

基于异质图卷积神经网络的论点对抽取模型

刘议丹, 朱小飞, 尹雅博
(重庆理工大学 计算机科学与工程学院, 重庆 400054)

摘要: 针对论点对抽取任务中存在着评审段和反驳段之间交互信息难以捕获以及忽略了对句子间的相对位置信息进行建模问题, 提出基于异质图卷积神经网络的论点对抽取模型. 该模型在评审段和反驳段中构建异质图, 定义 2 种不同类型的节点及 4 种不同类型的边, 通过关系图卷积神经网络来更新图中节点的表示. 提出位置感知的句子对生成器, 利用旋转位置编码来建模评审段和反驳段句子间的相对位置信息. 在 RR-passage 和 RR-submission-v2 数据集上进行实验, 实验结果表明, 提出模型的性能均优于所有的基线模型. 这表明通过构建异质图区分不同的节点类型和边的类型, 设计位置感知的句子对生成器, 能够提升论点对抽取模型的效果.

关键词: 论辩挖掘; 论点对抽取; 图神经网络; 旋转位置编码; 自然语言处理

中图分类号: TP 393 **文献标志码:** A **文章编号:** 1008-973X(2024)05-0001-09

Heterogeneous graph convolutional neural network for argument pair extraction

LIU Yidan, ZHU Xiaofei, YIN Yabo

(College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China)

Abstract: An argument pair extraction model based on heterogeneous graph convolutional neural network was proposed aiming at the issue of difficulty in capturing interactive information between review passage and rebuttal passage and neglecting to model relative positional information between sentences. Heterogeneous graphs were constructed within the review passage and rebuttal passage. Two types of nodes and four types of edges were defined. The relational graph convolutional neural network was utilized to update the representations of nodes within the graph. A position-aware sentence pair generator was introduced, and rotary position embedding was employed to model the relative positional information between sentences in review passage and rebuttal passage. Experimental evaluations on the RR-passage and RR-submission-v2 datasets demonstrate that the proposed model outperforms all baseline models. The performance of the argument pair extraction model can be enhanced by constructing heterogeneous graphs to distinguish between different types of nodes and edges and designing a position-aware sentence pair generator.

Key words: argument mining; argument pair extraction; graph neural network; rotary position embedding; natural language processing

论点对抽取 (argument pair extraction, APE) 属于对话式论辩, 目的是为了从 2 个相关的文章中抽取交互式论点对, 是论辩挖掘中的新兴任务^[1-5]. Cheng 等^[6] 首次在同行评审和反驳中提出论点对抽取任务, 因为同行评审和反驳中包含丰

富的论点. 它主要包含以下 2 个步骤: 1) 从评审段和反驳段中识别出论点; 2) 判断 2 个论点是否属于同一个论点对.

Cheng 等^[6] 将论点对抽取任务分解为序列标注任务和句子对分类任务, 通过多任务学习框架

收稿日期: 2023-07-03. 网址: www.zjujournals.com/eng/article/2024/1008-973X/202405000.shtml

基金项目: 国家自然科学基金资助项目 (62141201); 重庆市自然科学基金资助项目 (CSTB2022NSCQ-MSX1672); 重庆市教育委员会科学技术研究计划资助项目 (KJZD-M202201102); 重庆理工大学研究生教育高质量发展行动计划资助项目 (gzlxc20233230).

作者简介: 刘议丹 (1999—), 男, 硕士生, 从事论辩挖掘的研究. orcid.org/0009-0005-1101-8294. E-mail: lyd@stu.cqut.edu.cn
通信联系人: 朱小飞, 男, 教授. orcid.org/0000-0001-8239-7176. E-mail: zxf@cqut.edu.cn

来同时优化这 2 个子任务. 由于仅采用多任务学习的方式没有充分利用 2 个段落的独特特征以及段落之间句子的交互信息. Cheng 等^[7]提出注意力引导的多层多交叉模型, 隐式建模 2 个段落句子间的关系, 将句子对分类任务视为表格填充任务, 采用 2D-GRU 来更新表格的表示. Bao 等^[8]提出相互指导的框架, 通过评审段识别出的论点来指导反驳段论点的识别, 反之也可以通过反驳段识别出的论点来指导评审段论点的识别. Bao 等^[9]认为之前的工作建模的都是句子级别的交互, 忽略了对论点级别交互的建模, 提出具有两阶段的机器阅读理解 (machine reading comprehension, MRC) 框架.

尽管上述论点对抽取算法取得了一定的效果, 但任务存在以下 2 个问题. 1) 没有充分建模评审段和反驳段句子之间的交互信息. 2) 忽略了对评审段及反驳段内句子的相对位置进行建模. 为了解决上述 2 个问题, 本文提出基于异质图卷积神经网络的论点对抽取模型 (heterogeneous graph convolutional neural network for argument pair extraction, HGCN-APE). 在评审段和反驳段中构建异质图, 定义 2 种不同类型的节点、4 种不同类型的边, 通过关系图卷积神经网络来更新节点的表示. 除此之外, 设计位置感知句子对生成器, 建模评审段和反驳段内句子的相对位置信息.

总的来说, 本文的贡献主要可以分为以下几个方面.

(1) 针对以往基于图的论点对抽取工作, 难以充分建模评审段和反驳段句子间的交互关系问题, 构建异质图, 定义 2 种不同类型的节点及 4 种不同类型的边, 区分段落内和段落间句子的交互信息.

(2) 考虑到评审段和反驳段论点先后顺序有联系, 设计位置感知的句子对生成器, 采用旋转位置编码来建模段落内和段落间句子的相对位置信息.

(3) 在 RR-passage 和 RR-submission-v2 数据集上的实验结果表明, 本文提出的模型优于所有的基线模型. 为了促进论辩挖掘中论点对抽取的研究, 将代码进行开源: <https://github.com/ElevateSpirit/HGCN-APE>.

1 相关工作

1.1 论辩挖掘

论辩挖掘的目的是从具有辩论性的文本中自动提取出论点^[5], 近年来受到了越来越多的学者关注. 它在现实生活中有着广泛的应用, 如司法

庭审^[10]、AI 辩论^[11]、写作助手^[12]等. 目前对论辩挖掘的研究主要可以分为 2 个方面: 独白式论辩和对话式论辩. 以往的工作大多集中在独白式论辩的研究, 而相比于独白式论辩, 在现实生活中对话式论辩更常见. Ji 等^[13]基于在线辩论平台的数据构建数据集, 提出交互式论点对识别任务. Yuan 等^[14]提出使用外部知识来增强交互式论点对识别, 在上下文中对实体和路径进行编码, 以获得实体的嵌入和路径表示. Shi 等^[15]提出对比学习框架, 去除文档中与论点识别不相关的信息. Cheng 等^[6]在同行评审的反驳中提取论点对抽取任务, 将论点对抽取分解为序列标注和句子匹配任务, 通过多任务学习框架同时优化这 2 个子任务. Cheng 等^[7]提出注意力引导的多层多交叉模型, 隐式建模 2 个段落句子之间的关系. Bao 等^[8]提出互指导框架, 构建句间关系图, 显示建模 2 个段落句子之间的关系. Bao 等^[9]将论点对抽取视为两阶段阅读理解任务, 将识别出的论点作为问题, 在另一个段落中查询与之匹配的论点对.

1.2 图神经网络

图神经网络 (graph neural network, GNN) 包括图卷积网络 (graph convolutional network, GCN)^[16]、图注意力网络 (graph attention network, GAT)^[17]、关系图卷积网络 (relational graph convolutional network, RGCN)^[18] 等, 得益于其强大的表示能力, 在许多自然语言处理 (NLP) 任务中表现出优异的性能. Wang 等^[19]将关系图卷积网络应用于方面情感分析中. Hu 等^[20]提出基于异构图神经网络的方法, 用于半监督短文本分类.

最近, 图神经网络已经应用到了论点对抽取领域. Yuan 等^[14]提出利用外部知识来提升论点对抽取的效果, 在在线论坛数据集上构建论证知识图. Bao 等^[8]在评审反驳段落中构建句间关系图, 通过图卷积网络更新节点的表示. 与以往基于图的论点对抽取工作不同, 本文构建的图是异质的, 定义了不同节点的类型及边的类型, 通过关系图卷积神经网络建模句子之间的复杂关系.

2 模型简介

2.1 问题定义

给定评审段落 $V = \{v_1, v_2, \dots, v_m\}$ 和反驳段落 $B = \{b_1, b_2, \dots, b_n\}$, 其中 m 和 n 分别表示评审段中有 m 个句子, 反驳段中有 n 个句子. 在每个段落中, 可以将句子划分为论点和非论点, 评审段落中的论

点 $A^v = \{a_1^v, a_2^v, \dots, a_{l_v}^v\}$, 反驳段落中的论点 $A^b = \{a_1^b, a_2^b, \dots, a_{l_b}^b\}$, 其中 l_v 和 l_b 分别表示评审段中有 l_v 个论点, 反驳段落中有 l_b 个论点. 论点抽取的目的是从评审段落和反驳段落中匹配讨论同一个话题的论点 $P = \{p_1, p_2, \dots, p_{l_p}\}$, 其中 l_p 为评审反驳段中论点对的个数. 提出的 HGCN-APE 模型架构如图 1 所示.

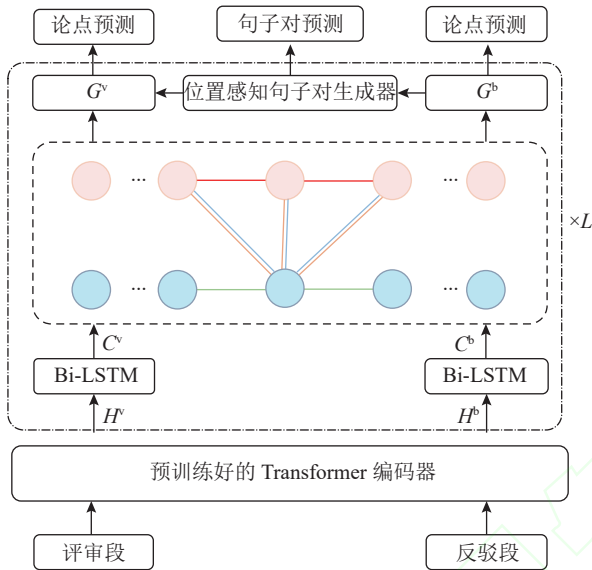


图 1 HGCN-APE 模型架构

Fig.1 Model architecture of HGCN-APE

2.2 句子编码器

对于评审段落中或者反驳段落中的每一个句子 $s_i = \{w_1, w_2, \dots, w_{|s_i|}\}$, 其中 $|s_i|$ 为句子中单词的个数, 在每个句子的开头和结尾分别拼接上 2 个特殊的字符 [CLS] 和 [SEP], 将拼接后的句子输入到预训练好的 Transformer^[21] 编码器 BERT^[22] 中, 得到句子中每个词的表示:

$$\mathbf{H}^{s_i} = [\mathbf{h}_0^{s_i}, \mathbf{h}_1^{s_i}, \dots, \mathbf{h}_{|s_i|+1}^{s_i}]^T = \text{BERT}(s_i). \quad (1)$$

式中: $\mathbf{H}^{s_i} \in \mathbf{R}^{(|s_i|+2) \times d}$, 其中 d 为隐藏层的维度.

对句子中每个词的表示进行加权求和, 得到最终句子的表示. 这样做的目的是充分利用句子中每个单词的语义信息:

$$\alpha = \text{softmax}(\mathbf{H}^{s_i} \mathbf{W} + \mathbf{b}), \quad (2)$$

$$\tilde{\mathbf{h}}^{s_i} = \sum_{j=1}^{|s_i|} \alpha_j \mathbf{h}_j^{s_i}. \quad (3)$$

式中: $\alpha \in \mathbf{R}^{|s_i|}$, $\mathbf{h}^{s_i} \in \mathbf{R}^d$ 为第 s_i 个句子的表示, \mathbf{W} 和 \mathbf{b} 为可训练的参数.

为了提升模式的泛化能力, 引入 Dropout^[23] 策略, 随机丢弃网络中部分神经元间的连边:

$$\mathbf{h}^{s_i} = \text{Dropout}(\tilde{\mathbf{h}}^{s_i}). \quad (4)$$

通过上述步骤, 可以得到评审段落和反驳段落中每个句子的表示: $\mathbf{H}^v = [\mathbf{h}_1^v, \mathbf{h}_2^v, \dots, \mathbf{h}_m^v]^T$, $\mathbf{H}^b = [\mathbf{h}_1^b, \mathbf{h}_2^b, \dots, \mathbf{h}_n^b]^T$.

为了捕获每个段落句子之间的上下文信息, 将评审段和反驳段中所有句子的表示输入到 BiLSTM^[24] 中, 把前项和后项的表示拼接在一起, 得到上下文感知的句子的表示:

$$\mathbf{C}^v = [\mathbf{c}_1^v, \mathbf{c}_2^v, \dots, \mathbf{c}_m^v]^T = \text{BiLSTM}(\mathbf{H}^v), \quad (5)$$

$$\mathbf{C}^b = [\mathbf{c}_1^b, \mathbf{c}_2^b, \dots, \mathbf{c}_n^b]^T = \text{BiLSTM}(\mathbf{H}^b). \quad (6)$$

式中: $\mathbf{C}^v \in \mathbf{R}^{m \times 2d}$, $\mathbf{C}^b \in \mathbf{R}^{n \times 2d}$.

2.3 异质图卷积层

在得到评审段落和反驳段落中所有句子的表示后, 为了建模段落内部和段落之间句子的复杂关系, 构建异质图, 如图 2 所示.

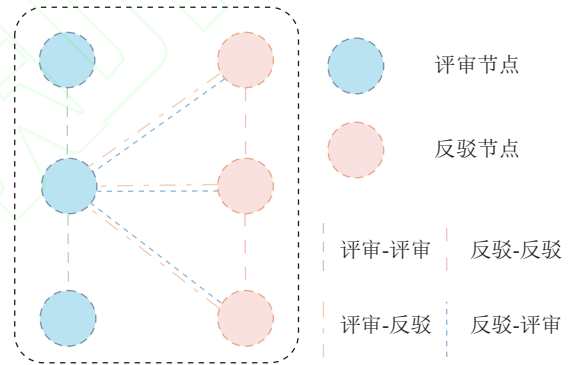
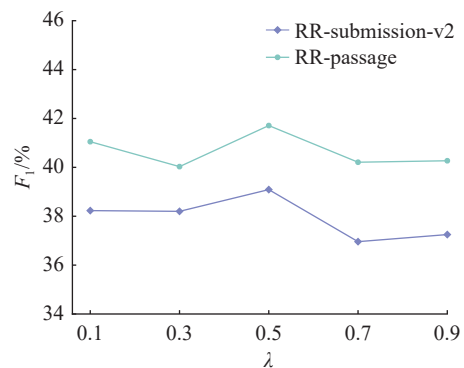


图 2 HGCN-APE 异质图构建

Fig.2 Heterogeneous graph construction of HGCN-APE

2.3.1 异质图构建 如图 3 所示, 在构建的异质图中, 包含了 2 种类型的节点, 分别是评审段落中的句子节点和反驳段落中的句子节点以及 4 种不同类型的边.

评审-评审边. 评审-评审边用于建模评审段内

图 3 HGCN-APE 超参数 λ 的影响Fig.3 Effect of HGCN-APE hyper-parameter λ

部句子之间的关系, 评审段中相邻 2 个句子之间有连边。

反驳-反驳边. 反驳-反驳边用于建模反驳段内部句子之间的关系, 反驳段中相邻 2 个句子之间有连边。

评审-反驳边. 评审-反驳边用于建模评审反驳段可能属于同一个论点对句子之间的关系, 用评审节点的表示来增强反驳节点的表示. 参考 Bao 等^[8]的工作, 利用 NLTK 对评审段和反驳段中的每个句子进行分词处理, 并去除停用词, 只有当去除停用词的 2 个句子之间的共现词数量大于 2 时, 才会有连边。

反驳-评审边. 反驳-评审边的目的是用反驳段节点的表示去增强评审段节点的表示。

2.3.2 异质图聚合 在构建好异质图后, 使用句子编码器得到的评审段落中每个句子的表示以及反驳段落中每个句子的表示, 分别初始化图中评审节点的表示和反驳节点的表示。

$$\mathbf{G}^{v_0} = \mathbf{C}^v = [c_1^v, c_2^v, \dots, c_m^v]^T, \quad (7)$$

$$\mathbf{G}^{b_0} = \mathbf{C}^b = [c_1^b, c_2^b, \dots, c_n^b]^T, \quad (8)$$

$$\mathbf{G}^0 = [g_1^0, g_2^0, \dots, g_{m+n}^0]^T. \quad (9)$$

使用关系图卷积网络 (relational graph convolutional network, RGCN)^[18] 来更新节点的表示:

$$g_i = \sigma \left(\sum_{r \in R} \sum_{j \in N_i^r} \frac{1}{|N_i^r|} \mathbf{W}_r g_j^0 + \mathbf{W}_0 g_i^0 \right). \quad (10)$$

式中: σ 为激活函数, R 为定义的所有关系类型, N_i^r 为节点 i 在关系 r 下的所有邻居集合, \mathbf{W}_r 和 \mathbf{W}_0 均为可训练的参数. 最终得到更新后的评审段落中句子的表示 $\mathbf{G}^v = [g_1^v, g_2^v, \dots, g_m^v]^T \in \mathbf{R}^{m \times 2d}$ 以及反驳段落中句子的表示 $\mathbf{G}^b = [g_1^b, g_2^b, \dots, g_n^b]^T \in \mathbf{R}^{n \times 2d}$.

2.4 位置感知句子对生成器

在论文投稿过程中, 当审稿人给出评审意见后, 作者会针对审稿人提出的意见进行回复. 事实上, 作者会按照审稿人提出评审意见的顺序进行回复, 然而以往的工作忽略对相对位置信息的建模。

采用旋转位置编码 (rotary position embedding, RoPE)^[25], 建模评审段和反驳段内句子间的相对位置关系:

$$\mathbf{PE}^v = \text{RoPE}(\mathbf{G}^v), \quad (11)$$

$$\mathbf{PE}^b = \text{RoPE}(\mathbf{G}^b). \quad (12)$$

式中: $\mathbf{PE}^v \in \mathbf{R}^{m \times 2d}$, $\mathbf{PE}^b \in \mathbf{R}^{n \times 2d}$ 分别为融合相对位置编码的评审段句子的表示和反驳段句子的表示。

将评审段和反驳段中两两句子的表示拼接在一起, 构成 $m \times n$ 的表格, 用于句子对预测:

$$T_{ij} = \text{ReLU} \left(\mathbf{W}^p [\mathbf{PE}_i^v; \mathbf{PE}_j^b]^T + \mathbf{b}^p \right). \quad (13)$$

式中: $T \in \mathbf{R}^{m \times n \times d}$ 为表格的表示, $\mathbf{W}^p \in \mathbf{R}^{d \times 4d}$, $\mathbf{b}^p \in \mathbf{R}^d$ 为可训练的参数, $[\cdot; \cdot]$ 表示拼接操作。

BiLSTM 层和异质图卷积层可堆叠多层来更新表格的表示, 不同层之间采用残差连接, 并进行层归一化:

$$\mathbf{G}^{v_l} = \text{LayerNorm}(\mathbf{G}^{v_l} + \mathbf{G}^{v_{l-1}}), \quad (14)$$

$$\mathbf{G}^{b_l} = \text{LayerNorm}(\mathbf{G}^{b_l} + \mathbf{G}^{b_{l-1}}), \quad (15)$$

$$\mathbf{T}^l = \text{LayerNorm}(\mathbf{T}^l + \mathbf{T}^{l-1}). \quad (16)$$

式中: l 为堆叠的层数, $\text{LayerNorm}(\cdot)$ 表示层归一化。

2.5 预测层

2.5.1 论点预测 给定评审段的句子序列 $V = \{v_1, v_2, \dots, v_m\}$ 和反驳段中的句子序列 $B = \{b_1, b_2, \dots, b_n\}$ 以及它们对应的标签 $Y^v = \{y_1^v, y_2^v, \dots, y_m^v\}$, $Y^b = \{y_1^b, y_2^b, \dots, y_n^b\}$, 论点预测的目的是给每一个句子分配一个标签, 是序列标注问题, 通过条件随机场 (conditional random field, CRF)^[26] 来实现。

具体来说, 以评审段的句子序列为例 (反驳段的预测类似), 计算观测序列和标签序列之间的得分:

$$\text{score}(V, Y^v) = \sum_{i=0}^m A_{y_i^v; y_{i+1}^v} + \sum_{i=1}^m F(V, y_i^v). \quad (17)$$

式中: $A_{y_i^v; y_{i+1}^v}$ 为标签 y_i^v 到 y_{i+1}^v 的状态转移分数, $F(V, y_i^v)$ 为通过神经网络得到的发射分数。

计算给定 V 下标签 Y^v 的条件概率:

$$P(Y^v|V) = \frac{\exp(\text{score}(V, Y^v))}{\sum_{i=0}^{m+1} \exp(\text{score}(V, y_i^v))}. \quad (18)$$

CRF 的损失函数可以定义为

$$L^v = -\ln P(Y^v|V). \quad (19)$$

通过同样的步骤, 可以得到反驳段论点识别的损失 L^b . 论点预测的损失函数由 2 部分组成, 分别是评审段的论点预测损失 L^v 和反驳段的论点

预测损失 L^b ,

$$L_{am} = L^v + L^b. \quad (20)$$

2.5.2 句子对预测 句子对预测的目的是判断评审段和反驳段中两两句子是否属于同一个论点对, 是二分类问题. 通过多层感知机 (multi-layer perceptron, MLP), 得到句子对概率分布:

$$P_{ij} = \text{MLP}(T_{ij}). \quad (21)$$

式中: T_{ij} 为评审段中第 i 个句子和反驳段中第 j 个句子在表格中的表示, P_{ij} 为这两个句子属于同一个论点对的概率。

使用交叉熵损失, 计算预测值与真实值的偏差:

$$L_{\text{pair}} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n - [y_{ij} \ln P_{ij} + (1 - y_{ij}) \times \ln(1 - P_{ij})]. \quad (22)$$

式中: $y_{ij} \in \{0, 1\}$ 为句子对的真实标签, 0 表示 2 个句子不属于同一个论点对, 1 表示 2 个句子属于同一个论点对。

2.6 模型优化及预测

模型最终的损失函数包含 2 个部分: 论点预测和句子对预测. 采用多任务学习的方式同时优化这 2 个损失, 最终的损失函数定义为

$$L = \lambda L_{am} + (1 - \lambda) L_{\text{pair}}. \quad (23)$$

式中: λ 为超参数, 用于调节 2 个子任务损失的权重。

参照 Cheng 等^[6-7] 的工作, 最终论点对的预测由 2 个子任务的结果组合而来. 具体来说, 只有当识别出的评审段论点 (v_{i_1}, \dots, v_{i_n}) 和反驳段论点 (b_{j_1}, \dots, b_{j_n}) 满足如下条件时, 这 2 个论点才属于同一个论点对。

$$\frac{\sum_{i=i_1}^{i_n} \sum_{j=j_1}^{j_n} I(P_{ij})}{(j_n - j_1 + 1)(i_n - i_1 + 1)} \geq 0.5. \quad (24)$$

式中: I 为指示函数。

3 实验分析

3.1 实验设置

3.1.1 数据集和评估指标 采用论点对抽取广泛应用的评审-反驳数据集 (review-rebuttal, RR)^[6] 来评估模型的性能, 该数据集收集了 ICLR 上的 4 764 篇评审反驳对. 它包含 2 个版本: RR-Passage 和 RR-Submission-v2. 每一个版本按照 8:1:1

的比例, 划分为训练集、验证集和测试集. 在 RR-Passage 中, 同一篇文章的论点对均包含在训练集、验证集或者测试集中的一个; 在 RR-Submission-v2 中, 同一篇文章的论点对可能同时分布在训练集、验证集和测试集中. 数据集的统计结果如表 1 所示, 表中, N_{vb} 为评审反驳对数, N_{tr} 为训练集数, N_{dev} 为验证集数, N_t 为测试集数, N_f 为句子总数, N_a 为论点总数, N_s 为论点句子总数。

表 1 论点对抽取数据集统计

Tab.1 Statistics of argument pair extraction dataset

评审反驳			评审				反驳		
N_{vb}	N_{tr}	N_{dev}	N_t	N_f	N_a	N_s	N_f	N_a	N_s
4764	3811	476	477	99 800	23 200	58 500	94 900	17 700	67 500

在论点预测、句子对预测、论点对预测上的评估指标均采用精确率 P (precision)、召回率 R (recall) 和 F_1 。

3.1.2 基线模型 为了验证提出模型的有效性, 与以下的基线模型进行比较。

1) MT-H-LSTM-CRF^[6]. MT-H-LSTM-CRF 将论点对抽取任务分解为序列标注任务和句子对预测任务, 通过多任务学习框架来同时优化这 2 个子任务。

2) MLMC^[7]. MLMC 提出多层多交叉编码器模型来充分建模 2 个段落句子之间的关系, 将句子对预测视为表格填充问题, 引入注意力损失来更好地预测论点对。

3) MGF^[8]. MGF 是互指导框架, 包含评审指导和反驳指导. 先识别出评审段中的论点, 再将该论点拼接到反驳段中, 找出与之匹配的论点, 反之亦然。

4) MRC-APE-Bert^[9]. MRC-APE-Bert 将论点对抽取任务视为两阶段阅读理解任务: 第 1 阶段先识别出评审段和反驳段中的所有论点, 第 2 阶段将每个识别的论点作为查询, 在另一个段落中找与之匹配的论点. 由于 MRC-APE^[9] 使用 Longformer^[27] 对句子进行编码, 为了公平比较, 本文采用其 Bert 版模型作为基线。

3.2 实验简介

HGCA-APE 模型基于 Pytorch 实现, 使用 BERT-Base-Uncased 作为基准模型, 在训练过程中冻结它们的参数. 采用 AdamW^[28] 优化器优化模型, 使用余弦退火策略在每个轮次后调整学习率. 训练轮数设置为 25, 批处理大小设置为 1. 在 2 个

数据集上的超参数 λ 均设置为 0.5, 堆叠层数 L 设置为 3, 图卷积的层数设置为 1, 丢弃率设置为 0.5. 所有的实验均用不同的随机种子运行 5 次, 最终结果为 5 次实验的平均值.

为了验证提出模型的有效性, 在 2 个基准数据集 RR-submission-v2 和 RR-passage 上进行实现, 与一系列的基线模型进行比较.

如表 2 所示为模型在论点对抽取及 2 个子任务上的结果. 表中, “—”表示这些基线模型没有将论点对抽取任务视为序列标注和句子对预测任务, 最好的结果用加粗表示, 次优的结果加“_”.

表 2 HGCN-APE 在 RR-passage 和 RR-submission-v2 数据集上的性能对比

Tab.2 Comparison of performance of HGCN-APE on RR-passage and RR-submission-v2 dataset

数据集	模型	论点对预测			句子对预测			论点对预测		
		P/%	R/%	F ₁ /%	P/%	R/%	F ₁ /%	P/%	R/%	F ₁ /%
RR-submission-v2	MT-H-LSTM-CRF	70.74	69.46	70.09	52.05	46.74	49.25	27.24	26.00	26.61
	MLMC	69.53	73.27	71.35	<u>60.01</u>	<u>46.82</u>	<u>52.60</u>	37.15	29.38	32.81
	MGF	70.40	71.87	71.13	—	—	—	34.23	<u>34.57</u>	34.40
	MRC-APE-Bert	73.36	68.35	70.77	—	—	—	<u>42.26</u>	34.06	<u>37.72</u>
	HGCN-APE	<u>71.86</u>	<u>71.80</u>	71.83	66.12	59.40	62.58	42.70	36.05	39.09
RR-passage	MT-H-LSTM-CRF	71.85	71.01	71.43	54.28	43.24	48.13	30.08	29.55	29.81
	MLMC	66.79	72.17	69.37	62.49	42.33	50.53	<u>40.27</u>	29.53	34.07
	MGF	73.62	70.88	72.22	—	—	—	38.03	<u>35.68</u>	<u>36.82</u>
	MRC-APE-Bert	66.81	69.84	68.29	—	—	—	34.70	35.53	35.51
	HGCN-APE	<u>72.50</u>	<u>71.61</u>	<u>72.05</u>	67.68	59.25	63.18	45.76	38.32	41.71

3.3 消融实验

为了探究构建的异质图及位置感知句子对生成器的有效性, 移除模型的某一部分进行消融实验.

w/o v2v 表示去掉异质图中的评审-评审边, 即图中不再建模评审段内句子之间的关系.

w/o b2b 表示去掉异质图中的反驳-反驳边, 即图中不再建模反驳段内句子之间的关系.

w/o v2b 表示去掉评审-反驳边, 即反驳段中的句子将不再建模评审段句子的信息.

w/o b2v 表示去掉反驳-评审边, 即评审段不再建模反驳段句子的信息.

w/o pos 表示去掉评审段和反驳段句子的相对位置信息.

如表 3 所示为去掉模型中某一模块后, 在 RR-Passage 数据集论点对抽取上的性能. 可以看出, 去掉任意一个模块, 模型的性能均有所降低, 证明了模型的有效性. 具体来说, 在异质图中, 去掉

从表 2 可以看出, 本文模型在 2 个基准数据集上的结果均达到了最优的结果, 在 RR-submission-v2 数据集上提升了 3.63%, 在 RR-passage 数据集上提升了 13.28%.

本文模型在论点预测性能方面不如 MGF, 在句子对预测子任务上的性能均优于其他基线模型. 一个可能的原因是通过共现词构建的异质图, 能够很好地建模潜在属于同一个论点对句子之间的关系, 也可能引入噪音, 即有连边的 2 个句子可能不属于同一个论点对, 这会影响模型在论点预测子任务上的性能.

评审-反驳边和反驳-评审边模型的性能, 比去掉评审-评审边及反驳-反驳边下降的性能更多, 进一步表明在论点对抽取任务中, 难点在于建模 2 个段落句子之间的关系. 在去掉位置编码之后, 模型性能有所下降, 说明有必要建模句子之间的相对位置.

表 3 HGCN-APE 在 RR-passage 数据集上的消融实验

Tab.3 Ablation study of HGCN-APE on RR-passage dataset

模型	论点对预测		
	P/%	R/%	F ₁ /%
HGCN-APE	45.76	38.32	41.71
w/o v2v	44.36	38.26	41.09
w/o b2b	44.69	37.63	40.86
w/o v2b	41.58	35.63	38.38
w/o b2v	42.68	32.84	37.12
w/o pos	44.56	37.68	40.83

3.4 参数敏感性实验

超参数 λ 控制2个子任务权重的大小,如图3所示为不同 λ 对模型性能的影响,随着 λ 的增大,论点预测子任务的权重越来越大,句子对预测的任务的权重越来越小.当 $\lambda=0.5$ 时,2个子任务的权重一样,同时模型在2个数据集上的性能均达到最佳,这说明论点预测和句子对预测在论点对抽取中是同等重要的.

为了探究不同堆叠层数 L 对模型性能的影响,将 L 设置为1~5,步长为1,在2个数据集上验证不同 L 对模型性能的影响,实验结果如表4所示.可以看出,随着 L 的增大,模型在论点对抽取任务上的性能不断提升,当 $L=3$ 时模型在2个数据集上的效果最佳,之后模型的性能不断降低.这是因为当 L 过小时,模型不能捕获句子之间的复杂关系;当 L 过大时,模型可能引入更多的噪音,导致性能降低.

表4 HGCN-APE 超参数 L 的影响

Tab.4 Effect of HGCN-APE hyper-parameter L

数据集	L	$P/\%$	$R/\%$	$F_1/\%$
RR-passage	1	33.47	25.94	29.23
	2	44.03	35.11	39.02
	3	45.76	38.32	41.71
	4	43.77	39.74	41.66
	5	44.49	38.11	41.05
RR-submission-v2	1	33.47	25.94	29.23
	2	40.47	32.50	36.05
	3	42.70	36.05	39.09
	4	42.02	35.68	38.59
	5	40.98	34.15	37.25

3.5 学习曲线实验

如图4所示为模型在训练过程中的收敛情况.随着训练轮数 N 的增加,模型在2个数据集上的

F_1 随着增大,损失不断减少.当训练轮数达到20的时候, F_1 到达峰值,之后随着训练轮数的增大, F_1 在峰值上下抖动.尽管20轮之后,损失不断降低,但是没有带来性能的提升,可能是因为模型在验证集上过拟合了.

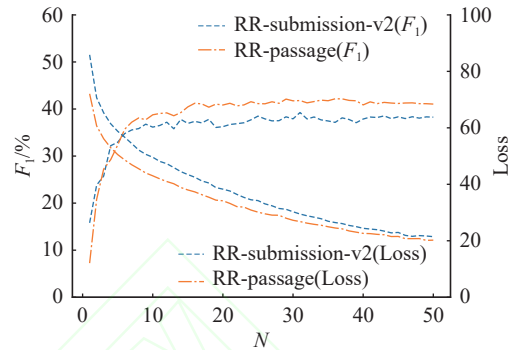


图4 HGCN-APE在RR-passage和RR-submission-v2数据集上的损失和性能对比

Fig.4 Comparison of loss and performance of HGCN-APE on RR-passage and RR-submission-v2 dataset

3.6 案例研究实验

为了更加直观地查看HGCN-APE模型在论点预测和论点对预测上的预测结果,在RR-submission-v2数据集的测试集上随机挑选3条数据进行展示,结果数据如表5所示.

从表5可以看出,在第1条样本中,HGCN-APE错误地将评审段中的第6、7句话识别为论点,这导致在论点对预测过程中,错误地将评审段中的6、7句话与反驳段中的第1~7句话识别为一个论点对.在第2条样本中,HGCN-APE将第17句话识别为论点,实际上第16、17句话是一个完整的论点,这表明HGCN-APE存在论点边界难以识别的问题,主要体现在2个方面:1)当一个论点包含多个句子时,只将部分句子识别成论点;2)多个连续的论点被识别为一个论点.第3条

表5 HGCN-APE 案例研究

Tab.5 Case study of HGCN-APE

样本	真实评审段论点	预测评审段论点	真实反驳段论点	预测反驳段论点	真实论点对	预测论点对
1	8-9	8-9	1-7	1-7	(8-9)-(1-7)	(6-7)-(1-7)
	10-15	10-15	8-13	8-13	(10-15)-(8-13)	(10-15)-(8-13)
	—	6-7	—	—	—	—
2	10-11	10-11	1-8	1-8	(10-11)-(1-8)	(10-11)-(1-8)
	12-12	12-12	9-11	9-11	(12-12)-(9-11)	(12-12)-(9-11)
	16-17	17-17	—	—	(16-17)-(1-8)	(17-17)-(1-8)
3	6-10	6-10	1-11	1-11	(6-10)-(1-11)	(6-10)-(1-11)
	12-13	12-13	14-18	14-18	(12-13)-(14-18)	(12-13)-(14-18)
	14-15	14-15	21-24	21-24	(14-15)-(21-24)	(14-15)-(21-24)

样本完全预测正确.

4 结 语

本文提出基于异质图卷积神经网络的论点对抽取模型, 定义 2 种不同类型的节点、4 种不同类型的边, 能够有效地建模评审段落和反驳段落内部句子之间的关系、段落之间潜在属于同一个论点对句子之间的关系. 设计位置感知的句子对生成器, 捕获段落内句子间的相对位置. 实验结果表明, 本文提出的模型优于现有的基线模型. 由于目前主要建模的是评审段落和反驳段落句子之间的交互, 模型在论点预测上不能取得最好的效果, 接下来将探究如何更好地建模论点级的交互.

参考文献 (References):

- [1] EGER S, DAXENBERGER J, GUREVYCH I. Neural end-to-end learning for computational argumentation mining [C]// **Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics**. Vancouver: ACL, 2017: 11-22.
- [2] KURIBAYASHI T, OUCHI H, INOUE N, et al. An empirical study of span representations in argumentation structure parsing [C]// **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics**. Florence: ACL, 2019: 4691-4698.
- [3] MORIO G, FUJITA K. End-to-end argument mining for discussion threads based on parallel constrained pointer architecture [C]// **Proceedings of the 5th Workshop on Argument Mining**. Brussels: ACL, 2018: 11-21.
- [4] CHAKRABARTY T, HIDEY C, MURESAN S, et al. AMPERSAND: argument mining for PERSuAsive oNline discussions [C]// **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing**. Hong Kong: ACL, 2019: 2933-2943.
- [5] 魏忠钰. 计算论辩技术: 迈向智能人类辩手之路 [J]. *世界科学*, 2023(5): 46-49.
WEI Zhongyu. Computational argumentation techniques: toward the path of intelligent human debaters [J]. *World Science*, 2023(5): 46-49.
- [6] CHENG L, BING L, YU Q, et al. APE: argument pair extraction from peer review and rebuttal via multi-task learning [C]// **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing**. [S. l.]: ACL, 2020: 7000-7011.
- [7] CHENG L, WU T, BING L, et al. Argument pair extraction via attention-guided multi-layer multi-cross encoding [C]// **Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing**. [S. l.]: ACL, 2021: 6341-6353.
- [8] BAO J, LIANG B, SUN J, et al. Argument pair extraction with mutual guidance and inter-sentence relation graph [C]// **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing**. Punta Cana: ACL, 2021: 3923-3934.
- [9] BAO J, SUN J, ZHU Q, et al. Have my arguments been replied to? argument pair extraction as machine reading comprehension [C]// **Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics**. Dublin: ACL, 2022: 29-35.
- [10] WALKER V, FOERSTER D, PONCE J M, et al. Evidence types, credibility factors, and patterns or soft rules for weighing conflicting evidence: argument mining in the context of legal rules governing evidence assessment [C]// **Proceedings of the 5th Workshop on Argument Mining**. Brussels: ACL, 2018: 68-78.
- [11] LE D T, NGUYEN C T, NGUYEN K A. Dave the debater: a retrieval-based and generative argumentative dialogue agent [C]// **Proceedings of the 5th Workshop on Argument Mining**. Brussels: ACL, 2018: 121-130.
- [12] ZHANG F, LITMAN D. Using context to predict the purpose of argumentative writing revisions [C]// **Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**. San Diego: ACL, 2016: 1424-1430.
- [13] JI L, WEI Z, LI J, et al. Discrete argument representation learning for interactive argument pair identification [C]// **Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**. [S. l.]: ACL, 2016: 5467-5478.
- [14] YUAN J, WEI Z, ZHAO D, et al. Leveraging argumentation knowledge graph for interactive argument pair identification [C]// **Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing**. [S. l.]: ACL, 2021: 2310-2319.
- [15] SHI L, GIUNCHIGLIA F, SONG R, et al. A simple contrastive learning framework for interactive argument pair identification via argument-context extraction [C]// **Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing**. Abu Dhabi: ACL, 2022: 10027-10039.
- [16] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks [EB/OL]. (2017-02-22). <https://arxiv.org/abs/1609.02907>.
- [17] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks [EB/OL]. (2017-10-30)[2023-06-20]. <https://arxiv.org/abs/1706.07538>.

arxiv.org/abs/1710.10903.

- [18] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks [C]//**The Semantic Web: 15th International Conference, Extended Semantic Web Conference**. Heraklion: Springer, 2018: 593-607.
- [19] WANG K, SHEN W, YANG Y, et al. Relational graph attention network for aspect-based sentiment analysis [C]// **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**. [S. l.]: ACL, 2020: 3229-3238.
- [20] HU L, YANG T, SHI C, et al. Heterogeneous graph attention networks for semi-supervised short text classification [C]//**Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing**. Hong Kong: ACL, 2019: 4821-4830.
- [21] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//**Proceedings of the 31st International Conference on Neural Information Processing Systems**. California: Curran Associates Inc., 2017: 6000-6010.
- [22] DEVLIN J, CHANG M W, LEE K, et al. Bert: pre-training of deep bidirectional transformers for language understanding [C]// **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics**. Minneapolis: ACL, 2019: 4171-4186.
- [23] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting [J]. **The Journal of Machine Learning Research**, 2014, 15(1): 1929-1958.
- [24] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. **Neural Computation**, 1997, 9(8): 1735-1780.
- [25] SU J, LU Y, PAN S, et al. Roformer: enhanced transformer with rotary position embedding [EB/OL]. (2021-04-09). <https://arxiv.org/abs/2104.09864>.
- [26] LAFFERTY J, MCCALLUM A, PEREIRA F C N. Conditional random fields: probabilistic models for segmenting and labeling sequence data [C]// **Proceedings of the 18th International Conference on Machine Learning**. San Francisco: Morgan Kaufmann, 2001: 282-289.
- [27] BELTAGY I, PETERS M E, COHAN A. Longformer: the long-document transformer [EB/OL]. (2020-05-10). <https://arxiv.org/abs/2004.05150>.
- [28] LOSHCHILOV I, HUTTER F. Decoupled weight decay regularization [EB/OL]. (2017-11-14). <https://arxiv.org/abs/1711.05101>.