基于能量理论的通用信息抽取框架

李祖超[1)] 彭天硕[2)] 张乐飞[2)] 赵 海[3)] 杜 博[2)]

[1)](武汉大学人工智能学院 武汉 430072)

[2)](武汉大学计算机学院 武汉 430072)

[3)](上海交通大学计算机学院 上海 200240)

摘 要 信息抽取(Information Extraction,IE)旨在从非结构化的自然文本中提取实体、关系配对、事件元素、情感极性等结构化信息。主流的基于跨度的方法通过联合建模目标标签的语义信息与抽取目标的边界分布,在不同任务上表现出强大性能,但仍存在局限:(1) 预测跨度起始和结束边界的策略忽略了跨度边界之间的配对关联性;(2)传统的微调损失使模型过度依赖目标跨度的精确边界;(3) 传统的 Transformer 结构注重全局表示,与 IE 任务中总是具有有限长度目标的情况不匹配。为解决上述问题,我们提出了基于能量的通用信息抽取(Energy Based Universal Information Extraction,EBUIE)框架。通过引入基于能量的模型的思想,我们设计了一种新颖的基于能量的配对专家(Energy-based Pairwise Expert,EPE),使用标量能量来量化跨度边界之间的配对关联性。受自然界中能量分布和转化的启发,我们提出了能量分布平滑损失(Energy-distribution Smoothing Loss,ESL)和能量场注意力(Energy Field Attention,EFA),以减轻模型对标签边界的过度依赖,并在决策过程中自适应调整注意力分布。在四个纯文本 IE 任务和三个多模态 IE 任务上均取得了优于最先进方法的结果。在纯文本场景下:命名实体识别、关系抽取、事件触发词抽取、情感观点三元组抽取上分别取得 0.69%、3.15%、1.6%、7.71%的最优提升;在多模态场景下:术语抽取、情感分类、联合抽取分别取得 1.8%、0.2%、2.8%的最优提升。

关键词 自然语言处理;通用模型;信息抽取;基于能量的模型

中图法分类号 TP18 **DOI 号** 10.11897/SP.J.1016.2025.02611

Universal Information Extraction Framework under Energy Based Theory

LI Zu-Chao[1)] PENG Tian-Shuo[2)] ZHANG Le-Fei[2)] ZHAO Hai[3)] DU Bo[2)]

[1)] School of Artificial Intelligence, Wuhan University, Wuhan 430072)

[2)](School of Computer Science, Wuhan University, Wuhan 430072)

[3)](Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240)

Abstract Information Extraction (IE) aims to extract structured information such as entities, relation pairs, event elements, and sentiment polarity from unstructured natural text. Mainstream span-based methods demonstrate strong performance on various tasks by jointly modeling the semantic information of target labels and the boundary distribution of extraction targets, but they still have limitations: (1) The strategy of predicting the start and end boundaries of spans ignores the pairing relationship between span boundaries. (2) Conventional fine-tuning loss causes the model to overly rely on the precise boundaries of target spans. (3) Traditional Transformer structures focus on global representations, which do not match the case in IE tasks where targets always have limited length. To address these issues, we propose the Energy Based Universal Information Extraction (EBUIE) framework. By introducing the concept of energy-based

收稿日期:2025-07-10;在线发布日期:2025-07-24。本课题得到国家自然科学基金青年科学基金项目(No. 62306216)、中央高校基本科研业务费专项资金(No. 2042025kf0026)、湖北省科技创新项目基金(No. 2024BAB043)资助。**李祖超**,博士,副教授,中国计算机学会(CCF)会员,主要研究领域为自然语言处理、人工智能。E-mail: zcli-charlie@whu.edu.cn。**彭天硕**,本科生,主要研究领域为自然语言处理,信息抽取以及多模态融合。**张乐飞**,博士,教授,主要研究领域为多模态信息处理。**赵海**,博士,教授,中国计算机学会(CCF)会员,主要研究领域为自然语言处理。**杜博(通信作者)**,博士,教授,主要研究领域为人工智能、医学图像处理。E-mail: dubo@whu.edu.cn。

models, we design a novel Energy-based Pairwise Expert (EPE) that uses scalar energy to quantify the pairing association between span boundaries. Inspired by the energy distribution and transformation in nature, we propose the Energy-distribution Smoothing Loss (ESL) and Energy Field Attention (EFA) to reduce the model's over-reliance on label boundaries and adaptively adjust the attention distribution during decision-making. We achieve results superior to state-of-the-art methods on four plain text IE tasks and three multimodal IE tasks. In plain text scenarios, we obtain optimal improvements of 0.69%, 3.15%, 1.6%, and 7.71% on named entity recognition, relation extraction, event trigger extraction, and sentiment-opinion triplet extraction, respectively; in multimodal scenarios, we obtain optimal improvements of 1.8%, 0.2%, and 2.8% on term extraction, sentiment classification, and joint extraction, respectively.

Keywords natural language processing; universal model; information extraction; energy based model

1 引 言

信息抽取(Information Extraction,IE)是自然语言处理(Natural Language Processing,NLP)领域中的一个关键任务,旨在从大量的自然语言文本中自动提取结构化信息。IE 的目标是从非结构化文本中识别和提取特定类型的实体、关系、事件和情感,以实现进一步的分析和应用。

由于 IE 任务在提取目标、任务格式和输入数据等方面存在多样性,针对不同的 IE 子任务训练不同的模型被认为是非常昂贵的。因此,将 IE 任务统一到一致的格式中具有重要意义。在构建 IE 任务的通用模型的各种方法中,基于跨度的通用信息抽取(span based Universal Information Extraction, span based UIE)由于对输入序列和目标标签的联合建模,展现了强大的能力,并在各种 IE 数据集上取得了广泛的成功。然而,基于跨度的通用信息抽取仍然存在一些局限性,主要体现在以下三个方面:

首先,现有的基于跨度的模型[1]在考虑某个位置作为起始或结束边界的可能性时独立处理,忽略了跨度边界之间的配对相关性,即"跨度边界应该在语义上相关"的先验知识。能量理论在建模不同输入的配对关系上具有极强的灵活性高效性,因此我们将跨度边界的配对稳定性视作一种能量。一方面通过关注配对稳定性而不是边界存在概率来指导模型学习跨度边界之间的语义联系;另一方面,通过标量能量的量化,可以更直观地理解和调整配对关系,进而提高模型的抽取性能。

其次,由于自然语言的复杂性,人工标注中存在着注释的歧义性。如图 1 所示,不同的标注跨度都可以被认为是合理的。然而,传统的微调损失函数使得模型依赖于训练数据中给出的精确跨度边界。这可能会由于注释的歧义性而导致性能瓶颈。同时,这也导致监督信息的利用不足,因为相对于距离真实跨度边界较远的位置,应该将靠近真实跨度边界的位置分配相对较高的准确度,但传统的微调损失函数将它们视为等价的负例。能量场中能量的平滑变化趋势能够很好地描述准确度分布以真实跨度边界为中心平滑衰减的状态,启发我们利用能量分布来构造监督信息缓解这种过度依赖。

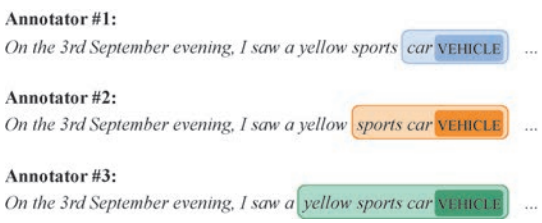

图 1 注释歧义性:以句子"On the 3rd September evening, I saw a yellow sports car drive past my house"中的"vehicle"实体为例

第三,Transformer 模型的主要目标是强调对输入文本的全局表示,而 IE 需要专注于用于确定跨度边界的特定文本段。Transformer 模型倾向于全局表示,而 IE 需要对局部文本进行分析,这种差异可能对模型的性能产生负面影响。当 Transformer 的架构与跨度表示的学习之间存在不匹配时,模型可能无法有效地利用 IE 中的先验知识。具体而言,当模型提供了标注跨度的起始边界(或结束边界)时,相应的结束边界(或起始边界)更有可能位于其之前或之后的特定范围内,而不是涵盖整个序列。

这个假设,即跨度具有有限的长度,构成了一种先验假设,在传统的 UIE 模型中被忽略了。更进一步地,我们将注意力分数也视作一种能量,并尝试利用能量分布平滑衰减及动态变化的性质实现注意力范围的动态调整,从而构造更适合信息抽取任务的注意力分布。

基于上述局限性,我们引入了能量模型的概念,以统一基于跨度的 IE 任务,并提出了能量模型通用信息抽取(Energy Based Universal Information Extraction,EBUIE)框架。据我们所知,这是首次将能量模型应用于信息抽取任务的尝试。具体而言,EBUIE 将跨度预测视为跨度起始和结束边界之间的配对任务。它利用基于能量的配对专家(Energy-based Pairwise Expert,EPE)将潜在跨度的配对稳定性映射到一个标量能量,定量描述了跨度存在的概率。通过这种方式,能量模型能够在考虑跨度边界配对相关性的同时,量化跨度的存在概率,直观地指导模型关注边界配对关系,从而提供更准确的预测。受现实世界能量场中平滑能量分布的启发,我们引入了能量分布平滑损失(Energy-distribution Smoothing Loss,ESL),描述了配对相关性超平面上的配对能量的平滑分布。这有助于减轻模型对精确跨度注释的过度依赖。此外,我们将自注意机制中的注意力分数视为一种能量形式,并尝试模拟现实世界能量场中能量的动态转换和平滑分布。我们提出能量场注意力(Energy Field Attention,EFA)来调整注意力分布,使其能够自适应地调整和平滑边界衰减序列上的注意力范围,从而便于学习跨度感知的表示,而不是全局表示。

总的来说,本文的贡献可以概括如下:

(1) 我们引入了能量模型的概念,以统一基于跨度的 IE 任务,首次利用标量能量来量化跨度边界之间的配对相关性,直观指导模型关注边界配对关系,进而提供更准确的预测。

(2) 我们设计了能量分布平滑损失,利用能量分布平滑衰减的特性有效地建模了边界配对准确性的平滑变化,首次使用能量模型的思想改进信息抽取领域的监督学习方式,有助于减轻模型对精确跨度注释的过度依赖。

(3) 我们提出了能量场注意力,利用能量分布平滑衰减和动态转换的特性调整注意力分布,首次为信息抽取领域设计独有的注意力调整机制,使序列上的注意力范围能够自适应地调整和平滑边界衰减,从而学习适用于 IE 任务的跨度感知表示。

接下来的文章组织结构如下:第 2 节回顾相关研究工作;第 3 节介绍我们提出的方法的技术细节;第 4 节呈现实验结果和分析;最后,第 5 节总结本论文的内容。

2　相关工作

2.1　信息抽取

信息抽取(Information Extraction,IE)任务是自然语言处理(Natural Language Processing,NLP)领域的关键任务,旨在从大量的自然语言文本中自动提取结构化信息,并涵盖了一系列子任务,旨在从文本数据中提取有价值的信息。这些子任务包括命名实体识别(Named Entity Recognition,NER)、关系抽取(Relation Extraction,RE)、事件抽取(Event Extraction,EE)、观点情感三元组抽取(Aspect Sentiment Triplet Extraction,ASTE)等。每个子任务都关注信息抽取的不同方面,有助于对文本数据进行全面的分析。

对于给定文本序列 T,命名实体识别任务旨在抽取 T 中属于特定类别的名词实体;关系抽取旨在识别 T 中符合特定关系的主体和客体;事件抽取旨在从 T 中识别特定事件类型对应的触发词和事件元素;情感观点三元组抽取则要求从文本中识别出实体及其情感倾向。

为了应对信息抽取的挑战,研究人员提出了各种方法和技术。传统的基于规则的方法[2]和基于模板的方法被广泛应用,这些方法涉及手动定义规则和模板来捕捉特定的信息模式。然而,由于它们依赖于显式规则,这些方法在处理大规模和多样化的文本数据时存在局限性。

近年来,基于深度学习的方法成为主流。针对特定任务提出的信息抽取模型不断涌现:命名实体识别中:大部分现有工作将 NER 视作序列标注任务,并引入条件随机场(Conditional random field,CRF)以取得更好的表现[3-4]。近期的工作中,Yu 等[5]将 NER 任务重新构想为依赖性解析问题,通过使用双仿射模型对句子中的所有可能跨度进行评分,从而实现对嵌套和平面实体的准确识别。Yuan 等[6]提出了一种新颖的三元机制,包括三元注意力和评分,用于嵌套命名实体识别。通过有效融合内部标记、边界、标签和相关跨度等异构因素,提高了嵌套 NER 的识别性能。Shen 等[7]首次将扩散模型应用于 NER 任务的研究,它将 NER 视为边界去噪

扩散过程,从噪声跨度生成命名实体,通过迭代去噪过程逐步细化实体边界,实现了高效灵活的实体生成能力。

关系抽取中:先前的方法将 RE 视作接收主语和宾语的文本向量后进行的分类任务[8-9],在这些实践中,实体通常不是直接可用的,因此这些方法需要额外的实体识别器来形成管道。近期的工作中,Ye 等[10]提出了一种新颖的跨度表示方法,称为打包悬挂标记(Packed Levitated Marker,PL-Marker),通过策略性地打包编码器中的标记来考虑跨度(成对)之间的相互关系。

事件抽取中:现有方法包括流水线分类[11-12]、多任务联合建模[13-14]、语义结构定位[15]以及(4)文本问答[16]。Hwang[17]等通过将每个事件表示为概率盒(box representation)来保证不同事件类型之间的关系具有一致性,而无需应用显式的约束。Lu[18]等提出了一个新型端到端事件抽取方法,该方法采用序列到结构的生成范式,直接从文本中提取事件,通过设计统一的事件抽取网络、约束解码算法和课程学习算法,实现了高效的模型学习和知识注入,显著提高了事件抽取的性能和数据效率。

情感分析也受到关注,开始的工作专注于抽取单独的情感元素,比如情感方面词抽取[19]以及基于给定方面词的情感分类[20-21],直到 Peng 等[22]提出观点情感三元组抽取,作为情感分析中最经典的任务。近期的工作中,Gou 等[23]通过从不同视角生成情感元素的序列,并采用投票机制选择最合理的三元组,从而提高了观点情感三元组抽取的性能。通过元素顺序提示引导语言模型生成多样的三元组,该方法有效解决了固定顺序生成中存在的问题,如不完整性、不稳定性以及错误累积。

在本文中,不同于为每个 IE 子任务设计专门的模块和孤立的训练方式带来巨大的模型训练成本以及应用复杂性 ,我们引入了基于能量的模型的思想来统一所有 IE 任务,缓解了 IE 中子任务的多样性给模型训练带来的不便,并利用不同 IE 子任务之间的语义互补性来提高信息抽取性能。

2.2 通用模型

近年来,针对各种自然语言处理任务,开发通用模型架构已成为一个活跃的研究领域。其目标是构建能够适应不同数据源、标签类型、语言和任务的模型结构。目前的研究者已经提出了许多通用模型来解决这一挑战。例如,有些模型学习深度上下文化的词表征[24-25],专门处理多种语言的模型[26-28],采用通用微调进行迁移学习的方法[29],学习跨语言句法依赖结构的模型[30-31],以及为各种信息抽取任务提供统一的文本到结构框架的模型[32]。

具体地说,对于信息抽取这一下游任务,Paolini 等[33]提出了一种新的框架,将多种结构化预测语言任务(如联合实体和关系抽取、嵌套命名实体识别等)统一框架为增强自然语言之间的翻译问题,通过设计特定的增强自然语言格式来编码结构化信息,从而实现了跨任务的模型架构和超参数的共享。Lu 等[32]首次提出了通用信息抽取(Universal Information Extraction,UIE)的概念,通过精心设计的“结构化提取语言”统一编码不同的信息抽取结构,并提出“结构化模式指导机制”,自适应地生成目标结构,从而实现从不同知识源中协作学习通用信息抽取能力。Lou 等[34]提出了一种新的框架,通过引入三种统一的标记链接操作来模拟结构化和概念化的能力,将信息抽取任务解耦为共享的抽取能力,并通过预训练获得了跨不同任务和架构的迁移能力。Wang 等[35]通过在任务不可知的语料库上预训练语言模型来生成文本结构,提出了一种结构预训练方法,使模型能够在零样本(zero-shot)情况下迁移到多个结构预测任务,并在 28 个数据集上实现了最先进的性能。Zhu 等[36]将信息抽取问题重新组织为统一的多槽元组,并提出了一种非自回归图解码算法来一步提取所有跨度,支持复杂的信息抽取任务,机器阅读理解和分类任务,展示了在少样本和零样本设置下的良好兼容性和性能。

与现有的用于信息抽取的通用模型相比,本文同样采用了提示词机制以指导模型抽取不同的结构化信息,并尝试利用不同信息抽取任务间的知识互补以提升抽取性能。与此不同的是 EBUIE 没有引入额外的图结构和元素链接操作构建更复杂的抽取结构,而使用能量模型理论来整合基于跨度的信息抽取任务,构建了基于能量的通用信息抽取框架,以在不同的抽取任务上取得一致的提升。

2.3 基于能量的模型

能量模型的概念最早由 LeCun 等[37]提出。其核心思想是建立不同变量配置与标量能量之间的映射关系,能够度量它们的兼容性,从而捕捉不同变量配置之间的依赖关系。模型学习的目标是找到一个有效的能量函数,将正确的变量配置映射到较低的能量值,而将不正确的变量配置映射到较高的能量值。模型推理的目标是找到使能量最小化的变量配置。训练过程中选择的损失函数可以用来衡量能量

函数的有效性。

能量模型提供了一种推断和训练的通用框架,能够通过能量函数和损失函数的多样性选项,让研究人员设计各种类型的统计模型。此外,它避免了估计归一化常数的问题,为模型设计带来更大的灵活性。

SPEECH[38]将能量模型和结构化预测任务相结合,提出了一种新型结构化预测模型,用于事件中心结构预测任务。SPEECH 利用能量网络的概念,通过定义输入输出对的能量函数来评估兼容性,有效地捕捉事件结构组件之间的复杂依赖性。模型结合了不同级别的能量表示——词元级别、句子级别和文档级别,分别用于触发词分类、事件分类和事件关系抽取。通过最小化能量函数来预测事件类别,从而实现了事件中心结构预测任务的有效建模和预测。

受到能量模型思想的启发,我们使用标量能量来量化潜在跨度的边界配对稳定性,并设计了基于能量的配对专家(Energy based Pairwise Expert,EPE),以基于配对稳定性来预测跨度。受自然界中能量分布和转化的启发,我们提出了能量分布平滑损失(Energy-distribution Smoothing Loss,ESL)和能量场注意力(Energy Field Attention,EFA),以减轻模型对标签边界的过度依赖,并在决策过程中自适应调整注意力分布。与 SPEECH 相比,EBUIE 使用能量函数描述抽取目标内部跨度边界的配对稳定性,而不是文本向量和不同的分类类别之间的配对稳定性,并将任务范围从事件中心的结构预测扩展到了广泛的信息抽取任务。据我们所知,这是将能量模型应用于通用信息抽取任务的首次尝试。

此外,先前工作中存在与能量分布平滑损失(Energy-distribution Smoothing Loss,ESL)和能量场注意力(Energy Field Attention,EFA)思想一致的工作:Hinton 等[39]提出改变 softmax 函数的温度以构造非 one-hot 的监督标签,从而减轻模型对精确标签的过度依赖。Sukhbaatar 等[40]提出动态地调整注意力机制作用的区间以减轻计算成本。我们在章节 4.5 中作出进一步比较分析。

3　方　法

在本节中,我们将详细介绍使用能量理论统一所有信息抽取任务的 EBUIE 框架。图 2 展示了 EBUIE 的框架。首先,我们为不同的信息抽取任务构建了不同的提示模板引导模型,具体的提示模板参考表 1。对于关系抽取任务、实体抽取任务和观点情感三元组抽取任务,我们使用具有相同结构但不同参数的多个模型进行流水线工作。

表 1　用于信息抽取任务的不同提示词模板

Task	Template
NER	Extract the "{entity-type}" entity in the sentence
RE-SE RE-OE	Extract the subjects in the sentence. Extract the object of relation "{subject-name} {relation-type}" in the sentence
ASTE-AE ASTE-OE ASTE-SE	Extract the aspect in the sentence Extract the opinion of "{aspect-name}" in the sentence Sentiment of "{opinion-name} {aspect-name}" is [positive, neutral, negative]
EE-TE EE-AE	Extract the trigger of "{event-type}" event in the sentence Extract the "{argument-type}" argument of event "{event-type}" with trigger "{event-trigger}" in the sentence

RE-SE(主语抽取)是用于在关系抽取任务中提取主语组件的提示。RE-OE(宾语抽取)是用于在关系抽取任务中提取与特定主语形成给定关系的宾语组件的提示,其中主语名称通过 RE-SE 过程获得,关系类型由具体数据集给出。

ASTE-AE(方面抽取)是用于提取 ASTE 任务中方面词组件的提示。ASTE-OE(观点抽取)是用于提取 ASTE 任务中描述特定方面词的观点组件的提示,其中方面词名称通过 ASTE-AE 过程获得。ASTE-SE(情感抽取)是用于在 ASTE 任务中提取特定观点—方面对的情感极性的提示,其中观点名称由 ASTE-OE 过程获得,方面名称由 ASTE-AE 过程获得。

EE-TE(触发词抽取)是用于提取 EE 任务中触发词组件的提示,事件类型由具体数据集给出。EE-AE(参数抽取)是用于提取与给定触发词对应的给定事件类型的参数组件的提示,事件触发词和事件类型由 EE-TE 过程获得,参数类型由具体数据集给出。

模型的文本嵌入层中,提示词和文本将转换为相应的词向量,并在输入文本编码器之前进行拼接。文本编码器包含堆叠的多头自注意力层,用于全局表征的文本编码;以及顶层一个单独的能量场注意力(EFA)层,用于进行跨度决策。基于能量的配对专家(EPE)将从文本编码器的输出序列 S_O 中获取

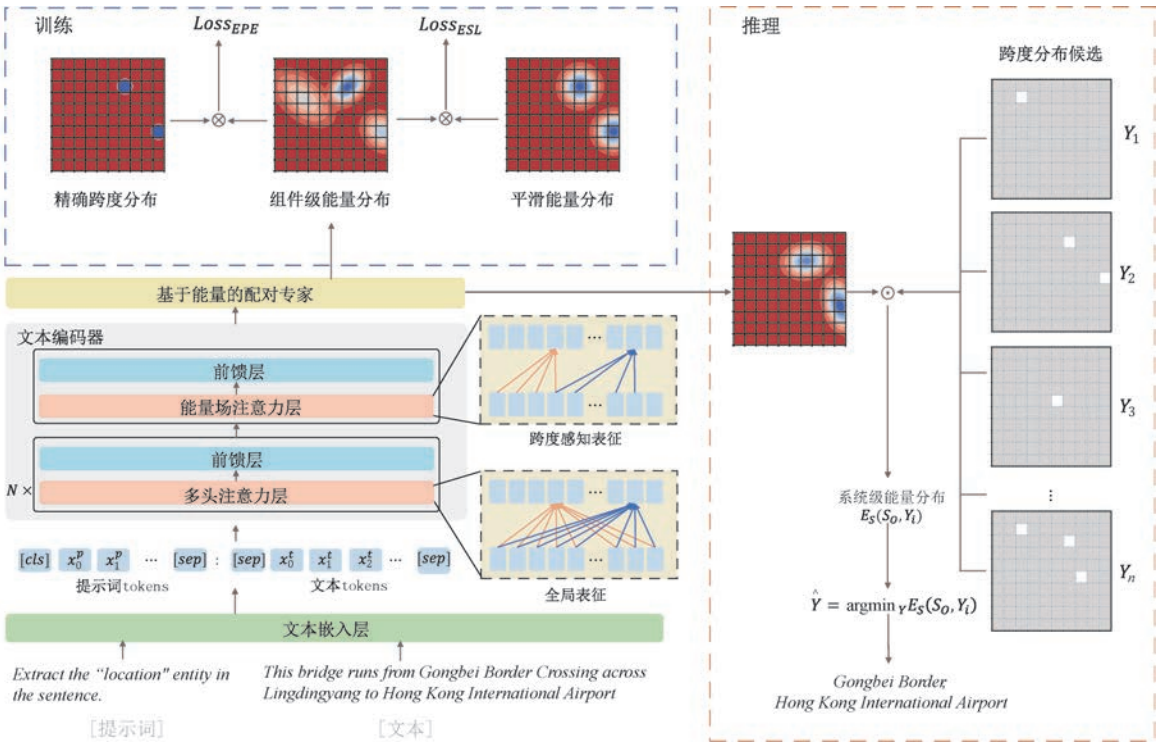

图 2 EBUIE 的模型框架总览

潜在目标跨度在序列上的配对稳定性分布，其中我们使用标量能量量化跨度边界的配对相关性，并输出表示配对相关性的超平面上的能量分布。

在训练阶段，输出的能量分布将分别与精确跨度分布和平滑能量分布一起用于计算损失。在推理阶段，我们将构建掩码来指示可能的跨度分布，并使用构建的跨度分布候选项和输出的能量分布来计算系统能量。最后，我们选择使系统能量最小化的候选项作为模型的预测结果。

3.1 基于跨度的通用信息抽取

为了展示我们提出的 EBUIE 框架的出色性能，我们 span based UIE 作为直接对比的基准。Span based UIE 将不同形式的信息抽取任务统一为基于提示的跨度预测任务。通过将预先设计的提示与输入文本拼接在一起，模型根据不同的信息抽取任务给出不同的跨度预测。

具体而言，span based UIE 包含一个文本编码器 ϕ 和两个边界投影层 Ψ_{start}，Ψ_{end}。给定提示词的词嵌入 $[x_0^p, x_1^p, x_2^p, \cdots, x_m^p]$ 和输入文本的词嵌入 $[x_0^t, x_1^t, x_2^t, \cdots, x_n^t]$，$m$ 和 n 表示提示词嵌入和文本嵌入的长度。Span based UIE 通过以下算法获得了目标跨度在序列上的潜在起始边界分布 D_{start} 和结束边界分布 D_{end}：

$$
\begin{aligned}
I_{\text{input}} &= [x_0^p, x_1^p, \cdots, x_m^p, \langle \text{sep} \rangle, x_0^t, x_1^t, \cdots, x_n^t], \\
I_{\text{output}} &= \phi(I_{\text{input}}), \\
D_{\text{start}} &= \text{Sigmoid}(\Psi_{\text{start}}(I_{\text{output}})), \\
D_{\text{end}} &= \text{Sigmoid}(\Psi_{\text{end}}(I_{\text{output}}))
\end{aligned} \tag{1}
$$

基于 D_{start}，D_{end} 和预定义的匹配规则，模型将给出最终的跨度预测结果。

3.2 多模态场景下基于跨度的通用信息抽取

为了验证 EBUIE 在多模态场景下的信息抽取性能，我们基于 span based UIE 构建了 Multi-modal UIE (MUIE)；基于 EBUIE-base 构建了 Multi-modal EBUIE-base (MEBUIE-base)。受 Li 等[41]启发，我们引入了一个双重查询器(Prompt as Dual Query，PDQ)模块作为可扩展的多模态信息集成器。该模块利用一种新颖的双查询机制，利用现有强大的单模态特征提取器的理解能力，同时根据不同的信息抽取任务差异性地关注多模态场景下中特定任务的信息。图 3 展示了 PDQ 的具体结构，每个层主要由两种交替堆叠的块组成：交叉注意力块(*Cross-Attention*)和自注意力块(*Self-Attention*)。

在每一个 PDQ 层中，每个图像都有对应的文本描述 $T_D \in \mathbb{R}^{L_D \times D}$ 和提示词 $T_P \in \mathbb{R}^{L_P \times D}$。来自图像编码器的图像特征表示为 $I \in R^{L_I \times D}$。L_D，L_P，L_I 分别表示描述、提示词和图像特征的序列长度，D 表示隐藏状态的维度。PDQ 的输入隐藏状态共三种，假定输入隐藏状态 $X \in \{T_P, T_D,$

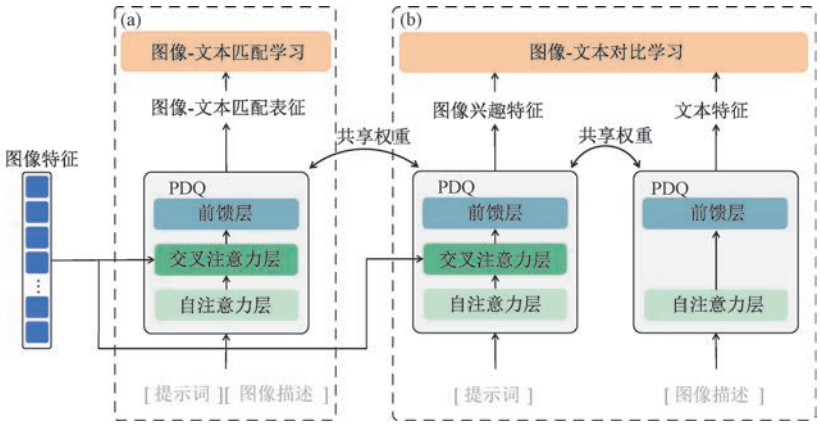

图 3　结构:双重查询器 (a)双重查询器执行图像－文本匹配学习的计算过程;(b)双重查询器执行图像－文本对比学习的计算过程

concat(T_P,T_D)}, concat 表示拼接操作。为了增强视觉信息和抽取目标之间的配对相关性,我们引入了图像－文本匹配(ITM)学习和同批次内图像－文本对比(ITC)学习,具体计算过程将在下文详述。当输入 $X \in \{T_P, T_D\}$ 时,PDQ 将执行图像－文本对比学习,两种情况下 PDQ 最后一层的输出分别为 F_{P} 和 F_{D},分别表示基于提示词抽取的视觉信息与经过编码的文本特征;当输入 $X = \mathrm{concat}(T_P, T_D)$ 时,PDQ 将执行图像－文本匹配学习,设最后一层的输出中的提示部分为 F_{P}',表示图像描述与视觉信息充分交互后的多模态特征。

具体地,我们将 *Self-Attention* 块定义如下:

$$\begin{aligned} Self\text{-}Attention(X) &= \mathrm{Softmax}\left(\frac{QK^{\top}}{\sqrt{D}}\right)V, \\ Q &= W^{Q}_{\mathrm{self}}X, \\ K &= W^{K}_{\mathrm{self}}X, \\ V &= W^{V}_{\mathrm{self}}X \end{aligned} \tag{2}$$

其中,$W^{\{Q,K,V\}}_{\mathrm{self}}$ 表示该块中的可优化参数。

对于 $X \in \{T_P, \mathrm{concat}(T_P, T_D)\}$,它将经历一个额外的 *Cross-Attention* 块。在此块中,X 首先被分割为 T_P 和 $T_{rest} \in \{T_D, \varnothing\}$。在那之后,该块中的计算定义如下:

$$\begin{aligned} &Cross\text{-}Attention(X, I) \\ &= concat\left(\mathrm{Softmax}\left(\frac{QK^{\top}}{\sqrt{D}}\right)V, T_{rest}\right), \\ &Q = W^{Q}_{\mathrm{cross}}T_P, \\ &K = W^{K}_{\mathrm{cross}}T_P, \\ &V = W^{V}_{\mathrm{cross}}T_P \end{aligned} \tag{3}$$

其中 $W^{\{Q,K,V\}}_{\mathrm{cross}}$ 表示该块中的可优化参数。

3.2.1　图像－文本对比(ITC)学习

当 PDQ 的输入为 $\{T_{\mathrm{P}}, T_{\mathrm{D}}\}$ 时,PDQ 进行同批次内 ITC 学习(见图 3 (b))。我们将 F_{P} 和 F_{D} 的第一个标记作为每个样本的整体表示。然后,我们将同批次内每个数据的整体表示组合在一起,形成 $T_{P\text{-}batch} \in \mathbb{R}^{B\times D}$ 和 $T_{D\text{-}batch} \in \mathbb{R}^{B\times D}$,其中 B 是批次大小。ITC 损失的计算方式为

$$\begin{aligned} Loss_{ITC} = -\frac{1}{2B}\Bigg(&\sum_{i=0}^{B-1}\boldsymbol{I}_i \log(S_i^{i2t}) \\ &+ \sum_{i=0}^{B-1}\boldsymbol{I}_i \log(S_i^{t2i})\Bigg), \end{aligned} \tag{4}$$

$$S_i^{i2t} = L2Norm(T_{P-batch}T_{D-batch}^{\top}),$$

$$S_i^{t2i} = L2Norm(T_{D-batch}T_{P-batch}^{\top})$$

其中,$\boldsymbol{I} \in R^{B\times B}$ 表示单位矩阵。

3.2.2　图像－文本匹配(ITM)学习

当 PDQ 的输入为 $\mathrm{concat}(T_P, T_D)$ 时,PDQ 进行 ITM 学习(见图 3 (a))。PDQ 通过线性投影 W_{ITM} 给出了预测 $Q \in R^2$,用于判断给定的图像和描述是否匹配:

$$Q = \mathrm{mean}(W_{\mathrm{ITM}}F_{\mathrm{P}}') \tag{5}$$

ITM 损失计算方式为

$$Loss_{\mathrm{ITM}} = \frac{1}{2}\sum_{i=0}^{1}P_i \log(Q_i) \tag{6}$$

其中,$P \in R^2$ 是表明给定图像和描述是否匹配的监督标签。

通过优化 $Loss_{\mathrm{ITM}}$ 和 $Loss_{\mathrm{ITC}}$,PDQ 被指导捕捉图像－文本对的配对相关性,从而为纯文本模型带来了提示词感知的视觉信息抽取能力。在推理阶段,PDQ 处理的视觉信息将被用作输入序列的前缀,以实现模型对多模态信息的整合。

3.3 基于能量的配对专家

基于跨度的抽取方法，例如 span based UIE，通过联合建模输入序列和抽取目标，充分捕捉了广泛的信息抽取任务中的知识相关性和互补性，但是基于跨度的方法忽略了配对目标的起始和结束边界之间的配对关系。因此，我们引入了能量模型（Energy Based Model，EBM）的思想，并设计了基于能量的配对专家（Energy based Pairwise Expert，EPE）帮助模型捕捉目标跨度的边界配对相关性，从而得到更准确的跨度评分。

EBM 通过将变量组合映射到标量能量来捕捉变量之间的相关性。在推理阶段，模型的目标是找到能量最小化的变量组合，而在训练阶段，模型被指导找到一个能量函数，将较低的能量值分配给配对的变量组合，而将较高的能量值分配给非配对的变量组合。

根据 EBM 的理论，EBUIE 的核心思想是将抽取目标的边界配对关系映射到标量能量，从而定量描述给定的跨度边界之间配对关系的强度，并给出更好的跨度预测分数。

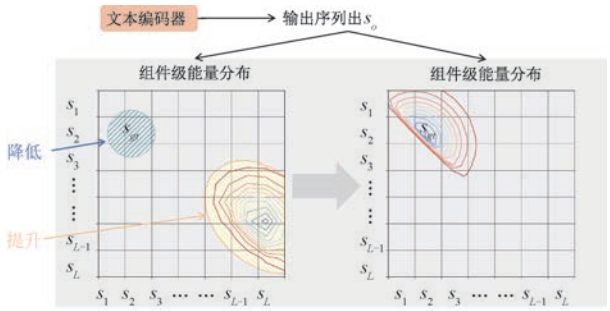

图 4 结构：基于能量的配对专家

图 4 展示了 EPE 的学习过程，其中我们降低了具有配对相关性的位置的能量，同时提高了其他位置的能量。具体而言，我们设计了组件级能量函数（Component-level energy function）E_C 和系统级能量函数（System-level energy function）E_S。设文本编码器的输出序列为 $S_O=\{s_1,s_2,\cdots,s_L\}$，$L$ 是 S_O 的长度。我们使用 E_C 表示边界为第 i 个和第 j 个标记的跨度的配对能量，记为

$$E_C(W_S,W_E,s_i,s_j)=-(s_i^{\top}W_{S_i}^{\top}(W_{E_j}s_j)) \quad (7)$$

其中，W_S 和 W_E 是待优化的参数。较小的 E_C 表示更稳定的配对关系，即 $S_O[i:j+1]$ 是目标跨度的概率更高。而所有坐标对之间的 E_C 构成了表示序列中潜在配对相关性的超平面。

对于整个序列上的预测跨度分布，我们使用 E_S 来衡量整个系统的能量。设跨度分布矩阵 $\boldsymbol{Y}\in \mathrm{R}^{\mathrm{L}\times\mathrm{L}}$ 为目标跨度的标签，其中 $Y_{ij}\in\{0,1\}$ 表示[i:j]是否为目标跨度。系统级能量可以表示为

$$E_S(S_O,\boldsymbol{Y})=\sum_{i=0}^{L-1}\sum_{j=i}^{L-1}(E_C(s_i,s_j)Y_{ij}) \quad (8)$$

在推理阶段，我们选择使系统能量最小化的跨度分布矩阵。在训练阶段，我们直接使用 E_C 来构造如下损失函数：

$$\begin{aligned}Loss_{EPE}=&-\frac{1}{\frac{L(L+1)}{2}}\sum_{i=0}^{L-1}\sum_{j=i}^{L-1}\\&+(1-Y_{ij})\log(1-x_{ij}),\\x_{ij}=&\mathrm{Sigmoid}(-E_C(s_i,s_j))\end{aligned} \quad (9)$$

通过优化 Loss_{EPE}，模型将学习到能量函数 E_C，该函数将较低的能量分配给配对的系统 $(S_O,\boldsymbol{Y})$，而将较高的能量分配给未配对的系统。

3.4 能量分布平滑损失

在 EBM 的理论框架中，能量被视为衡量参数配对稳定性的标量。而在自然界中，能量是一种基本的物理量，在不同形式和系统中经历连续的转化过程。需要强调的一点是，能量转化的物理过程始终是平滑的，没有能量分布的突然增加或减少。基于 EBUIE 的核心思想和自然界能量变化的平滑性，我们提出了能量分布平滑损失（Energy-distribution Smoothing Loss，ESL）作为模型训练的一种辅助损失。如图 5 所示，传统的微调损失通常使用 one-hot 标签 q 作为监督信息，在 EBUIE 框架中，它在序列的配对相关性超平面上表示为一个尖锐的能量峰。相比之下，ESL 使用真实跨度边界的概率分布 $\hat{q}$ 来表示目标跨度，更好地反映了与真实跨度相似的边界的准确性分布。在超平面上，它表示为围绕真实跨度位置的平滑能量峰。

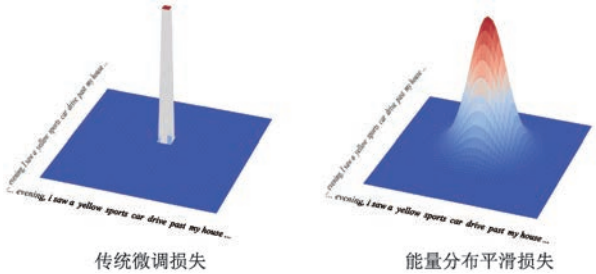

图 5 图例：能量分布平滑损失

具体而言，配对能量的平滑分布可以表示为

$$\hat{q}=\int_{S_{min}}^{S_{max}}\int_{E_{min}}^{E_{max}}Q(x,y)\mathrm{d}x\,\mathrm{d}y \quad (10)$$

其中，S_{min} 和 S_{max} 表示可能存在起始边界的范围，

$E_{\min}$ 和 $E_{\max}$ 表示可能存在结束边界的边界，$Q(x,y)$ 是超平面上坐标为 (x,y) 的位置的配对能量。在具体实现中,我们选择二维高斯分布的概率密度函数 $N(\mu_x,\mu_y,\sigma_x,\sigma_y,\rho)$ 作为 $Q(x,y)$，其具有连续和对称的特性,并且稳定、集中的分布能够很好地描述峰值周围区域能量的平滑下降。

为了将连续能量分布映射到离散序列坐标,我们设计了一个采样函数 $Q'(x,y)$ 来对超平面中的能量分布进行采样。配对能量的离散平滑分布可以表示为

$$\hat{q}=\sum_{x=0}^{L-1}\sum_{y=0}^{L-1}Q'(x,y)$$

$$Q'(x,y)=\begin{cases}\varepsilon, & \varepsilon\geqslant\theta\\ 0, & \varepsilon<\theta\end{cases} \tag{11}$$

$$\varepsilon=Q((x-x_g)S,(y-y_g)S)$$

其中，s 表示采样步长，x_g 和 y_g 表示目标跨度的起始和结束边界的坐标，θ 用作采样阈值,以过滤与相关位置无关的信息。

我们计算模型预测的配对能量分布相对于平滑分布的 KL 散度,作为一种补充损失,表示为

$$Loss_{\mathrm{ESL}}=\sum_{x=0}^{L-1}\sum_{y=0}^{L-1}Q'(x,y)(\log Q'(x,y)-\log(-E_C(s_x,s_y))) \tag{12}$$

最终的优化目标为两个损失的融合,表示为

$$Loss=Loss_{\mathrm{EPE}}+\lambda Loss_{\mathrm{EPE}} \tag{13}$$

其中，λ 是 ESL 的系数。

3.5　能量场注意力

为了解决 Transformer 结构对全局表征的侧重与信息抽取任务对跨度表征的侧重之间的不匹配,我们将序列的注意力分布视为一种特殊类型的能量场,并提出了能量场注意力(Energy Field Attention,EFA)。受到自然界中能量的动态转换和平滑分布的启发,EFA 实现了注意力范围的自适应变化,以及边界处注意力分数的平滑衰减。

在传统的多头注意力机制中,每个头计算序列中位置为 t 的标记与序列中其他标记之间的相似性矩阵。标记 t 和标记 r 之间的相似性可以

$$s_{tr}=y_t^TW_q^T(W_ky_r+p_{t-r}) \tag{14}$$

公式中的权重矩阵 W_k 和 W_q 分别表示转换矩阵“键(key)”和“查询(query)”。标记 t 和 r 的隐藏状态分别表示为 y_t 和 y_r，而 p_{t-r} 表示相对位置嵌入。通过应用 softmax 函数,可以计算与这些元素对应的注意力权重:

$$a_{tr}=\frac{\exp(s_{tr})}{\sum_{q=0}^{t-1}\exp(s_{tq})} \tag{15}$$

为了模拟能量场范围的动态变化和边界区域内能量的平滑衰减,我们设计了一个掩码函数 g_m 来控制注意力分数的计算,并旨在学习一个跨度感知的表示。假设可能的注意力跨度的最大长度为 L_{span}，新的注意力分数可以表示为

$$a_{tr'}=\frac{g_m(t-r)\exp(s_{tr})}{\sum_{q=t-L_{\mathrm{span}}}^{t-1}g_m(t-q)\exp(s_{tq})} \tag{16}$$

后续过程包括两个阶段:(1)确定能量场的能量衰减函数 g_a，以及(2)基于 g_a 构建用于跨度感知表示学习的掩码函数 g_m。考虑到能量场的特性,我们将 g_a 定义为一个单调递减的线性函数。在章节 4.2.1 中,我们对不同的能量衰减函数进行了消融实验,并根据实验结果最终选取线性函数作为模型设置。为了控制注意力跨度的长度,我们引入一个可学习的参数 $\delta\in[0,1]$。$g_a(x)$ 和相应的 $g_m(x)$ 的表达式如下所示:

$$g_a(z)=\frac{-z+1+d}{d},\quad 1=\delta L_{\mathrm{span}} \tag{17}$$

$$g_m(z)=\begin{cases}1, & g_a(z)>1\\ 0, & g_a(z)<0\\ g_a(z), & \text{otherwise}\end{cases} \tag{18}$$

结合公式 (17)与图 6 可见,参数 l 控制的是高能区域的长度,参数 d 控制的是注意力能量的衰减边界的宽度。模型通过优化 δ 进而控制 l 的变化,从而动态地调整注意力分布的范围。在反向传播中,模型可以学习特定任务的最佳注意力跨度分布。值得强调的是,每个头部独立地学习注意力跨度分布,使它们能够获得不同方面的最佳注意力分布。

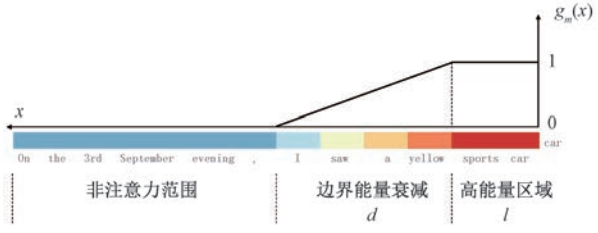

图 6　图例:注意力掩码函数 g_m

在我们的实现中,为了构建跨度感知的表示,EFA 我们在 Transformer 编码器之上引入了一个单独的 EFA 层,而不是使用多层 EFA。此外,由于 EFA 层不参与编码过程,尽管注意力的最大范围受

到 L_{span} 的限制，但其影响仅限于跨度决策，并不影响序列中标记的表征学习。

4 实 验

4.1 实验设置

4.1.1 任务定义

为了验证 EBUIE 在多模态情景下的性能，我们在五个数据集上进行了实验，涵盖了三个常见的信息抽取任务：命名实体识别（NER）、关系抽取（RE）、事件抽取（EE）和观点情感三元组抽取（ASTE）。我们实验中使用的数据集包括 ACE2004、ACE2005、ADE[42]（用于 NER 和 RE 任务）、ACE2005 和 CASIE（用于 EE 任务）以及 ASTE-Data-V2[43]（用于 ASTE 任务）。

为了验证 EBUIE 在多模态情景下的性能，我们选择了多模态基于方面的情感分析（MABSA）任务作为我们的基准。这个任务特别适合评估模型从视觉内容中提取和理解信息的能力，因为图像信息为识别实体的情感极性提供了有价值的补充线索。按照之前的工作，我们采用了 Twitter2015、Twitter2017[44]和 Political Twitter 数据集[45]来评估我们提出的 EBUIE。MABSA 是一个具有挑战性和细粒度的任务，由三个主要任务组成：给定一个图像文本对，多模态方面术语提取（MATE）任务旨在从句子中提取所有具有情感极性的方面术语，多模态基于方面的情感分类（MASC）任务旨在确定每个给定方面的情感极性，而联合多模态方面－情感分析（JMASA）任务要求模型共同提取方面－情感对。

4.1.2 度量指标

我们使用各种指标评估我们模型在这三个信息抽取任务中的性能。在 NER 任务中，我们使用实体 F1 分数，该分数根据实体的跨度和类型与参考实体的匹配情况确定实体预测的正确性。对于 RE 任务，我们采用严格关系 F1 分数，只有当关系类型和相关实体跨度都准确无误时，才认为关系是正确的。对于 EE 任务，我们报告事件触发器 F1 分数（如果事件类型和跨度都正确则为正确）和事件参数 F1 分数（如果事件类型、参数类型和跨度都正确则为正确）。至于 ASTE 任务，我们使用情感三元组 F1 分数，当方面、观点和情感极性都被准确识别时，认为三元组是正确的。

4.1.3 实现细节

我们训练了三种 EBUIE 的变体：基于 BERT-base 的 EBUIE-base，基于 BERT-base 的多模态 EBUIE-base（MEBUIE-base）以及基于 RoBERTa-large 的 EBUIE-large。此外，我们还训练了基于 BERT-base 的 span based UIE-base 和多模态 span based UIE-base（MUIE-base）作为基线模型，它们不使用 EPE、ESL 和 EFA 层。MEBUIE-base 和 MUIE-base 通过引入 PDQ 模块来获得多模态感知能力。

具体而言，EBUIE-large 具有 24 个 Transformer 层，每个层有 16 个头，模型的隐藏大小为 1024。MEBUIE-base、EBUIE-base、基于跨度的 UIE-base 和 MUIE-base 具有 12 个 Transformer 层，每个层有 12 个头，模型的隐藏大小为 768。MEBUIE-base 和 MUIE-base 使用 CLIP-ViT-bigG-14-laion2B-39B-b160k[46-47]作为冻结的图像编码器，该编码器包含 48 个 Transformer 层，每个层有 104 个头，图像编码器的隐藏大小为 1664。它们的 PDQ 模块基于 BERT-base[25]的架构和预训练参数，并且具有随机初始化的交叉注意力层。

在训练阶段的超参数选择中，我们将 ESL 中二维高斯分布 $N(\mu_x,\mu_y,\sigma_x,\sigma_y,\rho)$ 的参数设置为 $\mu_x=\mu_y=0,\sigma_x=\sigma_y=0.5,\rho=0$。对于 ESL 中的其他参数，我们将它们设置为 $\theta=0.5,s=0.1$，损失系数 λ 设置为 0.01。超参数 L_{span} 和 d 是通过分析训练数据中目标长度的统计信息来确定的。在我们的训练过程中，我们设置 $L_{\text{span}}=30,d=32$。我们首先将所有相关数据集聚合到一个大的预训练数据集中，并对模型进行一轮预训练。然后，我们利用预训练的参数进行后续的下游任务。对于每个实验，我们训练模型进行 50 个 epochs，并根据其在开发集上的表现选择最终模型。训练过程中使用的学习率为 1e-5。

4.2 超参数选择消融实验

为寻找合理的能量衰减函数 g_a 并证明本文提出的方法对超参数选择的鲁棒性，我们报告了不使用 EPE 的方法（EBUIE-base w/o EPE）在 ADE 数据集上 NER 任务的表现。基于实验结果，我们确定了章节 4.1.3 中使用的参数设置。

4.2.1 能量衰减函数 g_a

我们选取了多种可能的衰减函数变体进行初步研究。具体地说，我们选取了三种典型的衰减函数类型：线性衰减 g_a（公式 17）、截断式衰减 g_a'（公式 19）及非线性衰减 g_a''（公式 20）。g_a 将注意力分数随距离的衰减建模为线性下降，g_a' 直接抛弃能高能区

域 l 外的标记的信息，而 g_a'' 则使用缩放后的高斯曲线将注意力分数的衰减建模成典型的非线性变化，我们相信对这三种变体的研究能够完备地反映衰减函数的各种潜在候选的表现。实验结果参见表 2。

$$g_a'(z)=\begin{cases}1, z\leqslant l\\ 0, z>l\end{cases} \tag{19}$$

$$g_a''(z)=\begin{cases}1, & z\leqslant l\\ \exp\left(-\dfrac{(z-l)^2}{2\left(\dfrac{d}{3}\right)^2}\right), & z>l\end{cases} \tag{20}$$

与 UIE-base 相比，所有采用不同 g_a 函数的模型都获得了更好的结果，从而说明 EFA 的优越性。对比不同的 g_a 策略，EBUIE-base w/o EPE (g_a') 显示出最小的增强，这可能是因为注意力衰减的范围充分反映了真实的阅读环境，并使模型能够利用注意力范围内更丰富的信息。EBUIE-base w/o EPE (g_a)实现了最佳性能，这表明注意力在注意力跨度的边界处不应该衰减得太快，正如 g_a'' 的结果所证明的那样。

表 2　使用不同衰减函数的模型表现

	P	*R*	*F*1
UIE-base	87.85	91.56	89.67
EBUIE-base w/o EPE (g_a')	89.84	90.69	90.26
EBUIE-base w/o EPE (g_a'')	90.93	90.41	90.67
EBUIE-base w/o EPE (g_a)	91.17	92.17	92.49

4.2.2　超参数 L_{span}，d

针对 EFA 中使用的超参数 L_{span} 与 d，我们对多个可能的候选值进行了一系列消融实验，并选择分布在不同区间的代表性候选值结果在表 3 与表 4 展示，其余结果和表格中的趋势保持一致。结果表明，这些超参数的选择对模型的性能没有显著影响。

表 3　使用不同超参数 L_{span} 的模型表现

L_{span}	*P*	*R*	*F*1
16	91.65	93.83	92.73
30	92.58	92.40	92.49
48	92.25	92.69	92.47

表 4　使用不同超参数 d 的模型表现

d	*P*	*R*	*F*1
16	91.45	93.45	92.44
32	92.58	92.40	92.49
48	92.32	92.50	92.41

4.3　结果分析：纯文本场景

4.3.1　命名实体识别

命名实体识别(NER)任务的结果如表 5 所示。通过比较我们的基线模型 UIE-base 与其他方法的性能，可以明显看出 UIE-base 在与使用相同 BERT-base 架构的其他方法相比时取得了可比较的结果。通过引入 EPE、ESL 和 EFA，EBUIE 在两个广泛使用的基准数据集 ACE04 和 ACE05 上展现出了出色的性能，胜过使用更大的 PLM(如 ALBERT-xxlarge 和 T5-v1.1-large)的方法，并以更小的模型规模取得了更大的优势，实现了同规模模型中新的最先进结果，这充分证明了 EBUIE 框架的优越性。

与不考虑边界对之间关系的 FSUIE-large 和 EBUIE-large w/o EPE 相比，EBUIE-large 引入了 EPE 来对跨度识别任务进行 EBM 建模，并在更复杂的 ACE04 和 ACE05 数据集上显著提高了性能(分别提高了+3.25 和+2.35)，这反映了使用 EPE 来捕捉跨度边界之间的配对关联性的有效性。

我们观察到 EBUIE 在 ADE 数据集上表现相对竞争力较强，这可能是因为 ADE 数据集的实体类型较少(仅两个)，并且在每个句子中都存在，即该数据集中没有实体类型的负样本，因此无法利用目标跨度边界之间的配对关联的方法更容易适应这个简单的 NER 任务。然而，当处理像 ACE04 和 ACE05 这样具有多个类别(七个)和负样本的困难情况时，这些方法显示出明显的劣势，这反过来证明了 EPE 可以帮助模型捕捉不同目标跨度之间的配对关联性，从而实现更好的抽取性能。

与采用生成式模型进行 UIE 任务的方法相比，EBUIE 在结构预测方面更具优势。由于不需要生成复杂的 IE 结构，我们的 EBUIE-large 能够以远远少于使用大型生成模型(如 Flan-T5，GLM 和 Code-LLaMA-13B)的方法的参数实现更好的性能。其中 GoLLIE[60] 通过精心设计的注释指引将复杂的信息抽取任务转换成基于代码的插槽填空任务，以适应强大的 LLM 的预训练任务形式。然而使用 13B 参数量的模型作为主干只在 ACE05 上带来 0.14 的优势，这再次证明 EBUIE 由于对复杂生成结构的无依赖性，能够以更小的参数量实现具有竞争力甚至更优的抽取表现。同时，GoLLIE 的结果还表明：复杂的下游任务中 LLM 并不总能展现预期的优秀结果，更适合具体任务的模型结构和微调方法仍然等待研究者们探索。

4.3.2　关系抽取

在表 6 中，我们展示了关系抽取(RE)任务的结果。与不关注目标跨度边界之间配对关联性的

FSUIE-large 和 EBUIE-large w/o EPE 相比，EBUIE-large 在相似的模型规模下取得了更高的性能，并达到了新的最先进结果。这证明了我们提出的 EBUIE 框架从 EBM 的角度对跨度识别进行建模的策略在不同的 IE 任务上具有一致的有效性。与利用 NER 和 RE 任务之间的知识互补性来提高模型性能的表－序列编码器方法[61]相比，EBUIE 通过提示机制和 EPE 模块统一了所有 IE 任务，从而更高效地利用了 IE 任务中的知识互补性，从而达到更好的性能。与使用流水线策略处理关系抽取（例如 PL Marker）的模型相比，我们的 EBUIE 通过采用 EBM 理论来统一 IE 任务，直接提高了流水线的两个阶段的性能，从而整体改善了模型的性能，验证了 EBUIE 对不同形式的 IE 任务的普遍性性能提升。

表 5 命名实体识别实验结果（ACE04、ACE05、ADE 数据集）

Models	PLM	ACE04			ACE05			ADE		
		P	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1
BENSC[48]	BERT-base	85.80	84.80	85.30	83.80	83.90	83.90	—	—	—
TablERT[49]	BERT-large	—	—	—	—	—	88.60	—	—	—
SpERT[50]	BERT-base	—	—	—	—	—	—	88.69	89.20	88.95
SpERT-PL[51]	Bio-BERT	—	—	—	—	—	—	90.05	91.69	90.86
BoningKnife[52]	BERT-base	85.98	86.86	86.41	84.77	86.16	85.46	—	—	—
JCBIE[53]	Bio-BERT	—	—	—	—	—	—	—	—	87.80
GLOBAL POINTER[54]	BERT-base	—	—	—	—	—	—	—	—	90.10
BS[55]	RoBERTa-base	88.43	87.53	87.98	86.25	88.07	87.15	—	—	—
Triaffine[6]	BERT-large	87.13	87.68	87.40	86.70	86.94	86.82	—	—	—
Triaffine[6]	ALBERT-xxlarge	88.88	88.24	88.56	87.39	90.31	88.83	—	—	—
DiffusionNER[7]	BERT-large	88.11	88.66	88.39	86.15	87.72	86.93	—	—	—
PromptNER[56]	RoBERTa-large	88.64	88.79	88.72	88.15	88.38	88.26	—	—	—
Hashinge[57]	BERT-large	—	—	87.93	—	—	85.90	—	—	—
UniGEN[58]	BERT-large	87.27	86.41	86.84	83.16	83.38	84.74	—	—	—
Generative UIE (SEL only)[32]	T5-v1.1-large	—	—	86.52	—	—	85.52	—	—	—
Generative UIE[32]	T5-v1.1-large	—	—	86.89	—	—	85.78	—	—	—
DeepStruct[35]	GLM	—	—	—	—	—	86.90	—	—	—
TANL[33]	T5-base	—	—	—	—	—	84.90	—	—	—
InstructUIE[59]	Flan-T5	—	—	—	—	—	86.66	—	—	—
USM[34]	RoBERTa-large	—	—	87.62	—	—	87.14	—	—	—
Mirror[36]	DeBERTa-v3-large	—	—	87.16	—	—	85.34	—	—	—
GoLLIE[60]	Code-LLAMA-13B	—	—	—	—	—	**89.40**	—	—	—
UIE-base	BERT-base	88.25	80.31	84.09	86.19	83.12	84.63	87.85	91.56	89.67
FSUIE-base[1]	BERT-base	85.67	84.82	85.24	87.05	85.40	86.22	91.17	92.17	92.49
FSUIE-large[1]	BERT-large	88.15	86.17	86.16	88.06	85.79	86.91	93.82	92.21	**93.08**
EBUIE-large w/o EPE	RoBERTa-large	89.19	88.85	89.02	88.86	89.10	88.98	89.39	93.58	91.44
EBUIE-large	RoBERTa-large	89.47	89.34	**89.41**	89.93	88.59	89.26	92.36	91.48	91.92

与生成式方法相比，我们基于跨度的 EBUIE 模型与 IE 任务的固有结构一致，并且不需要额外的序列生成机制。因此，我们的方法在减少参数和简化架构设计的同时实现了优越的结果。即使使用 13B 规模的 LLM 作为主干模型，现有的人工定义的序列结构和训练数据依旧无法充分将 LLM 强大的推理能力迁移到信息抽取领域，因此甚至表现出远逊于判别式方法（－7.34）的抽取结果。

此外，我们的方法在 ADE 数据集上的表现优于使用 Bio-BERT 作为预训练语言模型的方法。Bio-BERT 是在生物医学语料库上预训练的领域特定的骨干模型。这个结果表明，通过将能量基模型理论应用于统一 IE 任务，EBUIE 能够从数据中提取普遍信息，增强其信息抽取能力，而不仅仅是对数据进行拟合。

4.3.3 事件抽取

表 7 展示了我们方法在事件抽取（EE）任务上的实验结果，其中“X-Tgg”表示事件触发器 F1 得分，“X-Arg”表示事件参数 *F*1 得分。总体而言，EBUIE 在相似规模的模型中取得了最佳性能，并且与使用更大模型（如 Flan-T5，Code-LLaMA-13B）作为骨干的方法相比仍然具有竞争力。与 EBUIE-large w/o EPE 相比，EBUIE-large 通过引入 EPE 模块来捕捉目标边界的配对关联性，从而在各种抽取目标上持续改进性能。

与使用生成式模型进行 UIE 任务的方法相比，

表 6　关系抽取实验结果(ACE04、ACE05、ADE 数据集)

Models	PLM	ACE04			ACE05			ADE		
		P	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1
SpERT[50]	BERT-base	—	—	—	—	—	—	77.77	79.96	78.84
SpERT-PL[51]	Bio-BERT	—	—	—	—	—	—	80.11	84.18	82.03
JCBIE[53]	Bio-BERT	—	—	—	—	—	—	—	—	74.18
TablERT[49]	BERT-base	—	—	—	—	—	66.10	—	—	—
TablERT[49]	BERT-large	—	—	—	—	—	68.10	—	—	—
Tabel-Sequence Encoder[61]	ALBERT-xxlarge	—	—	63.30	—	—	67.60	—	—	80.10
PL-Marker[10]	BERT-base	—	—	66.70	—	—	69.00	—	—	—
PL-Marker[10]	ALBERT-xxlarge	—	—	69.70	—	—	73.00	—	—	—
Generative UIE (SEL only)[32]	T5-v1.1-large	—	—	—	—	—	64.68	—	—	—
Generative UIE[32]	T5-v1.1-large	—	—	—	—	—	66.06	—	—	—
DeepStruct[35]	GLM	—	—	—	—	—	66.80	—	—	83.80
TANL[33]	T5-base	—	—	—	—	—	63.70	—	—	80.60
InstructUIE[59]	Flan-T5	—	—	—	—	—	—	—	—	82.31
USM[34]	RoBERTa-large	—	—	—	—	—	67.88	—	—	—
Mirror[36]	DeBERTa-v3-large	—	—	—	—	—	67.86	—	—	—
GoLLIE[60]	Code-LLAMA-13B	—	—	—	—	—	67.50	—	—	—
UIE-base	BERT-base	88.73	53.46	66.72	95.38	52.04	67.34	67.61	56.31	61.45
FSUIE-base[1]	BERT-base	91.78	58.79	71.82	96.79	57.69	72.29	91.10	75.87	82.79
FSUIE-large[1]	BERT-large	89.01	61.13	72.48	98.84	59.34	74.16	92.02	78.23	84.57
EBUIE-large w/o EPE	RoBERTa-large	71.77	74.21	72.97	70.16	76.20	73.05	84.55	88.01	86.24
EBUIE-large	RoBERTa-large	75.95	70.69	**73.23**	71.04	79.06	**74.84**	83.98	91.80	**87.72**

EBUIE 以更小的参数量取得了相媲美甚至更优秀的结果。并且我们注意到,使用 Code-LLaMA-13B 作为主干的 GoLLIE[60] 的表现在比较中处于劣势,这一现象说明现有的用于将信息抽取转化为结构化生成任务的预定义结构无法很好地将 LLM 的生成和推理能力迁移到信息抽取这一下游任务上,这一点在序列结构相对复杂的事件抽取上表现得更为明显。

同时需要指出,Mirror 等方法使用图解码的方式构建更复杂的结构序列,其优点在于能够一次生成所有类型的跨度。相比之下,EBUIE 需要枚举所有跨度类型进行抽取。因此元素关系复杂的抽取任务(如事件抽取)中的解码效率确实是 EBUIE 的不足之处。然而,EBUIE 由于没有构建复杂的结构序列,因此能够避免图解码过程中存在的由于关系链某一环节出错导致的连锁错误。同时我们的方法也能够避免人工设计的序列结构的复杂性对模型表现的影响。我们认为这种差异源自模型结构设计上对性能的取舍,并将把改进解码效率作为未来改进工作的重要方向。

值得强调的是,RE 和 NER 任务中使用的提示相对固定,而 EE 任务中的提示包含更多来自特定输入序列的内容,使得 EE 任务中使用的提示对于每组事件都不同。EBUIE 在 EE 任务上的出色表现验证了该框架具有强大的通用信息抽取和泛化能力,即模型有效地学习了不同类型信息的语义和内在连接,并能够推广到给定的细粒度文本数据。

表 7　事件抽取实验结果(ACE05、CASIE 数据集)

Models	PLM	ACE05-Tgg	ACE05-Arg	CASIE-Tgg	CASIE-Arg
Generative UIE (SEL only)[32]	T5-v1.1-large	72.63	54.67	68.98	60.37
Generative UIE[32]	T5-v1.1-large	73.36	54.79	69.33	61.30
DeepStruct[35]	GLM	69.80	56.20	—	—
TANL[33]	T5-base	68.40	47.60	—	—
InstructUIE[59]	Flan-T5	77.13	**72.94**	67.80	**63.53**
USM[34]	RoBERTa-large	72.41	55.83	71.73	63.26
Mirror[36]	DeBERTa-v3-large	**74.44**	55.88	71.81	61.27
GoLLIE[60]	Code-LLAMA-13B	70.90	67.80	—	—
EBUIE-large w/o EPE	RoBERTa-large	72.13	62.31	72.28	60.15
EBUIE-large	RoBERTa-large	73.73	65.19	**73.41**	61.28

4.3.4 观点情感三元组抽取

表 8 展示了我们在观点情感三元组抽取(ASTE)任务上进行实验的结果。鉴于 ASTE-Data-V2 数据集的规模有限,无需使用更大规模的模型。因此,本节仅使用 EBUIE-base 进行比较。表格显示,EBUIE-base 在四个数据集上取得了遥遥领先的表现(+4.25,+1.2,+7.71,+0.95),再次证明了我们提出的 EBUIE 框架的泛化能力和稳健性,以及它在信息抽取方面的强大性能,特别是在细粒度、强相关的情感分析等情况下。

表 8 观点情感三元组抽取实验结果(ASTEv2:14lap、14res、15res、16res 数据集)

Models	PLM	14lap			14res			15res			16res		
		P	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1
JET[43]	BERT-base	55.39	47.33	51.04	70.56	55.94	62.40	64.45	51.96	57.53	70.42	58.37	63.83
Dual-MRC[62]	BERT-base	57.39	53.88	55.58	71.55	69.14	70.32	63.78	51.87	57.21	68.60	66.24	67.40
ASTE-RL[62]	BERT-base	64.80	54.99	59.50	70.60	68.65	69.61	65.45	60.29	62.72	67.21	69.69	68.41
GAS[63]	BERT-base	—	—	60.78	—	—	72.16	—	—	62.10	—	—	70.10
Span-ASTE[64]	BERT-base	63.44	55.84	59.38	72.89	70.89	71.85	62.18	64.45	63.27	69.45	71.17	70.26
SBN[65]	BERT-base	65.68	59.88	62.65	76.36	72.43	74.34	69.93	60.41	64.82	71.59	72.57	72.08
MvP[23]	T5-base	—	—	65.30	—	—	76.30	—	—	69.44	—	—	73.10
Generative UIE (SEL only)[32]	T5-v1.1-large	—	—	63.15	—	—	73.78	—	—	66.10	—	—	73.87
Generative UIE[32]	T5-v1.1-large	—	—	63.88	—	—	74.52	—	—	67.15	—	—	75.07
USM[34]	RoBERTa-large	—	—	65.51	—	—	77.26	—	—	69.86	—	—	78.25
Mirror[36]	DeBERTa-v3-large	—	—	64.08	—	—	75.06	—	—	66.40	—	—	74.24
UIE-base	BERT-base	65.21	57.64	61.19	75.32	71.53	73.38	71.11	65.98	68.45	74.84	70.04	72.36
FSUIE-base	BERT-base	69.49	62.06	65.56	76.22	72.23	74.17	72.71	68.66	70.63	77.98	73.74	75.80
EBUIE-base	BERT-base	67.59	72.19	69.81	76.26	80.79	78.46	75.28	81.65	78.34	77.72	80.74	79.20

与同样使用 BERT-base 作为 PLM 的 FSUIE-base 相比,EBUIE 引入了 EPE 来捕捉成配对关联性,从而取得了显著的改进。与使用图解码进行通用信息抽取的 Mirror 相比,我们的方法更加注重目标元素之间的内部关联。通过使用不同的提示词,我们引导模型理解提取元素之间的关系。这减轻了图解码过程中可能出现的不完整图关系的问题。因此,我们的方法在具有多个元素的 ASTE 任务中表现出优越的性能。

与其他生成式方法相比,EBUIE 具有相对简单的架构,提高了性能而无需复杂的结构。与将 ASTE 任务分解为观点识别和情感分类子任务,并使用不同的结构处理它们的模型相比,EBUIE 使用统一且简单的结构取得了更好的性能。此外,EBUIE 的设计与 IE 任务的固有结构相一致,因此无需引入序列生成结构所需的额外参数。因此,在利用较少参数的情况下,EBUIE 优于生成式方法。

4.4 结果分析:多模态场景

4.4.1 联合多模态方面-情感分析

表 9 和表 10 显示了联合多模态方面-情感分析

表 9 联合多模态方面-情感分析结果(Twitter2015,Twitter2017 数据集)

	Methods	Twitter2015			Twitter2017		
		P	*R*	*F*1	*P*	*R*	*F*1
Text-based	SPAN[66]	53.7	53.9	53.8	59.6	61.7	60.6
	D-GCN[67]	58.3	58.8	59.4	64.2	64.1	64.1
	BART[68]	62.9	65.0	63.9	65.2	65.6	65.4
Multi-modal	UMT+TomBERT[44,69]	58.4	61.3	59.8	62.3	62.4	62.4
	OSCGA+TomBERT[44,70]	61.7	63.4	62.5	63.4	64.0	63.7
	OSCGA-collapse[70]	63.1	63.7	63.2	63.5	63.5	63.5
	RpBERT-collapse[71]	49.3	46.9	48.0	57.0	55.4	56.2
	UMT-collapse[69]	61.0	60.4	61.6	60.8	60.0	61.7
	JML[72]	65.0	63.2	64.1	66.5	65.5	66.0
	VLP-MABSA[73]	65.1	68.3	66.6	66.9	69.2	68.0
	CMMT[74]	64.6	68.7	66.5	67.6	69.4	68.5
	AoM[75]	67.9	69.3	68.6	68.4	71.0	69.7
	MUIE-base	66.2	72.2	69.1	66.1	67.5	66.8
	MEBUIE-base	71.7	72.0	**71.9**	71.1	70.2	**70.6**

表 10　联合多模态方面一情感分析结果
(Political-Twitter 数据集)

Methods	Political-Twitter		
	P	*R*	*F*1
RoBERTa[76]	63.1	62.1	62.6
UMT+collapse[69]	54.9	54.7	54.8
JML[72]	63.6	59.4	61.4
UMT-RoBERTa[69,76]	63.8	63.4	63.6
JML-PoBERTa[72,76]	63.0	60.2	61.6
CMMT[74]	65.3	65.7	65.5
MEBUIE-base	68.3	65.5	**66.9**

(JMASA)任务的结果。从结果可以看出,多模态模型总体表现高于基于文本的模型。这一现象表明对于 JMASA 任务,充分利用来自图像的视觉信息能够有效提升抽取表现。而在多模态方法的比较中,MUIE-base 在 Twitter2015 上,取得了比现有方法更好的结果,在 Twitter2017 上取得了具有竞争力的结果,这表明我们设计的 PDQ 模块有效地赋予极限模型 UIE 视觉感知能力。MEBUIE-base 通过进一步引入 EBM 的思想来统一 IE 任务,帮助模型捕捉目标跨度边界的配对关联性,从而提高了模型性能并取得了 SOTA 结果。具体地说,EBUIE 在 Twitter2015 及 Twitter2017 数据集上分别比先前的 SOTA 结果提升 3.3 及 0.9。这一观察结果证明了 EBUIE 框架在仅文本场景和多模态场景下的一致优势,验证了我们所提出的方法的强大可扩展性。

与使用合并标签策略或使用不同模型进行流水线处理的方法相比,EBUIE 统一了所有形式和类型的 IE 任务,完全消除了方面识别和情感分类之间的形式差异,并以更简洁的结构学习了两个 IE 子任务之间的语义信息交互,从而产生更好的性能。

与构造额外的掩码区域建模等视觉任务进行预训练的 VLP-MABSA 相比,EBUIE 使用统一的目标跨度抽取任务进行训练,在方法和结构上更加简洁且符合模型的推理场景,因此取得更优良的性能。

与需要使用外部工具预先抽取文本中名字作文方面词候选,进而进行复杂的文本特征融合的 AoM 相比,EBUIE 使用一套框架实现信息抽取中的多种任务,无需依赖外部工具,结构更加简洁高效。

4.4.2　多模态方面术语抽取与多模态基于方面的情感分类

表 11 和表 12 展示了多模态方面术语抽取(MATE)和多模态基于方面的情感分类(MASC)任务的结果。与 JMASA 任务的结果一致,DQPSA 在这两个子任务的性能上也取得了显著的改进或具有竞争力的表现。具体而言,MEBUIE-base 在 MATE 任务中表现优于次优方法,在 Twitter2015 提升了 1.5 个点,在 Twitter2017 提升了 2.0 个点;在 MASC 任务中,相比于次优方法,在 Twitter2015 提升了 5.2 个点。EBUIE 在 Twitter2017 数据集上没有和次优方法拉开差距,我们将其归因于 Twitter2017 文本数据包含大量无法解析和无法识别的符号,包括推特上常用的表情符号。我们的基础模型没有专门针对这一方面进行优化,这对 EBUIE 构成了相对严峻的挑战。由于 JMASA 任务由 MATE 与 MASC 两个子任务共同构成,因此 EBUIE 在 MATE 任务上的优势弥补了在 MASC 任务上的不足,使得 Twitter2017 数据集上的 JMASA 任务仍然与次优方法存在优势。

表 11　多模态方面术语抽取实验结果
(Twitter2015,Twitter2017 数据集)

Methods	Twitter2015			Twitter2017		
	P	*R*	*F*1	*P*	*R*	*F*1
RAN[77]	80.5	81.5	81.0	90.7	90.7	90.0
UMT[69]	77.8	81.7	79.7	86.7	86.8	86.7
OSCGA[70]	81.7	82.1	81.9	90.2	90.7	90.4
JML[72]	83.6	81.2	82.4	92.0	90.7	91.4
VLP-MABSA[73]	83.6	87.9	85.7	90.8	92.6	91.7
CMMT[74]	83.9	88.1	85.9	92.2	93.9	93.1
AoM[75]	84.6	87.9	86.2	91.8	92.8	92.3
MUIE-base	83.7	89.2	86.3	92.8	92.2	92.5
MEBUIE-base	88.3	87.1	**87.7**	95.1	93.5	**94.3**

表 12　多模态基于方面的情感分类实验结果
(Twitter2015,Twitter2017 数据集)

Methods	Twitter2015		Twitter2017	
	ACC	*F*1	ACC	*F*1
ESAFN[78]	73.4	67.4	67.8	64.2
TomBERT[44]	77.2	71.8	70.5	68.0
CapTrBERT[79]	78.0	73.2	72.3	70.2
JML[72]	78.7	—	72.7	—
VLP-MABSA[73]	78.6	73.8	73.8	71.8
CMMT[74]	77.9	—	73.8	—
AoM[75]	80.2	75.9	76.4	**75.0**
MUIE-base	80.9	80.9	73.1	73.1
MEBUIE-base	81.1	**81.1**	75.0	**75.0**

需要注意的是,MATE 任务等价于一种特殊类型的 NER 任务,因此我们通过采用 EBM 的思想将所有 IE 任务统一起来的方法可以帮助模型将在 NER 任务上学到的知识高效地转移到 MATE 任务上,这是通用模型相对于任务特定模型的巨大优势。

4.5　消融实验

为了深入了解 EPE、ESL 和 EFA 对提高模型

性能的影响，我们在 ACE04 数据集上对 NER 任务进行了消融实验。具体的实验结果如表 13 所示，其中 Base 表示去除了 EPE、ESL 和 EFA 的 EBUIE-large 模型。此外，"＋EFA"和"＋ESL"可以视作将相关工作 Sukhbaatar 等[40] 和 Hinton 等[39]的思想在信息抽取场景下进行实现的比较基线。

表 13 消融实验结果

Models	*P*	*R*	*F*1
EBUIE-large	89.47	89.34	89.41
＋ESL ＋EFA	89.19	88.85	89.02
＋EPE ＋EFA	88.86	89.08	88.97
＋EPE ＋ESL	87.70	89.36	88.52
＋EFA	86.62	88.32	87.46
＋ESL	87.54	88.83	88.18
＋EPE	88.49	88.17	88.33
Base	88.37	86.83	87.59

首先，四个使用/不使用 EPE 的并行对照结果显示，通过引入 EBM 的思想，EPE 改进了跨度识别，引导模型在目标跨度边界上利用配对关联性，从而在所有四种情况下实现了一致的性能提升。这表明，捕捉配对稳定性的 EPE 方法在直接预测跨度边界的传统方法上具有显著优势，对模型性能有直接贡献。

其次，在带和不带 ESL 的四个并行对照结果中，模型在添加 ESL 后也表现出一致的性能改进。作为传统微调损失的补充，ESL 通过表示精确边界周围潜在边界的准确度分布，帮助模型从训练数据中解析出更丰富的跨度分布知识，同时减轻了模型对精确跨度边界的过度依赖，从而实现更好的性能和更强的泛化能力。

然而，我们观察到仅引入 EPE 或 ESL 都能够提高基准模型的性能，但仅引入 EFA 却导致性能略微下降。我们分析认为，单独引入 EFA 使模型更专注于对序列的跨度感知表示，而不是全局表示，客观上导致了跨度外文本信息的损失，这可以解释 EFA 引起模型性能下降的原因。然而，观察结果也表明，特定跨度之外的序列信息对于 IE 任务的贡献非常有限。ESL 通过记录潜在边界周围的准确度分布，引入了更丰富的语义信息，EPE 通过描述跨度边界的配对稳定性，引入了更丰富的先验知识。而 EFA 对注意力分数的重新分配只能在具有足够丰富的可学习知识的情况下发挥作用。EFA 与 ESL 或 EPE 的结合将导致显著的性能提升，这验证了我们的分析。当这三个模块共同工作时，模型能够在广泛的语义信息范围内选择最适合 IE 任务的关键特征，从而实现最大的改进。

5 结 论

本文对单模态与多模态场景下的通用信息抽取问题进行了研究，针对信息抽取领域现有方法的局限性，使用基于能量的模型理论对信息抽取的任务多样性与目标多样性进行通用建模，提出了基于能量的通用信息抽取框架(EBUIE)。首先，EBUIE 利用能量理论在建模不同元素配对关系上的灵活性对抽取目标的边界配对关系进行量化描述，更直观地理解和学习配对关系；其次，受自然环境中能量场的平滑变化趋势启发，EBUIE 利用监督标签构造类似的平滑分布以缓解模型对标签的过度依赖；此外，EBUIE 将能量平滑衰减及动态变化的特性迁移至注意力模块，为信息抽取任务构造专注局部特征的自适应注意力模块，实现注意力分布的优化。在单模态与多模态场景下的实验结果证明，该框架在不同任务、不同领域、不同模态下具有一致的理想效果与泛化能力。

本文所提方法将不同的信息抽取任务建模为流水线工作的单元素抽取任务，对不同的信息抽取子任务形式具有较大的灵活性及适用性。与现有的使用图解码网络一次性生成抽取结构的工作相比，EBUIE 需要枚举所有的跨度类型进行抽取，将其推广到大量抽取候选类别的场景时的抽取效率需要进一步探讨。然而，EBUIE 由于没有构建复杂的结构序列，因此能够避免图解码过程中存在的由于关系链某一环节出错导致的连锁错误。同时我们的方法也能够避免人工设计的序列结构的复杂性对模型表现的影响。我们认为这种差异源自模型结构设计上对性能的取舍，并将把改进解码效率作为未来改进工作的重要方向。

参 考 文 献

[1] PENG T, LI Z, ZHANG L, et al. FSUIE: A novel fuzzy span mechanism for universal information extraction //Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Toronto, Canada, 2023: 16318-16333

[2] APPELT D E, ONYSHKEVYCH B. The common pattern specification language//TIPSTER TEXT PROGRAM PHASE III: Proceedings of a Workshop held at Baltimore, Maryland,

USA, 1998: 23-30

[3] AKBIK A, BERGMANN T, VOLLGRAF R. Pooled contextualized embeddings for named entity recognition//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis, USA, 2019,1: 724-728

[4] WANG X, JIANG Y, BACH N, et al. More embeddings, better sequence labelers? // Findings of ACL: EMNLP 2020 Findings of the Association for Computational Linguistics: EMNLP 2020, Online, 2020: 3992-4006

[5] YU J, BOHNET B, POESIO M. Named entity recognition as dependency parsing//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, 2020: 6470-6476

[6] YUAN Z, TAN C, HUANG S, et al. Fusing heterogeneous factors with triaffine mechanism for nested named entity recognition//Findings of the Association for Computational Linguistics. Dublin, Ireland, 2022: 3174-3186

[7] SHEN Y, SONG K, TAN X, et al. Diffusionner: Boundary diffusion for named entity recognition//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Toronto, Canada, 2023: 3875-3890

[8] VERGA P, STRUBELL E, MCCALLUM A. Simultaneously selfattending to all mentions for full-abstract biological relation extraction //Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. New Orleans, USA, 2018,1: 872-884

[9] WANG H, TAN M, YU M, et al. Extracting multiple-relations in one-pass with pre-trained transformers//Proceedings of the 57th Conference of the Association for Computational Linguistics. Florence, Italy, 2019,1: 1371-1377

[10] YE D, LIN Y, LI P, et al. Packed levitated marker for entity and relation extraction//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Dublin, Ireland, 2022,1: 4904-4917

[11] WANG X, WANG Z, HAN X, et al. HMEAE: hierarchical modular event argument extraction//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, 2019: 5776-5782

[12] ZHANG Z, KONG X, LIU Z, et al. A two-step approach for implicit event argument detection//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online, 2020: 7479-7485

[13] ZHANG T, JI H, SIL A. Joint entity and event extraction with generative adversarial imitation learning. Data Intelligence, 2019, 1(2):99-120

[14] ZHANG J, QIN Y, ZHANG Y, et al. Extracting entities and events as a single task using a transition-based neural model//Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. Macao, China, 2019: 5422-5428

[15] ZHANG H, WANG H, ROTH D. Unsupervised label-aware event trigger and argument classification. CoRR, 2020, abs/2012.15243. https://arxiv.org/abs/2012.15243

[16] LIU L, LIU M, LIU S, et al. Event extraction as machine reading comprehension with question-context bridging. Knowledge-Based Systems, 2024, 299:112041

[17] HWANG E, LEE J, YANG T, et al. Event-event relation extraction using probabilistic box embedding//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Dublin, Ireland, 2022: 235-244

[18] LU Y, LIN H, XU J, et al. Text2Event: Controllable sequence-tostructure generation for end-to-end event extraction//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Online, 2021: 2795-2806

[19] MA D, LI S, WU F, et al. Exploring sequence-to-sequence learning in aspect term extraction//Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, 2019,1: 3538-3547

[20] LEI Z, YANG Y, YANG M, et al. A multi-sentiment-resource enhanced attention network for sentiment classification//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia, 2018,2: 758-763

[21] LEI Z, YANG Y, YANG M, et al. A human-like semantic cognition network for aspect-level sentiment classification// The ThirtyThird AAAI Conference on Artificial Intelligence. The Thirty-First Innovative Applications of Artificial Intelligence Conference. The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, Honolulu, USA, 2019: 6650-6657

[22] PENG H, XU L, BING L, et al. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis//The Thirty-Fourth AAAI Conference on Artificial Intelligence. The Thirty-Second Innovative Applications of Artificial Intelligence Conference. The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence. New York, USA, 2020: 8600-8607

[23] GOU Z, GUO Q, YANG Y. Mvp: Multi-view prompting improves aspect sentiment tuple prediction//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Toronto, Canada, 2023: 4380-4397

[24] PETERS M E, NEUMANN M, IYYER M, et al. Deep contextualized word representations//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technolo-

gies New Orleans, USA, 2018,1: 2227-2237

[25] DEVLIN J, CHANG M W, LEE K, 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, USA, 2019,1: 4171-4186

[26] ARIVAZHAGAN N, BAPNA A, FIRAT O, et al. Massively multilingual neural machine translation in the wild: Findings and challenges. CoRR, 2019, abs/1907.05019

[27] AHARONI R, JOHNSON M, FIRAT O. Massively multilingual neural machine translation//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis, USA, 2019: 3874-3884

[28] CONNEAU A, KHANDELWAL K, GOYAL N, et al. Unsupervised cross-lingual representation learning at scale//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online, 2020: 8440-8451

[29] HOWARD J, RUDER S. Universal language model fine-tuning for text classification//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia, 2018,1: 328-339

[30] LI Z, HE S, ZHANG Z, et al. Joint learning of POS and dependencies for multilingual Universal Dependency parsing//Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies. Brussels, Belgium, 2018: 65-73

[31] SUN K, LI Z, ZHAO H. Cross-lingual universal dependency parsing only from one monolingual treebank. arXiv preprint arXiv:2012.13163, 2020

[32] LU Y, LIU Q, DAI D, et al. Unified structure generation for universal information extraction//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Dublin, Ireland,2022,1: 5755-5772

[33] PAOLINI G, ATHIWARATKUN B, KRONE J, et al. Structured prediction as translation between augmented natural languages//Proceedings of the 9th International Conference on Learning Representations. Virtual, Austria, 2021

[34] LOU J, LU Y, DAI D, et al. Universal information extraction as unified semantic matching//ThirtySeventh AAAI Conference on Artificial Intelligence. Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence. Thirteenth Symposium on Educational Advances in Artificial Intelligence. Washington, USA, 2023: 13318-13326

[35] WANG C, LIU X, CHEN Z, et al. DeepStruct: Pretraining of language models for structure prediction//Findings of the Association for Computational Linguistics. Dublin, Ireland, 2022: 803-823

[36] ZHU T, REN J, YU Z, et al. Mirror: A universal framework for various information extraction tasks//Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Singapore, 2023: 8861-8876

[37] LECUN Y, CHOPRA S, HADSELL R, et al. A tutorial on energy-based learning. Predicting structured data, 2006

[38] DENG S, MAO S, ZHANG N, et al. SPEECH: Structured prediction with energy-based event-centric hyperspheres//ROGERS A, BOYD-GRABER J L, OKAZAKI N. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1:Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023. Association for Computational Linguistics, 2023: 351-363. https://doi.org/10.18653/v1/2023.acl-long.21. DOI: 10.18653/V1/2023.ACL-LONG.21

[39] HINTON G E, VINYALS O, DEAN J. Distilling the knowledge in a neural network. CoRR, 2015, abs/1503.02531. http://arxiv.org/abs/1503.02531

[40] SUKHBAATAR S, GRAVE E, BOJANOWSKI P, et al. Adaptive attention span in transformers. CoRR, 2019, abs/1905.07799. http://arxiv.org/abs/1905.07799

[41] LI J, LI D, SAVARESE S, et al. BLIP-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. CoRR, 2023, abs/2301.12597

[42] GURULINGAPPA H, RAJPUT A M, ROBERTS A, et al. Development of a benchmark corpus to support the automatic extraction of drugrelated adverse effects from medical case reports. Journal of Biomedical Informatics, 2012, 45(5): 885-892

[43] XU L, LI H, LU W, et al. Position-aware tagging for aspect sentiment triplet extraction//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online: Association for Computational Linguistics, 2020: 2339-2349

[44] YU J, JIANG J. Adapting BERT for target-oriented multimodal sentiment classification//Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. Macao, China, 2019: 5408-5414

[45] YANG L, YU J, ZHANG C, et al. Fine-grained sentiment analysis of political tweets with entity-aware multimodal network//Lecture Notes in Computer Science: volume 12645 Diversity, Divergence, Dialogue-16th International Conference. Beijing, China, , 2021: 411-420

[46] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision//Proceedings of Machine Learning Research: volume 139 Proceedings of the 38th International Conference on Machine Learning. Virtual, 2021:8748-8763

[47] ILHARCO G, WORTSMAN M, WIGHTMAN R, et al. Openclip. Zenodo, 2021

[48] TAN C, QIU W, CHEN M, et al. Boundary enhanced neural span classification for nested named entity recognition//The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The

Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, USA, 2020: 9016-9023

[49] MA Y, HIRAOKA T, OKAZAKI N. Named entity recognition and relation extraction using enhanced table filling by contextualized representations. CoRR, 2020, abs/2010.07522

[50] EBERTS M, ULGES A. Span-based joint entity and relation extraction with transformer pre-training. CoRR, 2019, abs/1909.07755

[51] SANTOSH T, CHAKRABORTY P, DUTTA S, et al. Joint entity and relation extraction from scientific documents: Role of linguistic information and entity types. 2021

[52] JIANG H, WANG G, CHEN W, et al. Boningknife: Joint entity mention detection and typing for nested NER via prior boundary knowledge. CoRR, 2021, abs/2107.09429

[53] HE K, MAO R, GONG T, et al. Jcbie: A joint continual learning neural network for biomedical information extraction. BMC bioinformatics,2022, 23(1):1-20

[54] SU J, MURTADHA A, PAN S, et al. Global pointer: Novel efficient span-based approach for named entity recognition. CoRR, 2022, abs/2208.03054

[55] ZHU E, LI J. Boundary smoothing for named entity recognition//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Dublin, Ireland: Association for Computational Linguistics, 2022: 7096-7108

[56] SHEN Y, TAN Z, WU S, et al. Promptner: Prompt locating and typing for named entity recognition//ROGERS A, BOYD-GRABER J L, OKAZAKI N. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Toronto, Canada, 2023: 12492-12507

[57] WANG Y, UTIYAMA M. To be continuous, or to be discrete, those are bits of questions. CoRR, 2024, abs/2406.07812. https://doi.org/10.48550/arXiv.2406.07812. DOI: 10.48550/ARXIV.2406.07812

[58] YAN H, GUI T, DAI J, et al. A unified generative framework for various NER subtasks//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Online, 2021: 5808-5822

[59] WANG X, ZHOU W, ZU C, et al. Instructuie: Multi-task instruction tuning for unified information extraction. CoRR, 2023,abs/2304.08085

[60] SAINZ O, GARCÍA-FERRERO I, AGERRI R, et al. Gollie: Annotation guidelines improve zero-shot information-extraction//The Twelfth International Conference on Learning Representations. Vienna, Austria, 2024. OpenReview.net, https://openreview.net/forum?id=Y3wpuxd7u9

[61] WANG J, LU W. Two are better than one: Joint entity and relation extraction with table-sequence encoders//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online, 2020:1706-1721

[62] MAO Y, SHEN Y, YU C, et al. A joint training dual-mrc framework for aspect based sentiment analysis//Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence. Virtual, 2021:13543-13551

[63] ZHANG W, LI X, DENG Y, et al. Towards generative aspect-based sentiment analysis//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). Online, 2021: 504-510

[64] XU L, CHIA Y K, BING L. Learning span-level interactions for aspect sentiment triplet extraction//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Online, 2021: 4755-4766

[65] CHEN Y, CHEN K, SUN X, et al. A span-level bidirectional network for aspect sentiment triplet extraction//Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2022

[66] HU M, PENG Y, HUANG Z, et al. Open-domain targeted sentiment analysis via span-based extraction and classification//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy, 2019: 537-546

[67] CHEN G, TIAN Y, SONG Y. Joint aspect extraction and sentiment analysis with directional graph convolutional networks//Proceedings of the 28th International Conference on Computational Linguistics. Barcelona, Spain, 2020: 272-279

[68] YAN H, DAI J, JI T, et al. A unified generative framework for aspect-based sentiment analysis//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Online, 2021: 2416-2429

[69] YU J, JIANG J, YANG L, et al. Improving multimodal named entity recognition via entity span detection with unified multimodal transformer//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online, 2020: 3342-3352

[70] WU Z, ZHENG C, CAI Y, et al. Multimodal representation with embedded visual guiding objects for named entity recognition in social media posts//CHEN C W, CUCCHIARA R, HUA X, et al. MM '20:The 28th ACM International Conference on Multimedia. Virtual, Seattle, USA, 2020: 1038-1046

[71] SUN L, WANG J, ZHANG K, et al. Rpbert: A text-image

relation propagation-based BERT model for multimodal NER//Thirty-Fifth AAAI Conference on Artificial Intelligence, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, The Eleventh Symposium on Educational Advances in Artificial Intelligence, Virtual, 2021: 13860-13868

[72] JU X, ZHANG D, XIAO R, et al. Joint multi-modal aspect-sentiment analysis with auxiliary cross-modal relation detection//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics. 2021: 4395-4405

[73] LING Y, YU J, XIA R. Vision-language pre-training for multimodal aspect-based sentiment analysis//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Dublin, Ireland, 2022: 2149-2159

[74] YANG L, NA J, YU J. Cross-modal multitask transformer for end-to-end multimodal aspect-based sentiment analysis. Information Processing &Management, 2022, 59(5):103038

[75] ZHOU R, GUO W, LIU X, et al. AoM: Detecting aspect-oriented information for multimodal aspect-based sentiment analysis//Findings of the Association for Computational Linguistics: ACL 2023. Toronto, Canada, 2023: 8184-8196

[76] LIU Y, OTT M, GOYAL N, et al. Roberta: A robustly optimized BERT pretraining approach. CoRR, 2019, abs/1907.11692

[77] WU H, CHENG S, WANG J, et al. Multimodal aspect extraction with region-aware alignment network//Lecture Notes in Computer Science: volume 12430 Natural Language Processing and Chinese Computing-9th CCF International Conference. Zhengzhou, China, 2020: 145-156

[78] U J, JIANG J, XIA R. Entity-sensitive attention and fusion network for entity-level multimodal sentiment classification. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2020, 28:429-439

[79] KHAN Z, FU Y. Exploiting BERT for multimodal target sentiment classification through input space translation// MM '21: ACM Multimedia Conference. Virtual, China, 2021: 3034-3042

LI Zu-Chao, Ph. D., associate professor. His research interests include natural language processing and artificial intelligence.

PENG Tian-Shuo, undergraduate student, specializing in Natural Language Processing, Information Extraction, and Multi-modal Fusion.

ZHANG Le-Fei, Ph. D., professor. His research interests include multi-modal information processing.

ZHAO Hai, Ph. D., professor. His research interests include natural language processing.

DU Bo, Ph. D., professor. His research interests include artificial intelligence and medical image processing.

Background

The problem studied in this paper belongs to the field of universal information extraction in natural language processing. It involves extracting different structured information from unstructured text according to various extraction requirements. Existing span-based methods have shown strong performance by jointly modeling the input sequence and target labels. However, they overlook the pairwise relevance between span boundaries and tend to be over-reliant on precise target span boundaries. Moreover, the traditional Transformer structure, which emphasizes global representation, does not align well with the limited-length targets of information extraction (IE) tasks.

To overcome these limitations, we introduce the Energy Based Universal Information Extraction (EBUIE) framework. This framework incorporates the concept of Energy-Based Models and presents an innovative Energy-based Pairwise Expert that quantifies pairwise relevance between span boundaries using scalar energy. We also propose the Energy-distribution Smoothing Loss and Energy Field Attention mechanisms to reduce reliance on label boundaries and adaptively adjust attention distribution during decision-making.

Our experiments on various IE datasets, encompassing text-only and multi-modal scenarios, demonstrate the consistent and powerful performance of the EBUIE framework. This research contributes to the field by addressing limitations, providing an effective solution, and validating its versatility through extensive evaluations.