计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
62(10): 2611−2621, 2025

基于多粒度信息交互和数据增强的隐式篇章关系识别

姚 浩[1] 熊锦晖[1] 李春生[2] 邬昌兴[1]
[1](华东交通大学信息与软件工程学院 南昌 330013)
[2](宁夏回族自治区卫生健康委信息中心 银川 750004)
(1468869487@qq.com)

Implicit Discourse Relation Recognition Based on Multi-Granularity Information Interaction and Data Augmentation

Yao Hao[1], Xiong Jinhui[1], Li Chunsheng[2], and Wu Changxing[1]
[1](*School of Information and Software Engineering*, *East China Jiaotong University*, *Nanchang* 330013)
[2](*Information Center*, *Health Commission of Ningxia Hui Autonomous Region*, *Yinchuan* 750004)

Abstract Implicit discourse relation recognition aims at automatically identifying semantic relations (such as Comparison) between two arguments (sentence or clause) in the absence of explicit connectives. Existing methods have confirmed that the introduction of phrase information can effectively boost the performance. However, there are still the following shortcomings: 1) These models typically rely on syntactic parsers and do not fully capture the interactions among words, phrases, and arguments. 2) The problem of data sparsity often occurs during training when incorporating the phrase information. To address the above issues, we propose an implicit discourse relation recognition model based on multi-granularity information interaction (MGII) and develop a chain decoding-inspired data augmentation method (DAM). Specifically, our proposed model is designed to automatically acquire semantic representations of *n*-grams using a stacked convolutional neural network. It then explicitly models the interactions among words, phrases and arguments based on Transformer layers and ultimately predicts multi-level discourse relationships in a chain-decoding way. DAM proposed in our paper simultaneously pretrains both the encoding and decoding modules, enabling the effective utilization of massive explicit discourse data, which are naturally annotated by connectives, to mitigate the issue of data sparsity. The proposed method significantly outperforms recent benchmark models on PDTB datasets. Furthermore, it does not rely on syntactic parsers, demonstrating strong applicability.

Key words implicit discourse relation recognition; multi-granularity information interaction; data augmentation; explicit discourse relation data; discourse analysis

摘 要 隐式篇章关系识别旨在推导没有连接词的论元(句子或子句)之间的语义关系.现有研究证实引入短语信息能切实提升识别性能,但依然存在以下不足:1)通常依赖句法分析器且词、短语与论元之间的交互不充分;2)引入短语信息导致的数据稀疏性问题.为此,分别提出基于多粒度信息交互的隐式篇章关系识别模型MGII(multi-granularity information interaction)和基于链式解码的数据增强方法DAM(data augmentation method).所提模型通过卷积神经网络自动学习*n*-gram短语的语义表示,利用Transformer层

收稿日期:2024-06-11;**修回日期**:2025-01-09
基金项目:国家自然科学基金项目(62266017, 62166018);江西省自然科学基金项目(20232BAB202050, 20242BAB25117)
This work was supported by the National Natural Science Foundation of China (62266017, 62166018) and the Natural Science Foundation of Jiangxi Province (20232BAB202050, 20242BAB25117).
通信作者:邬昌兴(wuchangxing@ecjtu.edu.cn)

显式地建模词、短语和论元之间的交互,并通过链式解码进行多级篇章关系预测.提出的数据增强方法同时预训练编码模块和解码模块,从而能有效地利用大量显式篇章关系数据.所提方法在PDTB数据集上的性能显著优于近期的基准模型,且不依赖句法分析器,具有较强的适用性.

关键词 隐式篇章关系识别;多粒度信息交互;数据增强;显式篇章关系数据;篇章分析

中图法分类号 TP391

DOI: 10.7544/issn1000-1239.202440511 **CSTR:** 32373.14.issn1000-1239.202440511

隐式篇章关系识别(implicit discourse relation recognition, IDRR)旨在自动推导没有连接词的2个论元(句子或子句)之间的语义关系,例如,常见的“转折关系”和“因果关系”等.通常把篇章关系组织为层次结构,如图1中例子的第1级篇章关系为“对比”,更细粒度的第2级篇章关系为“对比.转折”,隐式连接词为“但是”.由于缺乏具有较强指示意义的连接词,隐式篇章关系识别需要理解2个论元的语义,是一项极具挑战性的任务.作为篇章结构分析的关键子任务之一,其性能的提高将直接推动篇章结构朝实用化的方向发展.篇章关系识别的结果还可直接用于情感分析[1]、信息抽取[2]、文本摘要[3]和问答系统[4]等自然语言处理上层应用.近年来,虽然ChatGPT等大模型在众多自然语言处理任务上表现出色,但在隐式篇章关系识别上的效果依然不理想,离实用尚有较大的差距[5-6].

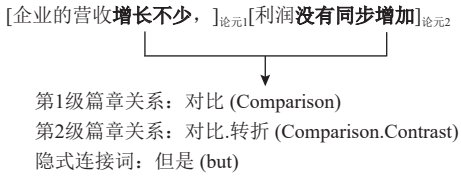


Fig. 1 An example of implicit discourse relation

图1 隐式篇章关系示例

现有研究通常基于预训练语言模型学习词在上下文中的表示,然后利用交叉注意力机制或图神经网络等建模论元间的语义交互[7-10].这类方法虽然取得了较好的效果,但没有显式地利用短语的信息.一般认为,短语是表达语义的一个基础且重要的语言单位,短语信息对于理解论元间的语义关系非常重要.如图1所示,该实例论元1中的短语“增长不少”和论元2中的短语“没有同步增加”是理解论元间“转折关系”的关键.为了利用短语信息,Wang等人[11]在短语句法分析结果上利用树结构神经网络编码每个论元的表示,拼接后用作分类模块的输入.最近,王秀利等人[12]首先使用句法分析器提取论元中的短语,然后构建短语级交互注意力机制用于隐式篇章关系识别.上述方法均取得了优于基准方法的效果,这充分说明了短语信息在该任务上的有效性.

然而,现有融合短语信息的隐式篇章关系识别方法仍存在3方面的不足:1)大多都依赖于句法分析器,这导致其实用性和适用性不强.一方面,使用句法分析器抽取论元中的短语会大幅增加时间消耗,不利于实际应用;另一方面,多数语言没有可用的句法分析器,例如,我国众多的少数民族语言,这较大地限制了其适用性.2)没有充分建模不同粒度信息之间的交互.例如,文献[11−12]强调短语之间的信息交互,而忽略了短语与词以及短语与论元之间的信息交互.3)需要更多的人工标注篇章关系数据用以训练.一方面,与建模词对信息相比,有效地建模短语对之间的信息需要更多的训练数据;另一方面,融合短语信息通常会导致神经网络模型的结构更复杂、参数量更多.然而,现有规模最大的PDTB 3.0语料中仅有约2万个隐式篇章关系实例,导致融合短语信息的方法在训练过程中常出现数据稀疏问题,从而影响模型的泛化性能.

本文提出一种基于多粒度信息交互和数据增强的隐式篇章关系识别方法.具体地:1)针对依赖句法分析器且没有充分建模不同粒度信息之间交互的问题,提出一种基于多粒度信息交互的隐式篇章关系识别模型MGII(multi-granularity information interaction).该模型通过多层卷积神经网络学习论元中n-gram短语(连续n个词构成的文本块)的语义表示;然后,基于Transformer层显式地充分建模词、短语和论元这3种不同粒度信息之间的交互;最后,利用链式解码模块同时识别输入实例的第1级和第2级篇章关系.模型的核心思想是:以n-gram短语代替语言学意义上的短语,从而消除对句法分析器的依赖,同时依托Transformer层强大的建模能力充分捕获多粒度信息之间的交互.2)针对人工标注训练数据不足的问题,提出了一种基于链式解码的数据增强方法DAM(data augmentation method),以有效地利用大量连接词自然标注的显式篇章关系数据.现有研究通常基于多任

务学习[13]或预训练方法[14]集成显式篇章关系数据，以缓解数据稀疏问题. 与现有方法仅基于显式篇章关系数据优化编码模块相比，提出的方法能同时优化编码模块和解码模块.

在常用数据集 PDTB 2.0 和 PDTB 3.0 上的实验结果表明，本文提出方法的性能明显好于近期的基准方法[8,12]. 另外，不依赖于句法分析器这一特性使得提出的方法具有较强的实用性和适用性.

1 相关工作

本节从基于深度学习的隐式篇章关系识别模型和显式篇章关系数据的利用这 2 个方面介绍相关研究工作.

1.1 基于深度学习的隐式篇章关系识别

就隐式篇章关系识别而言，基于深度学习的模型的性能早已全面超越基于人工特征的方法. 早期的研究基于浅层卷积神经网络[15]、循环神经网络[16]或树结构循环神经网络[11]分别学习 2 个论元的语义表示，并拼接它们用于分类. 随后，研究人员提出门控相关性网络[17]、多级注意力机制[18]或词交互模型[19]，以更好地建模论元之间的交互. 近年来，研究人员把 BERT 和 RoBERTa 等预训练语言模型作为基础编码模块，显著提升了识别的性能. 例如：Liu 等人[8]基于 RoBERTa 学习论元中词的表示，提出一种双向多视角匹配机制和门控融合机制用于捕获论元间的交互. Ma 等人[9]使用图神经网络同时建模论元内和论元间词的交互，取得了较好的效果. 王秀利等人[12]基于句法分析器输出的结果构建短语级交互注意力机制用于隐式篇章关系识别. 为了利用层次类别信息，Wu 等人[20]提出一种基于序列生成的多级隐式篇章关系识别方法，把类别依赖信息同时融入编码器和解码器中，取得了当时最好的效果. 随后，Long 等人[21]和 Jiang 等人[22]基于对比学习进一步利用层次类别信息，取得了更好的效果. 与文献 [8−9, 12, 20−21] 方法的主要不同之处在于，本文提出方法在不依赖句法分析器的前提下，充分建模了词、短语和论元这 3 种不同粒度的信息之间的交互.

基于提示学习的方法能有效地激发预训练模型的潜力，近 2 年来，在包括隐式篇章关系识别在内的众多自然语言处理任务上取得了非常好的效果. 例如，Xiang 等人[23]和 Zhou 等人[24]基于提示学习预测论元间的连接词，然后人工定义映射规则用于得到篇章关系类别. Zhao 等人[25]把层次类别信息融入基于提示学习的模型中，在仅优化较小参数的情况下取得了较好的性能. 准确地预测连接词是基于提示学习的模型的关键所在，为此，Wu 等人[26]基于知识蒸馏的方法利用篇章语料中额外标注的连接词信息. Liu 等人[27]提出一种 2 阶段的模型以模拟人类标注篇章关系的过程，其先基于提示学习的方法预测连接词，然后把连接词和论元输入到隐式篇章关系识别模型中. 虽然基于提示学习的方法取得了当前最好的性能，但 4.6 节的实验结果表明，结合本文提出的 MGII 模型可进一步提高模型的性能.

1.2 显式篇章关系数据的利用

大部分连接词（例如“但是”）具有较强的语义关系指示作用，因此，大量自然存在的显式篇章关系数据可潜在地用作隐式篇章关系识别的补充语料. 由于显式和隐式篇章关系数据在语言特性和数据分布等方面存在较大的差异，如何有效地利用这类数据一直是研究的热点. 例如，Liu 等人[13]基于多任务学习的方法，同时训练连接词分类和隐式篇章关系识别等 4 个相关任务. Wu 等人[28]基于连接词分类任务在大量显式篇章关系数据上学习任务相关的词向量，并用作隐式篇章关系识别模型的输入. Kishimoto 等人[14]则基于预训练的方法利用大量显式篇章关系数据，把连接词分类作为预训练任务，性能取得了较明显的提升. 近年来，Ru 等人[29]基于无监督的方法在大量显式篇章关系数据上学习连接词与篇章关系类别之间的语义关联. Wang 等人[30]通过基于提示学习的连接词预测任务，并结合基于互信息的代价函数，以把篇章相关知识注入预训练的大模型中. 与上述方法不同，本文提出一种基于链式解码的数据增强方法 DAM，其优点是可以利用大量显式篇章关系数据的同时优化模型的编码模块和解码模块.

2 MGII 模型

如图 2 所示，MGII 模型包括词编码模块、基于多层卷积的短语编码模块、基于 Transformer 层的多粒度信息交互模块和链式解码模块，并在多任务学习代价的引导下进行训练. 该模型具有 3 个方面的优点：1）以 *n*-gram 短语代替语言学上的短语，消除了对句法分析器的依赖；2）依托 Transformer 强大的建模能力充分捕获词、短语和论元这 3 种粒度信息之间的交互；3）引入的链式解码模块不仅能利用多级篇章关系类别之间的依赖性，还是本文所提数据增强方法的基础. 下面分别对相应模块进行介绍.

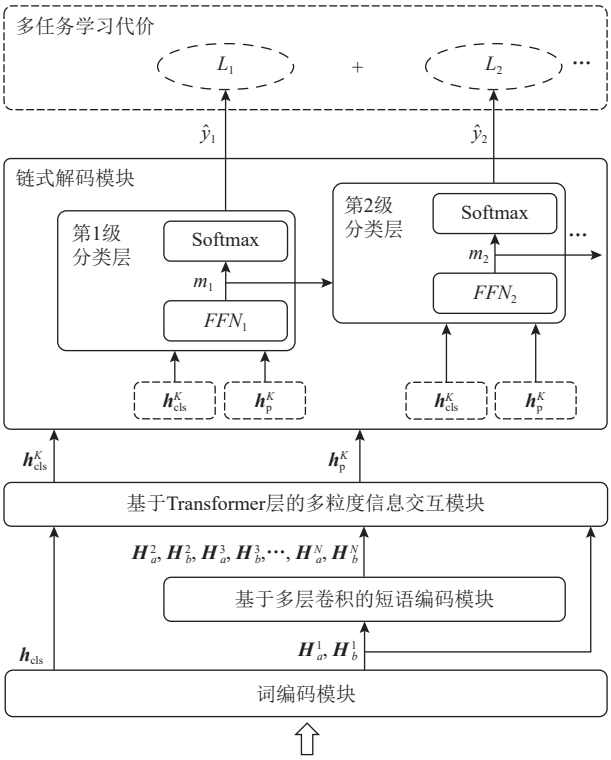


Fig. 2 MGII model

图 2 MGII 模型

2.1 词编码模块

词编码模块基于预训练语言模型学习论元和词的初始语义表示. 具体地，隐式篇章关系实例可表示为 $x=(arg_a, arg_b)$，其中 arg_a 和 arg_b 分别为该实例的第 1 论元和第 2 论元. 基于预训练语言模型学习论元和词的初始语义表示：

$$\boldsymbol{h}_{\text{cls}}, \boldsymbol{H}^1_a, \boldsymbol{H}^1_b = PLM(arg_a, arg_b), \tag{1}$$

其中 PLM 为预训练语言模型，2 个论元按照“[CLS]+arg_a+[SEP]+arg_b+[SEP]”的形式进行拼接，[CLS] 和 [SEP] 分别为预训练语言模型中常用的全局和分割占位符. $\boldsymbol{h}_{\text{cls}}$ 为 [CLS] 对应的输出向量，用作 2 个论元的初始语义表示，$\boldsymbol{H}^1_a$ 和 $\boldsymbol{H}^1_b$ 分别为由 arg_a 和 arg_b 中的词的初始语义表示构成的矩阵.

2.2 基于多层卷积的短语编码模块

该模块把论元中的 n-gram 看作短语，并基于多层卷积依次学习这些短语的向量表示. 以词的向量表示为基础，首先通过窗口为 2 的卷积层学习长度为 2 的短语表示，继而学习长度为 3 的短语表示，以此类推得到全部所需的短语表示. 具体地，给定词的初始语义表示矩阵 $\boldsymbol{H}^1_a$ 和 $\boldsymbol{H}^1_b$，基于多层卷积的短语编码模块的计算为：

$$\begin{aligned}
&\boldsymbol{H}^2_a, \boldsymbol{H}^2_b = CNN_2(\boldsymbol{H}^1_a), CNN_2(\boldsymbol{H}^1_b),\\
&\boldsymbol{H}^3_a, \boldsymbol{H}^3_b = CNN_3(\boldsymbol{H}^2_a), CNN_3(\boldsymbol{H}^2_b),\\
&\quad\vdots\\
&\boldsymbol{H}^N_a, \boldsymbol{H}^N_b = CNN_N(\boldsymbol{H}^{N-1}_a), CNN_N(\boldsymbol{H}^{N-1}_b),
\end{aligned} \tag{2}$$

其中 $CNN_n(n\in[2,N])$ 均为窗口大小为 2 的卷积神经网络，$\boldsymbol{H}^n_a(n\in[2,N])$ 和 $\boldsymbol{H}^n_b(n\in[2,N])$ 分别为由第 1 论元和第 2 论元中长度为 n 的短语表示构成的矩阵，N 为设定的最大短语的长度. 为了表述得简洁，把第 1 论元和第 2 论元中的全部短语表示分别记为 $\boldsymbol{H}_a$ 和 $\boldsymbol{H}_b$.

与基于句法分析器提取短语的方法相比，基于 n-gram 的方法具有 2 个优点：1)多数语言没有句法分析器或者其准确率不能满足实用的需求，因此，消除对句法分析器的依赖可以较大程度地增强所提模型的适用性. 2)可以避免句法分析所需的时间消耗，从而提高其实用性. 值得说明的是，句法分析本身是比较耗时的.

2.3 基于 Transformer 层的多粒度信息交互模块

给定词、短语和论元的语义向量表示，该模块堆叠多个 Transformer 层显式地建模它们之间的信息交互，以更好地捕获论元之间的语义关系. Transformer 层可以非常有效地建模任意距离的 2 个输入单元之间的交互，且常用于建模不同粒度、不同模态的信息之间的交互[31]. 具体地，首先构造位置信息(包括短语次序信息和短语所属论元指示信息)增强的输入，如下所示：

$$\boldsymbol{H}_{\text{p}} = (\boldsymbol{H}^1_a; \boldsymbol{H}^2_a; \cdots; \boldsymbol{H}^N_a; \boldsymbol{H}^1_b; \boldsymbol{H}^2_b; \cdots; \boldsymbol{H}^N_b), \tag{3}$$

$$\boldsymbol{H}^0_{\text{p}} = \boldsymbol{H}_{\text{p}} + \boldsymbol{P}_{\text{loc}} + \boldsymbol{P}_{\text{arg}}, \tag{4}$$

其中 $\boldsymbol{P}_{\text{loc}}$ 为短语次序信息矩阵，由可学习的位置向量 $\boldsymbol{p}_m(m\in\{1,2,\cdots,M\})$ 按需要拼接而成，$\boldsymbol{p}_m$ 为论元中第 m 个短语的次序向量，M 为论元中短语的最大次序；$\boldsymbol{P}_{\text{arg}}$ 为短语所属论元指示矩阵，由可学习的向量 $\boldsymbol{p}_a$ 和 $\boldsymbol{p}_b$ 按需要拼接而成，$\boldsymbol{p}_a$ 和 $\boldsymbol{p}_b$ 分别为短语属于第 1 论元和第 2 论元的指示向量. 引入短语次序信息和短语所属论元指示信息有利于 Transformer 层建模成对的序列数据.

然后，给定位置信息增强的输入，堆叠多个 Transformer 层显式地建模词、短语和论元这 3 种不同粒度的信息之间的交互，具体如下：

$$\boldsymbol{h}^k_{\text{cls}}, \tilde{\boldsymbol{H}}^k_{\text{p}} = Transformer_k(\boldsymbol{h}^{k-1}_{\text{cls}}; \tilde{\boldsymbol{H}}^{k-1}_{\text{p}}), \tag{5}$$

其中 $\boldsymbol{h}^k_{\text{cls}}, \tilde{\boldsymbol{H}}^k_{\text{p}}$ 为第 $k\in\{1,2,\cdots,K\}$ 个 Transformer 层的输出；在进行拼接操作时，把向量 $\boldsymbol{h}^{k-1}_{\text{cls}}$ 转换成相应的矩阵，并让 $\boldsymbol{h}^0_{\text{cls}} = \boldsymbol{h}_{\text{cls}}$.

最后，基于第K个 Transformer 层的输出得到输入的隐式篇章关系实例的最终表示$\boldsymbol{h}$：

$$\boldsymbol{h}_{\mathrm{p}}^{K}=Attention(\boldsymbol{w},\boldsymbol{H}_{\mathrm{p}}^{K}), \tag{6}$$

$$\boldsymbol{h}=(\boldsymbol{h}_{\mathrm{cls}}^{K},\boldsymbol{h}_{\mathrm{p}}^{K}), \tag{7}$$

其中$\boldsymbol{h}_{\mathrm{cls}}^{K}$和$\boldsymbol{h}_{\mathrm{p}}^{K}$可以看作该实例的不同视角的语义表示，$\boldsymbol{w}$为点积型注意力机制 Attention 中需要学习的参数向量.

一方面，基于 Transformer 层的多粒度信息交互模块在引入短语信息的基础上，充分建模了词、短语和论元之间的语义交互. 另一方面，与常用基于交叉注意力机制的方法仅建模论元间的信息交互相比，该模块同时建模了论元间和论元内部的多粒度信息的交互. 需要说明的是，基于多层卷积的短语编码模块中把论元中相邻的n个词看作长度为n的短语可能会引入大量的非法短语，即相邻的多个词并不能构成语言学意义上的短语. 引入 Transformer 层能够自动降低非法短语对应的权重，并提升合法短语对应的权重，从而有效地利用短语信息.

2.4 链式解码模块

篇章关系通常组织为多级层次结构，低级篇章关系是高级篇章关系的更细粒度表示. Wu 等人[20]基于循环神经网络从顶向下依次解码多级篇章关系，利用高级篇章关系的解码结果引导低级篇章关系的解码，取得了较好的效果. 类似地，我们采用一种简化的链式解码模块，以充分利用多级篇章关系类别之间的依赖信息. 具体地，给定隐式篇章关系实例的最终表示$\boldsymbol{h}$和第$j-1$级篇章关系解码的 logits 值$\boldsymbol{m}_j$，计算第$j\in\{1,2,\cdots,J\}$级解码的结果$\hat{\boldsymbol{y}}_j$：

$$\boldsymbol{m}_j=FFN_j(\boldsymbol{h};\boldsymbol{m}_{j-1}), \tag{8}$$

$$\hat{\boldsymbol{y}}_j=Softmax(\boldsymbol{m}_j), \tag{9}$$

其中FFN_j为多层前向神经网络，当$j=1$时，$\boldsymbol{m}_0$为零向量.

多任务学习训练方法通过共享参数在多个相关的任务之间共享信息，以达到相互促进、共同提高的目的. 把多级隐式篇章关系识别任务看作高度相关的多个任务，定义多任务代价函数L_{mt}：

$$L_j=-\frac{1}{|D|}\sum_{(x,\boldsymbol{y}_j)\in D}\sum_{z=1}^{Z_j}\boldsymbol{y}_j^z\ln\hat{\boldsymbol{y}}_j^z, \tag{10}$$

$$L_{\mathrm{mt}}=\sum_{j=1}^{J}L_j, \tag{11}$$

其中D为训练数据集，$\boldsymbol{y}_j$为实例x的第j级真实类别的独热编码表示，Z_j为第j级篇章关系类别的数量，L_j为第j级交叉熵代价.

3 DAM 方法

已有研究工作证实连接词自然标注的显式篇章关系数据可以缓解隐式篇章关系识别训练语料不足的问题. 例如，通过连接词“但是”可以知道显式篇章关系实例“经济开放度在逐步提高，但是开发的力度不够”表达的是“转折”关系. 基于显式篇章数据定义的连接词分类任务可以看作隐式篇章关系识别的相关任务（多任务学习中）或预训练任务. 例如，在预训练方法[27]中通常先基于显式篇章关系数据训练编码器和连接词分类层，然后把编码器用于隐式篇章关系识别. 现有方法的不足之处是仅基于显式篇章关系数据预训练编码模块.

本节提出一种基于链式解码的数据增强方法 DAM，其能同时预训练编码和解码模块，从而可以更充分地利用大量的显式篇章关系数据. 具体地，显式篇章关系实例可表示为$x_c=(arg_a,arg_b,c)$，c为 2 个论元间的连接词. 把连接词看作第$J+1$级篇章关系，基于链接解码预测连接词如下：

$$\boldsymbol{m}_c^{J+1}=FFN_c((\boldsymbol{h}_c;\boldsymbol{m}_c^{J})), \tag{12}$$

$$\hat{\boldsymbol{y}}_c=Softmax(\boldsymbol{m}_c^{J+1}), \tag{13}$$

其中$\boldsymbol{h}_c$为显式篇章关系实例的表示，$\boldsymbol{m}_c^{J}$为链式解码第J级篇章关系时对应的 logits 值，$\hat{\boldsymbol{y}}_c$为连接词预测结果.

给定显式篇章关系数据集D_c，定义连接词分类的代价函数为：

$$L_c=-\frac{1}{|D_c|}\sum_{(x_c,\boldsymbol{y}_c)\in D_c}\sum_{q=1}^{Q}y_c^q\ln\hat{y}_c^q, \tag{14}$$

其中$\boldsymbol{y}_c$为真实连接词的独热编码表示，y_c^q为$\boldsymbol{y}_c$的第q个分量，Q为连接词的数量.

DAM 是一种 2 阶段的预训练方法，其首先基于大量自然标注的显式篇章关系数据和式（14）所示代价函数进行预训练（连接词分类任务），然后基于隐式篇章关系数据和式（11）所示代价函数进行微调（隐式篇章关系识别任务）. 由式（12）可知，预训练阶段在计算连接词对应的 logits 值时，需要第J级篇章关系对应的 logits 值，也间接地需要所有其他级篇章关系对应的 logits 值. 因此，所提 DAM 方法不仅能预训练编码模块，还能预训练解码模块.

4 实 验

4.1 数 据

我们在常用的篇章关系数据集 PDTB 2.0[32] 和 PDTB 3.0[33] 上进行实验, 以方便与基准方法进行直接对比. 2 个数据集中的篇章关系类别都组织为层次结构, 第 1 级主要包括时序(temporal)、递进(contingency)、对比(comparison)和解释(expansion)四大类. 第 2 级和第 3 级是更细粒度的篇章关系类别, 例如, 在 PDTB 2.0 和 PDTB 3.0 中第 2 级分别包括 16 小类和 20 小类. 与大部分基准方法一样, 本文方法只在第 1 级和第 2 级篇章关系上进行实验(语料相当一部分实例没有标注第 3 级篇章关系); 忽略语料中标注数量较少的或测试集中没有的第 2 级篇章关系类别, 最终在 PDTB 2.0 和 PDTB 3.0 上分别执行 11 分类和 14 分类. 依照文献 [27, 34] 等的数据划分方法, 把第 2~20 节作为训练集, 第 0 节和第 1 节作为验证集, 第 21 节和第 22 节作为测试集. 具体实例数如表 1 所示. 从表 1 中可以看出, 目前最大的隐式篇章关系数据集中仅包含 2 万多个实例, 这对训练复杂的神经网络模型是不够的. 此外, 2 个数据集都是非常不平衡的, 例如, expansion 实例占比均超过 50%, 而 temporal 实例的占比都只有 10% 不到.

Table 1 Number of Instances in PDTB 2.0 and PDTB 3.0

表 1 PDTB 2.0 和 PDTB 3.0 中的实例数量

数据集	训练集	验证集	测试集
PDTB 2.0	12 632	1 183	1 046
PDTB 3.0	17 945	1 653	1 474

4.2 设 置

使用 RoBERTa-base[35] 作为词编码模块, 并在训练过程中进一步微调其参数. 考虑到数据的不平衡性, 使用准确率(accuracy, Acc)和宏平均 $F1$ 值 2 种指标评价模型的整体性能. 为了减少训练的不稳定性, 使用不同的随机数初始化模型, 取 5 次结果的平均值. 实验所有代码基于 PyTorch 实现, 并以默认的方式初始化模型参数. 基于 PDTB 2.0 验证集上的最优性能选择模型的超参数值在 PDTB 3.0 上同样取得了较好的效果, 如表 2 所示.

Table 2 Values of Hyperparameters

表 2 超参数的值

超参数	值
Transformer 层数	2
Transformer 层维度	768
最大短语长度 N	4
优化器	Adamw
学习率	0.000 01
批数据大小	32
训练轮次 Epoch	15

4.3 与基准方法的对比实验

为了验证所提方法在隐式篇章关系识别任务上的有效性, 我们与基于词交互的方法[8]、基于短语交互的方法[12]、在编码端和解码端同时利用类别关系的方法[20]、基于提示学习的方法[24-25,27] 和 ChatGPT[5-6] 等基准方法进行了对比, 具体如下:

1)BMGF[8]. 首先利用 RoBERTa 分别学习 2 个论元中词的表示, 然后基于一种双向多视角匹配机制建模词之间的交互.

2)PhraseINT[12]. 基于句法分析器提取论元中的短语, 然后构建短语级交互注意力机制用于隐式篇章关系识别.

3)LDSGM[20]. 首次把序列生成模型用于多级隐式篇章关系识别, 有效地在编码端和解码端同时利用类别之间的依赖关系.

4)PCP[24]. 较早地把基于提示学习的方法用于隐式篇章关系识别的研究, 其人工定义提示模板和映射规则.

5)PEMI[25]. 把篇章关系类别之间的层次信息融入到基于提示学习的模型中, 在仅精调少量参数的情况下, 取得了优于 LDSGM 的效果.

6)ADCG[27]. 模拟人工标注篇章关系语料的 2 阶段方法, 是当前效果较好的方法之一. 首先基于提示学习的模型先预测连接词, 然后把连接词和论元同时作为隐式篇章关系识别模型的输入, 最后联合训练 2 个模型.

7)ChatGPT[5-6]. 少样本和零样本的情况下, ChatGPT 在隐式篇章关系识别上的效果.

MGII 为第 2 节提出的基于多粒度信息交互的隐式篇章关系识别模型, MGII + DAM 为第 3 节基于链接解码的数据增强模型. 需要说明的是, 表 3 和表 4 中方法 BMGF 和 PCP 的结果引用自文献 [26], 方法 LDSGM 的结果引用自文献 [22]. 为了公平对比, 把 PhraseINT 的基础编码器替换为 RoBERTa, 并报告其在 2 个数据集上的结果.

从表 3 和表 4 的结果可以看出:

Table 3 Results Compared with the Baseline Methods on PDTB 2.0

表 3 PDTB 2.0 上与基准方法对比的结果 %

方法	第 1 级		第 2 级	
	Acc	*F*1	Acc	*F*1
ChatGPT	44.18	36.11	24.54	16.20
BMGF	69.06	63.39	58.13	-
PhraseINT	70.36	63.05	59.74	39.65
LDSGM	71.18	63.73	60.33	40.49
PCP	70.84	64.95	60.54	41.55
PEMI	71.13	64.05	60.66	41.31
ADCG	**74.59**	**68.64**	**62.79**	42.36
MGII（本文）	71.56	64.51	60.79	42.18
MGII + DAM（本文）	73.33	66.58	62.15	**43.70**

注：黑体数值表示最优结果.

Table 4 Results Compared with the Baseline Methods on PDTB 3.0

表 4 PDTB 3.0 上与基准方法对比的结果 %

方法	第 1 级		第 2 级	
	Acc	*F*1	Acc	*F*1
ChatGPT	52.84	47.53	36.98	25.77
BMGF	71.98	66.92	61.87	41.28
PhraseINT	72.81	68.10	61.56	50.39
LDSGM	73.18	68.73	61.33	53.49
PCP	73.81	69.82	63.63	49.87
PEMI	73.27	69.06	63.09	52.73
ADCG	76.23	71.15	**65.51**	55.26
MGII（本文）	75.40	70.63	63.12	54.22
MGII + DAM（本文）	**76.92**	**72.41**	65.21	**55.94**

注：黑体数值表示最优结果.

1)在隐式篇章关系识别上，少样本 ChatGPT 的效果还不理想，离有监督的模型还有较大的差距. 这充分说明，在大模型时代隐式篇章关系识别仍然值得深入研究.

2)与基于词交互或短语交互的模型相比，MGII 在 2 个数据集上都有较明显的提高. 例如，在 PDTB 2.0 的第 1 级和第 2 级关系上，和 PhraseINT 相比，MGII 的准确率分别提高了 1.2 个百分点和 1.06 个百分点，*F*1 值分别提高了 1.46 个百分点和 2.53 个百分点. 这些结果证实有效地建模词、短语和论元 3 种不同粒度的信息之间的交互，可以提高识别的性能.

3)与同样在解码端使用了类别之间的依赖关系的 LDSGM 相比，MGII 在 2 个数据集的多级关系上都取得了一致性的提高. 这再次证明在编码端建模多种粒度信息之间的交互是有效的. 需要说明的是，MGII 在 PDTB 2.0 上的效果与基于 RoBERTa 的同类基准模型 LDSGM 相比并没有明显的提高，主要原因是 PDTB 2.0 语料相对小，导致模型不能较好地学习短语的表示. 此外，LDSGM 还利用了语料中标注的连接词作为额外的监督信息.

4)加入 DAM 后，MGII 性能取得了较为明显的提升. 在 PDTB 2.0 的第 1 级和第 2 级关系上，准确率分别提高了 1.77 个百分点和 1.36 个百分点，*F*1 值分别提高了 2.07 个百分点和 1.52 个百分点；在 PDTB 3.0 上也有类似的提升.

5)ADCG 等基于提示学习的方法取得了较好的识别性能，是近年来研究的热点方法. 与之相比，所提 MGII 模型的性能尚有一定的差距. 从模型的角度来看，ADCG 和 MGII 具有较大的差异，本文 4.6 节的实验结果显示联合使用这 2 种方法可进一步提升隐式篇章关系识别的性能.

从以上分析可知，MGII 模型是有效的，DAM 能进一步提升识别的性能.

4.4 消融实验

本文提出的 MGII 模型的核心是建模词、短语和论元这 3 种粒度的信息之间的交互，以更好地捕获论元之间的语义关系. 为了验证每种粒度信息的必要性，我们在表 5 和表 6 中对比了以下方法的性能：1)w/o *n*-gram. 不利用短语的信息，直接把词编码模块的输出作为链式解码模块的输入，其中词的表示经式(6)中注意力机制融合后作为输入，也就是说模型中不包括短语编码模块和多信息交互模块. 2)w/o Word. 在多粒度信息交互模块中，不包括词的信息，即在式(3)中不包括 $\boldsymbol{H}_a^1$ 和 $\boldsymbol{H}_b^1$. 3)w/o CLS. 在多粒度信息交互模块中，不包括论元的信息，即该模块的输入

Table 5 Results of Ablation Experiments on PDTB 2.0

表 5 PDTB 2.0 上消融实验的结果 %

模型	第 1 级		第 2 级	
	Acc	*F*1	Acc	*F*1
MGII	71.56	64.51	60.79	42.18
w/o *n*-gram	70.23	63.25	59.73	41.26
w/o Word	70.57	63.96	60.05	41.42
w/o CLS	70.65	64.06	59.81	41.63
w/o 2-gram	70.68	64.06	60.25	41.53
w/o 3-gram	70.83	64.22	60.41	41.76
w/o 4-gram	71.20	64.41	60.62	42.03

Table 6 Results of Ablation Experiments on PDTB 3.0

表 6 PDTB 3.0 上消融实验的结果 %

模型	第 1 级		第 2 级	
	Acc	*F*1	Acc	*F*1
MGII	75.40	70.63	63.12	54.22
w/o *n*-gram	74.34	69.71	62.33	53.10
w/o Word	74.61	70.06	62.45	53.43
w/o CLS	74.73	70.22	62.52	54.07
w/o 2-gram	74.75	69.76	62.73	53.76
w/o 3-gram	74.93	70.23	62.89	54.02
w/o 4-gram	75.20	70.34	62.93	54.21

不包括 $\boldsymbol{h}_{\text{cls}}$. 4)w/o 2-gram, 3-gram, 4-gram. 在多粒度信息交互模块中, 不包括相应的短语信息.

从表 5、表 6 可以看出 2 个数据集上的性能变化倾向是类似的, 具体为: 1)交互过程中缺少词、短语和论元中的任何一种信息都会导致模型的性能有不同程度的下降, 这充分说明每种粒度的信息都是必要的. 2)在 3 种粒度的信息中, 去除短语信息导致性能下降的幅度最大. 上述实验结果与短语是理解论元语义关系的关键所在这一事实是相符的. 3)在 3 种 *n*-gram 中, 移除 2-gram 对模型性能的影响最大, 这也符合大部分句子中短语的分布情况, 即 2 个词的短语占的比重较大. 综上所述, 在隐式篇章关系识别中建模多粒度信息的交互是有必要的.

4.5 数据增强方法对比

为了验证所提基于链式解码的数据增强方法 DAM 的有效性, 我们将之与预训练方法[14]进行了对比. 具体地, 预训练方法首先在显式篇章关系数据上基于连接词分类任务(给定论元预测连接词)对编码端进行特定于任务的预训练, 然后基于人工标注的隐式篇章关系数据进行微调. 上述方法在训练过程中需要引入独立的连接词分类层, 可以认为仅增强了编码模块, 而所提基于链式解码的方法能同时增强编码和解码模块.

从文献 [36] 提供的显式篇章关系数据集中随机筛选 1×10^6 个实例用于实验, 其中包含 174 个英语篇章连接词. 具体地, 分别使用 5×10^4, 5×10^5, 1×10^6 个实例进行实验, 以对比不同数据量下 2 种数据增强方法的性能表现, 图 3 和图 4 显示了在 PDTB 2.0 和 PDTB 3.0 上不同数据增强方法的效果对比.

从图 3、图 4 可以发现: 1)在绝大多数情况下, DAM 的效果优于预训练的方法. 这些结果说明 DAM 基于显式篇章关系数据同时优化编码器和解码器的改进是有效的. 2)当数据量增加到 5×10^5 时, 2 种数据增强方法的性能都有较明显的提升, 继续增加到 1×10^6 时, 2 种方法性能提升的趋势都趋于平缓, 甚至在个别指标上略有下降. 导致这一实验现象的原因可能是显式和隐式篇章关系数据在语言特性和数据分布上存在较大的差异. 如何有效地利用更大量的显式篇章关系数据是值得继续探索的问题之一.

4.6 联合模型的效果

近年来, 基于提示学习的方法在隐式篇章关系识别等众多自然语言处理任务上取得了较好的性能. 本节实验把 MGII 与文献 [27] 提出的 ADCG 进行结合, 探索是否能进一步提高性能. 具体地, 第 1 阶段使用基于提示学习的方法预测合适的连接词, 第 2 阶段把论元和连接词输入本文 MGII 中, 最后采用文

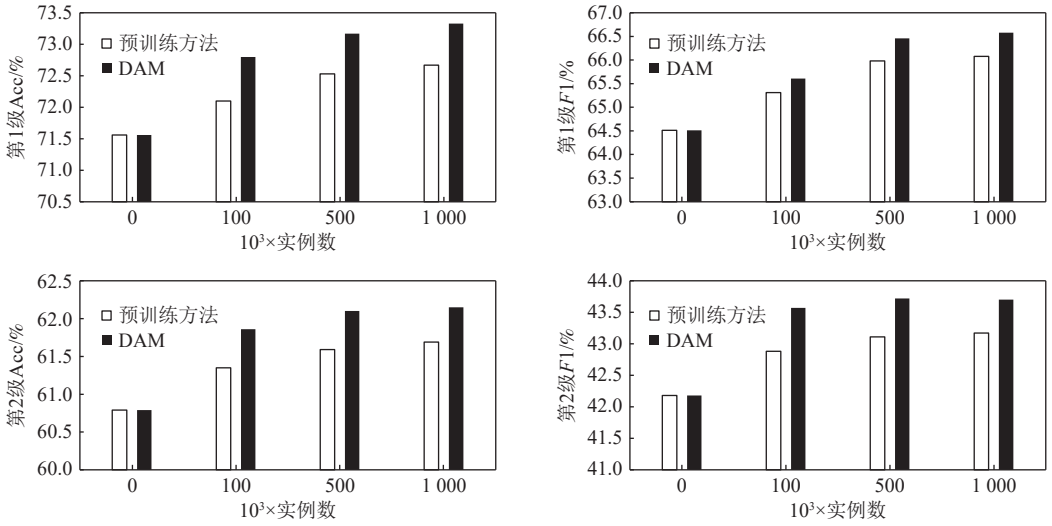


Fig. 3 Effectiveness comparison of different data augmentation methods on PDTB 2.0

图 3 PDTB 2.0 上不同数据增强方法的效果对比

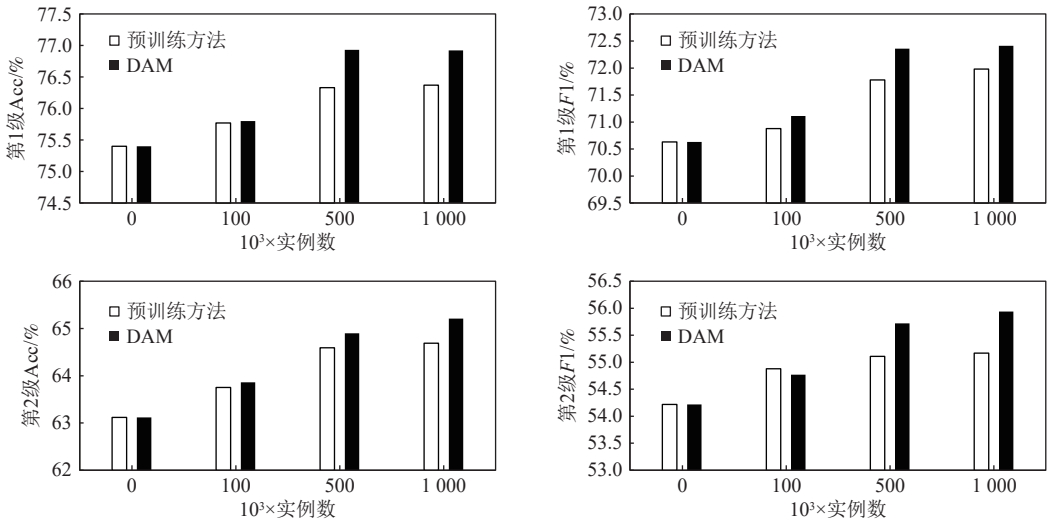


Fig. 4 Effectiveness comparison of different data augmentation methods on PDTB 3.0

图 4 PDTB 3.0 上不同数据增强方法的效果对比

献 [27] 中的 ADCG 方法联合训练 2 阶段的模型.

从表 7 和表 8 中的实验结果可以看出，虽然 ADCG 取得了较高的性能，但联合 MGII 模型仍然是有帮助的，进一步提高了模型的性能.

Table 7 Effect of Joint Modeling on PDTB 2.0

表 7 PDTB 2.0 上联合模型的效果 %

模型	第 1 级		第 2 级	
	Acc	*F*1	Acc	*F*1
ADCG	74.59	68.64	62.79	42.36
MGII+ADCG	75.31	69.54	63.47	43.43

Table 8 Effect of Joint Modeling on PDTB 3.0

表 8 PDTB 3.0 上联合模型的效果 %

模型	第 1 级		第 2 级	
	Acc	*F*1	Acc	*F*1
ADCG	76.23	71.15	65.51	55.26
MGII+ADCG	76.95	72.46	66.39	56.32

5 结 论

本文提出了一种基于多粒度信息交互和数据增强的隐式篇章关系识别方法. 主要贡献包括：1)在不依赖句法分析器的前提下显式地建模了词、短语和论元 3 种粒度的信息之间的交互；2)通过链式解码的数据增强方法更有效地利用了大量连接词自然标注的显式篇章关系数据. 与基准方法相比，所提方法不仅在性能上取得了较明显的提高，还具有更强的适用性，可进一步推动篇章分析朝实用化的方向发展.

所提方法的性能提升主要来自基于链式解码的数据增强方法，这充分说明如何有效地利用显式篇章关系数据依然是一个值得重点关注的问题. 在未来的工作中，一方面，我们将研究如何在目前性能最好的基于 Prompt 的方法中有效地利用显式篇章关系数据；另一方面，研究如何在更多的实际场景中利用显式篇章关系数据，例如，只有少量或者没有人工标注隐式篇章关系数据时. 此外，我们还将探索包括短语在内的多粒度信息交互在其他自然语言处理任务中的有效性.

作者贡献声明：姚浩提出算法思路和实验方案，完成主要实验并撰写论文；熊锦晖收集数据，协助完成部分实验；李春生提出论文修改意见；邬昌兴总体指导，提出算法思路并撰写并审核论文.

参 考 文 献

[1] Wang Chang, Wang Bang. An end-to-end topic-enhanced self-attention network for social emotion classification[C]//Proc of the Web Conf 2020. New York: ACM, 2020: 2210–2219

[2] Li Huifeng, Srihari R, Niu Cheng, et al. Location normalization for information extraction[C/OL]//Proc of the 19th Int Conf on COLING. Stroudsburg, PA: ACL, 2002[2024-01-03]. https://aclanthology.org/C02-1127/

[3] Cohan A, Dernoncourt F, Kim D, et al. A discourse-aware attention model for abstractive summarization of long documents[C]//Proc of the 2018 Conf of NAACL: Human Language Technologies. Stroudsburg, PA: ACL, 2018: 615–621

[4] Verberne S, Boves L, Oostdijk N, et al. Evaluating discourse-based answer extraction for why-question answering[C]//Proc of the 30th Annual Int ACM SIGIR Conf on Research and Development in

Information Retrieval. New York: ACM, 2007: 735–736

[5] Chan C, Cheng Jiayang, Wang Weiqi, et al. ChatGPT evaluation on sentence level relations: A focus on temporal, causal, and discourse relations[C]//Proc of Findings of ACL: EACL 2024. Stroudsburg, PA: ACL, 2024: 684−721

[6] Yung F, Ahmad M, Scholman M, et al. Prompting implicit discourse relation annotation[C]//Proc of the 18th Linguistic Annotation Workshop. Stroudsburg, PA: ACL, 2024: 150–165

[7] Ruan Huibin, Hong Yu, Xu Yang, et al. Interactively-propagative attention learning for implicit discourse relation recognition[C]//Proc of the 28th Int Conf on COLING. Stroudsburg, PA: ACL, 2020: 3168–3178

[8] Liu Xin, Ou Jiefu, Song Yangqiu, et al. On the importance of word and sentence representation learning in implicit discourse relation classification[C]//Proc of the 29th IJCAI. San Francisco, CA: Morgan Kaufmann, 2020: 3830–3836

[9] Ma Yuhao, Zhu Jian, Liu Jie. Enhanced semantic representation learning for implicit discourse relation classification[J]. Applied Intelligence, 2022, 52(7): 7700−7712

[10] Xiang Wei, Wang Bang, Dai Lu, et al. Encoding and fusing semantic connection and linguistic evidence for implicit discourse relation recognition[C]//Proc of Findings of ACL: ACL 2022. Stroudsburg, PA: ACL: 2022, 3247–3257

[11] Wang Yizhong, Li Sujian, Yang Jingfeng, et al. Tag-enhanced tree-structured neural networks for implicit discourse relation classification[C]//Proc of the 8th IJNLP. Stroudsburg, PA: ACL, 2017: 496–505

[12] Wang Xiuli, Jin Fangyan. Implicit discourse relation recognition integrating feature coding and phrase interaction perception[J]. Acta Electronica Sinica, 2024, 52(4): 1377−1388 (in Chinese)
(王秀利, 金方焱. 融合特征编码和短语交互感知的隐式篇章关系识别[J]. 电子学报, 2024, 52(4): 1377−1388)

[13] Liu Yang, Li Sujian, Zhang Xiaodong, et al. Implicit discourse relation classification via multi-task neural networks[C]//Proc of the 30th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2016: 2750–2756

[14] Kishimoto Y, Murawaki Y, Kurohashi S. Adapting BERT to implicit discourse relation classification with a focus on discourse connectives[C]//Proc of LREC 2020. Paris: ELRA, 2020: 1152–1158

[15] Zhang Biao, Su Jinsong, Xiong Deyi, et al. Shallow convolutional neural network for implicit discourse relation recognition[C]//Proc of the 2015 Conf on EMNLP. Stroudsburg, PA: ACL, 2015: 2230–2235

[16] Rutherford A, Demberg V, Xue Nianwen. A systematic study of neural discourse models for implicit discourse relation[C]//Proc of the 15th Conf of EACL. Stroudsburg, PA: ACL, 2017: 281–291

[17] Chen Jifan, Zhang Qi, Liu Pengfei, et al. Implicit discourse relation detection via a deep architecture with gated relevance network[C]//Proc of the 54th Annual Meeting of the ACL. Stroudsburg, PA: ACL, 2016: 1726–1735

[18] Liu Yang, Li Sujian. Recognizing Implicit discourse relations via repeated reading: Neural networks with multi-level attention[C]//Proc of the 2016 Conf on EMNLP. Stroudsburg, PA: ACL, 2016: 1224–1233

[19] Lei Wenqiang, Wang Xuancong, Liu Meichun, et al. SWIM: A simple word interaction model for implicit discourse relation recognition[C]//Proc of the 26th IJCAI. San Francisco, CA: Morgan Kaufmann, 2017: 4026–4032

[20] Wu Changxing, Cao Liuwen, Ge Yubin, et al. A label dependence-aware sequence generation model for multi-level implicit discourse relation recognition[C]//Proc of the AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2022, 11486–11494

[21] Long Wanqiu, Webber B. Facilitating contrastive learning of discourse relational senses by exploiting the hierarchy of sense relations[C]//Proc of the 2022 Conf on EMNLP. Stroudsburg, PA: ACL, 2022: 10704–10716

[22] Jiang Yuxin, Zhang Linhan, Wang Wei. Global and local hierarchy-aware contrastive framework for implicit discourse relation recognition[C]//Proc of Findings of ACL: ACL 2023. Stroudsburg, PA: ACL, 2023: 8048–8064

[23] Xiang Wei, Wang Zhenglin, Dai Lu, et al. ConnPrompt: connective-cloze prompt learning for implicit discourse relation recognition[C]//Proc of the 29th Int Conf on COLING. Stroudsburg, PA: ACL, 2022: 902–911

[24] Zhou Hao, Lan Man, Wu Yuanbin, et al. Prompt-based connective prediction method for fine-grained implicit discourse relation recognition[C]//Proc of Findings of ACL: EMNLP 2022. Stroudsburg, PA: ACL, 2022: 3848–3858

[25] Zhao Haodong, He Ruifang, Xiao Mengnan, et al. Infusing hierarchical guidance into prompt tuning: A parameter-efficient framework for multi-level implicit discourse relation recognition[C]//Proc of the 61st Annual Meeting of ACL. Stroudsburg, PA: ACL, 2023: 6477–6492

[26] Wu Hongyi, Zhou Hao, Lan Man, et al. Connective prediction for implicit discourse relation recognition via knowledge distillation[C]//Proc of the 61st Annual Meeting of ACL. Stroudsburg, PA: ACL, 2023: 5908–5923

[27] Liu Wei, Strube M. Annotation-inspired implicit discourse relation classification with auxiliary discourse connective generation[C]//Proc of the 61st Annual Meeting of ACL. Stroudsburg, PA: ACL, 2023: 15696–15712

[28] Wu Changxing, Shi Xiaodong, Chen Yidong, et al. Improving implicit discourse relation recognition with discourse-specific word embeddings[C]//Proc of the 55th Annual Meeting of ACL. Stroudsburg, PA: ACL, 2017: 269–274

[29] Ru Dongyu, Qiu Lin, Qiu Xipeng, et al. Distributed marker representation for ambiguous discourse markers and entangled relations[C]//Proc of the 61st Annual Meeting of ACL. Stroudsburg, PA: ACL, 2023: 5334–5351

[30] Wang Chenxu, Jian Ping, Huang Mu. Prompt-based logical semantics enhancement for implicit discourse relation recognition[C]//Proc of the 2023 Conf on EMNLP. PA, ACL: 2023: 687–699

[31] Fan Wei, Liu Yong. Social network information diffusion prediction based on spatial-temporal transformer[J]. Journal of Computer Research and Development, 2022, 59(8): 1757−1769 (in Chinese)
(范伟, 刘勇. 基于时空 Transformer 的社交网络信息传播预测[J]. 计算机研究与发展, 2022, 59(8): 1757−1769)

[32] Rashmi P, Nikhil D, Alan L, et al. The PENN discourse TreeBank

2.0[C]//Proc of the 6th Int Conf on LREC. Stroudsburg, PA: ACL, 2008: 2961–2968

[33] Webber B, Prasad R, Lee A, et al. The PENN discourse TreeBank 3.0 annotation manual[R]. Philadelphia, PA: University of Pennsylvania, 2019

[34] Ji Yangfeng, Eisenstein J. One vector is not enough: entity-augmented distributed semantics for discourse relations[J]. Transactions of the Association for Computational Linguistics 2015, 3: 329−344

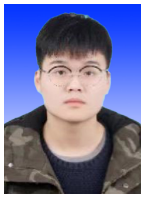
[35] Liu Yinhan, Ott M, Goyal N, et al. RoBERTa: A robustly optimized BERT pretraining approach[J]. arXiv preprint, arXiv: 1907.11692, 2019

[36] Sileo D Van-De-Cruys T, Pradel C, et al. Mining discourse markers for unsupervised sentence representation learning[C]//Proc of the 2019 Conf of NAACL: Human Language Technologies. Stroudsburg, PA: ACL, 2019: 3477–3486



Yao Hao, born in 1999. Master candidate. His main research interest includes nature language processing with a focus on discourse analysis.

姚　浩, 1999 年生. 硕士研究生. 主要研究方向为自然语言处理, 重点关注篇章分析.



Xiong Jinhui, born in 1999. Master candidate. His main research interest includes nature language processing with a focus on discourse analysis. (409119311@qq.com)

熊锦晖, 1999 年生. 硕士研究生. 主要研究方向为自然语言处理, 重点关注篇章分析.



Li Chunsheng, born in 1980. Associate professor of engineering. His main research interest includes information retrieval. (lcs_2002@163.com)

李春生, 1980 年生. 副高级工程师. 主要研究方向为信息抽取.



Wu Changxing, born in 1981. PhD, associate professor. Member of CCF. His research interests include nature language processing and information retrieval.

邬昌兴, 1981 年生. 博士, 副教授. CCF 会员. 主要研究方向为自然语言处理、信息抽取.