

任务协作表示增强的要素及关系联合抽取模型

刘小明^{1,4}, 王 杭^{1,2}, 杨 关^{1,2}, 刘 杰^{3,4}, 曹梦远^{1,2}

(1. 中原工学院计算机学院, 河南郑州 450007; 2. 郑州市文本处理与图像理解重点实验室, 河南郑州 450007;
3. 北方工业大学信息学院, 北京 100144; 4. 国家语委中国语言智能研究中心, 北京 100089)

摘要: 对文本中诸如实体与关系、事件及其论元等要素及其特定关系的联合抽取是自然语言处理的一项关键任务。现有研究大多采用统一编码或参数共享的方式隐性处理任务间的交互, 缺乏对任务之间特定关系的显式建模, 从而限制模型充分利用任务间的关联信息并影响任务间的有效协同。为此, 提出了一种基于任务协作表示增强的要素及关系联合抽取模型 (Task-Collaboration Representation Enhanced model for joint extraction of elements and relationships, TCRE)。该模型旨在从多个阶段处理任务间的特定关系, 帮助子任务进行更细致的调节和优化, 促进整体性能的提升。在三个关系抽取和一个事件抽取数据集上进行实验, TCRE 在实体识别和关系提取任务上平均性能分别提高 0.57% 和 0.77%, 在触发词识别和论元角色分类任务上分别提高 0.7% 和 1.4%。此外, TCRE 还显示出在缓解“跷跷板现象”方面的作用。

关键词: 关系表示; 联合抽取; 任务协作; 多任务学习; 跷跷板现象

基金项目: “新一代人工智能”国家科技重大专项 (No.2020AAA0109703); 国家自然科学基金 (No.62076167, No.61772020, No.U23B2029); 河南省高等学校重点科研项目 (No.23A520022)

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112(2024)06-1955-08

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20230623

Task-Collaboration Representation Enhanced Joint Extraction Model for Elements and Relationships

LIU Xiao-ming^{1,4}, WANG Hang^{1,2}, YANG Guan^{1,2}, LIU Jie^{3,4}, CAO Meng-yuan^{1,2}

(1. School of Computer Science, Zhongyuan University of Technology, Zhengzhou, Henan 450007, China;

2. Zhengzhou Key Laboratory of Text Processing and Image Understanding, Zhengzhou, Henan 450007, China;

3. School of Information Science, North China University of Technology, Beijing 100144, China;

4. Research Center for Language Intelligence of China, Beijing 100089, China)

Abstract: Jointly extracting elements like entities and their relationships, as well as events and their arguments, is a crucial natural language processing task. Current methods, primarily based on unified coding or parameter sharing, fail to explicitly model inter-task relationships. This limitation restricts the use of inter-task correlations and hinders effective collaboration. To address this, we propose a task-collaboration representation enhanced model for joint extraction of elements and relationships (TCRE). TCRE strategically captures and leverages specific inter-task relationship representations across multiple stages, facilitating precise tuning and optimization of subtasks, thereby enhancing overall model performance. In evaluations on three relation extraction and one event extraction datasets, TCRE demonstrated performance improvements of 0.57% in entity recognition, 0.77% in relation extraction, 0.7% in trigger word recognition, and 1.4% in argument role classification. Additionally, TCRE effectively mitigates the “seesaw phenomenon”.

Key words: relationship representation; joint extraction; task collaboration; multi task learning; seesaw phenomenon

Foundation Item(s): National Key Research and Development Program of China (No.2020AAA0109703); National Natural Science Foundation of China (No.62076167, No.61772020, No.U23B2029); Key Scientific Research Project of Higher Education Institutions in Henan Province (No.23A520022)

1 引言

关系抽取和事件抽取旨在从非结构化文本中识别和构造用户指定的信息^[1],是信息提取领域的关键任务.如图1所示,在实体及关系的抽取任务中,实体关系描述了文本中不同实体之间的关联关系.在事件及其论元的抽取任务中,论元角色描述了触发词和论元之间的特殊语义关系.因此,实体与关系抽取,及事件与论元抽取,都可以进一步统一为对文本中的要素识别和要素关系提取两个子任务.其中,要素识别和要素关系提取具有密切的相关性^[2],探索如何更有效地任务协作机制对于增强模型对文本语义理解的能力具有重要意义.

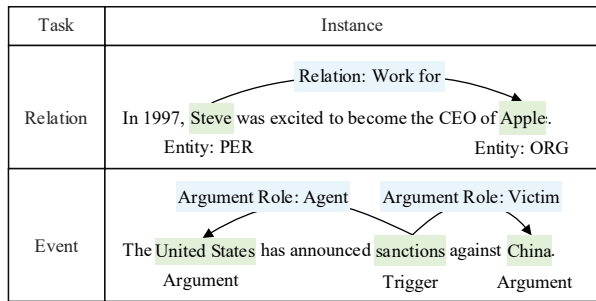


图1 要素抽取和要素关系抽取

在现有的要素及其关系联合模型中,研究者主要从联合解码^[3,4]和参数共享^[5,6]两个方面建模任务间的交互.联合解码的方法将子任务编码到统一的标签空间,学习统一的特征来同时表示要素及其关系.虽然这类方法增强了任务间的交互作用,但要素及其关系在本质上仍共享上下文表示,因此没有充分考虑到任务特征的特殊性^[7].基于参数共享的方法通常采用相互独立的网络分别为要素及其关系编码不同的特征表示,两个任务通过共享输入特征以及部分网络参数实现信息交互.参数共享在一定程度上能够捕捉任务间隐含的关联信息,但仍对任务间的复杂依赖关系挖掘和利用不足,且不合理的交互方式可能引发知识负迁移问题^[8].此外,由于任务间的冲突或竞争关系,多目标联合优化过程中可能会出现“跷跷板现象”(seesaw phenomenon)^[9].因此,联合解码和参数共享的方式各有其优势及局限性.值得注意的是,大模型已在信息抽取任务上展现了出色的多任务统一学习能力,近期的研究^[2,10]以大模型为底座,通过 Prompt 方式将不同任务统一建模为文本生成任务,但该过程中任务间知识的转移和利用是隐式归纳的^[11],这对于研究任务间的关系以及提升模型的可解释性带来了挑战.

尽管文本要素及其关系抽取的相关研究已经取得了显著进展,但目前仍面临多个亟待解决的问题.首先,针对任务间内在关联性的深度挖掘还不够充分^[7];

其次,在确保各个任务特征独特性的同时,如何恰当地权衡与整合其共享特性也是一大难题^[12];最后,如何有效地筛选并排除在任务交互过程中产生的冗余及干扰信息同样是一项挑战.针对上述问题,本文设计了一种任务协作表示增强的要素及关系联合抽取模型(Task-Collaboration Representation Enhanced joint extraction model for element and relationship, TCRE).主要贡献如下:(1)构建特征适应模块和特征交互模块,分别过滤干扰信息和充分挖掘任务间复杂的交互信息;(2)构建一种多任务协作框架,通过协调任务间的竞争与平衡关系,有效缓解了“跷跷板现象”;(3)在三个实体及其关系抽取和一个事件及论元抽取数据集上取得了良好的性能.

2 相关工作

在多任务学习的研究中,参数共享方式从初始的硬共享模式逐渐演变为软共享模式和共享-私有模式^[13].Yan等^[8]利用分区过滤网络来分离出任务的特定特征以及任务间的共享特征,缓解了特征交互过程中存在的特征冲突问题.Xu等^[5]通过事件触发词、论元和论元角色定义了事件关系三元组,以学习它们之间的相互依赖.Tang等^[9]提出 PLE 框架进一步考虑了不同专家网络的特征交互,缓解了“跷跷板现象”.随着大型模型的发展,Lu等^[2]提出的信息抽取框架 UIE 将多个信息任务的不同结构进行统一描述.Wei等^[10]基于 ChatGPT 提出的 ChatIE 通过多轮问答来解决零样本信息抽取任务.基于大型模型的方法在零样本和小样本场景中取得了明显的优势,但在全监督情况下相比传统模型仍在一些任务和数据集上存在不足.此外,当前研究中尚缺乏对这类方法在多任务学习中可能存在的负迁移问题及其应对策略的深入探讨.

上述研究表明,要素及其关系联合抽取利用任务之间的协同作用可以进一步优化子任务的特征表示.然而,当前关于联合抽取的研究正如 Zhong 等^[7]的观点:流水线模型在某种场景下相比联合模型仍然展现出可观的性能,有效的任务协作机制的还需要进一步探索.此外,当前方法大多将重点放在捕获任务之间的依赖关系,忽略了任务交互中知识负迁移和任务特征特殊性与共享性不平衡的问题^[12].

3 问题定义

对于文本序列 $X=(x_1, x_2, \dots, x_N)$,任务的目标是抽取文本中的要素及其关系,其中要素包括实体、论元或事件,关系描述了要素之间的语义关联.使用 $E=\{e_1, e_2, \dots, e_M\}$ 表示要素集合,其中, M 表示要素的个数,每个要素 e_i 都有其在本文本中的位置范围

$e_i=(start_i, end_i)$, 表示要素在文本序列中的起始位置和结束位置. 同时, 使用 $R=\{r_1, r_2, \dots, r_L\}$ 表示关系集合, 其中, L 表示关系的个数. 每个关系 r_j 包括两个要素的索引, 表示这两个要素之间的关联关系. 要素及其关系的联合抽取任务可以定义为: 在给定文本序列 X 的情况下, 找到最优的要素集合 E 和关系集合 R , 使得模型能够同时完成要素抽取和关系抽取两个子任务. 形式化地表示为

$$E^*, R^* = \underset{E, R}{\operatorname{argmax}} (P(E, R|X)) \quad (1)$$

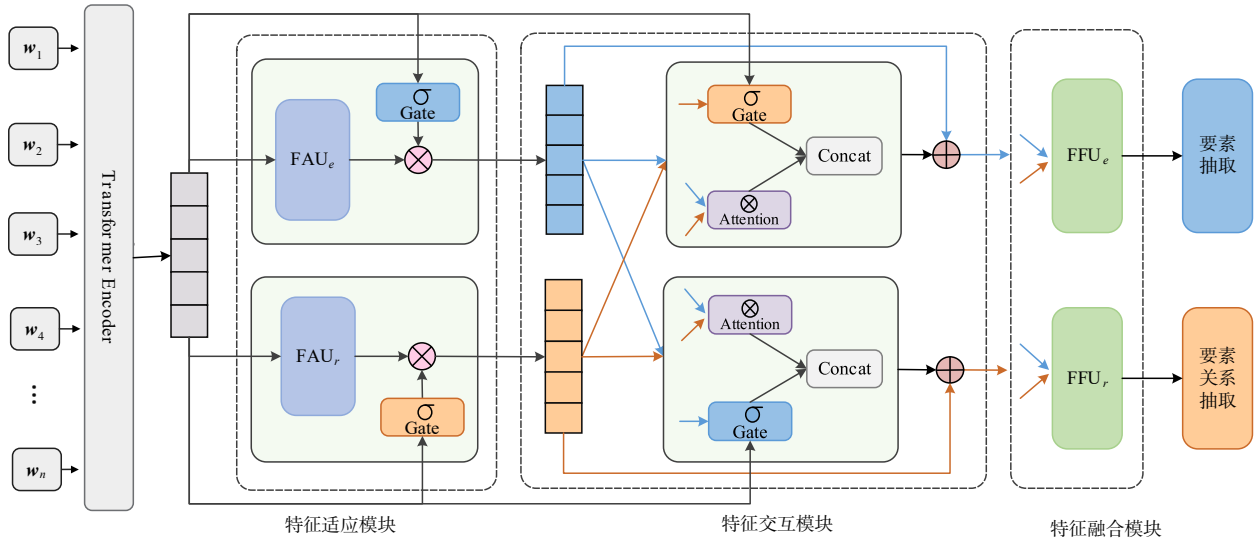


图2 模型框架图

4.1 特征适应模块

基于预训练的编码层在所有任务中共享, 然而, 不同任务关注不同的上下文信息, 每个任务都有其独特的特征需求. 因此, 在共享编码器之上引入特征适应模块. 该模块由特征适应单元 (Feature Adaptation Unit, FAU) 和任务门控单元组成, 分别用于提取任务初步适应特征和过滤共享特征中的干扰语义特征. 设 $H' \in \mathbb{R}^{N \times h'}$ 为 BERT 编码器提取的句子的上下文特征, 其中 h' 为 BERT 的隐藏层维度. 对于每个任务, 首先通过 FAU 计算初步适应特征, 然后将该特征输入到相应的任务门进行过滤和调整, 以进一步生成任务的适应特征. 每个任务的计算方法相似, 以要素识别任务为例, 具体的计算如下:

$$S'_e = \tanh(H'W_e + b_e) \quad (2)$$

$$g_e^{\text{adapt}} = \sigma(H'W_{e_{\text{adapt}}} + b_{e_{\text{adapt}}}) \quad (3)$$

$S'_e \in \mathbb{R}^{N \times h}$ 表示要素识别任务的初步适应特征, $g_e^{\text{adapt}} \in \mathbb{R}^{N \times h}$ 表示要素抽取的任务门, $W_e \in \mathbb{R}^{h' \times h}$ 和 $b_e \in \mathbb{R}^h$ 表示可训练参数, h 是本文模型的隐藏层维度, $\tanh(\cdot)$ 表示 \tanh 的激活函数, $\sigma(\cdot)$ 表示 sigmoid 激活函

其中, E^* 和 R^* 表示最优的要素集合和关系集合, $P(E, R|X)$ 表示在给定文本序列 X 的情况下, 要素集合 E 和关系集合 R 的联合概率.

4 本文模型

本文提出一种多任务协作表示增强的要素关系联合抽取模型, 如图 2 所示, 模型利用特征适应模块、特征交互模块以及特征融合模块, 根据不同任务需求从多个阶段处理任务间的依赖关系, 以帮助子任务获取协作增强的特征表示.

数. 初步适应特征通过任务门进行过滤, 可以选择性地消除可能对特定任务产生干扰的语义特征, 从而形成每个任务所需的适应特征:

$$S_e = g_e^{\text{adapt}} \odot S'_e \quad (4)$$

其中, \odot 为哈达玛积运算, $S_e \in \mathbb{R}^{N \times h}$ 表示要素识别任务的适应特征, 要素关系提取任务的适应特征 $S_r \in \mathbb{R}^{N \times h}$ 与 S_e 的计算方式相同.

4.2 特征交互模块

为建模任务间的信息交互并充分挖掘其依赖关系, 本文设计了一个由注意力单元和交互门控单元组成的特征交互模块. 注意力单元聚焦于最相关的信息以建模全局关系, 交互门控单元来弥补注意力机制对弱相关特征的关注不足. 通过结合两个单元, 以获得更全面的交互特征表示. 图 3 展示了注意力单元和门控单元协作捕获任务相关信息的方式.

注意力采用多头模式, 利用任务的适应特征来加权其他任务特征. 以要素识别任务为例, 该单元的输入是要素识别任务和要素关系提取任务的适应特征, 其中查询矩阵 Q 要素识别任务的适应特征 S_e , 键矩阵 K 和

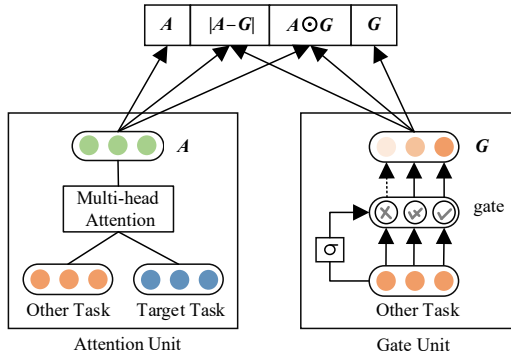


图3 注意力单元和门控单元协作捕获任务相关信息

值矩阵 V 是要素关系提取任务的适应特征 S_r , 注意力单元可以形式化如下:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (5)$$

$$\text{head}_i = \text{Attention}(S_e W_i^Q, S_r W_i^K, S_r W_i^V),$$

$$A_e = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_n) W^O$$

$W_i^Q \in \mathbb{R}^{h \times d_i}$, $W_i^K \in \mathbb{R}^{h \times d_i}$ 和 $W_i^V \in \mathbb{R}^{h \times d_i}$ 表示可训练的投影参数, $d_k = h/n$, 其中 n 为注意力的头数, $W^O \in \mathbb{R}^{h \times h}$ 表示可训练参数, $A_e \in \mathbb{R}^{N \times h}$ 表示要素识别任务通过注意力单元与要素关系提取任务交互生成的特征.

$$g_e^{\text{inter}} = \sigma(H_e W_{e_{\text{inter}}} + b_{e_{\text{inter}}}) \quad (6)$$

$$G_e = g_e^{\text{inter}} \odot S_r \quad (7)$$

其中, $g_e^{\text{inter}} \in \mathbb{R}^{N \times h}$ 表示要素识别任务的交互门, $G_e \in \mathbb{R}^{N \times h}$ 表示要素识别任务通过门控单元从要素关系提取任务中学习到的相关特征, $W_{(e)} \in \mathbb{R}^{h' \times h}$ 和 $b_{(e)} \in \mathbb{R}^h$ 表示可训练参数. 当注意力单元和交互门控单元作用之后, 每个单元从不同角度生成了有利于要素识别任务的特征表示 A_e 和 G_e , 然后采用绝对差值和元素级乘积^[14]对两个特征进行整合来获取交互特征. 最后, 通过任务适应特征和交互特征进行连接, 以获取整个特征交互模块的任务特征:

$$I_e' = \text{MLP}([A_e; |A_e - G_e|; A_e \odot G_e; G_e]) \quad (8)$$

$$I_e = [S_e; I_e'] \quad (9)$$

其中, $[\cdot; \cdot]$ 表示向量的拼接操作, $\text{MLP}(\cdot)$ 表示多层感知机, $I_e' \in \mathbb{R}^{N \times h}$ 表示要素识别任务的交互特征, $I_e \in \mathbb{R}^{N \times 2h}$ 表示要素识别任务通过两个任务交互后生成的协作特征, 要素关系提取任务的协作特征 $I_r \in \mathbb{R}^{N \times 2h}$ 与 I_e 的计算方式相同. 该模块综合考虑任务特征的特异性和共享性, 通过任务间的交互, 外部特征与自身的适应特征相结合, 形成更加丰富的表示.

4.3 特征融合模块

该模块由特征融合单元 (Feature Fusion Unit, FFU) 组成, 它计算了任务协作特征的融合表示. 对于来自两个任务同一位置的字符特征 $I_e^i, I_r^i \in \mathbb{R}^{2 \times 2h}$, 某一

任务使用 Softmax 归一化获得每个任务特征的融合分数, 然后把融合分数作为 I_e^i 和 I_r^i 的权重来组合生成该任务的融合特征. 具体计算方法如下:

$$f_e^\Delta(I_e^i) = \text{Softmax}(I_e^i W_\Delta^i) \quad (10)$$

$$H_\Delta^i = f_e^\Delta(I_e^i) \cdot I_e^i + f_r^\Delta(I_r^i) \cdot I_r^i \quad (11)$$

其中, $\Delta \in \{e, r\}$ 表示要素识别或要素关系抽取中的某个任务, $W_{(e)}^i \in \mathbb{R}^{2h \times 2}$ 表示可训练参数, $H_e^i \in \mathbb{R}^{1 \times 2h}$ 和 $H_r^i \in \mathbb{R}^{1 \times 2h}$ 分别表示两个任务字符级特征的最终表示. $f_e^\Delta(\cdot)$ 和 $f_r^\Delta(\cdot)$ 分别代表特征 I_e^i 和 I_r^i 在新融合的特征 H_Δ^i 中的融合分数, 该分数决定了一个任务对另一个任务的最终表示的贡献程度. 由于任务通过特征交互模块生成的特征已经足够丰富, 在特征融合过程对其他任务关注过多, 可能会对自身的特异性特征造成损害. 因此, 本文引入一个辅助损失来迫使任务融合过程中更加侧重任务自身特征:

$$\mathcal{L}_f = - \sum_{i=1}^N \log(f_e^\Delta(I_e^i) \cdot f_r^\Delta(I_r^i)) \quad (12)$$

4.4 解码模块

在该模块中以实体识别和关系提取任务选用的解码方式进行介绍. 本文将两个任务都视为表填充问题, 为每个实体类型 $m \in E$ 分别构造一个 $N \times N$ 的表, 其第 i 行和第 j 列的元素 e_{ij}^m 表示词 w_i 和词 w_j 是类型为 m 的实体起始和结束位置的概率. 利用词 w_i 的特征 H_e^i 和词 w_j 的特征 H_e^j 拼接后通过线性层, 并使用 ELU 及 sigmoid 函数激活, 得到该实体的特征表示 $H_{i,j}^e$ 和实体概率分布:

$$H_{i,j}^e = \text{ELU}(\text{Linear}([H_e^i; H_e^j])) \quad (13)$$

$$e_{i,j}^m = \sigma(\text{Linear}(H_{i,j}^e)) \quad (14)$$

关系提取任务中为每个关系类型 $l \in R$ 分别构造一个 $N \times N$ 的表中, 其第 i 行和第 j 列的元素 r_{ij}^l 表示词 w_i 和词 w_j 同时成为主体和客体起始词的概率. 结合实体识别任务中对实体边界的预测可以获得实体与实体之间的关系特征 $H_{i,j}^r$ 以及关系概率分布:

$$H_{i,j}^r = \text{ELU}(\text{Linear}([H_r^i; H_r^j])) \quad (15)$$

$$r_{i,j}^l = \sigma(\text{Linear}(H_{i,j}^r)) \quad (16)$$

4.5 目标优化

训练过程中每个子任务的损失函数都采用二元交叉熵损失 (Binary CrossEntropy Loss):

$$\mathcal{L}_e = \sum_{i=1}^N \sum_{j=1}^N \sum_{m \in E} \text{BCELoss}(e_{i,j}^m, \hat{e}_{i,j}^m) \quad (17)$$

$$\mathcal{L}_r = \sum_{i=1}^N \sum_{j=1}^N \sum_{l \in R} \text{BCELoss}(r_{i,j}^l, \hat{r}_{i,j}^l)$$

$\hat{e}_{i,j}^m$ 和 $\hat{r}_{i,j}^l$ 分别是实体和关系的真实标签. 由于不同任务的损失可能具有不同的量级, 设置不同的权重和会给任务的训练速度和梯度带来不同的影响. 为了更好地平衡两个任务在训练过程中的差异, 本文采用

动态加权的方法.

$$\begin{aligned} w_{\Delta}(t-1) &= \frac{L_{\Delta}(t-1)}{L_{\Delta}(t-2)} \\ \lambda_{\Delta}(t) &= \frac{2\exp(w_{\Delta}(t-1))}{\sum_{\tau \in \{e, r\}} \exp(w_{\tau}(t-1))} \end{aligned} \quad (18)$$

其中, t 代表训练轮数 epoch, L_{Δ} 为任务 Δ 在整个 epoch 的平均损失, $w_{\Delta}(t-1)$ 代表任务 Δ 前一轮损失的更新速率, 通过计算各个任务的损失速率占比得到对应的权重 $\lambda_{\Delta}(t)$. 因此模型在当前轮数 t 的整体损失函数可以表示为

$$\text{Loss}(t) = \lambda_e(t)\mathcal{L}_e + \lambda_r(t)\mathcal{L}_r + \mathcal{L}_f \quad (19)$$

5 实验

5.1 数据集

本文在 ACE2005^[15]、ACE2004^[16] 和 SciERC^[17] 三个包含丰富实体类别和关系类别的数据集进行实验, 具体统计数据如表 1 所示. 其中 ACE2005 和 ACE2004 数据来自多个领域, 如新闻文章和在线论坛, SciERC 数据集来源于 500 篇人工智能领域的论文摘要. 与 Yan 等^[8] 和 Zhong 等^[7] 的方法一致, 本文使用 ACE2004 数据集 15% 的训练集构建其验证集.

表 1 数据集信息统计

数据集	类型数		句子数		
	实体	关系	训练集	验证集	测试集
ACE2005	7	6	10 051	2 424	2 050
ACE2004	7	6	8 683	8 686(5-fold)	
SciERC	6	7	1 861	275	55

表 2 对比实验结果

单位: %

方法	方法类型	ACE2005		ACE2004		SciERC	
		NER	RE	NER	RE	NER	RE
GPT-NER ^[21]	LLM+Prompt	73.0	—	73.6	—	—	—
UIE ^[2]	LLM+Prompt-tuning	85.8	66.0	86.9	—	—	36.5
PURE ^[7]	PLM+Fine-tuning	89.7	65.6	88.8	60.2	66.6	35.6
PFN ^[8]	PLM+Fine-tuning	89.0	66.8	89.3	62.5	66.8	38.4
USM ^[11]	PLM+Fine-tuning	87.1	67.9	87.6	—	—	37.4
TCRE	PLM+Fine-tuning	90.1	67.4	89.8	63.7	67.5	39.5

6 实验分析

6.1 消融实验

为验证各个模块的有效性, 在 ACE2005 和 SciERC 数据集上进行消融实验, 结果如表 3 所示.

其中, w/o ALL 表示把特征适应、特征交互和特征融合三个模块删除, 并替换为一个独立的双向 LSTM 进行对比. 结果显示, 在两个任务上效果平均降低了

5.2 评价指标和参数设置

本文使用 Micro-F1 评分作为实体识别任务和关系抽取任务的评价指标. 最后在验证集上选择具有两个任务最佳平均 F1 的模型. 本文模型、PURE^[7] 和 PFN^[8] 在 ACE 系列的数据集上使用 Albert^[18] 作为预训练语言模型, 在 SciERC 数据集上使用 Scibert^[19], USM^[11] 在所有数据集上使用 Roberta^[20] 作为预训练语言模型. 对于数据集 ACE2005 和 ACE2004, 批次大小设置为 16, 学习率为 2×10^{-5} , 对于数据集 SciERC, 批大小设置为 4, 学习率为 1×10^{-5} . 三个数据集的其他实验参数都相同, 优化器选择 AdamW, 隐藏维度为 300, 序列最大长度为 128, 迭代次数为 100, 为防止过拟合将 dropout 设置为 0.1.

5.3 实验结果

实验结果如表 2 所示, GPT-NER 在 ACE 数据集命名实体识别任务上的效果还不能令人满意, 而 UIE 同时建模多个任务, 在命名实体识别和关系抽取任务的效果都取得了更好的效果. 在数据集 ACE2005 的实体识别任务中, PFN 相比于流水线模式的 PURE 降低了 0.7%, 而 TCRE 相比 PURE 和 PFN 提升了 0.3% 和 1.1%, 在该数据集的关系抽取任务上, TCRE 相比 PURE 和 PFN 提升了 1.8% 和 0.6%. 另外在 ACE2004 和 SciERC 数据集上, 两个任务分别提升了 0.2%、0.7% 和 1.2% 和 1.1%. USM 分析并处理了 UIE 在建模任务知识转移方面的难以解释性问题, 在三个数据集上的效果相比 UIE 有全面的提升, 并在 ACE2005 数据集的关系抽取任务上比本文模型效果领先 0.5%, 然而, USM 在其他任务上相比本文模型或其它基于预训练模型的基线方法仍存在不足. 总体来看, TCRE 在建模任务间的协作方面, 在三个数据集上都取得了明显的效果.

1.55% 和 2.8%, 这证明了 TCRE 的整体有效性. FA-GATE 和 FA-FAU 分别表示特征适应模块的门控单元和特征适应单元, FI-GATE 和 FI-ATT 分别表示特征交互模块的门控单元和注意力单元. 在特征适应模块中消去特征适应单元后, 效果下降明显, 表明保持任务特殊性特征的重要性. 在特征交互模块中发挥了较大作用的是注意力单元, 其对促进任务知识正向迁移产生了

表3 消融实验结果 单位:%

方法	ACE2005		SciERC	
	NER	RE	NER	RE
Ours	90.1	67.4	67.8	39.5
w/o ALL	89.3	63.7	65.5	37.6
w/o FA-GATE	88.7	66.6	66.8	38.8
w/o FA-FAU	88.5	65.5	66.1	38.0
w/o FI-GATE	89.5	65.9	67.2	38.2
w/o FI-ATT	89.1	65.4	67.0	38.5
w/o FF	89.6	67.1	67.2	38.9
w/o DWL	89.9	67.5	67.5	39.3

显著的贡献. FF表示特征融合模块,消除该模块后,两个任务的平均性能分别降低了0.5%和0.35%.消融实验表明三个模块对整体模型都有不同程度的积极作用. DWL表示动态加权,与直接固定两个任务的损失权重为1:1相比,整体性能有一定的提升.

6.2 跷跷板现象分析

为分析本文模型对缓解“跷跷板现象”的作用,将TCRE和PFN在ACE2005数据集上进行对比.如图4和图5所示,在实体识别任务上,PFN前期取得了较好的效果,最高可达89.70%.然而,随着关系提取任务效果的提升,实体识别任务的效果开始下降,最后取得的最佳值仅为89.0%,出现了“跷跷板现象”.相比之下,TCRE在两个任务上的效果总体都呈上升趋势.

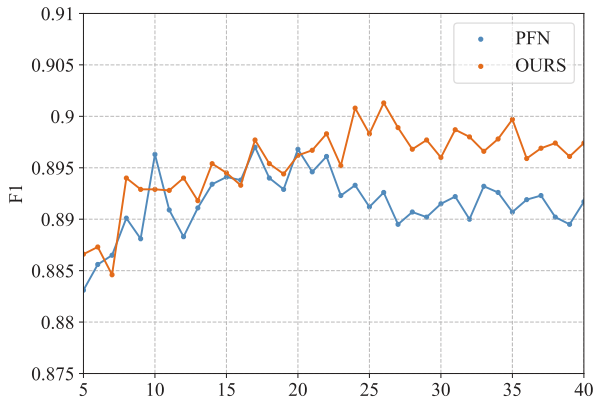


图4 实体识别任务模型效果对比

对于序列标注多任务,传统方法大多利用一个共享空间捕获跨任务的共享特征表示,而共享参数之间可能存在冲突或竞争关系,本文认为这是导致出现“跷跷板现象”的关键原因.因此本文从共享底层逐步调整任务的协作机制,旨在有效提取任务共性和特性的同时协调任务间的竞争与平衡.具体来说,特征适应模块为每个任务在预训练特征与下游任务特征的转换之间形成过渡,分担共享底层参数调整的压力,减少底层参数之间的竞争.其次在特征交互模块中,分别为每个任

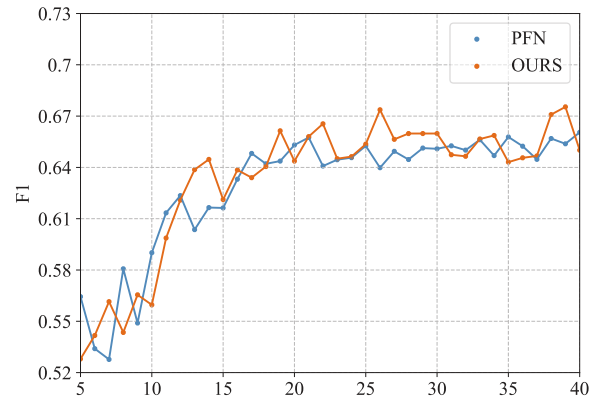


图5 关系提取任务模型效果对比

务设计一个更加注重任务自身需求的交互空间来实现共享,该空间从多任务相互作用中做出对自身任务更有利的参数调整,以减少任务直接交互可能产生的冲突或竞争.

7 拓展实验

为验证模型的拓展性,在DuEE1.0数据集进行事件识别和论元角色分类联合学习实验.采用与基线模型JEEDG^[6]相同的标注策略和解码方式,将本文模型任务协作模块与JEEDG对应部分进行替换,实验验证了在事件抽取领域对要素及其关系协作的能力,实验结果由表4所示.

表4 事件抽取实验结果 单位:%

方法	Paradigm	TRIGGER	ROLE
BERT+BiGRU+CRF	Pipeline	86.1	65.8
BERT+BiGRU+CRF	Joint	84.9	66.3
BERT+MMOE+CRF	Joint	85.7	66.2
JEEDG	Joint	86.6	66.9
TCRE	Joint	87.3	68.3

本文模型对比基线模型在两个任务上性能分别提高0.7%和1.4%,与流水线的方式相比,TCRE和JEEDG都给两个任务带来了增益.基于BiGRU和MMOE的联合抽取,触发词识别任务效果均受到影响,结果相比流水线方式分别下降1.2%和0.4%,原因可能是交互过程对特征的选择过滤机制不够完善,论元角色分类任务提供给触发词识别任务的知识更多是干扰特征,导致触发词识别任务的特定特征不如流水线方式.总体来看,TCRE能够通过积极的任务协作机制和缓解“跷跷板现象”给事件抽取任务带来性能提高.

8 结语

本文提出一种基于任务协作表示增强的要素关系联合抽取模型TCRE,其核心是获取利用任务间特定关

系来增强其特征表示. 在实体和关系的联合抽取任务和事件及论元联合抽取任务的结果都优于现有方法, 证实任务间关系可有效增强对要素及其关系抽取联合协作学习的效果. 未来研究将探索多任务协作表示增强在更广泛领域的联合抽取潜力.

参考文献

- [1] 隗昊, 唐焕玲, 周爱, 等. 基于双路分段注意力神经张量网络的临床文本关系抽取[J]. 电子学报, 2023, 51(3): 658-665.
WEI H, TANG H L, ZHOU A, et al. Clinical relation extraction via dual piecewise attention neural tensor network[J]. Acta Electronica Sinica, 2023, 51(3): 658-665. (in Chinese)
- [2] 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 (Volume 1: Long Papers). Stroudsburg: Association for Computational Linguistics, 2022: 5755-5772.
- [3] FU T J, LI P H, MA W Y. GraphRel: Modeling text as relational graphs for joint entity and relation extraction[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2019: 1409-1418.
- [4] REN F L, ZHANG L H, YIN S J, et al. A novel global feature-oriented relational triple extraction model based on table filling[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2021: 2646-2656.
- [5] XU N, XIE H H, ZHAO D Y. A novel joint framework for multiple Chinese events extraction[C]//China National Conference on Chinese Computational Linguistics. Cham: Springer, 2020: 174-183.
- [6] 冯兴杰, 赵新阳, 冯小荣. 基于软参数共享的事件联合抽取方法[J]. 计算机应用研究, 2023, 40(1): 91-96.
FENG X J, ZHAO X Y, FENG X R. Joint event extraction method based on soft parameter sharing[J]. Application Research of Computers, 2023, 40(1): 91-96. (in Chinese)
- [7] ZHONG Z X, CHEN D Q. A frustratingly easy approach for entity and relation extraction[C]//Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: Association for Computational Linguistics, 2021: 50-61.
- [8] YAN Z H, ZHANG C, FU J L, et al. A partition filter network for joint entity and relation extraction[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2021: 185-197.
- [9] TANG H Y, LIU J N, ZHAO M, et al. Progressive layered extraction (PLE): A novel multi-task learning (MTL) model for personalized recommendations[C]//Proceedings of the 14th ACM Conference on Recommender Systems. New York: ACM, 2020: 269-278.
- [10] WEI X, CUI X Y, CHENG N, et al. Zero-shot information extraction via chatting with ChatGPT[EB/OL]. [2023]. <http://arxiv.org/abs/2302.10205>.
- [11] LOU J, LU Y J, DAI D, et al. Universal information extraction as unified semantic matching[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(11): 13318-13326.
- [12] XIONG X, LIU Y, LIUA, et al. A multi-gate encoder for joint entity and relation extraction[C]//Proceedings of the 21st Chinese National Conference on Computational Linguistics. Nanchang: Chinese Information Processing Society of China, 2022: 848-860.
- [13] LIU X D, HE P C, CHEN W Z, et al. Multi-task deep neural networks for natural language understanding[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2019: 4487-4496.
- [14] MOU L L, MEN R, LI G, et al. Natural language inference by tree-based convolution and heuristic matching[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Stroudsburg: Association for Computational Linguistics, 2016: 130-136.
- [15] CHRISTOPHER WALKER. ACE 2005 multilingual training corpus[EB/OL]. [2023]. <https://doi.org/10.35111/mwxc-vh88>.
- [16] DODDINGTON G, MITCHELL A, PRZYBOCKI M, et al. The automatic content extraction (ACE) program - Tasks, data, and evaluation[C]//Proceedings of the Fourth International Conference on Language Resources and Evaluation. Lisbon: European Language Resources Association (ELRA), 2004: 837-840.
- [17] LUAN Y, HE L H, OSTENDORF M, et al. Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Com-

putational Linguistics, 2018: 3219-3232.

- [18] LAN Z Z, CHEN M D, GOODMAN S, et al. ALBERT: A lite BERT for self-supervised learning of language representations[EB/OL]. [2023]. <http://arxiv.org/abs/1909.11942>.
- [19] BELTAGY I, LO K, COHAN A. SciBERT: A pretrained language model for scientific text[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Stroudsburg: Association for Computational Linguistics, 2019: 3615-3620.
- [20] LIU Y H, OTT M, GOYAL N, et al. RoBERTa: A robustly optimized BERT pretraining approach[EB/OL]. [2023]. <http://arxiv.org/abs/1907.11692>.
- [21] WANG S H, SUN X F, LI X Y, et al. GPT-NER: Named entity recognition via large language models[EB/OL]. [2023]. <http://arxiv.org/abs/2304.10428>.

作者简介



刘小明 男, 1979年6月出生于河南省许昌市. 现为中原工学院计算机学院讲师、硕士生导师. 主要研究方向为自然语言、中文信息处理和机器学习.

E-mail: ming616@zut.edu.cn



王杭 男, 1997年11月出生于河南省商丘市. 现为中原工学院计算机学院硕士研究生, 研究方向为自然语言处理.

E-mail: 2021107253@zut.edu.cn

杨关 男, 1974年6月出生于陕西省西安市. 现为中原工学院计算机学院副教授、硕士生导师. 研究领域为机器学习与图像处理.

刘杰 男, 1970年5月出生于江苏省徐州市. 现为北方工业大学信息学院教授、博士生导师. 研究领域为机器学习、人工智能和知识图谱.

曹梦远 男, 1998年5月出生河南省信阳市. 现为中原工学院计算机学院硕士研究生, 研究方向为自然语言处理.