融合MoE的多任务学习文档级企业新闻事件抽取

郑傲泽，张坤丽*，王影，袁颂瑞，田怡豪，咎红英
郑州大学计算机与人工智能学院，河南省 郑州市，450001
zhengaoze@gs.zzu.edu.cn，ieklzhang@zzu.edu.cn，zzu_wying@163.com
379343216@qq.com，tianyihao@gs.zzu.edu.cn，iehyzan@zzu.edu.cn

摘要

企业新闻事件抽取是支撑企业动态分析与产业决策的关键技术。企业新闻事件抽取具有文本篇幅较长，内容多元化的特点，面临多事件抽取和论元分散等核心挑战。大语言模型（Large Language Model，LLM）虽然具有强大的长距离依赖建模和语义关联能力，但通用大语言模型难以满足企业级应用对专业性与资源效率的需求。本文提出了融合MoE的多任务学习企业新闻事件抽取模型（MoE-Enhanced Multi-Task Learning for Corporate News Event Extraction，MoE-ML-CNEE）。通过构建统一微调数据集与多任务联合训练范式，将事件检测与论元抽取构建为结构化语言模板，增强模型全局建模能力。设计MoELoRA模块，利用动态路由机制实现多专家网络在低秩空间的知识共享与特征解耦，进一步提升模型事件抽取性能。实验表明，MoE-ML-CNEE模型在ChiFinAnn和DuEE-fin公共数据集和自建企业新闻数据集的事件检测、事件论元抽取结果均优于现有基线模型。

关键词： 事件抽取；事件检测；事件论元抽取；混合专家模型；大语言模型

Document-Level Multi-Task Learning for Corporate News Event Extraction via Mixture-of-Experts

Aoze Zheng，Kunli Zhang*, Ying Wang, Songrui Yuan, Yihao Tian, Hongying Zan
School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou Henan, 450001
zhengaoze@gs.zzu.edu.cn，ieklzhang@zzu.edu.cn，zzu_wying@163.com
379343216@qq.com，tianyihao@gs.zzu.edu.cn，iehyzan@zzu.edu.cn

Abstract

Corporate news event extraction is a key technology supporting enterprise dynamics analysis and industrial decision-making. This task is characterized by long-text inputs and diverse content, posing core challenges such as multi-event extraction and dispersed arguments. Although Large Language Models (LLMs) demonstrate strong capabilities in long-range dependency modeling and semantic association, general-purpose LLMs often fall short in meeting the requirements for domain specificity and resource efficiency in enterprise-level applications. To address these issues, this paper proposes a Mixture-of-Experts Enhanced Multi-Task Learning framework for Corporate News Event Extraction (MoE-ML-CNEE). By constructing a unified fine-tuning dataset and adopting a multi-task joint training paradigm, we formulate event detection and argument extraction as structured language templates, thereby enhancing the model's global representation capabilities. A novel MoELoRA module is designed to leverage dynamic routing mechanisms, enabling knowledge sharing and feature disentanglement

第二十四届中国计算语言学大会论文集，第448页-第459页，济南，中国，2025年8月11日至14日。

among expert networks within a low-rank parameter space. Experimental results show that the proposed MoE-ML-CNEE model consistently outperforms existing baselines on both public datasets (ChiFinAnn and DuEE-fin) and a self-constructed corporate news dataset across tasks of event detection and argument extraction.

Keywords: Event Extraction , Event Detection , Event Argument Extraction , Mixture-of-Experts Model , Large Language Model

1 引言

企业新闻作为直观反映企业运营状况与战略动向的重要信息载体，其价值传导机制已形成“企业—产业链—行业”三级扩散效应。在企业层面，新闻事件不仅影响公司声誉，还可能对产品销售和股价造成显著影响；在产业链层面，上游供应商与下游经销商的业务运营亦会受到波及；在消费者层面，新闻信息成为消费者认知企业的重要渠道，进而影响其产品选择行为。然而，海量新闻的非结构化特征导致人工提取关键事件时存在效率低下、主观偏差等问题，难以支撑企业动态的精准追踪与产业链趋势的量化分析。因此，利用计算机技术对新闻文本中的事件信息进行结构化处理具有重要的现实意义。这种模型不仅能提高信息提取效率，还能为后续数据分析奠定基础，驱动企业决策智能化，辅助产业趋势分析和风险预警等。

事件抽取（Event Extraction，EE）作为信息抽取领域的关键任务，旨在识别特定类型的事件，并把事件中担任既定角色的要素以结构化的形式呈现出来(Li et al., 2022)。如图 1所示，事件抽取任务通常涵盖事件检测（Event Detection，ED）和事件论元抽取（Event Argument Extraction，EAE）两个核心环节。事件检测通过分析文本内容判别事件类别，事件论元抽取基于已确定的事件类型，从文档中提取该事件相关的各类参与要素。

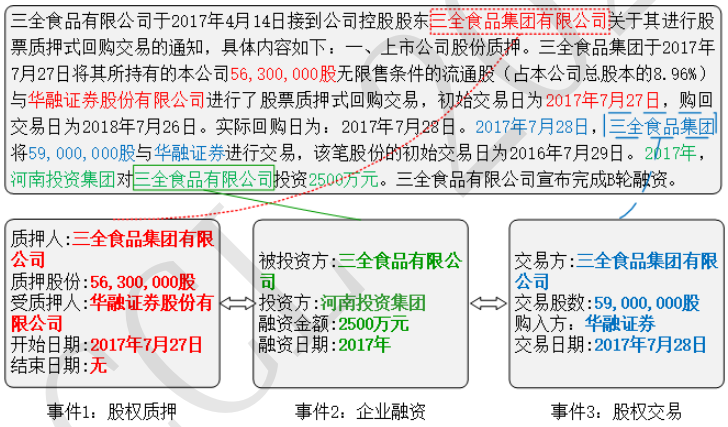

图 1: 事件抽取任务示例

事件抽取任务根据研究对象不同分为句子级事件抽取和文档级事件抽取(Hogenboom et al., 2011)，鉴于企业新闻文本普遍篇幅较长，且单篇新闻往往包含多个独立事件，本文将重点聚焦于文档级事件抽取。相较于句子级事件抽取，文档级事件抽取面临更为复杂的任务场景，其核心挑战是多事件抽取和论元分散等(Li et al., 2021)。首先，新闻文本可能包含多个事件，且不同事件的论元可能存在嵌套现象，确保事件类型的准确识别是事件论元抽取的基础；其次，由于事件的不同论元可能分布在文档的多个句子中，若仅基于单一句子进行事件抽取，往往难以获得完整的事件信息(赵庆珏et al., 2024)。此外，事件论元还存在论元重叠的现象，一个论元可能是多个事件的论元。这就要求模型具有良好的长文本建模能力，从而有效整合分散论元并解决复杂事件关联问题。

近年来，大语言模型(Wu et al., 2025)（Large Language Model，LLM）凭借其强大的长距离依赖建模和语义关联能力，为自然语言处理任务提供了新的解决思路，例如针对性地

基金项目：国家自然科学基金重点项目（U23A20316）

优化模型结构、调整训练策略以及引入领域知识增强，以提升模型在特定场景下的适用性和效果。然而，即使通用大语言模型（如GPT-4(Achiam et al., 2023)、Qwen2.5(Yang et al., 2024)、DeepSeek(Guo et al., 2025)）在信息抽取任务中被直接应用，但因事件抽取任务的复杂性，其直接使用的准确度并不理想。直接应用于企业新闻场景时面临显著挑战，千亿级参数规模导致模型训练与推理成本高昂，难以满足企业级应用对实时性与资源效率的需求，且企业新闻文本本身的专业性与复杂性、事件抽取的高精度要求都制约了大语言模型在垂直应用领域的落地。这就要求在使用大语言模型的同时要考虑模型训练和推理的成本。并且大模型幻觉可能导致其生成上下文之外的论元，直接使用大模型进行论元生成可能会导致严重的精度问题。

混合专家模型(Gormley and Frühwirth-Schnatter, 2019)（Mixture of Experts，MoE）通过整合多个专家网络来处理不同的任务或数据特征，从而提高模型的效率和性能。其核心机制是动态路由，即通过门控网络根据输入数据的特征选择性地激活部分专家进行计算。这种稀疏激活的设计显著减少了计算量和内存占用，同时保持了模型的高性能。将MoE引入大语言模型的研究中，可以缓解参数数量庞大导致的计算和存储成本高昂问题，同时提升模型在特定任务上的泛化能力。

综上所述,本文的主要贡献如下：

（1）设计MoELoRA模块，结合MoE对大语言模型进行改进和微调，通过动态门控机制协同多专家网络，在低秩参数空间实现任务间知识共享与特征解耦，提升事件抽取的多任务学习效果，实现企业新闻的联合事件抽取。

（2）针对错误传播问题和任务独立性等问题，构建统一的多任务事件抽取微调训练集，将事件检测、论元抽取等子任务标准化为语言模板，强化模型对事件结构的整体感知。将事件检测与论元抽取构建为为结构化语言模板，增强模型全局建模能力。

2 相关工作

事件抽取作为信息抽取领域的核心任务之一，其技术演进历程与自然语言处理方法的发展具有显著关联性(王人玉et al., 2023)。自20世纪80年代概念提出以来，该领域的技术路径可划分为四个主要发展阶段，基于模式匹配的事件抽取、基于机器学习的事件抽取、基于深度学习的事件抽取和基于大语言模型的事件抽取。

得益于深度学习技术的发展，事件抽取研究大多在此基础上进行探讨。在文档级事件抽取领域，Zheng等人(2019)进一步提出了Doc2EDAG模型，一种端到端的解决方案，通过构建基于实体的有向无环图来完成文档级事件抽取任务，为复杂文档的事件信息提取提供了新的技术路径。Xu等人(2021)提出了基于异构图的事件交互模型GIT，该模型通过构建句子与实体之间的异质图，结合图神经网络GNN和事件追踪器技术，有效实现了多事件检测。Wang等人(2023)进一步提出了基于图神经网络的ProCNet模型，该模型通过强化多事件间的交互机制，显著提升了复杂场景下的事件抽取性能。深度学习方法凭借其强大的非线性建模能力，能够有效捕捉特征间的复杂依赖关系，从而显著降低了传统特征工程的复杂度。

随着大语言模型的发展，其能够更好地理解复杂的自然语言文本，展现出强大的上下文能力和泛化能力。LLM已被广泛的应用于各类下游任务，如对话、摘要生成、信息抽取等。Gao等人(2023)测试了ChatGPT在ACE2005数据集上进行事件抽取的效果，结果发现ChatGPT在通用领域数据集上表现并不突出。Whitehouse C等人(2023)通过使用多种LLM（如ChatGPT和GPT-4）增强数据集，从而检测事件触发词进而预测事件类型，并通过微调LLM来提高模型事件抽取的性能。Dai H等人(2025)提出了一种基于ChatGPT的文本数据增强方法AugGPT，该方法通过将训练样本中的每个句子重述为多个概念上相似但语义上不同的样本，从而生成数据并用于下游模型训练，克服了当前文本数据增强方法在标签正确性和数据多样性上的不足问题。Chen R等人(2024)通过将LLM作为专家标注者用于事件抽取校验，纠正传统模型的错误，提高了事件抽取的准确率。鲍彤等人(2023)测试了ChatGPT在金融事件抽取数据上的效果，发现ChatGPT能够在长文本上通过上下文语义建模理解语义获得更好的事件检测效果，但是由于生成式模型输出元素不可控，存在输出边界模糊的问题。

然而，因为企业新闻文档中可能包含多个事件，其文本本身的专业性与复杂性，LLM的幻觉可能导致在上下文之外生成论元，直接使用LLM进行论元生成可能会导致严重的精度问题。因此，直接应用大语言模型来提取论元效果并不理想。综上所述，本文通过将大语言模型结合MoE的多任务学习能力和LoRA高效微调技术。在低秩参数空间实现任务间知识共享与特征

解耦，提升事件抽取的多任务学习效果，实现企业新闻的联合事件抽取。

3 MoE-ML-CNEE 方法

MoE-ML-CNEE通过构建多任务专家网络，将事件检测、论元抽取、事件抽取等任务统一整合至端到端架构，借助多专家低秩适应机制实现任务间的知识共享与参数隔离。

3.1 模型架构

本研究设计了一种基于多专家机制的低秩自适应（MoELoRA）模块，通过将其嵌入大型语言模型的稠密层中，在保持参数高效性的同时显著提升了模型的多任务学习能力。其中，任务驱动门控模块作为MoELoRA的核心组件，通过动态调节专家网络的贡献度，实现了模型复杂度的精确控制，有效避免了参数冗余问题。这种设计不仅保留了LoRA方法的参数效率优势，还通过多专家机制显著增强了模型处理多样化任务的能力。MoE-ML-CNEE的整体架构如图2所示，包含三个关键组成部分，分别是MoELoRA模块结构设计、任务驱动门控模块和微调与推理流程。

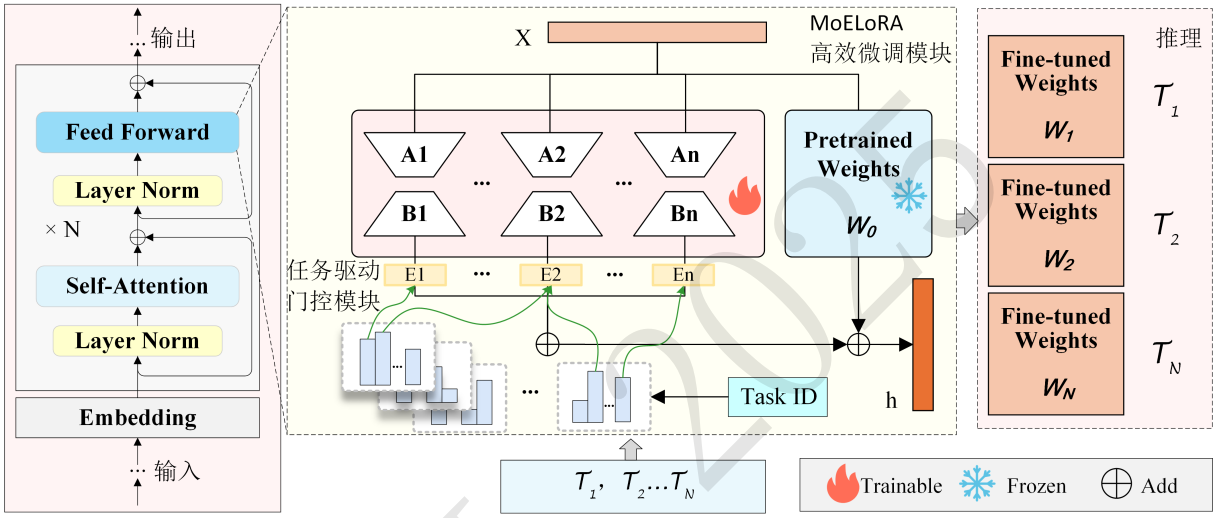

图 2: MoE-ML-CNEE架构图

（1）MoELoRA结构设计：在LLM的稠密层中嵌入MoELoRA模块,通过多专家机制，MoELoRA在保持参数高效性的同时，增强了任务多任务学习能力。

（2）任务驱动门控模块：该模块提出了一种基于语义意图的动态路由策略，通过构建可训练任务嵌入矩阵和轻量级投影网络生成任务感知向量。设计了基于任务特征的自适应门控机制，动态调节各专家在不同任务场景下的贡献度。通过全局门控函数实现细粒度知识选择，有效控制模型复杂度，避免参数冗余，为事件处理提供有效的特征表示。

（3）微调与推理流程：在微调阶段，冻结预训练大语言模型的参数，仅优化MoELoRA模块，通过多任务联合损失函数整合事件类型分类和论元跨度检测任务。训练过程中采用动态损失权重平衡机制，确保各任务均衡学习。推理阶段，MoELoRA根据任务特征生成定制化参数，实现高效的事件抽取，如从企业新闻中识别事件类型及标注论元角色。

3.2 MoELoRA 结构设计

MoELoRA是一种基于低秩适应原理的模型结构，该方法将事件抽取所需的增量参数分解为多专家协同机制，充分利用了LoRA的参数效率优势。LoRA将大语言模型参数微调过程重新表述为低秩分解，具体通过方程$\Delta W \in R^{d_{\text{in}} \times d_{\text{out}}}$来体现。其中，$W_0 \in R^{d_{\text{in}} \times d_{\text{out}}}$代表预训练大语言模型的参数矩阵，$W \in R^{d_{\text{in}} \times d_{\text{out}}}$为微调期间更新的矩阵，而$B \in R^{d_{\text{in}} \times r}$和$A \in R^{r \times d_{\text{out}}}$则是低秩且可训练的矩阵。在LoRA层与线性层结合的前向过程中，表达式如公式所示1。

$$h = W_0 x + \frac{\alpha}{r} \cdot \Delta W x = W_0 x + \frac{\alpha}{r} \cdot BAx \tag{1}$$

其中x为d_{in}维输入向量，h为d_{out}维输出向量。可训练低秩矩阵的秩r决定了可训练参数的数量，而超参数α则用于调整r的大小(Hu et al., 2022)。在LoRA微调过程中，大语言模型中的参数如W_q、$\mathbf{A}W_k$、W_v保持冻结状态，仅低秩矩阵A和B进行微调。由于$r \ll d_{\text{in}}$且$r \ll d_{\text{out}}$，矩阵A和B中的参数数量远少于原始矩阵W_0，从而实现微调过程的高参数效率。

在事件抽取任务中，触发词识别、事件检测和事件论元抽取被整合为一个多任务学习框架。传统的LoRA方法将所有任务的微调参数共享，这导致模型在特定子任务上的学习能力受限。为了克服这一挑战，通过借鉴多专家网络的理念，设计了MoELoRA模块，旨在融合LoRA的参数效率与MoE的任务特定学习能力，从而更好地适应事件抽取任务的多任务需求。

MoELoRA引入一组专家$\{E_i\}_{i=1}^{N}$来学习更新后的矩阵W，这些专家通过微调所有任务的数据，能够捕获任务间的共享知识。为了保持模型的参数紧凑性，每个专家由两个低秩矩阵构成。对于任务T_j的样本，MoELoRA层的前向传播过程如公式2所示。

$$\begin{aligned} \boldsymbol{h}_j &= W_0 x_j + \frac{\alpha}{r} \cdot \Delta W_j x_j \\ &= W_0 x_j + \frac{\alpha}{r} \cdot \sum_{i=1}^{N} \omega_{ji} \cdot E_i(x_j) \\ &= W_0 x_j + \frac{\alpha}{r} \cdot \sum_{i=1}^{N} \omega_{ji} \cdot B_i A_i x_j \end{aligned} \tag{2}$$

其中h_j和x_j分别表示任务T_j样本的输入输出，专家E_i由矩阵$B_i \in R^{d_{\text{in}} \times r_N}$和$A_i \in R^{r_N \times d_{\text{out}}}$构成专家，$N$为专家数量，每个专家的秩为$r_N$。任务特定权重$\omega_{ji}$由门函数决定，确保不同子任务学习不同参数，从而增强任务特定学习能力。

在可训练参数数量方面，传统LoRA的矩阵B和A包含$r \times (d_{\text{in}} + d_{\text{out}})$个参数，而MoELoRA的$N$个专家各含$\frac{r}{N} \times (d_{\text{in}} + d_{\text{out}})$个参数，总参数数量仍为$r \times (d_{\text{in}} + d_{\text{out}})$，与LoRA相同，这表明MoELoRA在保持参数高效性的同时，通过多专家机制显著增强了任务特定学习能力。

3.3 任务驱动门控模块

本模块设计了基于语义意图的动态路由策略。构建可训练任务嵌入矩阵$E \in R|T| \times d_T$，其中d_T为隐式任务表征维度，$|T|$涵盖事件检测、事件论元抽取等各类子任务。

当给定输入序列时，通过轻量级投影网络生成任务感知向量e_j，并采用稀疏门控函数计算专家权重，如公式3所示。

$$\omega_j = \text{Softmax}\left(\text{Top}(W_T e_j, K)\right) \tag{3}$$

其中$W_T \in \mathbb{R}^{N \times d_T}$为参数矩阵，Top-K操作保留最相关的$K$个专家，例如当$K = 2$时，会选择事件类型判别与实体关系建模专家。通过Softmax归一化确保权重具有可解释性。该机制使得模型在无需显式阶段划分的情况下，能够自主协调事件结构与论元要素的联合推理。

为提升模型的多任务适应能力，特别设计了基于任务特征的自适应门控模块。该模块通过可学习的权重分配策略，动态调节各专家在不同任务场景下的贡献度，生成任务专属的参数更新方案。如果希望恢复任务T_j的微调参数，其过程如公式4所示。

$$W_j = W_0 + \frac{\alpha}{r} \cdot \sum_{i=1}^{N} \omega_{ji} B_i A_i \tag{4}$$

通过门控模块实现了细粒度的知识选择，为后续事件处理提供了丰富且有效的特征表示。此外，本架构采用全局门控函数而非分层门控设计。这种跨层共享的调控方式具有双重优势：一方面通过参数复用有效控制模型复杂度，另一方面避免了传统多门控架构中常见的参数冗余问题。

第二十四届中国计算语言学大会论文集，第448页-第459页，济南，中国，2025年8月11日至14日。

3.4 微调与推理流程

在微调阶段，本文采用多任务混合训练策略，旨在同时训练多个相关任务，以提升模型的泛化能力和整体性能。对于每个训练批次B中的样本，执行前向传播过程。通过公式2计算大语言模型与MoELoRA模块的联合输出。MoELoRA模块通过低秩分解的专家模块$\{E_i = B_iA_i\}_{i=1}^{N}$动态更新模型参数，生成任务特定的增量参数$\Delta W$。同时，结合门控机制，根据任务特征动态调整各专家模块的贡献权重ω_j，生成任务专属的参数更新方案。

在前向传播完成后，根据公式2计算多任务联合损失函数。损失函数整合了事件类型交叉熵损失与论元跨度检测损失，如公式5所示。

$$\mathcal{L} = \lambda_1 \sum_{c \in C} y_c \log p_c + \lambda_2 \sum_{(s,e) \in S} \mathrm{BCE}(s, e) \tag{5}$$

其中为C事件类别集合，S为论元边界集合，BCE表示二元交叉熵。

在损失计算完成后，通过反向传播更新MoELoRA 模块的参数，包括低秩分解矩阵$\{A_i, B_i\}_{i=1}^{N}$和门控函数的参数$\{E, W_T\}$。采用梯度下降法优化这些参数，同时保持预训练大语言模型的参数不变。在整个训练过程中，采用多任务混合训练策略，随机采样包含不同事件类型与论元结构的样本批次，确保模型能够同时学习多个任务的特征。通过动态损失权重平衡机制，避免某些任务主导训练过程，提升模型的整体性能。

在推理阶段，MoELoRA 根据任务特征生成定制化参数，实现高效的事件抽取。对于每个任务T_j，通过公式3计算各专家模块的贡献权重ω_j。门控机制根据任务特征动态选择最相关的专家模块，生成任务专属的参数更新方案。接着，恢复MoELoRA微调后的参数，生成任务特定的模型参数。最后，对于特定任务T_j，应用相应的模型参数进行预测。

3.5 微调数据集构建

针对于事件抽取的不同子任务和不同数据集，本文通过定义一致的输入和输出模式来标准化事件抽取任务，从而简化模型设计过程。以事件检测任务为例，针对事件检测任务，传统模型通常处理为I_M表示的事件文本，以产生触发词或事件类型；针对事件论元抽取，传统模型通常也是通过表示学习直接产生事件论元。但是利用大语言模型会产生独特的范式，大语言模型的输入输出通常都是语言性的，因此需要重新定制与大语言模型兼容的事件抽取任务。

在事件抽取任务中，传统方法通常直接处理原始文本（记为I_M）以识别结构化的事件类型及关联实体角色，而大语言模型的引入需将任务重新适配为自然语言交互范式。为此，首先需对输入和输出进行重构。在输入侧，通过嵌入指令模板将原始文本转化为引导大语言模型执行事件抽取的自然语言指令，例如“请从以下文本中提取事件及其论元：[文本内容]”，其中占位符“[文本内容]”替换为实际待分析的文本I_M，此类输入模板标记为TPQ_{EE}；在输出侧，模型不再直接生成结构化标签，而是将事件类型和实体角色信息转化为符合自然语言逻辑的描述，例如“文本包含以下事件：文本包含以下事件：[事件类型：[事件类型1]，事件论元：角色：[角色1],论元：[论元1]...”，此类输出模板标记为TPA_{EE}，其通过分号、括号等符号隐式表达事件的多实例性和角色层级关系。完整流程可形式化表示为

$$I_M \rightarrow TPQ_{EE}(I_M) | LLM \rightarrow TPA_{EE}(E_{type}, E_{args}) \rightarrow E_{type}, E_{args} \tag{6}$$

其中E_{type}为事件类型，E_{args}为事件论元。微调数据集构建模板如图3所示，以事件抽取任务为例。通过此类适配，大语言模型可基于纯文本数据进行微调，生成符合语言模板的规范化输出，从而兼顾事件要素的结构化提取需求与模型的自然语言生成能力，同时保留结果的可解释性和多任务扩展性。 实验数据集是由事件检测任务（ED）、事件抽取任务（EE）、触发词识别任务(TD)、事件论元抽取任务（EAE）4个任务的14个数据集共同组成。

4 实验和分析

4.1 数据集

本文选用两个公开的中文数据集ChFinAnn(Zheng et al., 2019)、DuEE-Fin(Han et al., 2022)作为实验数据集。ChFinAnn包含5个事件类型，24种不同的论元角色。DuEE-Fin包含13个事件类型，61种不同的论元角色。

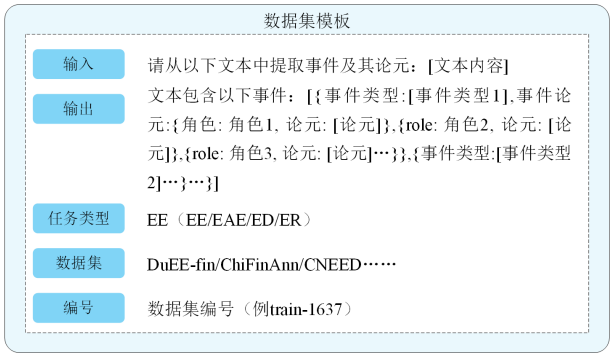

图 3: 微调数据集构建模板

由于当前缺乏公开的企业新闻事件抽取数据集，为验证模型在该领域的性能，针对食品领域构建了企业新闻事件抽取标注数据集（Corporate News Event Extraction Dataset，CNEED），构建流程如图4所示。

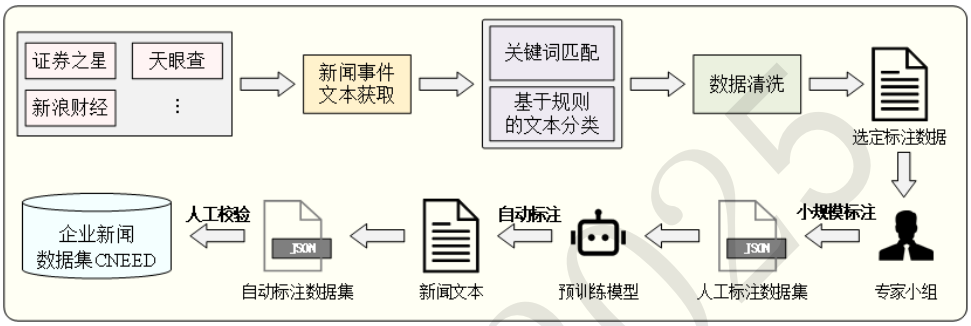

图 4: CNEED数据集构建流程图

数据来源覆盖证券之星、天眼查、新浪财经、搜狐新闻等权威财经平台。通过正则表达式过滤广告、促销及非新闻内容并去重，然后对文本进行Unicode规范化处理，统一编码格式并剔除异常字符与无效标点，以提升文本质量与事件相关性。为了提高数据的可用性，预定义了一套企业事件框架，用于指导后续的数据标注工作。在事件框架设计阶段，首先基于文献调研与实际企业新闻数据的分析，初步构建事件类别体系与相应的事件结构。随后，组织专家小组对小规模数据集进行人工标注，并基于标注结果进行迭代优化，包括调整事件框架、补充或精简事件类别、明确论元角色的定义与范围等。在正式标注过程中，利用实验室数据标注平台[60]采用半自动化标注策略实现标注。基于前期的小规模人工标注数据集，训练事件分类模型和事件论元抽取模型，并利用模型对大规模新闻文本进行自动化标注。随后，标注团队对自动标注结果进行人工校验，确保数据质量，对标注错误或边界不清晰的样本进行人工修正。

通过以上步骤，最终得到了6200条企业新闻事件抽取标注数据，可用于企业文档级事件抽取任务的研究。该数据集共包含产品发布、企业融资、企业合作等12个事件类型，以及83个论元角色。三个数据集的详情及划分如表 1所示。

表 1: 公共数据集划分数量统计

数据集	训练集	验证集	测试集	事件类型数量	论元角色数量
ChiFinAnn	25,632	3,204	3,204	5	24
DuEE-fin	6,011	1,004	1,171	13	61
CNEED	4,340	620	1,240	12	83

4.2 评价指标和实验设置

为验证本模型的有效性和可靠性，本文在论元级别计算预测事件与真实事件标签的微平均

的精确率P、召回率R和F1。本文的所有实验均在单张RTX 4090 GPU 上进行。实验参数设置如表2所示。

表 2: 实验参数设置

参数	数值
Pytorch	1.12.0
CUDA Version	12.1
Batch size	8
Gradient accumulation steps	16
Learning rate	1.0e-5
LoRA rank	16
LoRA alpha	32

4.3 对比基线模型

（1）DCFEE(Yang et al., 2018)：该模型从文档的中心句抽取论元，针对单事件和多事件抽取任务提出了两个变体DCFEE-O和DCFEE-M。

（2）Doc2EDAG(Zheng et al., 2019)：该模型利用基于实体的有向无环图实现文档级多事件抽取。

（3）GIT(Xu et al., 2021)：该模型利用异质图建模句子和实体提及之间的关系，并提出了一种跟踪机制捕获事件间的依赖。

（4）PTPCG(Zhu et al., 2021)：该模型是一个轻量级模型，构造了伪触发词的修剪完全图用于非自回归解码。

（5）ReDEE(Liang et al., 2022)：该模型介绍了一种建立在Transformer框架上的多尺度关系增强转换器

（6）ProCNet(Wang et al., 2023)：该模型结合了代理节点，使用图模型将句子和论点信息聚合到这些节点中。使用Hausdorff最小距离法进一步优化模型。

（7）CAINet(Pan et al., 2024)：该模型通过事件关系图来建模各种事件之间的关系，参数关联图来建模参数之间的相关性，以有效地聚合跨句参数。

（8）BERT(Lee and Toutanova, 2018)，该模型是由Google开发的开源中文预训练语言模型；

（9）BERT-WWM(Cui et al., 2021)，该模型采用了全词掩码策略，相较于原始BERT的单字掩码，更符合中文的语言特性，能够有效提升模型对中文语义的理解能力；

（10）Ernie(Zheng et al., 2019)，该模型通过引入知识图谱和多任务学习机制，在语义理解方面具有显著优势；

（11）FinBERT(Liu et al., 2021)，该模型能够更好地捕捉金融领域的专业术语和特定语义，在金融文本处理任务中表现出色；

（12）BERT_BiLSTM_CRF(Dai et al., 2019)，该模型采用BERT作为特征提取器，结合BiLSTM和CRF进行序列标注。

4.4 实验

4.4.1 微调数据集

如表 3所示，该表统计了微调数据集构建所采用的数据集所包含的任务类型以及实验中用到的数据量和事件类型数量。针对包含不同子任务的数据集，将其根据任务模板拆分为不同的数据集，共同用于本文的微调任务。

4.4.2 实验结果和分析

为了验证MoE-ML-CNEE模型在事件抽取任务中的性能提升，本文在公共数据集和自建企业新闻数据集上进行了对比实验。实验主要针对事件检测和事件论元抽取两个子任务展开。事件检测实验结果如表4所示。MoE-ML-CNEE 在事件检测任务中的表现优异，展示了

表 3: MoE-ML-CNEE模型微调使用数据集

数据集	任务类型	使用量	事件类型数量
ACE2005	ED/EAE/TD/EE	559	33
DuEE	ED/EAE/TD/EE	5,000	65
DuEE-fin	ED/EAE/TD/EE	10,700	13
ChiFinAnn	ED/EAE/EE	32,040	5
CCKS2021	ED/EAE/TD/EE	7,000	59
CCKS2020	ED/EE	20,000	27
MAVEN	ED/EAE/TD/EE	4,480	168
WIKIEVENTS	ED/EE	3,241	67
RAMS	ED/EE	5,000	139
DCFEE	ED/EAE/TD/EE	2,976	9
MUC4	ED/EAE/TD/EE	1,700	4

表 4: 各模型在ChiFinAnn和CNEED数据集上的事件检测对比实验结果

模型	**ChiFinAnn**	**CNEED**
	F1	F1
BERT(Lee and Toutanova, 2018)	90.1	89.6
BERT-WWM(Cui et al., 2021)	90.2	88.3
Ernie(Zheng et al., 2019)	89.9	87.2
FinBERT(Dai et al., 2019)	90.4	88.9
BERT_BiLSTM_CRF(Dai et al., 2019)	91.3	90.3
GIT(Xu et al., 2021)	94.2	93.6
CAINet(Pan et al., 2024)	95.4	95.1
GPT-4(Achiam et al., 2023)	91.4	90.3
MoE-ML-CNEE（Ours）	**98.7**	**97.6**

其在不同事件检测任务中的适应能力。由于加入了MoE模块，本文模型在多个数据集上的性能均实现了显著提升。 事件论元抽取实验结果如表5所示。根据实验结果可以看到，当前主流的文档级事件抽取模型在ChiFinAnn与DuEE-fin两个数据集上表现略有差异。相较之下，本文提出的MoE-ML-CNEE模型在两个公共数据集和自建新闻数据集上均取得最佳性能，在ChiFinAnn和DuEE-fin上分别达到88.5和81.9的F1值，显著优于现有方法。该模型通过引入多专家机制融合不同类型的语义专家知识，有效增强了模型对多类事件类型及复杂参数结构的表达能力。同时，大规模参数配置进一步提升了模型的拟合能力与泛化表现。

综合对比结果表明，MoE-ML-CNEE在文档级事件抽取任务中具有优秀的性能与强鲁棒性，验证了其设计的有效性与实用价值。

4.4.3 消融实验

为了全面验证本文所提出模型各个模块的有效性，在两个金融领域事件抽取数据集——ChiFinAnn 和DuEE-fin 上，分别在Qwen2-7B、DeepSeek-R1上进行探讨各模块的有效性的表现。LoRA 微调后的大模型（即Qwen2-7B+LoRA、DeepSeek-R1+LoRA）、以及最终提出的MoE-ML-CNEE 模型进行了系统性的对比实验。实验结果如表 6 所示。对于基础大模型，本文通过构建提示模板构建指令数据集获得结果。从实验结果来看，得益于Qwen2-7B、DeepSeek-R1优秀的信息抽取能力，在ChiFinAnn和DuEE-fin上即使未经过微调仍然可以取得75%以上的分数。

在进行LoRA微调之后，两个大模型在两个数据集上的性能均有所提升，其中在Qwen2-7B+LoRA ChiFinAnn 上的F1 值提升至86.0%，在DuEE-fin 上提升至79.0%。这一结果表明，在已有的强大语义理解能力基础上，LoRA 微调能够进一步加强模型对特定任务的适应性，提

表 5: 各模型在ChiFinAnn和DuEE-fin数据集上论元抽取的表现

模型	ChiFinAnn			DuEE-fin			CNEED		
	P	R	F1	P	R	F1	P	R	F1
DCFEE-O(Yang et al., 2018)	68.0	63.3	65.6	59.8	55.5	57.6	53.6	51.2	52.3
DCFEE-M(Yang et al., 2018)	63.0	64.6	63.8	50.2	55.5	52.7	49.6	45.2	47.3
Doc2EDAG(Zheng et al., 2019)	82.7	75.2	78.8	67.1	60.1	63.4	70.9	69.4	70.2
GIT(Xu et al., 2021)	83.6	76.9	80.1	72.1	65.9	67.8	65.4	61.3	63.3
PTPCG(Zhu et al., 2021)	83.2	74.9	78.8	71.3	61.7	66.0	71.8	68.3	70.0
ReDEE(Liang et al., 2022)	83.9	79.9	81.9	74.1	72.0	74.4	78.3	75.6	76.9
ProCNet(Wang et al., 2023)	84.4	80.9	82.7	75.3	72.6	75.5	72.9	70.4	71.6
CAINet(Pan et al., 2024)	84.3	82.9	83.6	76.5	72.9	75.8	78.6	77.2	77.9
MoE-ML-CNEE（Ours）	**89.0**	**88.0**	**88.5**	**83.0**	**80.8**	**81.9**	**79.9**	**77.5**	**78.7**

升其事件抽取的精准度。

MoE-ML-CNEE模型旨在实现不同子任务间的能力解耦与协同优化。实验结果显示，MoE-ML-CNEE使用Qwen2-7B大模型为基座模型时，性能更加优越，因此本文采用Qwen2-7B为基座模型。MoE-ML-CNEE在ChiFinAnn 上的F1 值达到88.5%，比Qwen2-7B+LoRA 进一步提升了2.5%，而在DuEE-fin 上则达到了81.9%，较微调模型提升了2.9%。该结果充分验证了引入MoE 结构对于模型专精能力提升的显著效果，以及逻辑约束机制在增强事件结构一致性方面的重要作用。

表 6: MoE-ML-CNEE模型在公共数据集上的事件论元抽取消融实验

模型	ChiFinAnn			DuEE-fin		
	P	R	F1	P	R	F1
Qwen2-7B	82.2	84.5	83.8	74.3	76.5	75.0
Qwen2-7B+LoRA	86.5	85.5	86.0	78.5	79.5	79.0
DeepSeek-R1	81.3	83.6	82.4	75.3	77.5	76.4
DeepSeek-R1+LoRA	82.6	84.8	83.7	79.8	80.2	80.0
MoE-ML-CNEE（DeepSeek）	88.5	87.4	87.9	81.5	79.4	80.4
MoE-ML-CNEE（Qwen2）	**89.0**	**88.0**	**88.5**	**83.0**	**80.8**	**81.9**

4.4.4 超参数分析实验

本文进一步探讨了超参数对MoELoRA性能的影响。研究分析了专家数量M和LoRA秩r的变化对结果的影响，结果如图5所示。从图（a）中可以看到，随着专家数量从0增加到4，MoELoRA的性能逐步提升。这种提升主要归因于更多专家能够学习更广泛的知识。然而，当专家数量增加到8时，性能出现轻微下降。这一现象的原因是专家数量的增加导致每个专家的LoRA秩变小，从而削弱了低秩矩阵的学习能力。因此，超参数选择的优化不仅需要考虑性能提升的潜力，还应综合考虑训练效率和计算资源的平衡。从图（b）中可以看出，虽然r的增加持续地提高了性能，但同时导致了可训练参数大小的成比例激增。考虑到需要在效率和绩效之间取得平衡，本文r的实际选择为8。

5 总结

本文提出了一种融合MoE的多任务学习企业新闻事件抽取模型，该模型采用多任务训练策略，将事件检测、论元抽取等子任务标准化为统一语言模板，强化模型对事件结构的整体感知；同时引入MoELoRA模块，通过动态门控机制协同多专家网络，在低秩参数空间实现任务间知识共享与特征解耦，实现联合事件抽取。实验表明，MoE-ML-CNEE模型

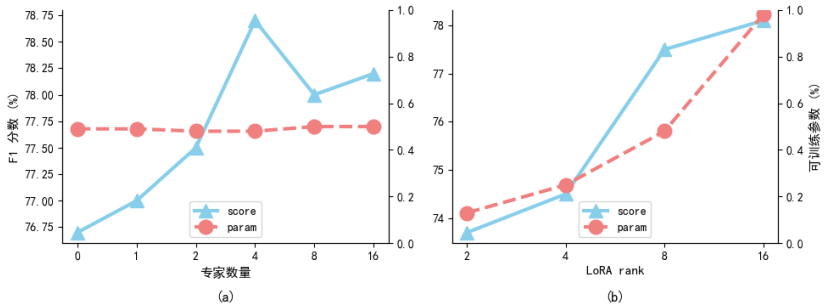

图 5: 专家数M和LoRA秩r的超参数实验结果

在ChiFinAnn和DuEE-fin数据集的事件论元抽取任务性能均优于最优基线模型。下一步将进一步研究更灵活、更高效的混合专家模块融合方法，以集成不同任务的特征表示来提高模型性能，同时降低模型复杂度。

参考文献

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Ruirui Chen, Chengwei Qin, Weifeng Jiang, and Dongkyu Choi. 2024. Is a large language model a good annotator for event extraction? In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 17772–17780.

Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, and Ziqing Yang. 2021. Pre-training with whole word masking for chinese bert. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:3504–3514.

Zhenjin Dai, Xutao Wang, Pin Ni, Yuming Li, Gangmin Li, and Xuming Bai. 2019. Named entity recognition using bert bilstm crf for chinese electronic health records. In *2019 12th international congress on image and signal processing, biomedical engineering and informatics (cisp-bmei)*, pages 1–5. IEEE.

Haixing Dai, Zhengliang Liu, Wenxiong Liao, Xiaoke Huang, Yihan Cao, Zihao Wu, Lin Zhao, Shaochen Xu, Fang Zeng, Wei Liu, et al. 2025. Auggpt: Leveraging chatgpt for text data augmentation. *IEEE Transactions on Big Data*.

Jun Gao, Huan Zhao, Changlong Yu, and Ruifeng Xu. 2023. Exploring the feasibility of chatgpt for event extraction. *arXiv preprint arXiv:2303.03836*.

Isobel Claire Gormley and Sylvia Frühwirth-Schnatter. 2019. Mixture of experts models. In *Handbook of mixture analysis*, pages 271–307. Chapman and Hall/CRC.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Cuiyun Han, Jinchuan Zhang, Xinyu Li, Guojin Xu, Weihua Peng, and Zengfeng Zeng. 2022. Duee-fin: A large-scale dataset for document-level event extraction. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 172–183. Springer.

Frederik Hogenboom, Flavius Frasincar, Uzay Kaymak, and Franciska De Jong. 2011. An overview of event extraction from text. *DeRiVE@ ISWC*, pages 48–57.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. 2022. Lora: Low-rank adaptation of large language models. *ICLR*, 1(2):3.

JDMCK Lee and K Toutanova. 2018. Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 3(8).

Sha Li, Heng Ji, and Jiawei Han. 2021. Document-level event argument extraction by conditional generation. *arXiv preprint arXiv:2104.05919*.

Qian Li, Jianxin Li, Jiawei Sheng, Shiyao Cui, Jia Wu, Yiming Hei, Hao Peng, Shu Guo, Lihong Wang, Amin Beheshti, et al. 2022. A survey on deep learning event extraction: Approaches and applications. *IEEE Transactions on Neural Networks and Learning Systems*.

Yuan Liang, Zhuoxuan Jiang, Di Yin, and Bo Ren. 2022. Raat: Relation-augmented attention transformer for relation modeling in document-level event extraction. *arXiv preprint arXiv:2206.03377*.

Zhuang Liu, Degen Huang, Kaiyu Huang, Zhuang Li, and Jun Zhao. 2021. Finbert: A pre-trained financial language representation model for financial text mining. In *Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence*, pages 4513–4519.

Bangze Pan, Yang Li, Suge Wang, Xiaoli Li, Deyu Li, Jian Liao, and Jianxing Zheng. 2024. Document-level event extraction via information interaction based on event relation and argument correlation. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 5156–5166.

Xinyu Wang, Lin Gui, and Yulan He. 2023. Document-level multi-event extraction with event proxy nodes and hausdorff distance minimization. *arXiv preprint arXiv:2305.18926*.

Chenxi Whitehouse, Monojit Choudhury, and Alham Fikri Aji. 2023. Llm-powered data augmentation for enhanced cross-lingual performance. *arXiv preprint arXiv:2305.14288*.

Junchao Wu, Shu Yang, Runzhe Zhan, Yulin Yuan, Lidia Sam Chao, and Derek Fai Wong. 2025. A survey on llm-generated text detection: Necessity, methods, and future directions. *Computational Linguistics*, pages 1–66.

Runxin Xu, Tianyu Liu, Lei Li, and Baobao Chang. 2021. Document-level event extraction via heterogeneous graph-based interaction model with a tracker. *arXiv preprint arXiv:2105.14924*.

Hang Yang, Yubo Chen, Kang Liu, Yang Xiao, and Jun Zhao. 2018. Dcfee: A document-level chinese financial event extraction system based on automatically labeled training data. In *Proceedings of ACL 2018, System Demonstrations*, pages 50–55.

An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, et al. 2024. Qwen2. 5 technical report. *arXiv preprint arXiv:2412.15115*.

Shun Zheng, Wei Cao, Wei Xu, and Jiang Bian. 2019. Doc2edag: An end-to-end document-level framework for chinese financial event extraction. *arXiv preprint arXiv:1904.07535*.

Tong Zhu, Xiaoye Qu, Wenliang Chen, Zhefeng Wang, Baoxing Huai, Nicholas Jing Yuan, and Min Zhang. 2021. Efficient document-level event extraction via pseudo-trigger-aware pruned complete graph. *arXiv preprint arXiv:2112.06013*.

王人玉, 项威, 王邦, and 代璐. 2023. 文档级事件抽取研究综述. 中文信息学报, 37(6):1–14.

赵庆玨, 余正涛, 王剑, 黄于欣, and 朱恩昌. 2024. 融入文档图和事件图的新闻核心事件检测. 中文信息学报, 38(5):99–106.

鲍彤. 2023. Chatgpt中文信息抽取能力测评——以三种典型的抽取任务为例. 数据分析与知识发现, 7(09):1–11.