

基于多特征和跨模态知识蒸馏的鱼病命名实体识别

沈志成^{1,2}, 陈明^{1,2}

(1. 上海海洋大学信息学院, 上海 201306; 2. 农业农村部渔业信息重点实验室, 上海 201306)

摘要: 为解决多模态鱼病知识缺乏合理安排的问题, 同时降低知识蒸馏过程的冗余数据, 从而部署存储低、样本小、精度高的识别模型, 提出一种基于多特征协同预测-跨模态多头蒸馏的方法, 命名为FSFDAI-TMRD。在多特征协同预测方面, 重点改进了原多任务多特征协同预测架构。首先使用更细粒度的BMES(Begin-middle-end-single)法代替原工作中BIO(Begin-inside-outside)法的粗略标注, 其次修改原架构的联合概率分布计算公式, 使得模型可以更好地识别嵌套名词实体。在跨模态多头蒸馏方面, 本文运用了跨模态注意力机制。首先计算合并、拆分和点积后的多头关系矩阵, 其次利用相对熵进行知识蒸馏, 使得模型可以更好地对齐异构师生间的中间特征。同时, 本文还应用了双仿射注意力机制及对抗性权重扰动函数等方法, 加强学习语义语音和字形词义等多特征知识。与主流模型相比, 本文方法的精确率、召回率和 F_1 值分别提升了0.45%、3.96%和2.28%, 并且存储优化比例提高3.01%, 模型参数规模缩小94.86%。

关键词: 知识蒸馏; 鱼病命名实体识别; 对抗训练; 双仿射注意力机制; 模型压缩

中图分类号: TP311 **文献标志码:** A

Named Entity Recognition of Fish Disease Based on Multi-feature and Cross-Modal Knowledge Distillation

SHEN Zhicheng^{1,2}, CHEN Ming^{1,2}

(1. School of Information, Shanghai Ocean University, Shanghai 201306, China; 2. Key Laboratory of Fishery Information, Ministry of Agriculture and Rural Affairs, Shanghai 201306, China)

Abstract: In order to solve the lack of reasonable arrangement of multi-modal fish disease knowledge, and at the same time reduce the redundant data in the knowledge distillation process, so as to deploy a recognition model with low storage, small samples, and high accuracy, this paper proposes a new method, named as FSDAI-TMRD. In terms of multi-feature collaborative prediction, this paper focuses on improving the original multi-feature collaborative multi-feature prediction architecture of multi-tasks. Firstly, the finer-grained begin-middle-end-single (BMES) method is used instead of the rough labeling of the begin-inside-outside (BIO) method in the original work. Secondly, the formula for calculating the joint probability distribution of the original architecture is modified, so that the model can better recognize the nested noun entities. In terms of cross-modal multi-head distillation, this paper proposes to employ a cross-modal attention mechanism. Firstly, it calculates the multi-head relationship matrix after merging, splitting, and dot product, and secondly, it utilizes the relative entropy for knowledge distillation, so that

the model can better align the intermediate features between heterogeneous teachers and students. Meanwhile, this paper also applies the biaffine attention and adversarial weight perturbation function to enhance the learning of multi-feature knowledge such as semantic phonology and word form word meaning. Compared with the mainstream model, the precision, recall and F_1 value of the FSFDAI-TMRD method are improved by 0.45%, 3.96% and 2.28%, respectively. The storage optimization ratio is improved by 3.01% and the model parameter size is reduced by 94.86%.

Key words: knowledge distillation; fish disease named entity recognition; adversarial training; biaffine attention; model compression

引 言

中国是世界公认的渔业大国。据联合国粮食及农业组织报道,2020年以来水产养殖产品占全球总产量的49%,占人类鱼类食用量的56%。在未来10年内,养殖鱼类将占据更多的消费比例和贸易份额^[1]。为确保中国渔业资源安全,智慧鱼病专家系统建设是促进渔业高质量发展的迫切需要,是智慧渔业应用的新方向^[2]。鱼病文本数据是构建领域专家知识库的核心支撑,是推动鱼病知识图谱、智慧专家系统和决策方案辅助等传统下游任务不可或缺的要素来源。

命名实体识别任务(Named entity recognition,NER)作为挖掘海量文本事实知识的关键步骤,旨在准确、全面地从多源异构文本中抽取出实体概念^[3],诸如鱼病症状、病原菌特征及寄生虫特征等。近几年,由Google AI研究院提出无监督基于转换器的双向编码表征(Bidirectional encoder representation from transformers,BERT)^[4]和一系列大规模预训练语言模型(Large language models,LLMs)大放异彩,FGM^[5]、SMART^[6]及AWP^[7]等对抗权重干扰方法引入和应用^[8-10],成为目前自然语言处理的研究新范式^[11-14]。为充分提取汉字的多模态知识,综合增强模型识别效果,Zhang等^[15]侧重应用汉字的部首偏旁和语调语音特征,主体设计概率分布模型,通过特征相关序列捕捉汉语内在知识,融入中文字词嵌入。Sun等^[16]侧重使用汉字的拼音字形特征,重点结合预训练语言模型BERT、混合字词掩码策略和卷积神经网络(Convolutional neural network,CNN)池化层,通过改进叠加底层融合嵌入,提高重音消歧能力。陈梦莹等^[17]侧重利用汉字语素的词义特征,核心构造双仿射注意力机制增强关联前后向语素语义和融合学习跨实体词级表征,进而缓解训练梯度消失缺陷,解决语素标签预测难题。

NER作为LLMs和专家系统的核心任务,通常模型存储较大,为消减其训练资源开销极大^[18-19]。如何针对鱼病领域,本地构建和部署存储空间低、样本规模小且精度识别高的NER模型,将成为当下亟待解决的挑战性难题^[20]。

知识蒸馏(Knowledge distillation,KD)作为复杂模型压缩和训练架构迁移的崭新举措,其教师-学生结构充分应用“温度”调控及“标签”判断等超参调整思想,形象化地将特征知识从复杂神经网络模型“蒸馏”到规模小且存储低的网络中。为充分蒸馏教师的多任务知识^[21],综合增强学生模型识别效果,胡滨等^[22]侧重改进KD结构的损失函数,面向生物医药领域紧密依托Transformer模型,动态更新权重,构建FastBioNer模型高效拟合重要知识。在此基础上,Jiao等^[23]设计二段式TinyBERT学习框架,前者定义映射函数实现师生模型相互索引,后者精调参数矩阵线性更迭网络隐藏状态,为特定任务蒸馏下运用细粒度知识、拓宽数据差异性提供有力保障。Liang等^[24]构建数据增强实例选择,引入多轮驱动搭建基于强化学习的迭代知识蒸馏(Reinforced iterative KD,RIKD),有力缓解师生蒸馏过程的冗余数据。Zhou等^[25]注重解决学生汲取知识能力不强的问题,优化K维特比算法,改善底层输出矩阵,综合压实压紧KD,构筑了更紧凑、更高效的BiLSTM学生模型。Moscatto等^[26]优化KD架构迁移学生多任务知识,

确保高精确度并极大减少内存容量。

然而,目前结合KD方法的鱼病垂直领域NER模型面临以下两种难题:(1)特征表述多,识别精确度低。鱼病垂直领域NER模型存在实体识别精确度偏低、特征表述能力不强及实体边界鉴定困难等诸多痛点。例如,“斑点叉尾鮰病毒病”属于常见鱼病类型,而“斑点叉尾鮰病毒”属于病原菌实体。为解决上述问题,传统NER模型或局限于字符向量的视觉特征研究,或局限于字符向量的拼音特征研究,或局限于字符向量的语义特征研究,使得细粒度的鱼病字符向量特征知识缺乏合理安排及统筹规划,多任务的鱼病NER模型学习能力、实体分类精度存在进一步的提升空间^[27-30]。(2)泛化性能弱,蒸馏损失大。传统KD方法多数着眼于教师-学生模型的“代沟”难题,通过自蒸馏、相互蒸馏和知识合并等微调KD架构,以此整体收缩教师模型差异,减少蒸馏损失规模,从而提高学生模型的泛化性能。但是上述方法增加了部署模型的冗余数据,抑制了模型运算的迁移能力,不利于大幅提高鱼病NER低存储模型的鲁棒性和泛化性^[31-34]。

针对上述挑战,本文工作归纳如下:(1)面向鱼病垂直领域标注数据稀疏,构建了细粒度、高质量的鱼病本体库。由于目前缺乏成熟、全面且高水平的鱼病中文标注数据集,本研究根据上海海洋大学专家指导意见,广泛收集高质量鱼病文本,由专家进行人工标注,高质量地保障生成数据集的科学性和权威性。(2)面向鱼病领域识别精度低、泛化性能弱的痛点难点,提出了一种融合多特征协同预测框架、对抗权重干扰函数(Fusion of synergistic feature distillation and anti-weight interference,FSFDAI)和跨模态多头蒸馏方法(Teaching assistant multi-head relationship distillation,TMRD)。为此,本文主体设计了多特征协同预测-跨模态多头蒸馏方法,命名为FSFDAI-TMRD。

FSFDAI-TMRD方法既改进了原多任务多特征协同预测架构,修改了原架构下的联合概率分布计算方式,由更细粒度的BMES(Begin-middle-end-single)法代替原工作中BIO(Begin-inside-outside)法的粗略标注,又构建了跨模态多头蒸馏,运用跨模态注意力机制,额外蒸馏通过合并、拆分和点积后的多头关系矩阵。同时,利用对抗性权重扰动函数(Adversarial weight perturbation,AWP)和双仿射注意力机制等框架,结合字音、字形、词义、词性及语义等多特征知识预测结果,微调学生模型的蒸馏误差,以字符嵌入和特征权重为双扰动机制,实现蒸馏主体映射双向长短期记忆轻量级网络(Bi-directional long short-term memory,BiLSTM)。

1 FSFDAI-TMRD 方法

如图1所示,FSFDAI-TMRD方法分为3个阶段:单任务学习阶段、多任务学习阶段和知识蒸馏阶段。

1.1 单任务学习阶段

1.1.1 单任务语义和词义特征学习

一方面,针对鱼病领域词义特征学习,架构了双仿射注意力机制(Biaffine attention,BA)。作为中文语素间级联交互的崭新举措,BA拼接可分类头尾表征,分析依存句法关系,提高增强词义知识,将有力解决识别精确低、实体嵌套多及边界鉴定难等诸多痛点。首先,BA机制经由两层多层感知机(Multilayer perceptron,MLP)处理BERT生成矢量,嵌合头(Head)、尾(Tail)实体Token形成索引对Spans,为充分运用注意力手段、加强语素间词义联系提供关键支撑。其次,Biaffine矩阵依次相乘头Token层、尾Token层,生成式词义关系矩阵 φ 形式为 $[\text{Batch}_{\text{size}}, \text{Seq}_{\text{len}}, \text{Seq}_{\text{len}}, \text{Num}_{\text{label}}]$,紧密依赖嵌入隐层特征和句法分类标签,计算区间索引对Span得分。最后,人工定义全面整理、排序候选区间功能,引入交叉熵损失函数返回分数矩阵 r^{Span} ,实现词义特征学习,计算公式为

$$h_{\text{start}} = \text{MLP}_{\text{head}}(x_{\text{start}}) \quad (1)$$

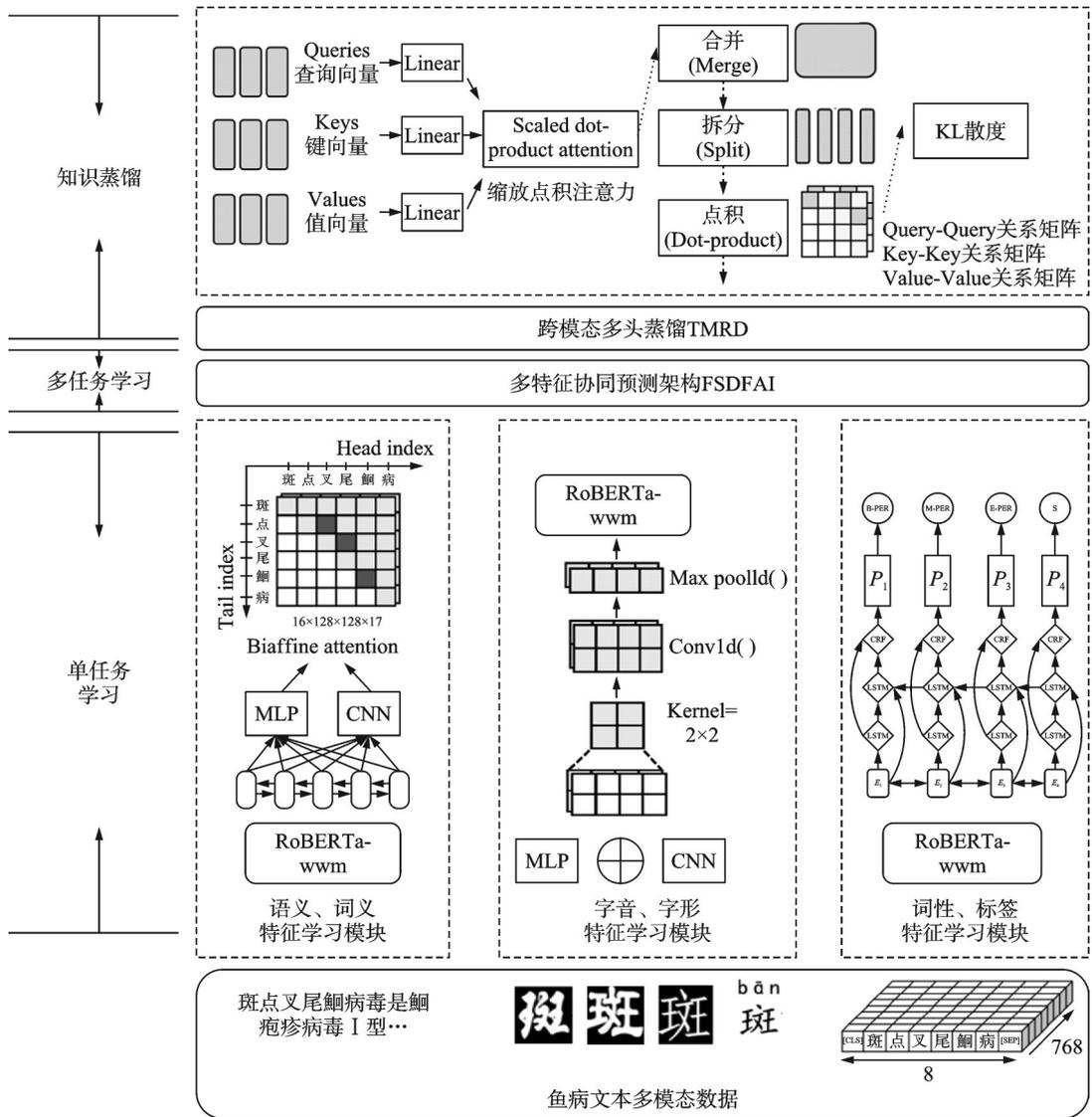


图1 FSFDAI-TMRD方法框架

Fig.1 Framework of FSFDAI-TMRD method

$$h_{end} = \text{MLP}_{tail}(x_{end}) \tag{2}$$

$$r_m^{Span} = h_{start}^T U^{(1)} h_{end} + (h_{start} \oplus h_{end}) U^{(2)} + b \tag{3}$$

式中: x_{start} 、 x_{end} 表示为第 i 个索引对 Span 的头和尾; h_{start} 、 h_{end} 表示该索引对 Span 的头尾表征; $U^{(1)}$ 表示 $2d \times 2d \times 2d$ 的张量, $U^{(2)}$ 表示为 $4d \times 2d$ 的矩阵, d 为样本输入维度; \oplus 是异或运算; b 为偏置。图 2 为双仿射注意力机制算法流程。

另一方面,针对鱼病领域语义特征学习,本文采用 RoBERTa-wwm-ext-large 结构,其优势在于以下两方面:(1)更贴切的全词掩码策略。由于中文不存在词根现象,相比英语 WordPiece 切词法,中文语素更适合动态全词掩码(Whole word masking, WWM),切实满足少量 Token 聚焦词级别鱼病语义。(2)更出色的鲁棒模型性能。RoBERTa-wwm 模型参数多、批量规模大、训练数据广,其结构微调 BERT 超

参,裁剪语句预测(Next sentence prediction, NSP)任务,突出更高数量级的文本编码,堆叠 24 层 Transformers 实现泛化更强、性能更好的鱼病语义学习能力。图 3 为 RoBERTa-wwm-ext-large 算法流程。

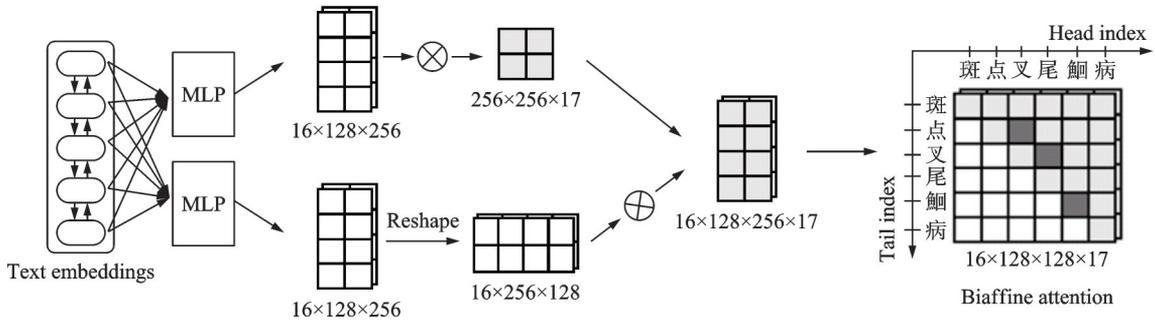


图 2 双仿射注意力机制算法流程

Fig.2 Flow of the biaffine attention algorithm

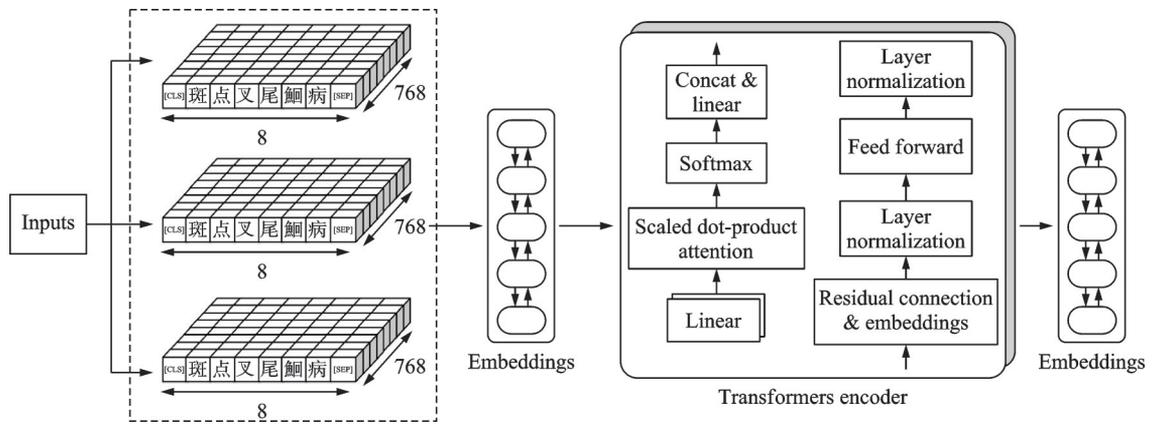


图 3 RoBERTa-wwm-ext-large 算法流程

Fig.3 Flow of the RoBERTa-wwm-ext-large algorithm

1.1.2 单任务字音和字形特征学习

在鱼病垂直领域语料文本中,典型的有“斑点叉尾鲷病毒病”。“鲷”字不仅拥有丰富的字形象征意义,例如“鱼”部首偏旁指代字符标识;还具备罗马拼音规则,并且汉字四声语调“hui”作为补充标记。

一方面,为深入理解鱼病的字形视觉信息,本文运用 MLP 压实展平嵌合华文仿宋、方正古隶和华文行楷等多种字体字形向量。由于官方字库特征鲜明,首先,调整字形图像规模为 24×24 ,限定浮点像素范围为 $0 \sim 255$ 。其次,经由全连接层铺平 $24 \times 24 \times 3$ 维矩阵,线性映射样本标识空间。最后,通过全连接层(Full connect, FC)计算 $21\ 128 \times 1\ 728$ 大小的张量,导出字形嵌入向量 W_0 。图 4 为鱼病字形学习算法流程。

另一方面,为充分学习鱼病的语音语调知识,本文采用开源的、结合字典启发式规则的第三方库 pypinyin 注释汉语拼音序列,消歧一字多义现象。首先,推理字符在鱼病上下文的切实含义,定义消息函数更新拼音序列,添加特殊符号表示四声语调。其次,在序列基础上叠加

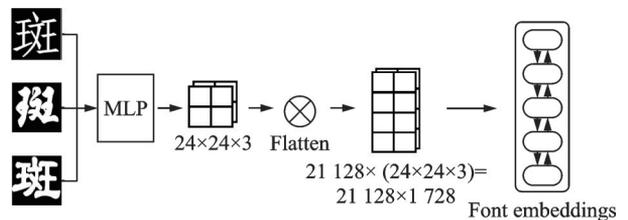


图 4 鱼病字形学习算法流程

Fig.4 Flow of the fish disease font learning algorithm

CNN,全面提取语音特征。最后,经由池化层导出固定8长度的语音嵌入向量 W_1 。图5为鱼病字音学习算法流程。

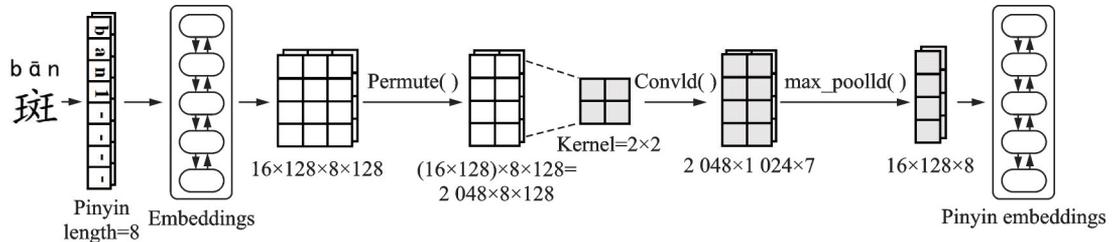


图5 鱼病字音学习算法流程

Fig.5 Flow of the fish disease pinyin learning algorithm

1.1.3 单任务词性和标签特征学习

一方面,针对鱼病领域词性特征学习,构建了双向长短期记忆网络BiLSTM。鱼病词性特征紧密联系语义上下文,某一时刻经由基于前向LSTM_F、后向LSTM_B组成的BiLSTM层,按照时间顺序捕获、交流、加强正逆序语义知识,互补输出当前词性特征。图6为BiLSTM内部结构。图中:Forget_t、Input_t、Cell_t、Output_t分别表示遗忘门、输入门、细胞状态、输出门在某一时刻t的输出值;x_t表示当前时刻登录字词;h_{t-1}、h_t分别表示前一时刻和当前时刻的隐藏状态;ĉ_t表示临时细胞状态;σ、tanh分别表示Sigmoid和Tanh激活函数。

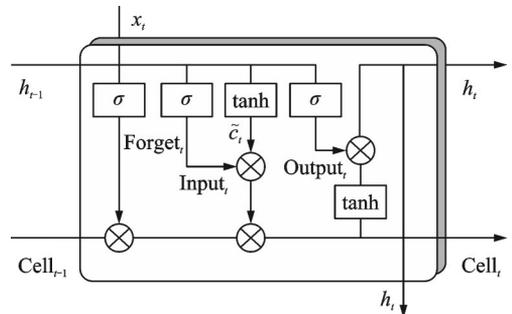


图6 BiLSTM网络内部结构

Fig.6 Internal structure of BiLSTM network

另一方面,针对鱼病领域标签特征学习,建模了条件随机场(Conditional random field, CRF)。该结构既全局优化标签约束关系、学习训练转移概率矩阵A,充分联系实体上下文语义解决标签偏置,又整体采用对数极大似然估计L(θ|X)、Softmax归一化运算预测概率序列P,综合运用概率最大分数、维特比算法(Viterbi)动态选取最优路径、规划输出最佳序列。得分公式为

$$Score(X, Y) = \sum_{i=0}^n A_{y_i, y_{i+1}} + \sum_{i=1}^n P_{i, y_i} \tag{4}$$

式中:A_{y_i, y_{i+1}}为转移概率矩阵A中标签y_i和标签y_{i+1}之间的线性映射;P_{i, y_i}为预测概率序列P中字符i和标签y_i之间的对应关系。Score(X, Y)是结合鱼病文本X与标签序列Y,计算得出的概率分数。

1.2 多任务学习阶段

受Moscato等^[26]工作的启发,证明了多任务、多特征协同预测架构在迁移学生知识和优化KD架构的同时,拥有极大的减少内存存储量、确保高精确度识别的能力。而在鱼病垂直领域语料文本中,为更好地解决边界鉴定难及特征表述弱的嵌套名词实体,本文转由更细粒度的BMES法标注,生成多教师预测序列,代替原工作中的BIO法粗略标注,并依据此修改了联合概率分布计算方式,改进了多教师多任务协作预测架构的整体识别精度。为便于理解,表1给出了本节使用的符号代表及其释义。

(1)问题描述

为规范抽象多任务多特征协同预测任务,假设鱼病垂直领域的第i个鱼病实体e_i隶属于实体集合Ē,域值为i∈{1, 2, ..., |Ē|}。同时规定鱼病语料库数据集为Θ, X_i表示针对鱼病文本输入句子集X第i个

字符嵌入 Tokens, 域值为 $X_i \in X$ 且 $i \in \{1, 2, \dots, |\omega|\}$, ω 表示为鱼病文本输入句子集 X 的最大长度。 Y_i 表示 Tokens 对应实体标签, 定义如下

$$\Theta = \{(x, y) \in X_i \times Y_i\} \quad (5)$$

BMES 法的四位标注方式拥有更好地表述细粒度的鱼病实体边界的能力。 B 标识词首位置、 M 标识中间位置、 E 标识词尾位置、 S 标识单独字词。 例如面向鱼病文本“病原是斑点叉尾鲷病毒”, 其标注序列即为 $\{S, S, S, B, M, M, M, M, M, E\}$ 。 多任务多特征协同预测架构的目的在于建立自适应的函数学习模型 $f(\cdot)$, 参照句子集 X 中的 X_i 直接映射 Y_i 表示相应实体标签, 其 NER 样本输入形式为

$$T^{\text{Mutill}} = \left\{ \begin{array}{l} B - e_1, M - e_1, E - e_1, \\ B - e_2, \dots, E - e_{|\bar{E}|}, S \end{array} \right\} \quad (6)$$

具体来讲, 假定有常见鱼病 CFD、病原菌名称 PBN 实体, 则 $\bar{E} = \{e_1, e_2\} = \{\text{CFD}, \text{PBN}\}$, 那么 FSFDAI 对应样本输入标识形式为

$$T^{\text{Mutill}} = \{B\text{-CFD}, M\text{-CFD}, E\text{-CFD}, B\text{-PBN}, M\text{-PBN}, E\text{-PBN}, S\} \quad (7)$$

(2) 解决方法

多任务学习、多特征协同和多教师协作预测的基础要义是单任务教师各自面向独立分布的 Θ 进行文本标签预测并计算其概率分布。 为更好地规范定义其概率预测过程, 需形象化描述单任务教师的离散分布空间, 表达式为

$$T_i = \{Y_i = k | X; \vartheta_T^1, \vartheta_T^2, \dots, \vartheta_T^{\bar{E}}\} \quad k \in \{B, M, E, S\} \quad (8)$$

式中: T_i 表示为第 i 个单任务教师预测概率分布; $\vartheta_T^1, \vartheta_T^2, \dots, \vartheta_T^{\bar{E}}$ 表示单任务教师学习参数 ϑ_T 与实体集合 \bar{E} 的隶属关系。

由于多教师的预测概率具备独立性和互斥性, 其既满足联合概率分布的运算条件, 又契合权重动态更迭的过程需要。 假定鱼病文本输入 Tokens 的 $X_i \in X$ 且 $i \in \{1, 2, \dots, |\omega|\}$, 那么为聚合该 Tokens 的实际概率, 以式(8)计算 T_i 和 T_j 的单任务教师相互交集形成式(9~12)。

$$P_B = P(B - e_i) = P((T_i \in B - e_i) \cap_{i \neq j} (T_j \notin B - e_i)) \quad (9)$$

$$P_M = P(M - e_i) = P((T_i \in M - e_i) \cap_{i \neq j} (T_j \notin M - e_i)) \quad (10)$$

$$P_E = P(E - e_i) = P((T_i \in E - e_i) \cap_{i \neq j} (T_j \notin E - e_i)) \quad (11)$$

$$P_S = P(S) = P((T_i \in S) \cap_{i=j} (T_j \in S)) \quad (12)$$

式中 P_B, P_M, P_E, P_S 分别表示 BMES 法标注实体的多教师预测概率。

(3) 基于 BMES 法的联合概率分布计算公式

为充分学习异构的多教师知识, 聚合其预测概率输出分布, 相较于传统标准方法, 多任务多特征协同预测架构的关键在于生成的融合教师结构能够学习多特征, 聚合离散分布空间的、异构个性化的教师不同实体类型预测, 转向平面的、统一的且用于后续跨模态知识蒸馏的联合概率分布计算。 聚合公式如下

$$\text{Union}_{i,j} (y_j = B - e_j | X_i; \vartheta_T^1, \vartheta_T^2, \dots, \vartheta_T^{\bar{E}}) = P_B^i \prod_{i \neq j} (P_M^j + P_E^j + P_S^j) \quad (13)$$

表 1 本节使用的符号代表及其释义

Table 1 Symbols used in this subsection represent and interpret

符号	释义
\bar{E}	鱼病实体集合
e_i	第 i 个鱼病实体
Θ	鱼病语料库、数据集
X	鱼病文本输入句子集
X_i	第 i 个字符嵌入 Tokens
Y_i	Tokens 对应实体标签
ω	输入句 X 的最大长度
$f(\cdot)$	函数学习映射模型
T_i	第 i 个单任务教师预测概率分布
ϑ_T	单任务学习参数
$\vartheta_T^{\bar{E}}$	ϑ_T 参数与 \bar{E} 实体的隶属关系

$$\text{Union}_{i,j}(y_j = M - e_j | X_i; \vartheta_T^1, \vartheta_T^2, \dots, \vartheta_T^E) = P_M^i \prod_{i \neq j} (P_B^i + P_S^i) \quad (14)$$

$$\text{Union}_{i,j}(y_j = E - e_j | X_i; \vartheta_T^1, \vartheta_T^2, \dots, \vartheta_T^E) = P_E^i \prod_{i \neq j} (P_B^i + P_S^i) \quad (15)$$

$$\text{Union}_{i,j}(y_j = S | X_i; \vartheta_T^1, \vartheta_T^2, \dots, \vartheta_T^E) = \prod_i P_S^i \quad (16)$$

(4) 举例说明

假设鱼病 NER 任务识别 {CFD, PBN} 两大实体类型, 则面向文本“水霉病菌”, 其序列是 {B-CFD, M-CFD, E-CFD, S}, 也可能是 {B-PBN, M-PBN, M-PBN, E-PBN}。针对字符嵌入“水”, 教师 A 预测为 CFD 实体, 其概率序列为 {0.7, 0.1, 0.1, 0.1}; 教师 B 预测为 PBN 实体, 其概率序列为 {0.1, 0.05, 0.05, 0.8}。根据式 (13~16), 多教师概率聚合为 {0.63, 0.09, 0.09, 0.03, 0.04, 0.04, 0.08}, 其实体标签路径是 {B-CFD, M-CFD, E-CFD, B-PBN, M-PBN, E-PBN, S}。换言之, “水”63% 概率的实体类型标签为 {B-CFD}, 其作为监督信号蒸馏、教导后续训练。

1.3 知识蒸馏阶段

TMRD 为进一步结合鱼病名词、文本特征和权重分配, 运用跨模态注意力和更细粒度监督方式, 减少可能引入的噪音。前向计算时, 添加多头自注意力层, 取代原师生最后隐层的传统蒸馏方式, 转由计算查询向量 Query-Query 关系矩阵 Q 、键向量 Key-Key 关系矩阵 K 、值向量 Value-Value 关系矩阵 V , 通过 KL 散度进行知识蒸馏。反向传播时, 通过获取运算梯度变化、更迭运行参数和优化器、动态更新模型权重以及清空梯度迭代过程数据, 以期减少残差错误, 反馈软目标, 促使学生模型精度更佳。同时利用对抗性权重扰动函数 AWP, 攻击前后向传播计算过程, 以字符嵌入和特征权重为双扰动机制, 扩充对抗样本, 以更好地微调训练任务。图 7 为跨模态多头蒸馏算法流程。

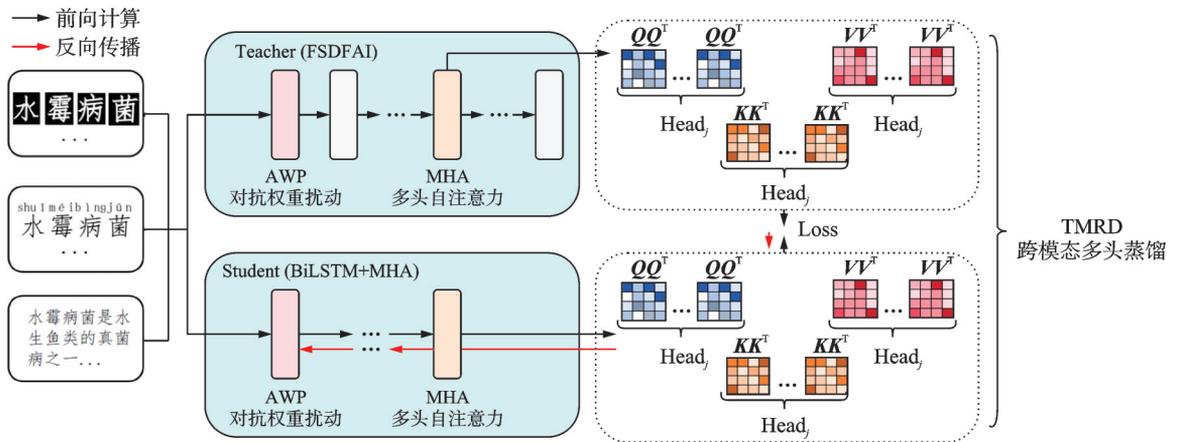


图 7 跨模态多头蒸馏算法流程

Fig.7 Flow of the cross modal multi-head distillation algorithm

跨模态注意力是指, Q 、 K 、 V 三维度叠加处理的输入矩阵经由 Softmax 归一化、多头组合拆分及线性投影空间等运算步骤, 捕捉语句间的字词权重。按照 FSDFAI 头部数 h 设置, 人为定义关系矩阵总头部数 H_R (Relation_heads), 并依据此拆分 TMRD 的生成向量, 分别形成新的 Q 、 K 、 V 矩阵, 确保关系矩阵维度相同。根据式 (17~19) 可计算 Query-Query、Key-Key、Value-Value 关系矩阵。其中, $Q_{i,j}$ 、 $K_{i,j}$ 、 $V_{i,j}$ 对应 TMRD 新生成的 Q 、 K 、 V 矩阵, d_R 表示关系矩阵的维数。

$$\mathbf{R}_{i,j}^{\text{QQ}} = \text{Softmax} \left(\frac{\mathbf{Q}_{i,j} \mathbf{Q}_{i,j}^{\text{T}}}{\sqrt{d_{\text{R}}}} \right) \quad (17)$$

$$\mathbf{R}_{i,j}^{\text{KK}} = \text{Softmax} \left(\frac{\mathbf{K}_{i,j} \mathbf{K}_{i,j}^{\text{T}}}{\sqrt{d_{\text{R}}}} \right) \quad (18)$$

$$\mathbf{R}_{i,j}^{\text{VV}} = \text{Softmax} \left(\frac{\mathbf{V}_{i,j} \mathbf{V}_{i,j}^{\text{T}}}{\sqrt{d_{\text{R}}}} \right) \quad (19)$$

TMRD 基于跨模态注意力,根据式(20~24)计算彼此间的 KL 散度以及相应损失函数。其中: L_{KL} 为 KL 散度, Loss 为总损失函数, Z^{T} 和 Z^{S} 分别对应教师和学生模型输出的逻辑值, $\mathbf{R}_{\text{Student},j}^{\text{KK}}$ 表示学生第 j 个头部计算的 Key-Key 关系矩阵, $\mathbf{R}_{\text{Student},j}^{\text{QQ}}$ 表示学生第 j 个头部计算的 Query-Query 关系矩阵, $\mathbf{R}_{\text{Student},j}^{\text{VV}}$ 表示学生第 j 个头部计算的 Value-Value 关系矩阵。同理, $\mathbf{R}_{\text{Teacher},j}^{\text{KK}}$ 表示老师第 j 个头部计算的 Key-Key 关系矩阵, $\mathbf{R}_{\text{Teacher},j}^{\text{QQ}}$ 表示老师第 j 个头部计算的 Query-Query 关系矩阵, $\mathbf{R}_{\text{Teacher},j}^{\text{VV}}$ 表示老师第 j 个头部计算的 Value-Value 关系矩阵。

$$L_{\text{KL}} = \text{KL} \left(\log(\text{Softmax}(Z^{\text{S}})), \text{Softmax}(Z^{\text{T}}) \right) \quad (20)$$

$$L^{\text{KK}} = \frac{1}{H_{\text{R}}} L_{\text{KL}}(\mathbf{R}_{\text{Student},j}^{\text{KK}}, \mathbf{R}_{\text{Teacher},j}^{\text{KK}}) \quad (21)$$

$$L^{\text{QQ}} = \frac{1}{H_{\text{R}}} L_{\text{KL}}(\mathbf{R}_{\text{Student},j}^{\text{QQ}}, \mathbf{R}_{\text{Teacher},j}^{\text{QQ}}) \quad (22)$$

$$L^{\text{VV}} = \frac{1}{H_{\text{R}}} L_{\text{KL}}(\mathbf{R}_{\text{Student},j}^{\text{VV}}, \mathbf{R}_{\text{Teacher},j}^{\text{VV}}) \quad (23)$$

$$\text{Loss} = L^{\text{KK}} + L^{\text{QQ}} + L^{\text{VV}} \quad (24)$$

2 实验与分析

2.1 实验准备

本文实验使用英伟达 GPU 针对鱼病 NER 任务进行硬件加速,其实验配置如表 2 所示。

由于鱼病垂直领域缺乏成熟、全面且高水平的中文标注数据集,本实验根据上海海洋大学专家指导意见,紧密围绕斑点叉尾鲷这一典型鱼种鱼病作为本体,广泛收集高质量鱼病文本,包括知网论文、推荐书籍和百科数据等,通过 Linux 虚拟机充分运用 BRAT 软件,并交由专家进行人工标注,极大地保障了生成数据集的科学性和权威性。

为减少过拟合现象,提高样本鲁棒性,该数据集全面整合数据增强 BMES 标注方法,共涵盖 17 种实体类型、18 622 个实体、244 978 个字符、4 585 条样本,充分囊括细粒度的鱼病实体类型。训练集、验证集、测试集划分比为 8:1:1。鱼病数据集不同实体类型和统计分析如表 3 所示。

2.2 衡量指标与参数设置

本实验衡量指标采用 F_1 指标 (F_1 score)、精确率 (Precision) 和召回率 (Recall),如式 (25~27) 所示^[3,28]。

表 2 实验软硬件配置

Table 2 Experimental hardware and software configuration

类型	配置	参数
软件	语言	Python 3.6
	环境	Pytorch 1.13.1
	开发	Pycharm
	显卡	RTX 3080
硬件	处理器	i7-10780H
	内存	32 GB
	硬盘	SSD 2 TB

表3 鱼病数据集实体类型及数理统计

Table 3 Entity types and mathematical statistics of fish disease dataset

实体类别	训练集	验证集	测试集	总计
常见鱼病	1 065	132	135	1 332
病原菌名称	1 638	205	205	2 048
病原菌特征	310	38	40	388
寄生虫名称	200	25	25	250
寄生虫特征	188	23	24	235
分布区域	985	123	124	1 232
疾病特征	687	85	87	859
病因分析	758	94	96	948
病鱼行为	1 865	233	234	2 332
解剖病鱼	1 166	145	147	1 458
实验名称	400	50	51	501
药品名称	2 105	263	268	2 636
实验方式	141	18	17	176
给药方式	931	116	117	1 164
给药频次	625	78	78	781
给药剂量	798	100	100	998
预防措施	1 027	128	129	1 284
合计	14 889	1 856	1 877	18 622

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \quad (25)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \quad (26)$$

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (27)$$

式中:TP为预测正确的正样本标签数;FP为预测正确的负样本标签数;FN为预测错误的负样本标签数。

本实验训练过程紧密依托超参微调,不断优化Adam算法,参照线性热身(Linear warmup)进行迭代,以期实现更好的实验结果,设置参数如表4所示。

表4 实验参数设置

Table 4 Experimental parameter settings

参数	值	参数	值
Learning-rate(RoBERTa-wwm)	2e-5	train_epochs	15
Learning-rate(BiLSTM、MLP)	2e-3	eval_batch_size	32
T	3	dropout	0.3
max_grad_norm	1	dropout_probability	0.1
warmup_proportion	0.1	max_seq_len	128
weight_decay	0.01	swa_start	3
adam_epsilon	1e-8	lstm_hidden	128
train_batch_size	16	lstm_num_layers	1

本文参考文献[15-17, 28-30]的工作,通过反复实验从而确定最佳参数设置。实验中存在4种对FSFDAI-TMRD识别效果影响较大的参数。(1)学习率(Learning-rate)。学习率过高,模型难收敛;学习率过低,模型收敛慢且梯度会损失。根据文献[15-17],RoBERTa-wwm预训练模型学习率设置为 $2e-5$,这是因为过高的学习率会导致权重震荡。FSFDAI-TMRD模型的学习率如图8(a)所示,选取Learning-rate= $2e-3$ 是因为模型识别效果较好,并且过低的学习率会造成梯度损失。(2)批处理尺寸(Train_batch_size)。批处理尺寸越高,模型识别越好。但是批处理尺寸过高,模型可能会导致内存溢出或陷于局部最优而非全局最优。如图8(b)所示,Train_batch_size=16时,FSFDAI-TMRD模型识别效果较好。(3)训练轮数(Train_epochs)。训练轮数越多,模型识别越好。但是训练轮数达到临界值后,模型识别会趋于过拟合,效果反而下降。如图8(c)所示,当Train_epochs=15时,FSFDAI-TMRD模型 F_1 值较高且趋于收敛,证明模型识别效果较好。(4)温度超参数(T)。温度参数越高,标签预测越平滑,学生模型学习效果越好。但是温度参数过高,标签预测趋同,效果反而下降。如图8(d)所示,当 $T=3$ 时,FSFDAI-TMRD模型精确率、召回率、 F_1 值相较于其他温度情况,证明其概率分布趋近平滑,能最大化地保留细节特征。

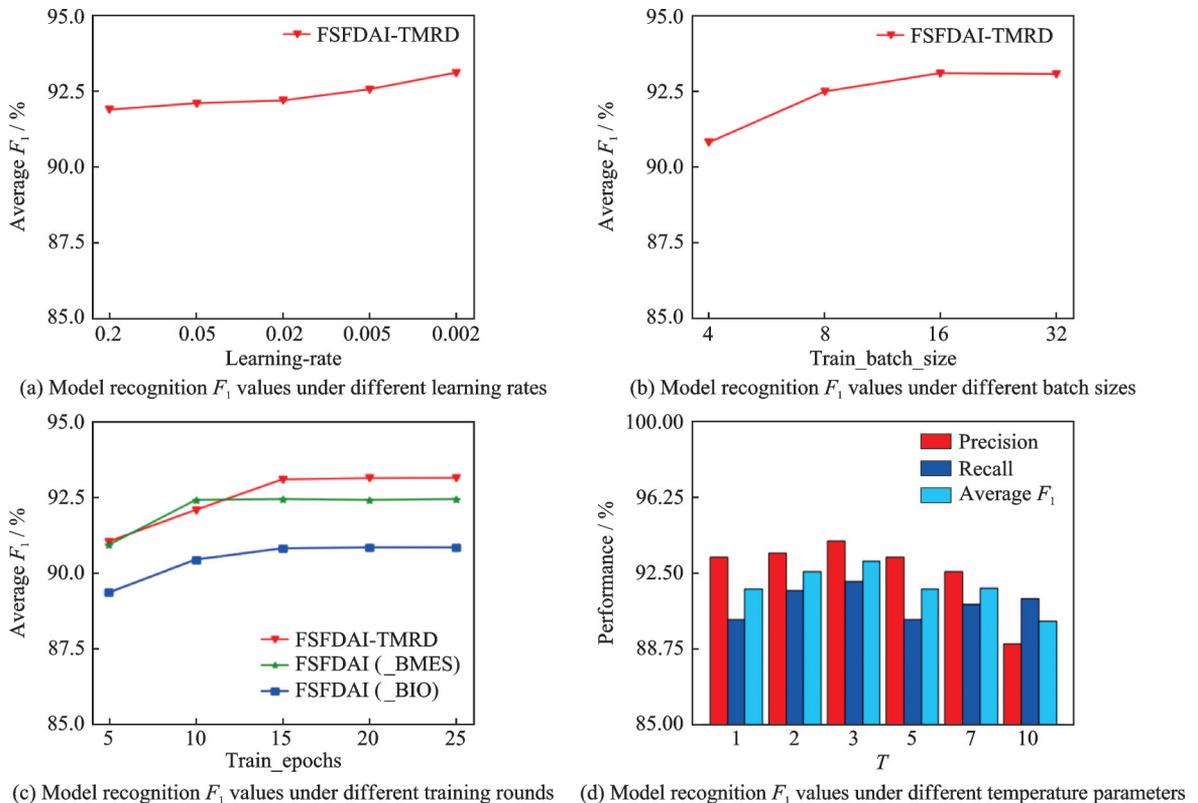


图8 实验参数对比结果

Fig.8 Comparison results of experimental parameters

2.3 教师模型识别结果与分析

由于鱼病垂直领域数据集呈现多重重叠和易混淆特点,本文以BERT及CRF为基线模型,考察FSFDAI教师算法的实用性。对比结果如表5所示。由表5可以看出:(1)与未充分运用鱼病上下文的字音、字形、词义等多特征学习的识别模型对比。BiLSTM-CRF受限于网络结构和表征匮乏,精确率、召回率和 F_1 值整体效果不佳,分别为76.58%、72.63%和74.65%。BERT-BiLSTM-CRF模型以一定的识

别精度为代价,改善鱼病嵌套实体识别,精确率为81.14%,相较提高8.51%; F_1 值为78.09%,相较提高3.44%。RoBERTa-wwm的精确率、召回率和 F_1 值分别为86.74%、88.24%和87.48%,相较提高11.48%、7.1%和9.39%。(2)与未充分运用鱼病的语义关联、词性捕捉及标签依赖等多知识学习的模型对比。ChineseBERT模型通过学习鱼病文本的字音和字形特征,极大地增强了实体识别精度,精确率、召回率和 F_1 值分别为83.71%、78.45%和81.00%。RoBERTa-wwm-MLP-Biaffine模型另辟蹊径,有力地解决了识别精确低、实体嵌套多及边界鉴定难等诸多痛点,其精确率为90.72%,相较提高7.01%, F_1 值为87.85%,相较提高6.85%。但RoBERTa-wwm-MLP-Biaffine模型未能全面利用鱼病上下文的语义关联等多知识学习,整体召回率相对偏低,为85.16%。

整体来看,本文FSFDAI模型既综合采样鱼病上下文的字音、字形及词义等多特征学习,又充分实现语义关联、词性捕捉和标签依赖等多知识学习, F_1 值分别提高了17.82%、14.38%、11.47%、4.99%和4.62%,证明了本文方法在鱼病垂直领域优于主流对比模型的性能,多教师、多特征协作学习可以更有效且更合理地捕捉鱼病文本上下文信息。

2.4 学生模型识别结果与分析

为顺利推动小规模样本的高精度鱼病NER低存储模型部署,构建更紧凑高效的学生模型,考察BiLSTM结构,其对比实验如表6所示。由表6可见:鱼病NER低存储模型的识别结果取决于BiLSTM学生结构。BiLSTM模型作为整体实验参照,精确率低、召回率差、 F_1 值整体较低,分别为76.38%、71.50%和73.86%。FSFDAI-TMRD整合了鱼病文本中语义和词义学习、字音和字形学习、词性和标签学习的多特征优点,并且

相较于传统蒸馏方法, F_1 值分别提高了19.22%、14.64%、7.17%、6.57%和2.28%,证明了本文方法增加的跨模态多头蒸馏以及蒸馏训练过程中的对抗学习在鱼病垂直领域实体识别的有用性。

一方面,FSFDAI-TMRD蒸馏后的BiLSTM学生模型在识别嵌套名词实体表现较好。学生模型部分识别结果对比如图9所示。其中,图9(a)作为参照组,即未蒸馏的BiLSTM学生模型只能识别出地点(美国),但是无法识别出鱼病实体、病原菌实体;图9(b~d)作为对比组,即BERT-BiLSTM-CRF、ChineseBERT、RoBERTa-wwm-BiLSTM-CRF蒸馏后的BiLSTM学生模型,面对嵌套名词实体,容易错误将“病原菌实体”(斑点叉尾鲷病毒)识别为“鱼病实体”(斑点叉尾鲷病);图9(e~g)作为实验组,即RoBERTa-wwm-MLP-Biaffine-KD、FSFDAI-KD、FSFDAI-TMRD蒸馏后的BiLSTM学生模型,可以准确识别并标注出地点、鱼病及病原菌等实体。

表5 本文方法与现有实验对比结果

教师模型	精确率	召回率	F_1 指标	%
BiLSTM-CRF	76.58	72.63	74.65	
BERT-BiLSTM-CRF	75.26	81.14	78.09	
ChineseBERT(pinyin+font)	83.71	78.45	81.00	
RoBERTa-wwm-BiLSTM-CRF	86.74	88.24	87.48	
RoBERTa-wwm-MLP-Biaffine	90.72	85.16	87.85	
FSFDAI(ours)	91.62	93.35	92.47	

表6 学生模型BiLSTM实验对比结果

学生模型	精确率	召回率	F_1 指标	%
BiLSTM(MHA)	76.38	71.50	73.86	
BERT-BiLSTM-CRF-KD	79.62	77.28	78.44	
ChineseBERT-KD	89.55	82.56	85.91	
RoBERTa-wwm-MLP-Biaffine-KD	89.66	83.57	86.51	
RoBERTa-wwm-BiLSTM-CRF-KD	93.64	88.13	90.80	
FSFDAI-KD(ours)	93.21	91.60	92.40	
FSFDAI-TMRD(ours)	94.09	92.09	93.08	



图9 学生模型部分识别结果对比

Fig.9 Comparison of some students' model recognition results

另一方面,FSFDAI-TMRD在实验名称和药品名称等原数据集数据稀疏分布、标注序列不规范的类别预测下仍有不足:中英文混杂名称较多、药品俗名歧义较复杂,例如“三黄粉(中草药)”“三黄散(主要成分大黄、黄芩、黄柏)”“Trics-HCL(PH8.8)”等,预测类别 F_1 值相对偏低,需要在后续实验加强研究,寻找更好的方法。

为进一步验证FSFDAI-TMRD模型的实用性,考察主流模型与FSFDAI-TMRD经过蒸馏实验后的存储比和优化比,结果如图10所示。学生存储比越低、实验优化比越高,证明该模型蒸馏效果越好。从图10可以看出,FSFDAI-TMRD拥有良好的蒸馏能力,具备本地部署内存敏感设备的实用性能力。

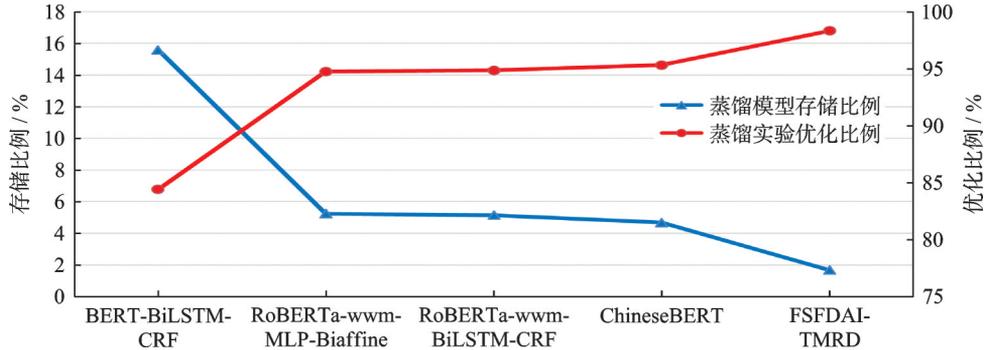


图10 不同模型存储优化对比

Fig.10 Comparison of storage and optimization of different models

2.5 消融实验结果与分析

为进一步论证FSFDAI-TMRD各模块的合理性,消融实验各步骤具体为:(1)Model 1裁剪FSFDAI模块,参考Wang等^[35]的工作,代替为传统的、参数硬共享的多任务学习;(2)Model 2替换多任务、多特征协同预测架构中的BMES标注法,转由Moscato等^[34]工作的BIO标注法;(3)Model 3裁剪跨模态多头蒸馏模块TMRD,替换为普通蒸馏;(4)FSFDAI-TMRD保持原结构模型,作为消融实验的结果参照。对比实验结果如表7所示。根据消融实验结果,FSFDAI-TMRD与未使用FSFDAI模块的Model 1相比, F_1 值提高了

表7 消融实验结果对比

Table 7 Comparison of ablation results				%
模型	精确度	召回率	F_1 指标	
Model 1	89.02	82.25	85.50	
Model 2	93.64	88.13	90.80	
Model 3	93.21	91.60	92.40	
FSFDAI-TMRD	94.09	92.09	93.08	

8.27%;与未使用多任务多特征协同预测架构中的BMES标注法 Model 2相比, F_1 值提高了2.97%;与未使用跨模态多头蒸馏的 Model 3相比, F_1 值提高了0.68%,表明了FSFDAI-TMRD各模块的合理性。原因可归纳为:(1)改进的多任务、多特征协同预测架构FSFDAI。相较于原工作BIO法的粗略标注和过度简化,面向鱼病领域的嵌套名词实体,更细粒度的BMES四位标注法将更好地识别边界和表述特征。为此本文修改了联合概率分布计算方式,改进整体识别精度。(2)提出的跨模态多头蒸馏TMRD。为更好地对齐师生中间特征,减少因参数规模相差过大的“代沟”匹配问题,构建了跨模态多头蒸馏。为此运用了跨模态多头注意力,取代了原师生依赖最后隐层的传统蒸馏方式,从而合理分配权重关系,减少可能引入的噪音。

2.6 鱼病知识图谱可视化

通过FSFDAI-TMRD方法可以更好地从多源、异构知识中提取鱼病实体,从而构建鱼病知识图谱的数据层。最终涵盖17种鱼病实体类型、39种鱼病关系类型、18 622个实体、244 978个字符以及4 585条样本,以三元组存储本地Neo4j数据库中,有力推动鱼病专家系统应用落地。部分鱼病知识图谱可视化如图11所示。鱼病知识图谱和鱼病专家系统的整体技术路线如图12所示。

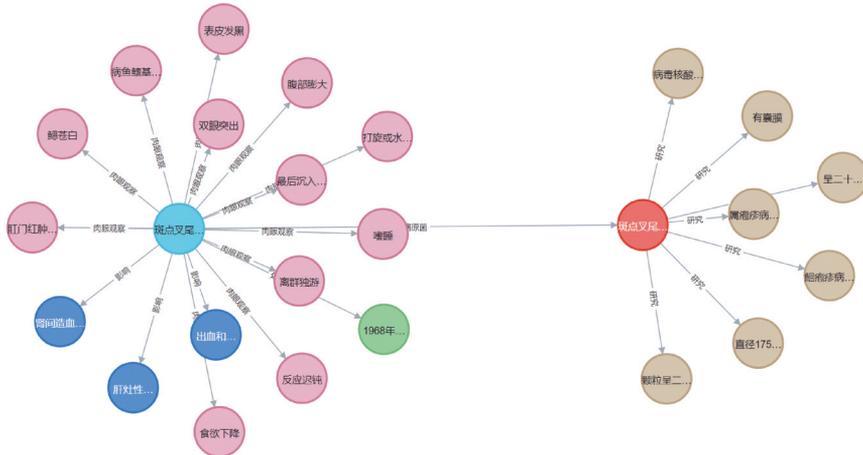


图 11 鱼病知识图谱(部分)

Fig.11 Fish disease knowledge graph (part)

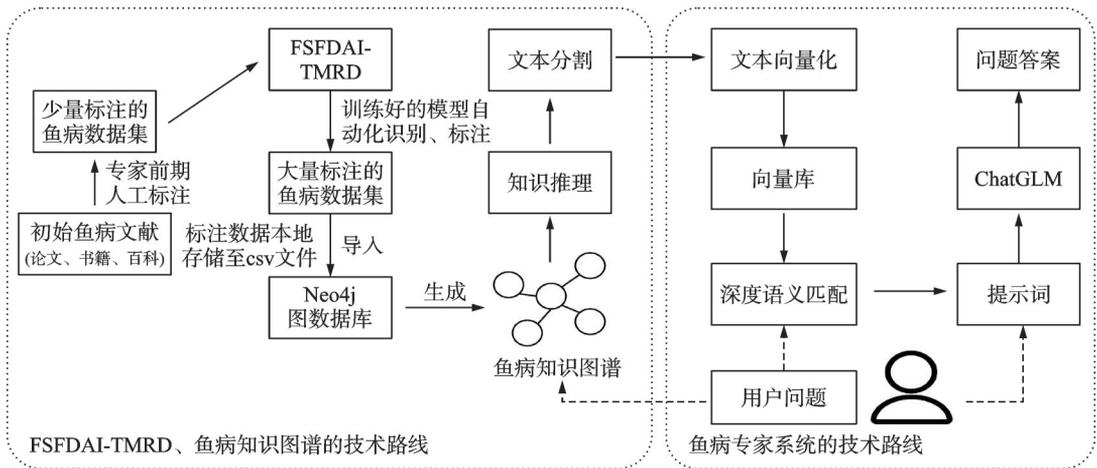


图 12 鱼病知识图谱以及专家系统的整体技术路线图

Fig.12 Overall technology road map for fish disease knowledge graph and expert systems

3 结束语

低存储、高精度、小样本的鱼病命名实体识别研究是知识图谱、专家系统落地应用的实际需求。面向鱼病垂直领域独有的痛点难点:一是特征表述多、识别精确度低,二是泛化性能弱、蒸馏损失大,本文提出一种基于多特征协同预测、跨模态多头蒸馏的方法FSFDAI-TMRD。一方面,该方法综合采样鱼病上下文的字音、字形和词义等多特征学习,重点应用多特征协同预测架构和双仿射注意力机制,核心运用联合概率分布,动态调整特征权重、整合优化模型压缩,缓解鱼病数据固有噪声,解决过拟合及精度低难题。另一方面,该方法充分提取语义关联、词性捕捉和标签依赖等多知识学习,紧密依托对抗权重干扰措施、大规模预训练语言模型,主体利用AWP以字符嵌入及特征权重为双扰动机制,收缩权重损失。下一步将结合加强学习和提示学习等新技术,围绕鱼病专家系统部署应用研究更高效的多模态、多标识模型。

参考文献:

- [1] 联合国粮食及农业组织. 2022世界渔业和水产养殖现状[R/OL].(2022-06-29)[2023-03-20].<http://www.fao.org/3/cc0461zh/cc0461zh.pdf>.
- [2] 中华人民共和国农业农村部.“十四五”渔业科技发展规划[R/OL].(2021-12-09)[2023-03-20].<http://www.moa.gov.cn/govpublic/YYJ/202201/P020220113593490778905.ofd>.
- [3] LIU P, GUO Y, WANG F, et al. Chinese named entity recognition: The state of the art[J]. *Neurocomputing*, 2022, 473: 37-53.
- [4] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis, Minnesota, USA: Association for Computational Linguistics, 2019: 4171-4186.
- [5] MIYATO T, DAI A M, GOODFELLOW I. Adversarial training methods for semi-supervised text classification[EB/OL]. (2021-11-16). <https://arxiv.org/abs/1605.07725>.
- [6] JIANG H, HE P, CHEN W, et al. SMART: Robust and efficient fine-tuning for pre-trained natural language models through principled regularized optimization[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. [S.l.]: Association for Computational Linguistics, 2020: 2177-2190.
- [7] WU D, XIA S T, WANG Y. Adversarial weight perturbation helps robust generalization[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 2958-2969.
- [8] CHEN H, YUAN S, ZHANG X. Rose-NER: Robust semi-supervised named entity recognition on insufficient labeled data [C]//Proceedings of the 10th International Joint Conference on Knowledge Graphs. [S.l.]: ACM, 2021: 38-44.
- [9] MA H, ZHANG Y, WANG J. Pretrained models with adversarial training for named entity recognition in scientific text[C]// Proceedings of 2022 International Conference on Asian Language Processing (IALP). Singapore: IEEE, 2022: 259-264.
- [10] 吴炳潮,邓成龙,关贝,等.动态迁移实体块信息的跨领域中文实体识别模型[J].*软件学报*, 2022, 33(10): 3776-3792.
WU Bingchao, DENG Chenglong, GUAN Bei, et al. Cross-domain Chinese entity recognition model with dynamic migration of entity block information[J]. *Journal of Software*, 2022, 33(10): 3776-3792.
- [11] HUANG Z, XU W, YU K. Bidirectional LSTM-CRF models for sequence tagging[EB/OL]. (2015-08-05). <https://arxiv.org/abs/1508.01991>.
- [12] ZHANG D, XIA C, XU C, et al. Improving distantly-supervised named entity recognition for traditional Chinese medicine text via a novel back-labeling approach[J]. *IEEE Access*, 2020, 8: 145413-145421.
- [13] LIU F, YOU C, WU X, et al. Auto-encoding knowledge graph for unsupervised medical report generation[J]. *Advances in Neural Information Processing Systems*, 2021, 34: 16266-16279.
- [14] WU Y, ZHU X, ZHU Y. An improved approach to the construction of Chinese medical knowledge graph based on CTD-BLSTM model[J]. *IEEE Access*, 2021, 9: 74969-74976.
- [15] ZHANG Y, LIU Y, ZHU J, et al. FSPRM: A feature subsequence based probability representation model for Chinese word

- embedding[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2021, 29: 1702-1716.
- [16] SUN Z, LI X, SUN X, et al. ChineseBERT: Chinese pretraining enhanced by glyph and pinyin information[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. [S.l.]: Association for Computational Linguistics, 2021: 2065-2075.
- [17] 陈梦萱,陈艳平,扈应,等.基于词义增强的生物医学命名实体识别方法[J/OL].计算机工程:1-9[2023-03-28]. DOI:10.19678/j.issn.1000-3428.0065880.
CHEN Mengxuan, CHEN Yanping, HU Ying, et al. Biomedical named entity recognition method based on word sense enhancement[J/OL]. Computer Engineering: 1-9 [2023-03-28]. DOI: 10.19678/j.issn.1000-3428.0065880.
- [18] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[EB/OL]. (2015-03-09). <https://arxiv.org/abs/1503.02531>.
- [19] TANG R, LU Y, LIU L, et al. Distilling task-specific knowledge from BERT into simple neural networks[EB/OL]. (2019-03-28). <https://arxiv.org/abs/1903.12136>.
- [20] 浦通,黄书剑,张洋铭,等.TCS:一种用于跨语言文本标签预测的“老师-课程-学生”学习框架[J].计算机学报,2022,45(9): 1983-1996.
PU Tong, HUANG Shujian, ZHANG Yangming, et al. TCS: A “Teacher-Course-Student” learning framework for cross-language text label prediction[J]. Journal of Computer, 2022, 45(9): 1983-1996.
- [21] GOU J, YU B, MAYBANK S J, et al. Knowledge distillation: A survey[J]. International Journal of Computer Vision, 2021, 129: 1789-1819.
- [22] 胡滨,耿天玉,邓赓,等.基于知识蒸馏的高效生物医学命名实体识别模型[J].清华大学学报(自然科学版),2021,61(9): 936-942.
HU Bin, GENG Tianyu, DENG Geng, et al. Efficient biomedical named entity recognition model based on knowledge distillation[J]. Journal of Tsinghua University (Science and Technology), 2021, 61(9): 936-942.
- [23] JIAO X, YIN Y, SHANG L, et al. TinyBERT: Distilling BERT for natural language understanding[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings. [S.l.]: Association for Computational Linguistics, 2020: 4163-4174.
- [24] LIANG S, GONG M, PEI J, et al. Reinforced iterative knowledge distillation for cross-lingual named entity recognition[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. [S.l.]: ACM, 2021: 3231-3239.
- [25] ZHOU X, ZHANG X, TAO C, et al. Multi-grained knowledge distillation for named entity recognition[C]//Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. [S.l.]: Association for Computational Linguistics, 2021: 5704-5716.
- [26] MOSCATO V, POSTIGLIONE M, SANSONE C, et al. TaughtNet: Learning multi-task biomedical named entity recognition from single-task teachers[J]. IEEE Journal of Biomedical and Health Informatics, 2023, 27(5): 2512-2523.
- [27] 吴智妍,金卫,岳路,等.电子病历命名实体识别技术研究综述[J].计算机工程与应用,2022,58(21): 13-29.
WU Zhiyan, JIN Wei, YUE Lu, et al. Research review on named entity recognition technology in electronic medical records [J]. Computer Engineering and Applications, 2022, 58(21): 13-29.
- [28] 王颖洁,张程焯,白凤波,等.中文命名实体识别研究综述[J].计算机科学与探索,2023,17(2): 324-341.
WANG Yingjie, ZHANG Chengye, BAI Fengbo, et al. Review of Chinese named entity recognition[J]. Journal of Computer Science and Exploration, 2023, 17(2): 324-341.
- [29] 姜丽华,赵瑞雪,董春岩,等.基于深度学习的水产病害可视化知识图谱构建与验证[J/OL].农业工程学报:1-9[2023-09-14]. <http://kns.cnki.net/kcms/detail/11.2047.S.20230908.1424.024.html>.
JIANG Lihua, ZHAO Ruixue, DONG Chunyan, et al. Construction and verification of visualized knowledge graph for aquatic animal diseases based on deep learning[J/OL]. Transactions of the Chinese Society of Agricultural Engineering: 1-9 [2023-09-14]. <http://kns.cnki.net/kcms/detail/11.2047.S.20230908.1424.024.html>.
- [30] 刘巨升,杨惠宁,孙哲涛,等.面向知识图谱构建的水产动物疾病诊治命名实体识别[J].农业工程学报,2022,38(7): 210-217.
LIU Jusheng, YANG Huining, SUN Zhetao, et al. Named entity recognition for diagnosis and treatment of aquatic animal diseases oriented to knowledge graph construction[J]. Transactions of the Chinese Society of Agricultural Engineering, 2022, 38(7): 210-217.

- [31] 黄震华,杨顺志,林威,等.知识蒸馏研究综述[J].计算机学报,2022,45(3): 624-653.
HUANG Zhenhua, YANG Shunzhi, LIN Wei, et al. Review of knowledge distillation[J]. *Journal of Computer*, 2022, 45(3): 624-653.
- [32] XU C, MCAULEY J. A survey on model compression for natural language processing[EB/OL]. (2015-02-15). <https://arxiv.org/pdf/2202.07105v1>.
- [33] 刘媛媛,王定坤,鄢雷,等.基于知识蒸馏和模型剪枝的轻量化模型植物病害识别[J/OL].浙江农业学报:1-18[2023-09-14].
<http://kns.cnki.net/kcms/detail/33.1151.S.20230817.1412.008.html>.
LIU Yuanyuan, WANG Dingkun, WU Lei, et al. Lightweight model for plant disease recognition based on knowledge distillation and model pruning[J/OL]. *Journal of Zhejiang Agricultural Sciences*: 1-18 [2023-09-14]. <http://kns.cnki.net/kcms/detail/33.1151.S.20230817.1412.008.html>.
- [34] MIRZADEH S I, FARAJTABAR M, LI A, et al. Improved knowledge distillation via teacher assistant[C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2020: 5191-5198.
- [35] WANG X, ZHANG Y, REN X, et al. Cross-type biomedical named entity recognition with deep multi-task learning[J]. *Bioinformatics*, 2019, 35(10): 1745-1752.

作者简介:

沈志成(1996-),男,硕士研究生,研究方向:知识图谱、专家系统, E-mail: 992536989@qq.com。



陈明(1966-),通信作者,男,博士、教授,研究方向:知识图谱、数据挖掘, E-mail: mchen@shou.edu.cn。

(编辑:张黄群)