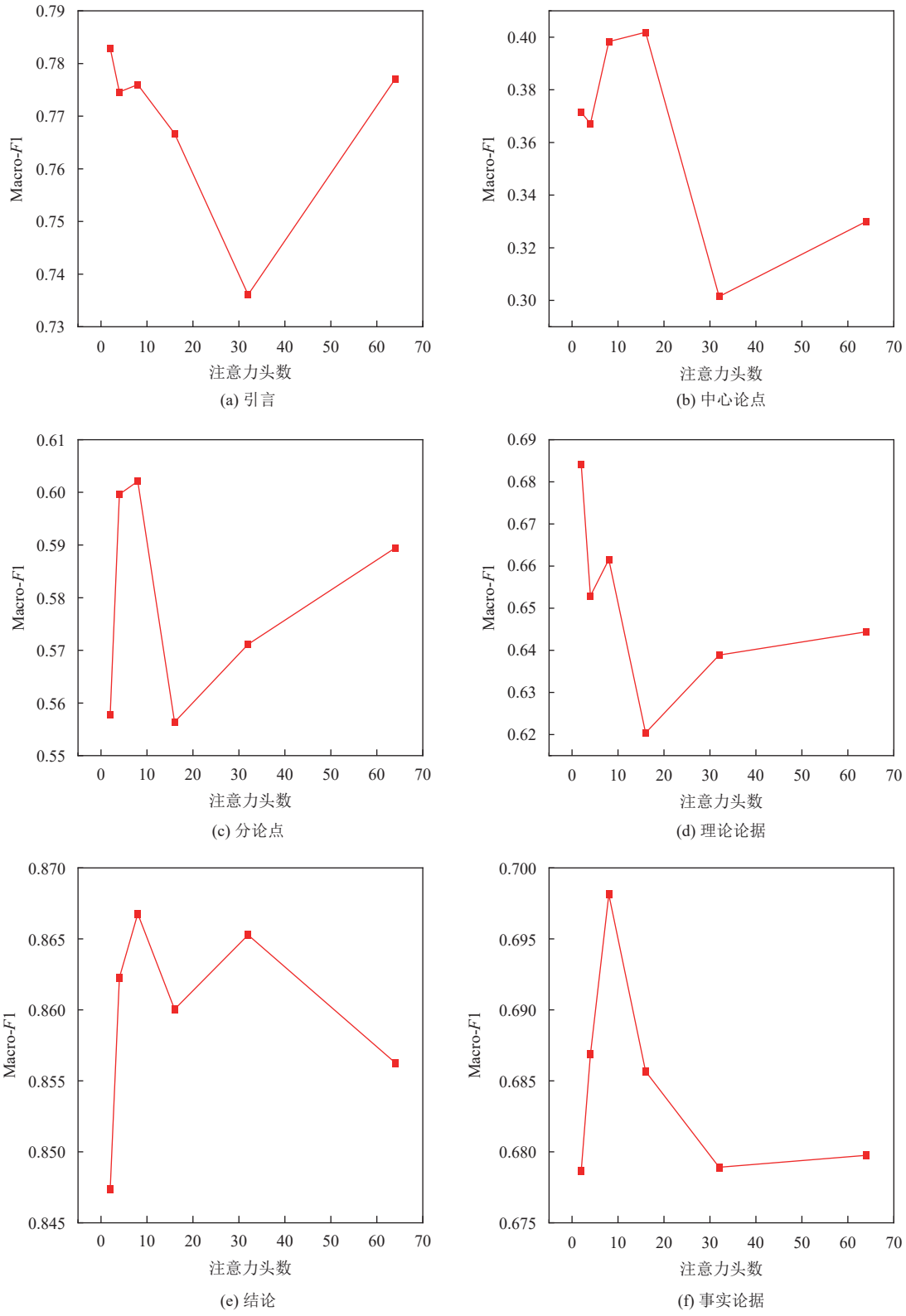
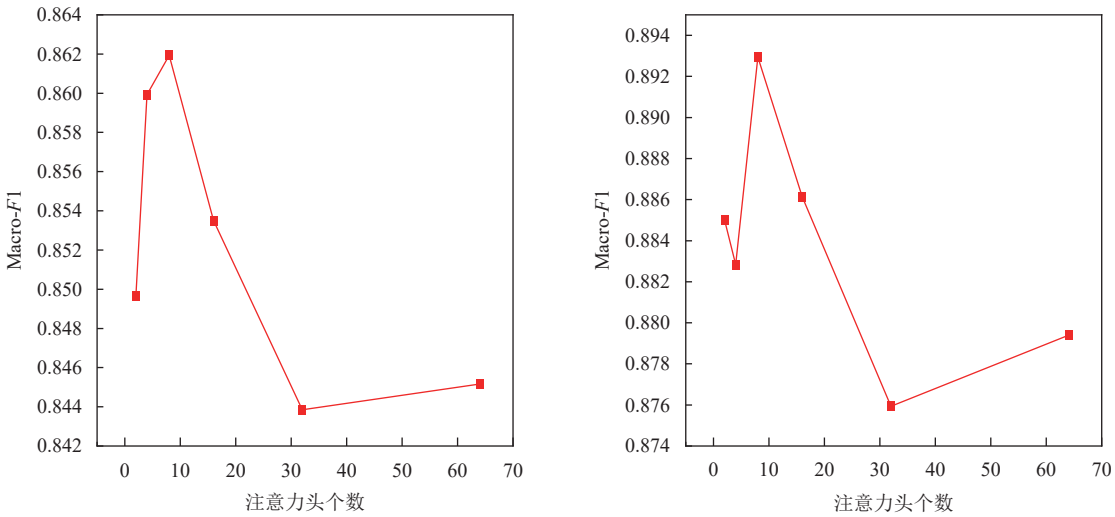


图 7 识别特定篇章要素的对比分析

图 8 所示是英文数据集上的识别特定篇章要素的分析对比图. 由图 8 中结果显示, 当注意力头个数为 8 时, 其整体效果是最好的. 当 n 为 2 时, 论点的准确率最低; 当 n 为 4 时, 论点要素准确率有明显提升, 但主旨要素准确率较之前比有所下降; 当 n 为 16 时, 论点和主旨要素的准确率都有所下降; 当 n 为 32 时, 论点和主旨要素的准确率最低; 当 n 为 64 时, 较之前相比有些提升.



(a) 论点 (b) 主旨

图 8 识别特定篇章要素的对比分析

3 结论与展望

本文提出基于 BiLSTM-Attention 的议论文篇章要素识别模型, 采用注意力机制, 可以自动学习文本中的重要部分, 进一步提高篇章要素识别准确性. 句间多头自注意力模块能捕捉不同的关注点和特征表示, 将多个不同角度的注意力进行融合, 提供更全面和多样化的模型表达能力. 实验结果显示, 在中文数据集 Acc 提高了 1.3%, 在英文数据集 Acc 提高了 3.6%. 这表明改进的方法在中、英文数据集上都具有有效性和通用性. 改进后的方法在准确精度上有一定的提升, 为未来的作文评分工作提供良好的参考价值. 但对于英文提高效果优于中文提高效果, 表明该改进方法更加适用于英文数据集, 未来可以侧重中文研究, 考虑中文的语法等因素. 此外, 未来还可以从提高准确性和加快收敛速度、提高模型的语言理解和适应能力等方面进行, 在提高准确精度的同时提高收敛速度.

参考文献

1 樊子鹏, 张鹏, 高珲. 量子自然语言处理: 历史演变与新进展. 中文信息学报, 2023, 37(1): 1–15.

2 褚晓敏, 朱巧明, 周国栋. 自然语言处理中的篇章主次关系研究. 计算机学报, 2017, 40(4): 842–860. [doi: 10.11897/SP.J.1016.2017.00842]

3 蒋峰, 范亚鑫, 褚晓敏, 等. 英汉篇章结构分析研究综述. 软件学报, 2023, 34(9): 4167–4194. [doi: 10.13328/j.cnki.jos.006650]

4 Li PF, Zhou GD, Zhu QM, *et al*. Employing compositional semantics and discourse consistency in Chinese event extraction. Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Jeju Island: ACL, 2012. 1006–1016.

5 Chen C, Ng V. Joint modeling for Chinese event extraction with rich linguistic features. Proceedings of the 24th International Conference on Compatational Linguistics (COLING 2012). Mumbai: The COLING 2012 Organizing Committee, 2012. 529–544.

6 Li JW, Hovy E. A model of coherence based on distributed sentence representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha: ACL, 2014. 2039–2048. [doi: 10.3115/v1/D14-1218]

7 Song W, Liu LZ. Representation learning in discourse parsing: A survey. Science China Technological Sciences, 2020, 63(10): 1921–1946. [doi: 10.1007/s11431-020-1685-2]

8 Song W, Wang D, Fu RJ, *et al*. Discourse mode identification in essays. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver: ACL, 2017. 112–122. [doi: 10.18653/v1/P17-1011]

9 Liakata M, Dobnik S, Saha S, *et al*. A discourse-driven content model for summarising scientific articles evaluated in a complex question answering task. Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Seattle: ACL, 2013. 747–757. [doi: 10.18653/v1/d13-1070]

10 张迎, 张宜飞, 王中卿, 等. 基于主次关系特征的自动文摘方法. 计算机科学, 2020, 47(S1): 6–11. [doi: 10.11896/JsJkx.191000007]

11 Zou BW, Zhou GD, Zhu QM. Negation focus identification with contextual discourse information. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Baltimore: ACL, 2014. 522–530.

12 Chistova E, Smirnov I. Discourse-aware text classification for argument mining. Proceedings of the 2022 Computational Linguistics and Intellectual Technologies, Papers from the Annual International Conference “Dialogue”. 2022. 93. [doi: 10.28995/2075-7182-2022-21-93-105]

13 刘海顺, 王雷, 孙媛媛, 等. 基于预训练语言模型的案件要素识别方法. 中文信息学报, 2021, 35(11): 91–100.

14 Mim FS, Inoue N, Reisert P, *et al*. Corruption is not all bad: Incorporating discourse structure into pre-training via corruption for essay scoring. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2021, 29: 2202–2215. [doi: 10.1109/TASLP.2021.3088223]

15 Song W, Fu RJ, Liu LZ, *et al*. Discourse element identification in student essays based on global and local cohesion. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon: ACL, 2015. 2255–2261.

16 Fu J, Liu J Tian HJ, *et al*. Dual attention network for scene segmentation. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019. 3141–3149. [doi: 10.1109/CVPR.2019.00326]

17 张周彬, 邵党国, 马磊, 等. 一种循环互作用注意力的属性级情感分类模型. 计算机应用与软件, 2020, 37(5): 140–144, 150.

18 程艳, 尧磊波, 张光河, 等. 基于注意力机制的多通道CNN和BiGRU的文本情感倾向性分析. 计算机研究与发展, 2020, 57(12): 2583–2595. [doi: 10.7544/issn1000-1239.2020.20190854]

19 Jianlin Su JL, Ahmed M, Lu Y, *et al*. RoFormer: Enhanced Transformer with Rotary Position Embedding. Neurocomputing, 2024, 568(C): 127063. [doi: 10.1016/j.neucom.2023.127063]

20 Daxenberger J, Eger S, Habernal I, *et al*. What is the essence of a claim? Cross-domain claim identification. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen: ACL, 2017. 2055–2066. [doi: 10.18653/v1/D17-1218]

21 肖琳, 陈博理, 黄鑫, 等. 基于标签语义注意力的多标签文本分类. 软件学报, 2020, 31(4): 1079–1089. [doi: 10.13328/j.cnki.jos.005923]

22 Song W, Song ZY, Fu RJ, *et al*. Discourse self-attention for discourse element identification in argumentative student essays. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. ACL, 2020. 2820–2830. [doi: 10.18653/v1/2020.emnlp-main.225]

23 Song W, Song ZY, Liu LZ. Hierarchical multi-task learning for organization evaluation of argumentative student essays. Proceedings of the 29th International Joint Conferences on Artificial Intelligence. Yokohama, 2021. 536. [doi: 10.24963/ijcai.2020/536]

24 Sun ZH, Jiang F, Li PF, *et al*. Macro discourse relation recognition via discourse argument pair graph. Proceedings of the 9th CCF International Conference on Natural Language Processing and Chinese Computing. Zhengzhou: Springer, 2020. 108–119.

25 Wang SJ, Zhang ZW, Dou Y, *et al*. Discourse component recognition via graph neural network in Chinese student argumentative essays. Proceedings of the 15th International Conference on Knowledge Science, Engineering and Management. Singapore: Springer, 2022. 358–373. [doi: 10.1007/978-3-031-10983-6_28]

26 余正涛, 樊孝忠, 郭剑毅, 等. 基于潜在语义分析的汉语问答系统答案提取. 计算机学报, 2006, 29(10): 1889–1893. [doi: 10.3321/j.issn:0254-4164.2006.10.021]

27 Zhang Y, Kamigaito H, Okumura M. A language model-based generative classifier for sentence-level discourse parsing. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Punta Cana: ACL, 2021. 2432–2446. [doi: 10.18653/v1/2021.emnlp-main.188]

28 Burstein J, Marcu D, Knight K. Finding the WRITE stuff: Automatic identification of discourse structure in student essays. IEEE Intelligent Systems, 2003, 18(1): 32–39. [doi: 10.1109/MIS.2003.1179191]

29 Stab C, Gurevych I. Parsing argumentation structures in persuasive essays. Computational Linguistics, 2017, 43(3): 619–659. [doi: 10.1162/COLI_a_00295]

30 Devlin J, Chang MW, Lee J, *et al*. BERT: Pre-training of deep bidirectional Transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: ACL, 2019. 4171–4186.

(校对责编: 张重毅)