引用格式：刘万里, 雍新有, 曹开臣, 等. 基于提示学习的 ERNIE-BiLSTM-PN 通用信息抽取方法研究 [J]. 电子科技大学学报, 2025, 54(3): 411-423.
LIU W L, YONG X Y, CAO K C, et al. Universal information extraction method based on prompt learning with ERNIE-BiLSTM-PN[J]. Journal of University of Electronic Science and Technology of China, 2025, 54(3): 411-423.

基于提示学习的 ERNIE-BiLSTM-PN 通用信息抽取方法研究

刘万里[1]，雍新有[2]，曹开臣[1]，陈俞舟[1]，刘禄波[1]，蔡世民[2*]

（1. 西南电子技术研究所 第二实验室，成都 610036；2. 电子科技大学 大数据研究中心，成都 611731）

摘要：随着大数据时代的到来，信息抽取已成为自然语言处理领域的重要研究方向。信息抽取涉及多项任务，包括命名实体识别、关系抽取和事件抽取等，每项任务通常需要依靠专用模型来应对其特定的挑战。该文提出一种基于提示学习的 ERNIE-BiLSTM-PN 通用信息抽取方法（EBP-UIE），结合预训练语言模型（ERNIE）、双向长短期记忆网络（BiLSTM）和指针网络（PN），旨在通过一个统一的框架解决信息抽取任务的复杂性，并实现跨任务知识的共享。ERNIE 优化了对文本的深层理解和上下文分析，BiLSTM 的应用加强了对序列特征的捕捉及长距离依赖关系的解析，PN 则提高了对文本中信息元素起止位置的精确标定，提示学习机制灵活实现多个信息抽取任务的统一建模。实验结果显示：在命名实体识别任务，EBP-UIE 在 MSRA 和 PeopleDaily 数据集上的 F1 分数比 UIE 模型分别高出 7.12% 和 0.53%；在关系抽取任务，EBP-UIE 在 DuIE 数据集上的 F1 分数超过 UIE 模型 6.84%；对于事件抽取任务，EBP-UIE 在 DuEE 数据集上的触发词和论元抽取 F1 分数分别比 UIE 模型高出 4.49% 和 0.95%。

关键词：通用信息抽取；深度学习；指针网络；提示学习
中图分类号：TP391 **文献标志码：**A DOI: 10.12178/1001-0548.2024106

Universal information extraction method based on prompt learning with ERNIE-BiLSTM-PN

LIU Wanli[1], YONG Xinyou[2], CAO Kaichen[1], CHEN Yuzhou[1], LIU Lubo[1], and CAI Shimin[2*]

(1. Second Laboratory, Southwest Institute of Electronic Technology, Chengdu 610036, China;
2. Big Data Research Center, University of Electronic Science and Technology of China, Chengdu 611731, China)

Abstract: With the advent of the big data era, information extraction has become a significant research direction in the field of natural language processing. Information extraction involves multiple tasks, including named entity recognition, relation extraction, and event extraction, each typically relying on specialized models to address its specific challenges. This paper proposes a universal information extraction method based on prompt learning (EBP-UIE), enhanced representation through knowledge integration (ERNIE), bi-directional long short-term memory networks (BiLSTM), and pointer networks (PN), aimed at resolving the complexities of information extraction tasks through a unified framework and facilitating cross-task knowledge sharing. The introduction of the ERNIE model enhances deep text understanding and contextual analysis, the application of BiLSTM strengthens the capture of sequential features and the parsing of long-distance dependencies, and the pointer network improves the precise identification of start and end positions of information elements in text. The experimental results show that on named entity recognition, the F1 scores of EBP-UIE on the MSRA and PeopleDaily datasets are respectively 7.12% and 0.53% higher than those of the UIE models; on relation extraction, the F1 score of EBP-UIE on the DuIE dataset exceeded that of the UIE model by 6.84%; And on the event extraction, the F1 score of EBP-UIE outperforms the UIE model by 4.49% and 0.95% in trigger word and argument extraction performance on the DuEE dataset, respectively.

Key words: universal information extraction; deep learning; pointer network; prompt learning

收稿日期：2024−05−09
基金项目：国家自然科学基金（T2293771, 11975071）
作者简介：刘万里，高级工程师，主要从事深度学习和自然语言处理方面的研究。
*通信作者 E-mail: shimincai@uestc.edu.cn

信息抽取（information extraction, IE）[1] 是指从非结构化文本中提取结构化事实信息的过程。在这个过程中，通常涉及将实体、关系和事件分别转换为片段、三元组、记录等异构结构形式。面对各种复杂多样的 IE 任务时，往往需要为每种任务单独设计和构建不同的模型，以满足不断变化的复杂需求。这种操作受到了多样抽取目标、不同的复杂结构以及多变的领域需求的制约，导致 IE 模型难以实现统一建模，极大地阻碍了 IE 系统的高效架构开发、有效知识共享和快速跨域适配。因此，开发一种通用的 IE 模型具有重大意义。

在当前的通用信息抽取领域，序列到序列架构（sequence to sequence, Seq2Seq）[2] 因其高度的灵活性而得到广泛应用。这种架构能够处理输入和输出序列长度不等的情况，理论上适用于所有自然语言处理（natural language processing, NLP）问题。基于 Seq2Seq 架构，文献 [3] 提出了一个统一文本到结构的生成框架 UIE（universal information extraction）。UIE 设计了统一的结构生成网络，通过结构抽取语言将异构 IE 编码为统一的表示，并通过结构模式指导机制形成需要发现和关联的抽取要素。然而，生成式模型的高自由度也可能导致一些与预期不符的输出，尤其是在复杂的 IE 任务中，其决策过程往往难以解释。

抽取式方法直接从原始文本中识别并提取出具有特定信息的片段。这类方法不对原始文本进行任何修改或重写，仅仅是从文本中挑选出关键信息，所以比生成式模型准确度更高，更易于解释和验证，计算效率更高，鲁棒性也更好。最常用的抽取文本片段的方法是基于 BIO 模式和条件随机场（conditional random field, CRF）的序列标注模型[4]。在这类模型中，深度学习网络[5] 通常被用于提取文本的特征表示，而 CRF 层则被用于对这些特征进行解码，输出最终的标注序列。然而，序列标注模型抽取信息的最大弊端在于无法解决元素重叠问题，并且 CRF 的维特比解码往往更耗时。与传统的序列标注方法不同，指针标注通过直接指向文本中实体的起始和结束位置来识别和提取信息。它不依赖于预定义的标签集合来标注整个序列，而是利用指针网络（pointer network, PN）[6] 来确定实体的边界，从而能够更灵活地处理各种长度的实体，并有效应对嵌套实体和重叠实体的挑战。

因此，利用统一文本到结构生成的生成式机制，同时考虑 PN 在文本片段抽取中的优势，本文提出了基于提示学习的 ERNIE-BiLSTM-PN 通用信息抽取方法（universal information extraction based on prompt learning with ERNIE-BiLSTM-PN, EBP-UIE）。它主要由抽取子任务和分类子任务组成，随后引入提示学习机制，灵活组合多个子任务，实现对多种信息抽取任务的统一建模。在抽取子任务中，预训练语言模型 ERNIE（enhanced representation through knowledge integration）[7] 提供强大的文本理解和上下文分析支持，长短期记忆网络 BiLSTM（bidirectional long short-term memory）进一步分析文本中的序列特征和长距离依赖关系，最后使用 PN 显著提高了对文本元素起止位置的识别能力。在分类子任务中，EBP-UIE 利用 ERNIE 强大的语义表示能力实现文本的多分类，以精准地分类各类信息。通过对多个公开数据集的测试，本文验证了 EBP-UIE 在命名实体识别（named entity recognition, NER）、关系抽取（relation extraction, RE）和事件抽取（event extraction, EE）等核心 IE 任务上相比于现有常见方法的优越性。

1 相关工作

1.1 通用信息抽取

通用信息抽取任务是 NLP 领域的一项重要且充满挑战的任务，其核心目标是开发一个能够高效处理多种文本抽取需求的统一模型框架。

近年来，预训练语言模型如 GPT（generative pre-trained transformer）[8]、BERT（bidirectional encoder representations from transformers）[9] 和 ERNIE 迅速在 NLP 界引起广泛关注并被采纳，使得基于预训练语言模型的微调成为了处理多种 NLP 任务的通用方法。文献 [10] 首先提出了基于 Seq2Seq 架构的转化框架 T5（text-to-text transfer transformer），其通过为输入序列增加前缀，将所有 NLP 任务统一转化为文本到文本的格式，为整个 NLP 预训练模型领域提供了一个通用框架。进一步地，文献 [3] 在 T5 基础上提出了一个统一的文本到结构生成框架 UIE，将结构化的模式提示器和文本内容作为输出，直接生成结构化提取语言，从而实现对 IE 任务的统一建模。

另一方面，为了解决 Seq2Seq 架构的黑盒特性问题，即无法预测跨任务或跨模式迁移成功或失败的情况，文献 [11] 提出了基于统一语义匹配的通用框架 USM （unified semantic matching），将 IE 任务解耦为结构化和概念化两个基础任务。它通过

结构化操作重建目标结构与文本的语义关系，并通过概念化操作抽取文本或文本对，与目标语义标签进行匹配，从而有效地完成 IE 任务。基于大语言模型通过多任务训练和统一编码所展现出巨大的泛化潜力，文献 [12] 提出了用于通用抽取的端到端框架 InstructUIE（multi-task instruction tuning for unified information extraction），该框架利用自然语言指令来指导大模型完成各类 IE 任务。文献 [13] 设计了基于跨度提取的通用框架 UniEX（unified information extraction via a span-extractive perspective），通过将 IE 任务转化为跨度检测、分类和关联的统一问题，实现了对不同 IE 任务的高效处理，同时利用自编码器语言模型和流量关注机制协同编码不同元素，有效地提升了提取目标的准确性和泛化能力。

1.2　提示学习

提示学习（prompt learning）是 NLP 领域的一种新兴范式，特别是在预训练语言模型的使用背景下受到广泛关注。尽管传统的“预训练加微调”方法提升了多种 NLP 任务中的性能[14]，但预训练阶段与微调阶段不一致的优化目标可能导致性能降低，同时随着模型体量增大，训练和微调过程也变得更加复杂。提示学习通过将输入文本按照提示模板的要求进行转换，重构信息抽取任务以更有效地利用预训练模型的处理能力。这不仅实现了预训练和微调阶段优化目标的统一，还显著提升了预训练语言模型的性能。此外，提示学习通过在模型输入中加入特定的文本提示，扩大了预训练模型的适用范围，解决了不同任务间统一建模的挑战，有效增强了预训练模型的实用性和灵活性。

目前，提示学习的研究尚处于探索阶段，面临许多挑战。提示模板的设计是影响提示学习效果的关键因素，结合领域知识是优化提示模板设计的有效方法[15]。针对 NER 任务，文献 [16] 通过引入提示学习策略，旨在提高细粒度实体类型标注的准确性和效率，从而更好地利用预训练语言模型的能力，以适应特定的实体识别需求。文献 [17] 提出的 LightNER 是一个针对低资源环境下命名实体识别任务设计的轻量级生成式框架，该框架特别引入提示引导的注意机制，通过增强模型在处理关键信息时的聚焦能力，有效提升了在训练数据较少的情况下识别的准确性和效率。

面向 RE 任务，文献 [18] 将提示设计为可微分元素，使预训练语言模型更有效地适应于关系抽取任务，即使是在仅有极少量训练样本的情况下。文献 [19] 使用提示来增强在特定关系没有可用训练数据情况下的关系抽取任务，通过提示语言模型生成合成数据集，这些数据集随后作为训练材料，以提升模型在未经训练的相似例子中抽取关系三元组的能力。

面向 EE 任务，文献 [20] 通过使用提示学习改善模型对不同语言间实体关系的识别和抽取，进一步提升了模型在少数样本和零样本学习环境下的性能。文献 [21] 采用了提示学习的方法设计了一种策略，使得模型能主动生成查询或问题，从而精准地从文本中识别和提取事件相关的论元。文献 [22] 通过内部知识增强的提示学习来生成事件数据，这种方法利用预训练语言模型内嵌的知识，并通过设计的提示来指导模型更准确地识别和分类文本中的事件元素。

本文提出的 EBP-UIE 通过 ERNIE、BiLSTM 和 PN 的组合，综合利用统一文本到结构生成的生成式机制和 PN 文本片段抽取，相较于以往的工作，在语义理解、上下文处理、细节特征捕捉和精准定位上具有明显优势，特别是通过提示学习机制，实现了对复杂信息抽取任务的高效处理和统一建模。然而，EBP-UIE 模型结构复杂，训练和推理资源消耗较大，且高度依赖预训练语言模型，对于特定领域或语言需要进一步优化和调整。总体而言，EBP-UIE 在提高信息抽取准确性和复杂实体识别能力上表现突出，但在实际应用中仍需克服模型复杂性和任务分解的局限性。

2　模型

EBP-UIE 的总体框架如图 1 所示，针对不同类型的 IE 任务（如 NER、RE 和 EE），该框架把这些任务统一视为一个将文本中的抽取片段与分类标签组成多元组集合的过程。

具体而言，根据不同的 IE 需求，模型将复杂任务分解为若干抽取和分类的子任务，并引用提示学习机制灵活组合不同任务的多个子任务，实现对不同 IE 任务的统一建模。以关系抽取为例，包括主体抽取、客体抽取以及关系分类 3 个子任务，其操作流程为：首先，主体抽取模块处理输入句子，识别出所有潜在主体，即完成子任务 1；随后，这些主体与原句共同输入客体抽取模块，以识别与每个主体相关的客体，形成（主体、客体）映射对，对应子任务 2；最后，将这些映射对与原文一并输

入关系分类模块，以确定它们之间的具体关系，形成完整的关系三元组，完成子任务 3。

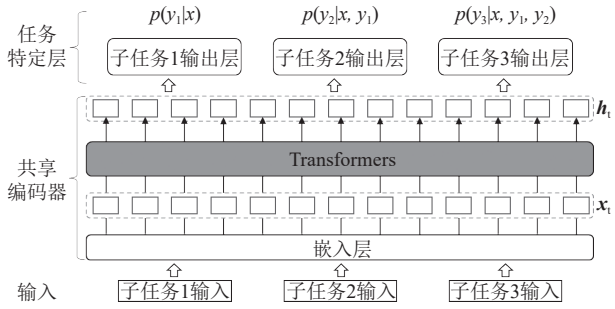

图 1 基于提示学习的 EBP-UIE 模型框架

模型的输入层接收各子任务的输入数据，其中可能包括特定的提示信息，帮助模型更好地理解每个子任务的需求。紧接着是嵌入层，负责将文本数据转换为向量形式$\boldsymbol{x}_t$，作为模型后续处理的基础。在嵌入层之上是共享编码器层，由一系列 Transformer 编码器组成，这些编码器共同捕获包含了丰富的上下文和语言结构信息的隐层向量 $\boldsymbol{h}_t$。特别地，本模型采用 ERNIE 中文预训练语言模型作为基础，针对中文数据提供更加精准的处理能力。

每个子任务在任务特定层都配备了专属的输出层，从共享的 Transformer 编码器接收特征并生成任务相关的结果。整个过程从输入原始文本及其相关提示开始，最终输出一系列精准的信息片段及其分类标签，通过精确提取和分类文本中的关键信息，模型能够有效组织这些信息成结构化地输出，提供一种既高效又可扩展的 IE 方案。

2.1 面向 EBP-UIE 的抽取子任务

抽取子任务致力于从输入文本中提取关键信息片段（span），本文采用 PN 来应对文本中 SPAN 的抽取挑战。这种方法不仅有效应对了实体的嵌套问题，还能够依据数据分布自适应调节阈值，实现精确度与召回率之间的优化平衡。如图 2 所示，抽取模块的核心策略是针对特定的 IE 任务，设计输入文本的构造方式，随后通过 ERNIE 模型进行编码，然后利用 BiLSTM 层进一步萃取文本特征，最终借助 PN 定位并抽取所需的 span。该模块主要包括输入层、特征提取层和输出层。

图 2 基于 EBP 的抽取模型结构图

1）输入层

输入层由输入文本和嵌入层组成。首先对输入文本进行预处理，包括分词、添加特殊标记符号 [CLS] 和 [SEP] 等。[CLS] 标记表示句子的开头，用于文本分类任务中，其对应的嵌入被用作分类的聚合表示，而 [SEP] 表示句子的结尾，其对应着输入文本中最后一个词的词向量，它的作用是用来分割不同的句子。之后，这些处理过的文本被送入嵌入层转换为数值型向量，以符合 ERNIE 模型的输入格式。ERNIE 的输入向量由 3 部分组成：词向量、段向量、位置向量。

生成词向量首先依赖于细致的分词过程。对于英文，ERNIE 使用 Wordpiece 算法进行基于贪婪算法的最长匹配分词；中文文本则按字符切分，每个汉字独立成为一个词元。分词后，每个词元对应到词汇表中的唯一标识符 ID。这些 ID 被进一步转化为 one-hot 编码，再通过与训练好的词嵌入矩阵相乘，产生最终的词向量$\boldsymbol{E}^{\text{token}}$。段向量的设计旨在区分句子对，特别是在处理句子对或段落级任务时显得尤为重要。ERNIE 通过识别特殊的 [SEP] 标记来区分句子对，其中位于第 1 个 [SEP] 标记之前的所有词元被编码为段向量 **0**，而第 1 个 [SEP] 标记

之后的词元则被编码为段向量 **1**。位置向量则用于捕捉词元在文本中的位置信息。由于 Transformer 架构并不直接处理输入的位次信息，ERNIE 通过位置向量来补充每个词元在句子中的相对或绝对位置信息，使模型能够理解词序和句子结构。最终的输入结果如式 (1) 所示：

$$\boldsymbol{E}^{\text{input}} = \boldsymbol{E}^{\text{token}} + \boldsymbol{E}^{\text{seg}} + \boldsymbol{E}^{\text{pos}} \tag{1}$$

2）特征提取层

特征提取层采用针对中文数据优化的 ERNIE。它在预训练阶段引入更细粒度的知识掩蔽策略，如实体级和短语级的 MASK 机制，使其不仅能够学习到基于字符的语言规律，还能够更好地理解和表示更复杂的语言结构。ERNIE 由 12 个 Transformer 编码器层堆叠而成，其中每个编码器层都包含多头自注意力机制、层级归一化以及前馈神经网络。模型编码器架构如图 3 所示，在处理流程中输入向量会通过多头自注意力层进行初步处理。

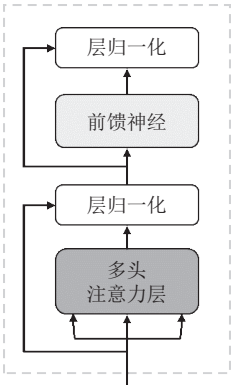

图 3　Transformer 编码器结构图

在多头自注意力层中，采用点积缩放的方式来执行自注意力计算。这种机制通过计算查询、键和值之间的关系来实现注意力的分配，其中查询、键和值通常是输入数据的不同表示。具体而言，它使用点积来计算查询和每个键之间的相似度，然后通过缩放操作调整这些相似度的范围，接着应用 softmax 函数将相似度转换为概率分布，最后用这个分布来加权值，产生加权和作为输出。带缩放的点积注意力机制的公式为：

$$\text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\text{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{2}$$

式中，$\boldsymbol{Q}$、$\boldsymbol{K}$ 和 $\boldsymbol{V}$ 分别代表查询（query）、键（key）和值（value）矩阵；$\sqrt{d_k}$是键向量的维度。除以 $\sqrt{d_k}$是为了防止点积在维度较高时值过大，导致 softmax 函数的梯度过小，从而影响模型的学习和性能。在多头自注意力层中，每个注意力头都独立进行计算，之后将所有头的输出结果进行拼接，形成该层的最终输出，该输出随后被送入下一层继续进行处理。该层的计算过程如下所示：

$$\text{MultiHead}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \text{Concat}(\text{head}_1,\text{head}_2,\cdots,\text{head}_i)$$

$$\text{head}_i = \text{self\_attention}_i(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n) \tag{3}$$

式中，$\boldsymbol{X} = \boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$为输入序列；$\text{head}_i$为第 i 个自注意力头的输出。

随着模型网络层数的增加，梯度在反向传播过程中往往会逐渐减弱，产生梯度消失。尤其是在深层网络结构中，梯度消失会导致模型难以训练。为了缓解这一问题，模型中引入了残差连接（residual connection）和层归一化（layer norm）机制。残差连接通过将输入直接添加到网络某层的输出上，形成一条“捷径”，从而使得深层网络中的信号能够更直接地传递。这种结构有效地缓解了深度网络在训练过程中遇到的优化难题，提高了网络的训练效率和稳定性。层归一化则是在网络的每一层内部对输入数据进行归一化处理，它通过调整数据分布来改善梯度消失或梯度爆炸的问题，从而有助于模型的稳定和快速训练。层归一化在处理每一层的输入时，会对该层内所有神经元的激活值进行归一化，确保数据分布保持在一个合理的范围内，从而维持梯度在合适的大小，促进模型的学习过程。计算过程如下所示：

$$\boldsymbol{Z} = \text{LayerNorm}(\boldsymbol{X} + \text{multi\_head\_layer}(\boldsymbol{X})) \tag{4}$$

残差连接和层归一处理后的输出作为输入进入前馈神经网络（feed-forward neural network, FFNN）层，该层通常包含两个线性变换，中间夹杂一个非线性激活函数，目的在于增强模型的非线性拟合能力，具体的计算过程如下：

$$\text{FNN}_1(\boldsymbol{Z}) = \sigma(\boldsymbol{W}_1\boldsymbol{Z} + \boldsymbol{b}_1)$$

$$\text{FNN}_2(\boldsymbol{Z}) = \boldsymbol{W}_2 \times \text{FNN}_1(\boldsymbol{Z}) + \boldsymbol{b}_2 \tag{5}$$

BiLSTM 层是由两个单向的 LSTM 组合而成的，它们分别从文本的前向和后向对序列进行编码。BiLSTM 层能够有效地处理序列数据，捕捉长距离的依赖关系。在处理 ERNIE 输出的隐层向量时，BiLSTM 层不仅进一步提取特征，而且实现了抽取要素与原始文本特征的深度融合。具体而言，原始文本通过 ERNIE 进行编码后，得到一系列的特征向量$\boldsymbol{X} = \boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$，这些特征向量为 BiLSTM

层提供了丰富的语境信息和初步的语义表示。随后，BiLSTM 层对这些向量进行进一步的特征提取，分别从序列的前向和后向进行编码，并提取特征$\overrightarrow{\boldsymbol{h}_t}$和$\overleftarrow{\boldsymbol{h}_t}$。通过将这两个方向上的特征进行拼接，BiLSTM 层生成了一个新的文本编码$\boldsymbol{Z}=\boldsymbol{z}_1,\boldsymbol{z}_2,\cdots,\boldsymbol{z}_n$，其中每个 $\boldsymbol{z}$包含了从两个方向融合而来的信息。计算过程如下所示：

$$\overrightarrow{\boldsymbol{L}}=\{\overrightarrow{\boldsymbol{l}_1},\overrightarrow{\boldsymbol{l}_2},\cdots,\overrightarrow{\boldsymbol{l}_n}\}=\overrightarrow{\mathrm{LSTM}}(\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\})$$

$$\overleftarrow{\boldsymbol{L}}=\{\overleftarrow{\boldsymbol{l}_1},\overleftarrow{\boldsymbol{l}_2},\cdots,\overleftarrow{\boldsymbol{l}_n}\}=\overleftarrow{\mathrm{LSTM}}(\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\})$$

$$\boldsymbol{Z}=\mathrm{concat}\left(\overrightarrow{\boldsymbol{L}}\overleftarrow{\boldsymbol{L}}\right) \tag{6}$$

式中，$\overrightarrow{\boldsymbol{L}}$和$\overleftarrow{\boldsymbol{L}}$分别为前后向 LSTM 提取的文本特征，它们由系列向量$\boldsymbol{l}_1,\boldsymbol{l}_2,\cdots,\boldsymbol{l}_n$组成，每个输入向量$\boldsymbol{l}_i$通常包含了关于输入序列中相应元素及其上下文的信息。

3）输出层

输出层采用的是 PN，它主要由两层 softmax 分类网络组成，专门设计用于预测文本片段的开始和结束位置。文本通过 ERNIE 进行编码之后，通常取 ERNIE 最后一层的输出作为文本的高级编码表示。为了从这些高级编码中准确抽取目标信息，模型引入了两个新的参数矩阵$\boldsymbol{W}_s$和$\boldsymbol{W}_e$。这些参数矩阵分别用于对文本编码进行线性变换，以生成针对抽取目标的开始位置和结束位置的预测。第 t 类要素的起止位置计算公式如下：

$$P_{\mathrm{ts}}(i)=\mathrm{softmax}(e_i\boldsymbol{W}_{\mathrm{ts}}+b_{\mathrm{ts}})$$

$$P_{\mathrm{te}}(i)=\mathrm{softmax}(e_i\boldsymbol{W}_{\mathrm{te}}+b_{\mathrm{te}}) \tag{7}$$

式中，e_i代表第i个字的 ERNIE 编码；$\boldsymbol{W}_{\mathrm{ts}}\in\boldsymbol{R}^{H\times T}$、$\boldsymbol{W}_{\mathrm{te}}\in\boldsymbol{R}^{H\times T}$分别是用于计算第 t 类要素开始和结束位置的权重矩阵，其中H表示隐层向量的维度，$T=t+1$表示考虑的要素类型总数；b_{ts}和b_{te}是偏置参数。权重矩阵与偏置参数都是在训练过程中学习得到的。计算得到概率分布后，利用 argmax 函数确定抽取要素的开始和结束位置的序列。抽取要素起始位置计算公式为：

$$\mathrm{start}_t=\mathrm{argmax}(P_{\mathrm{ts}})$$

$$\mathrm{end}_t=\mathrm{argmax}(P_{\mathrm{te}}) \tag{8}$$

在 IE 任务的最终阶段，模型通过对之前获得的开始和结束位置的概率分布进行解码，以精确定位特定信息片段在文本中的具体区间。这种基于开始和结束位置序列解码的策略核心优势在于，它使得模型能够独立地确定每个实体的界限，而不需依赖任何预设的实体结构或其序列。这种方法的灵活性允许模型在相同文本段落中重复执行解码过程，每一次都能集中识别不同的起始和终止点，从而有效地识别和提取文本中可能存在的嵌套或相交的实体。

4）损失函数

PN 本质上通过将抽取任务转化为两个关联的多分类任务来工作，即对于文本中的每一个 token，模型需要预测该 token 作为抽取目标 span 的开始位置和结束位置的概率。这种方法的优势在于它能够直接针对每个 token 进行分类，从而精确地识别出信息片段的边界。

由于这一任务本质上是多分类问题，因此自然而然地选用了交叉熵损失函数（cross entropy loss, CEloss）作为模型的损失函数。交叉熵损失函数能够衡量模型预测的概率分布与真实标签之间的差异，是分类任务中常用的损失函数之一。具体计算如下式所示：

$$H(qp)=-\sum_{i=1}^{n}p(x)\log(q(x)) \tag{9}$$

式中，$p(x)$为正确的抽取要素标签；$q(x)$为模型预测的抽取要素标签。

整体的损失函数$\mathcal{L}_{\mathrm{t}}$由开始位置损失函数$\mathcal{L}_{\mathrm{ts}}$和结束位置损失函数$\mathcal{L}_{\mathrm{te}}$相加得到，计算过程如下所示：

$$\mathcal{L}_{\mathrm{ts}}=\mathrm{CELoss}(pq)$$

$$\mathcal{L}_{\mathrm{te}}=\mathrm{CELoss}(pq)$$

$$\mathcal{L}_t=\mathcal{L}_{\mathrm{ts}}+\mathcal{L}_{\mathrm{te}} \tag{10}$$

2.2 面向 EBP-UIE 的分类子任务

分类子任务使用 ERNIE 进行文本分类，模型结构如图 4 所示，也是由输入层、编码层、解码输入层组成。其中，输入层和编码层与抽取模块类似，在分类模块的输入层中，首先对原始文本进行一系列预处理操作，以便适配后续的模型处理流程，然后使用 ERNIE 进行编码。

在解码输出阶段，将文本经过嵌入处理后的上下文向量$\boldsymbol{E}_{\mathrm{context}}$传递至全连接层。为了避免过拟合现象，模型引入了 dropout 机制。此层输出的是一

个 k 维的特征向量，其中 k 的值等于分类类别数目加 1。随后，模型通过 Sigmoid 激活函数来计算每个类别的概率，得到每个类别的 0～1 标签的概率分布。模型的损失函数采用 Sigmoid 交叉熵，用于评估输出层预测标签与实际标签之间的误差，其定义如下式所示：

$$\mathrm{Loss}(y\widehat{y}) = -\frac{1}{C}\sum_{i=1}^{m} y^i \log\left(\frac{1}{1+\exp(-\widehat{y}^i)}\right) + (1-y^i)\log\left(\frac{\exp(-\widehat{y}^i)}{1+\exp(-\widehat{y}^i)}\right) \tag{11}$$

式中，C代表分类的类别数量；y^i是实际的标签序列；而$\widehat{y}^i$则是模型输出的、尚未通过激活函数处理的预测标签序列。此损失函数确保了模型可以在训练过程中通过最小化预测误差来优化其参数。该模型通过其高效的学习机制，不仅能够在 IE 任务中准确地执行关系分类，而且同样适用于复杂的事件分类，极大地增强了在各种文本语料中识别和理解复杂语义关系的能力。

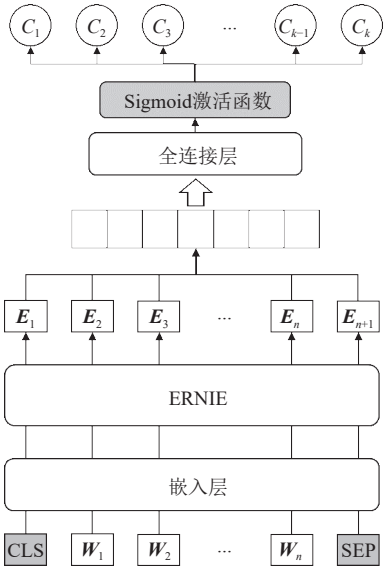

图 4　基于 ERNIE 的分类模型结构图

2.3　基于 EBP-UIE 的通用信息抽取任务实现

2.3.1　基于 EBP-UIE 的命名实现识别

命名实体识别是 NLP 领域的核心任务之一，其目的在于从文本中辨识出具有实际意义的实体，诸如人名、地点、组织机构和时间等。在 EBP-UIE 框架下，NER 任务可以通过单一的抽取模块来实现。相较于传统的基于序列标注的 NER 方法，该框架能够有效地处理实体识别中常见的诸多挑战，包括固定窗口大小导致的上下文限制、实体嵌套、实体间的重叠以及对标签体系的过度依赖等问题。

2.3.2　基于 EBP-UIE 的关系抽取

关系抽取的核心目的在于从文本资料中挖掘并标识出各种实体（如个人姓名、地点名称、机构等）及它们之间的语义联系，并将此信息以结构化的格式表示出来。

使用 EBP-UIE 进行 RE 的流程如图 5 所示，该模型采纳了提示学习技术来增强任务处理的效率和准确性。具体而言，模型将命名实体识别子任务的输出结果作为当前任务的提示信息输入，以此来增进对语言微观细节的洞察力及加强对复杂文本内容的处理能力。

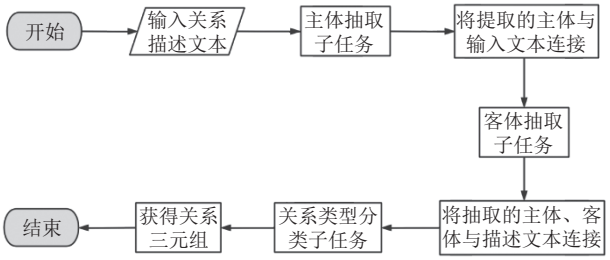

图 5　基于 EBP-UIE 模型的关系抽取流程图

模型将整个 RE 过程细化为 3 个关键步骤：主体抽取、客体抽取以及关系分类，其中包含两个抽取子任务和一个分类子任务，以下部分将进一步阐释各个步骤的具体操作流程。

主体抽取由面向 EBP-UIE 的抽取子任务完成，旨在从提供的句子中精确地识别出所有潜在的主体，这些主体通常指的是文本中的实体。为了实现这一目标，主体抽取任务需要经过输入层和特征提取层的处理，最终在输出层得到结果。随后，在客体抽取阶段，继续使用抽取子任务针对已经识别出的主体，进一步从句子中提取相应的客体，形成主体−客体对。这一过程同样经历输入层和特征提取层，最终在输出层完成抽取。作为关系抽取流程的最后阶段，关系分类的核心任务是明确已识别主体和客体之间的具体关系。该过程通过框架内的分类模块完成，该模块将结合主体−客体对和原始文本作为输入，输出它们之间的关系标签，形成完整的关系三元组（主体、关系、客体）。与分类子任务一样，关系分类也要通过输入层和编码层，最终在解码输出层得到分类结果。

2.3.3　基于 EBP-UIE 的事件抽取

事件抽取旨在识别文本中的关键事件信息，并

将这些信息转化为结构化格式。典型的 IE 系统会识别事件的触发词（即标志事件发生的词语），事件类型（如事故、选举、攻击等），以及事件的相关论元和属性（如时间、地点、参与者等）。使用 EBP-UIE 模型进行事件抽取的流程如图 6 所示，初始阶段通过事件类型检测模块对文本进行深度分析，旨在准确辨识文中所述的一个或多个事件实例。此后，依据预定义的事件模式（schema），模型为识别到的每个事件构建一组标签集合，这些集合详尽地描述了事件的本质、触发词及其相关论元，为整个抽取过程提供了清晰的框架和指导。随着流程的推进，模型采用提示学习策略，以先前生成的标签集为导向，精确地提取与事件密切相关的触发词和论元角色。在处理描述多重事件的复杂文本时，模型能够灵活地将各独立事件的结果综合汇总，输出一份完整的多事件结构化信息，确保了信息的全面性与准确性。总的来说，事件抽取主要由事件类型检测、事件触发词与事件论元角色抽取 3 个模块组成。

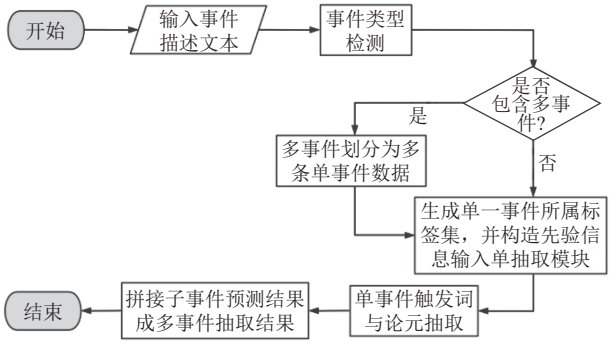

图 6 基于 EBP-UIE 模型的事件抽取流程图

事件类型检测模块由 EBP-UIE 框架中的分类子任务实现，其主要任务是从输入的事件描述文本中准确识别并分类出各种事件类型。鉴于一段文本可能同时涉及多个事件，该模块采用了 ERNIE 预训练语言模型执行多标签分类，以全面捕捉文本中潜在的所有事件大类及其子类别。事件触发词抽取由框架中的抽取子任务实现，事件触发词是指那些能够明确表示某一事件发生的词汇，如“签署”可能表示一个“协议签署”事件的触发词。以检测出的事件类型作为提示信息构建输入文本，抽取对应的触发词。最后，论元抽取任务在 EBP-UIE 框架内也是通过抽取子任务实现。核心思路是根据文本中的事件类型，利用 schema 中相应的描述来构造针对该事件类型的论元抽取问题。接着，通过 ERNIE 对文本进行编码并结合 BiLSTM 来进一步提取特征，最后应用一系列二分类器来识别事件的相关论元。

3 实验设计

3.1 数据集

为了适应各种 IE 任务，本文选取了专门的数据集进行模型训练和评估。每个数据集针对其对应的抽取任务提供了丰富且专业的标注信息，为模型的优化和精确度提供了坚实的基础。

面向 NER 任务，本文使用了两个不同的中文实体数据集，分别是 MSRA（https://github.com/InsaneLife/ChineseNLPCorpus/tree/master/NER/MSRA）和 PeopleDaily（https://github.com/OYE93//Chinese-NLP-Corpusblob/master/NER/People's%20Daily/readme.md）。MSRA 数据集是由微软亚洲研究院发布的一个新闻领域 NER 数据集，实体类型主要包括 LOC（地名）、ORG（机构名）、PER（人名）。MSRA 数据集并没有验证集，本文实验采用随机从训练集中选择 10% 的样本作为验证集的方法。PeopleDaily 数据集则是基于人民日报语料进行标注的，同样标注了地名（LOC）、机构名（ORG）和人名（PER）3 种实体类型，且未进行人工分词标注。两个数据集的相关信息如表 1 所示。

表 1 NER 数据集统计信息 个

数据集	训练集	验证集	测试集	总计
MSRA	46 364	4 365	4 364	55 093
PeopleDaily	20 864	4 636	2 318	27 818

针对关系抽取，本实验选用了百度发布的用于信息抽取的大规模中文数据集 DuIE（https://hyper.ai/datasets/16618）。它汇集了超过 21 万条中文句子，涵盖 45 万个实例，包括 24 万个实体和 34 万个主谓宾（subject-predicate-object, SPO）三元组，源自百度百科、百度新闻和百度贴吧等多样化的文本来源。DuIE 数据集的统计信息如表 2 所示。

表 2 DuIE 数据集统计信息 个

数据类型	训练集	验证集	测试集	总计
句子	173 108	21 639	19 992	214 739
实例	364 218	45 577	48 389	458 184

DuEE 数据集是百度发布的一个专为中文事件抽取任务设计的大规模数据集（https://github.

com/zhoujx4/DuEE）。DuEE 共包含 19 640 篇事件描述文本和 41 520 个事件论元，覆盖 121 种论元角色，是目前最大规模的中文事件抽取数据集。该数据集经过严格的人工审核和标注，标注准确率超过 95%。DuEE 数据集的统计信息如表 3 所示。

表 3　DuEE 数据集统计信息　　个

数据类型	训练集	验证集	测试集	总计
实例	11 958	1 498	3 500	16 956
事件	13 478	1 790	4 372	19 640
论元	29 052	3 696	8 772	41 520

3.2　评价指标

通用信息抽取的评价策略是通过比较抽取系统从输入文本中抽取出来的记录与标准答案中的多元组来进行的。这个过程涵盖了将系统输出的结果转换为多元组形式，并在去重后进行评价。

1）将抽取任务的输出结果统一表示为多元组形式，这些多元组根据任务的不同可能是二元组或三元组。如命名实体识别（实体提及、实体类型）表示为实体二元组，关系抽取（关系类型、主体 span、客体 span）表示为关系三元组，事件抽取（事件类型、论元角色、论元 span）表示为事件论元三元组。多元组中的基本元素包括文本块抽取结果（span）、类型标签和关联关系标签。

2）对于输入文本中重复出现的相同信息，评价时将进行去重处理。即模型输出了多个相同的多元组，评价脚本会将它们视为一个多元组进行评价。评价主要基于正确预测的多元组数量。对于每个与标准答案匹配的多元组，正确预测数增加 1。这包括准确匹配的实体抽取结果、关系类型、事件类型、论元角色等。

最终的评价指标采用精确率（Precision）、召回率（Recall）和 F1 分数，这些都是基于 TP（真正例）、TN（真负例）、FP（假正例）和 FN（假负例）来计算的。计算公式分别如下：

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}}$$

$$\text{F1} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{12}$$

3.3　参数设置

本文首先对数据集中的文本长度进行了统计，据此将模型能够接受的最大输入文本长度（max_seq_len）设定为 512。对于超出这一长度的文本，进行截断处理，而长度不足的部分则用 0 进行填充。采用了预训练模型 ERNIE 3.0，该模型的词嵌入维度设置为 768，并使用了 12 层 Transformer Encoder 层进行叠加。在训练过程中，选择 Adam 作为优化器。对于命名实体识别任务，batch_size 设为 64，学习率定为 3×10^{-5}；而关系抽取和事件抽取任务的 batch_size 设为 32，学习率同样为 3×10^{-5}。训练共进行 20 个轮次（epochs），每完成一个训练轮次便执行一次验证，对在验证集上性能最优的模型参数进行保存，供最后的测试使用。所有的池化操作均采用平均池化方法。

4　实验结果与分析

4.1　基于 EBP-UIE 的命名实现识别实验结果与分析

为验证 EBP-UIE 模型在 NER 任务的有效性，分别在 PeopleDaily 和 MSRA 数据集上进行实验，并将实验结果与对比模型进行比较，如表 4 所示。

表 4　面向 NER 任务的算法对比实验结果　　%

Models	MSRA			PeopleDaily		
	Precision	Recall	F1	Precision	Recall	F1
BiLSTM-CRF	88.06	83.82	85.88	84.65	82.04	83.32
BiLSTM-CNN-CRF	89.86	84.95	87.33	86.16	85.59	85.88
BERT	93.77	92.42	93.09	91.31	90.89	91.09
BERT-BiLSTM-CRF	94.57	93.62	94.09	92.89	91.05	91.96
UIE	94.87	83.45	88.80	**96.87**	93.82	95.32
EBP-UIE(本文)	**96.37**	**95.47**	**95.92**	96.28	**95.42**	**95.85**

1）BiLSTM-CRF[4]：结合了 BiLSTM 与 CRF，有效地捕获了序列数据的前后文依赖关系，并在序列标注任务中实现了高精度。

2）BiLSTM-CNN-CRF[23]：在 BiLSTM-CRF 的基础上引入 CNN 来提取更丰富的文本特征，进一步提高了模型对文本特征的学习能力。

3）BERT[9]：作为一个代表性的预训练语言模型，BERT 通过大量无标注文本的预训练，学习到了丰富的语言表示，能够直接应用于 NER 等下游任务，显著提高了识别的准确性。

4）BERT-BiLSTM-CRF[24]：将 BERT 的强大语义表示能力与 BiLSTM 的序列处理能力结合，并通过 CRF 层进行序列标注，有效融合了 BERT 的深度语义学习和 BiLSTM-CRF 的序列建模优势。

5）UIE[3]：统一文本到结构生成的通用框架，

不局限于传统的 NER 任务，通过提示学习等先进技术，能够灵活适应各种实体类型和复杂关系的抽取，展现了良好的通用性和适应性。

通过分析表 4 可以发现，EBP-UIE 在 NER 任务上取得了最高的 F1 分数，在 MSRA 和 PeopleDaily 数据集上分别达到了最高的 95.92% 和 95.85%，表明它在识别正确实体的整体效能上具有显著优势。同时，它的召回率也是所有模型中最高的，分别为 95.47% 和 95.42%，说明模型在覆盖所有正确实体方面都表现优异。尽管 EBP-UIE 在 PeopleDaily 数据集上，其精确率略低于 UIE，但是其在两个数据集上平均精确率优于 UIE，验证模型在正确识别实体方面具有较高适应性。

与传统的 BiLSTM-CRF 和 BiLSTM-CNN-CRF 模型相比，EBP-UIE 的 F1 提升显著。相比于 BERT 和 BERT-BiLSTM-CRF 这两个结合了预训练语言模型的方法，EBP-UIE 仍然展现出更强的性能，尤其是在保持高召回率的同时实现了更高的精确率，这意味着 EBP-UIE 不仅能识别出更多正确的实体，而且减少了错误识别。 UIE 作为生成式模型，能够直接生成标注实体和其类型的序列，不需要单独的实体边界识别和分类步骤，但是由于生成式架构的黑盒特性，它在不同类型的数据集上平衡精确率和召回率的能力还有改进空间。

这些结果表明 EBP-UIE 在 NER 任务上具有较强的泛化能力和优越的性能。通过整合 ERNIE 和 PN，EBP-UIE 优化了实体的边界识别，使其在处理重叠实体和复杂文本结构方面尤为有效。

4.2 基于提示学习的 EBP-UIE 关系抽取实验结果与分析

为验证 EBP-UIE 模型在 RE 任务上的有效性，在 DuIE 数据集上进行实验，且与对比模型比较。对比模型（UIE 不再赘述）描述如下所示。

1）BERT-CRF[25]：结合了 BERT 的强大上下文编码能力和 CRF 的序列优化特性，使得模型在预测实体关系时能够考虑到标签之间的依赖性，从而提高整体的预测准确性。

2）BERT-Sigmoid[26]：独立地预测文本中每对实体之间可能存在的多种关系类型，对于识别并分类实体间的复杂关系网络尤为有用。

3）CasRel[27]：CasRel 采用层叠 PN，会先识别所有可能的主语（头实体）；然后在给定类别关系的条件下，再去识别与主语相关的宾语（尾实体）。

从表 5 中可以看出，EBP-UIE 模型在 RE 任务上表现卓越，尤其在 DuIE 数据集上以 81.60% 的最高 F1 分数领先于其他模型。EBP-UIE 与其他经典模型相比，在 F1 分数上达到了最高的 81.60%，表明它在识别关系的整体效能上具有显著优势。它的精确率和召回率分别为 80.95% 和 82.27%，表明其不仅在辨识正确关系上准确无误，而且在避免错过关系的同时，也确保了广泛的覆盖。

表 5 面向 RE 任务的算法对比实验结果 %

Models	DuIE		
	Precision	Recall	F1
BERT-CRF	69.54	70.38	69.96
BERT-Sigmoid	72.14	72.57	72.35
CasRel	**81.29**	80.54	80.91
UIE	73.35	76.22	74.76
EBP-UIE(本文)	80.95	**82.27**	**81.60**

相较于 BERT-CRF 和 BERT-Sigmoid 这两种基线模型，EBP-UIE 在精确率、召回率以及 F1 分数上均有显著提升，证明其在关系抽取精度上的优势。UIE 在关系抽取任务中需要同时理解文本、确定实体边界并预测关系，这一连串的任务导致错误的累积和放大。CasRel 采用级联标签的方式，逐步细化抽取结果，从而提高了关系抽取的精确率。尽管 CasRel 在精确率上略占优势，但 EBP-UIE 的高召回率更符合关系抽取的任务要求，这表明了 PN 与使用提示学习引入先验信息的有效性。

综上所述，EBP-UIE 在关系抽取任务中展现出了显著的优势。相较于现有较为先进的模型，EBP-UIE 通过 PN 和提示学习，提高了关系抽取的精确率和召回率，表现出了更高的泛化能力和鲁棒性，有效解决了关系抽取中的重叠实体和复杂关系挑战，克服了生成式模型的黑盒特性。

4.3 基于提示学习的 EBP-UIE 事件抽取实验结果与分析

为了评估 EBP-UIE 模型在 EE 任务的性能，本文选择在 DuEE 数据集上进行了一系列的实验。事件抽取主要分为触发词抽取和论元抽取两个模块，通过与当前领域内的先进模型进行比较，本文验证 EBP-UIE 在处理复杂事件抽取任务时的有效性和优越性。下面列出了参与对比的模型（UIE 不再赘述），以便于全面评估 EBP-UIE 的性能表现。

1）CNN-BiLSTM-CRF[28]：采用 CNN 进行特征提取，BiLSTM 捕获上下文信息，CRF 层建模标签序列依赖关系，适用于序列标注任务如事件抽

取，提升了标注准确性。

2）MTL-CRF[29]：通过联合标注触发词和论元，针对每种事件类型训练专门的 CRF 模型，利用多任务学习共享不同事件之间的知识，缓解数据不平衡和稀疏性问题。

3）BMPN[30]：一种基于 BERT 的多层标签指针网络模型，采用“尺取法”策略来识别文本中事件元素的起始和结束点。此外，该模型通过部署多个二元分类网络，精确地判断元素的事件类型和角色，从而有效应对事件元素及其角色重叠的问题。

通过分析表 6 可以发现，EBP-UIE 在触发词抽取任务中的精确率高达 87.41%，F1 分数为 87.96%，均为最高，召回率仅次于 BMPN 模型，达到了 88.52%。在角色抽取任务中的精确率为 83.45%，F1 分数为 82.84%，均优于上述提到的其他模型，召回率仅次于 UIE，达到了 82.25%。

表 6　面向 EE 任务的算法对比实验结果　%

Models	Trigger Extraction			Role Extraction		
	Precision	Recall	F1	Precision	Recall	F1
CNN-BILSTM-CRF	79.31	81.58	80.43	70.26	71.18	70.72
MTL-CRF	82.33	85.20	83.74	82.75	76.44	79.47
BMPN	85.29	**90.13**	87.64	82.39	81.98	81.72
UIE	83.58	83.37	83.47	80.08	**83.78**	81.89
EBP-UIE(本文)	**87.41**	88.52	**87.96**	**83.45**	82.25	**82.84**

在对比模型中，基于 skip-gram 词嵌入和传统 CNN 技术的 CNN-BiLSTM-CRF 模型主要用于学习句子级特征，但其性能未能超越基于预训练模型的方法。MTL-CRF 虽在精确率方面展现了一定的优势，但由于分词精度的限制，其召回率较低。BMPN 在提高触发词和论元抽取的召回率方面取得显著成效，但其精确率仍有提升空间，这与模型在解决元素重叠问题时所融合信息的不足有关。UIE 在触发词抽取和论元抽取的召回率较为平衡，但是精确性弱于 EBP-UIE 和 BMPN。

综上所述，本文所提出的 EBP-UIE，通过整合 ERNIE 的深层语义理解能力和 PN 的精确定位能力，显著提高了 EE 任务的鲁棒性和泛化性。特别是在事件类型检测、触发词定位及论元识别等方面，相比其他传统方法，本模型展现出了明显的性能优势，表明融入事件类别先验知识的重要性。

4.4　基于提示学习的 EBP-UIE 关系抽取消融实验分析

进一步地，本文将进一步讨论 EBP-UIE 各个模块对通用信息抽取性能的影响。鉴于关系抽取的复杂性，在关系抽取任务上进行了消融实验，并得到了表 7 所示的结果。

通过对表 7 的分析可知，移除 ERNIE 模块后，模型性能显著下降，F1 分数下降了 13.02%，这说明 ERNIE 模块的作用尤为显著，其对模型性能的贡献远超过其他任何单一模块。这主要是因为 ERNIE 模块能够利用预训练语言模型来理解文本中的复杂语义信息，通过在大规模语料库上的预训练，能够捕捉到丰富的语言特征和先验知识，这对于理解文本内容，尤其是在实体和关系抽取任务中，是极其有价值的。

表 7　基于提示学习的 EBP-UIE 消融实验结果　%

Models	DuIE		
	Precision	Recall	F1
-ERNIE	62.56	73.47	67.58
-BiLSTM	78.52	81.24	79.85
-PN	77.34	80.63	78.95
-Prompt	76.34	78.25	77.28
EBP-UIE(本文)	**80.95**	**82.27**	**81.60**

注：-代表去除模型中相应的模块，EBP-UIE作为消融实验的基线模型

相比之下，移除 BiLSTM 模块后，F1 分数下降了 1.75%，这说明虽然 BiLSTM 的文本特征抽取能力和词嵌入的效果不如 ERNIE，但是能够处理序列数据，进一步捕获文本中的长距离依赖信息。在实体和关系抽取任务中，理解实体之间的上下文关系是至关重要的，BiLSTM 通过其双向结构帮助模型捕获前后文信息，从而提升模型对实体和实体关系的理解能力。

PN 模块的移除也导致模型 F1 分数下降了 2.65%，这表明 PN 模块能有效地定位抽取要素的起始位置，它通过精确定位实体和关系的起始点，进一步提升了模型在关系抽取任务中的表现。PN 模块对于提高模型整体的精确率和召回率至关重要，特别是对处理具有复杂结构的文本。

另外，去除提示信息模块后，模型的 F1 分数下降了 4.32%，说明提示信息模块能够为模型提供额外的上下文或背景知识，帮助模型更好地理解和处理复杂的文本信息。这种额外的信息可以是关于实体类型的提示、实体间关系的可能性，或是其他有助于模型理解的指导信息。提示信息模块的存在极大地增强了模型的鲁棒性和适应性，使其能够更好地应对多样化的文本场景。

因此，上述的各个模块在模型中都发挥着重要

作用，共同促进了模型的整体性能，移除任何一个模块都会导致性能的下降，也表明 EBP-UIE 通过这些模块的有效组合，较好地实现 IE 任务。

5 结束语

通用信息抽取方法通过统一的模型架构，旨在从非结构化文本中识别并提取关键信息，包括实体、关系和事件等，以便支持各种下游应用。尽管现有方法取得了一定的进展，但仍面临多样化任务处理的复杂性、对新实体类别的适应性不足。鉴于此，本文提出基于提示学习的 ERNIE-BiLSTM-PN 通用信息方法，通过统一模型架构，克服 IE 任务的复杂性和跨任务知识共享的局限性，实现高效准确的信息提取。该方法巧妙地融合了 ERNIE 的深度语义解析能力、BiLSTM 的高效序列分析功能，以及 PN 的精确定位机制。在此框架中，ERNIE 以其卓越的文本理解和上下文分析能力为基础，为复杂的 IE 任务打下坚实的基石；BiLSTM 则通过其对序列数据的敏锐洞察，捕捉并理解文本中的时序特征和远程依赖关系；PN 的引入则进一步增强了框架对文本信息起始和结束位置的精准识别，特别是在处理实体嵌套这一挑战性问题上具有较突出的优势。这种综合应用不仅为通用信息提供了一种全面而高效的解决方案，也极大地提升了 IE 的效率和准确性。

本文虽然对通用信息抽取方法展开了研究，但是仍存在着一些需要改进的地方。首先，EBP-UIE 为每一个实体、关系和事件都生成相应的样本来进行推理，在未来的工作中，有必要优化其推理效率以应用于实际场景。其次，本文集中探讨了句子级 IE 任务，其中输入通常由单个句子或几句话构成。相比之下，文档级 IE 涉及的输入范围更广，可能包括整个段落或完整文档。最后，在文档级 IE 中，IE 元素的识别需要跨越句子乃至跨越段落，这显著增加了 IE 的复杂度。鉴于现实世界中许多文本数据以文档形式存在，深入研究文档级 IE 不仅具有理论意义，也对实际应用场景极具价值。

参 考 文 献

[1] PISKORSKI J, YANGARBER R. Information extraction: Past, present and future[M]//POIBEAU T, SAGGION H, PISKORSKI J, et al. Multi-source, Multilingual Information Extraction and Summarization. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012: 23-49.

[2] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks[J]. Advances in Neural Information Processing Systems, 2014, 4: 3104-3112.

[3] LU Y J, LIU Q, DAI D, et al. Unified structure generation for universal information extraction[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2022: 5755-5772.

[4] HUANG Z H, XU W, YU K. Bidirectional LSTM-CRF models for sequence tagging[EB/OL]. (2015-08-09) [2024-05-01]. https://arxiv.org/abs/1508.01991.

[5] NGUYEN T H. Deep learning for information extraction[EB/OL]. [2024-05-01]. https://dblp.uni-trier.de/pid/17/19407.html.

[6] VINYALS O, FORTUNATO M, JAITLY N. Pointer networks[J]. Advances in Neural Information Processing Systems, 2015, 28: 2692-2700.

[7] ZHANG Z Y, HAN X, LIU Z Y, et al. ERNIE: Enhanced language representation with informative entities[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2019: 1441-1451.

[8] RADFORD A, WU J, CHILD R, et al. Language models are unsupervised multitask learners[J]. OpenAI Blog, 2019, 1(8): 9.

[9] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[EB/OL]. (2018-10-11) [2024-05-01]. https://arxiv.org/abs/1810.04805.

[10] COLIN R, NOAM S, ADAM R, et al. Exploring the limits of transfer learning with a unified text-to-text transformer[J]. Journal of Machine Learning Research, 2020, 21(140): 1-67.

[11] LOU J, LU Y J, DAI D, et al. Universal information extraction as unified semantic matching[EB/OL]. [2024-05-01]. https://arxiv.org/pdf/2301.03282.

[12] WANG X, ZHOU W K, ZU C, et al. InstructUIE: Multi-task instruction tuning for unified information extraction[EB/OL]. (2023-04-17) [2024-05-01]. https://arxiv.org/abs/2304.08085.

[13] YANG P, LU J Y, GAN R Y, et al. UniEX: An effective and efficient framework for unified information extraction via a span-extractive perspective[EB/OL]. (2023-05-17) [2024-05-01]. https://arxiv.org/abs/2305.10306.

[14] QIU X P, SUN T X, XU Y G, et al. Pre-trained models for natural language processing: A survey[J]. Science China Technological Sciences, 2020, 63(10): 1872-1897.

[15] LIU P F, YUAN W Z, FU J L, et al. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing[J]. ACM Computing Surveys, 2023, 55(9): 1-35.

[16] DING N, CHEN Y L, HAN X, et al. Prompt-learning for

fine-grained entity typing[EB/OL]. (2021-08-24) [2024-05-01]. https://arxiv.org/abs/2108.10604.

[17] CHEN X, LI L, DENG S M, et al. LightNER: A lightweight tuning paradigm for low-resource NER via pluggable prompting[EB/OL]. (2021-08-31) [2024-05-01]. https://arxiv.org/abs/2109.00720.

[18] ZHANG N Y, LI L Q, CHEN X, et al. Differentiable prompt makes pre-trained language models better few-shot learners[EB/OL]. (2021-08-30) [2024-05-01]. https://arxiv.org/abs/2108.13161.

[19] CHIA Y K, BING L D, PORIA S, et al. RelationPrompt: Leveraging prompts to generate synthetic data for zero-shot relation triplet extraction[EB/OL]. (2022-03-17) [2024-05-01]. https://arxiv.org/abs/2203.09101.

[20] HSU C, ZAN C T, DING L, et al. Prompt-learning for cross-lingual relation extraction[C]//Proceedings of the International Joint Conference on Neural Networks. New York: IEEE, 2023: 1-9.

[21] YE H B, ZHANG N Y, BI Z, et al. Learning to ask for data-efficient event argument extraction[EB/OL]. (2021-10-01) [2024-05-01]. https://arxiv.org/abs/2110.00479.

[22] SONG H T, ZHU Q M, YU Z P, et al. Generative event extraction via Internal knowledge-enhanced prompt learning[M]//ILIADIS L, PAPALEONIDAS A, ANGELOV P, et al. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2023: 90-102.

[23] MA X Z, HOVY E. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF[EB/OL]. (2016-03-04) [2024-05-01]. https://arxiv.org/abs/1603.01354.

[24] DAI Z J, WANG X T, NI P, et al. Named entity recognition using BERT BiLSTM CRF for Chinese electronic health records[C]//Proceedings of the 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics. New York: IEEE, 2019: 1-5.

[25] SOUZA F, NOGUEIRA R, LOTUFO R. Portuguese named entity recognition using BERT-CRF[EB/OL]. (2019-09-23) [2024-05-01]. https://arxiv.org/abs/1909.10649.

[26] YIN X Y, GOUDRIAAN J, LANTINGA E A, et al. A flexible sigmoid function of determinate growth[J]. Annals of Botany, 2003, 91(3): 361-371.

[27] WEI Z P, SU J L, WANG Y, et al. A novel cascade binary tagging framework for relational triple extraction[EB/OL]. (2019-09-07) [2024-05-01]. https://arxiv.org/abs/1909.03227.

[28] HUANG K H, HSU I H, NATARAJAN P, et al. Multilingual generative language models for zero-shot cross-lingual event argument extraction[EB/OL]. (2022-03-15) [2024-05-01]. https://arxiv.org/abs/2203.08308.

[29] 贺瑞芳, 段绍杨. 基于多任务学习的中文事件抽取联合模型[J]. 软件学报, 2019, 30(4): 1015-1030.
HE R F, DUAN S Y. Joint Chinese event extraction based multi-task learning[J]. Journal of Software, 2019, 30(4): 1015-1030.

[30] 王炳乾, 宿绍勋, 梁天新. 基于 BERT 的多层标签指针网络事件抽取模型[J]. 中文信息学报, 2021, 35(7): 81-88.
WANG B Q, SU S X, LIANG T X. BERT based multi-layer label pointer network for event extraction[J]. Journal of Chinese Information Processing, 2021, 35(7): 81-88.

编　辑　张　莉