

基于知识增强自适应原型网络的小样本关系分类

张河萍¹, 方志军^{1*}, 卢俊鑫², 高永彬¹

(1. 上海工程技术大学电子电气工程学院, 上海 201620; 2. 华东师范大学计算机科学与技术学院, 上海 200241)

摘要: 小样本关系分类(FSRC)是指在任务中使用少量标注实例对各种关系进行分类,可快速适用于对全新的类别进行归类。然而,当测试域与训练域之间存在分布差异时,现有的小样本分类算法泛化能力有限,导致分类性能下降。针对该问题,提出一种适用于领域适应任务的知识增强自适应原型网络。通过探索实例之间的联系以提高模型的鲁棒性,同时学习关于关系的先验知识和内在语义以获得可解释原型。通过引入交互注意力机制来捕捉支持实例与查询实例之间的相关性,突出关键实例,并生成交互实例。同时,自适应原型融合机制以关系信息为锚点生成自适应混合系数,通过特征融合将实例与关系信息相结合,从而生成混合原型。在公开数据集 FewRel 1.0 和 FewRel 2.0 上的实验结果验证了该网络的有效性。实验结果表明,与基线模型相比,所提网络模型的分类准确率取得了显著提升,具有更好的分类效果与稳定性。

关键词: 关系分类;小样本学习;小样本关系分类;原型网络;知识增强

中图分类号: TP391.1

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0068855

Few-Shot Relation Classification Based on Knowledge-Enhanced Adaptive Prototype Networks

ZHANG Heping¹, FANG Zhijun^{1*}, LU Junxin², GAO Yongbin¹

(1. School of Electronics and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China;

2. School of Computer Science and Technology, East China Normal University, Shanghai 200241, China)

【Abstract】 Few-Shot Relation Classification (FSRC) refers to the use of a small number of labeled instances to classify various relations within a task, and it can be quickly applied to categorize completely new classes. However, existing few-shot classification algorithms show limited generalization capabilities when distributional differences exist between the test and training domains, which leads to significant performance degradation. Therefore, a knowledge-enhanced adaptive prototyping network is proposed for domain adaptation tasks, which helps improve the robustness of the model by exploring connections between instances while learning a priori knowledge about relations and intrinsic semantics to obtain interpretable prototypes. Specifically, the correlations between supporting and query instances are captured by introducing an interactive attention mechanism to highlight key instances and generate interactive instances. The adaptive prototype fusion mechanism generates adaptive hybrid coefficients using relational information as anchors and combines instances with relational information through feature fusion to generate hybrid prototypes. Experiments are conducted on FewRel 1.0 and FewRel 2.0 datasets, and the effectiveness of the method is demonstrated. The experimental results show that the classification accuracy of the proposed network model is significantly higher compared to that of the baseline model. The proposed model has better classification performance and stability.

【Key words】 Relation Classification (RC); few-shot learning; Few-Shot Relation Classification (FSRC); prototype networks; knowledge-enhanced

0 引言

关系分类(RC)是信息抽取任务和构建知识图谱中的一项重要任务。给定一个包含两个目标实体(Bob Gaudet 和 ice hockey)的语句([Bob Gaudet] is an American [ice hockey] coach who has served as the head coach at Dartmouth since 1997.),通过

RC模型得到实体对的语义关系。RC模型旨在所有候选关系类中区分给定语句中的语义关系。

RC是一个广泛研究的领域。基于深度学习的有监督方法^[1-2]使用大量具有广泛关系的标注数据来训练神经网络,从而使模型能够在测试阶段识别这些关系,并在性能上取得了显著的进展。然而,这些模型的性能严重依赖于大规模的标注数据^[3],数

收稿日期: 2023-11-16 修回日期: 2024-01-02

基金项目: 科技创新 2030—“新一代人工智能”重大项目(2020AAA0109300)。

通信作者 E-mail: * zjfang@sues.edu.cn

据标注需要耗费大量人力和时间。传统的 RC 模型在训练过程中难以快速适应新的关系类别,泛化性能不足。相比之下,人们可以通过很少的例子来识别新的关系。因此,当只有极少量带标注数据可用时,研究小样本关系分类(FSRC)变得至关重要。

FSRC 利用先验知识训练模型,使之具有自适应学习的能力,并利用模型所学的知识来迅速适应全新的学习任务。当前,已有大量工作^[4-9]被提出,其中一种简单而高效的元学习方法是基于度量学习的原型网络^[8],该网络从少量实例中生成每个类别的原型表示,并通过最近邻^[10]的方式对查询表示进行分类。然而,这些方法都假定测试数据里的未知类别样本与训练数据属于相同的领域。在现实生活中,目标样本可能来自不同的领域,例如,医疗记录可能包含医学术语和缩写。文献^[11]指出,当基础类和新类之间存在较大的领域偏移时,现有 FSL 算法的性能会显著下降。为此,研究人员提出许多方法,其中对抗训练已被证明能有效获得领域不变特征。然而,基于对抗性领域适应的方法在相同领域的实验中表现不佳,难以很好地捕捉实例中关系的一致性。

为解决上述问题,本文提出一种知识增强自适应原型(KEAP)学习方法。该方法通过改进原型网络的表示来解决跨域小样本关系分类问题,为每个关系获得更好的原型。KEAP 引入了一个交互注意力模块,利用实例之间的交互信息来突出与查询实例相关的支持实例,从而增强样本在不同领域中的空间表示。此外,本文设计了一个自适应原型融合模块,利用关系信息来融合支持实例,使其能够代表该类别。本文在 FewRel 1.0^[6](通常场景)和 FewRel 2.0^[12](跨领域场景)上的实验证明了所提 KEAP 的有效性。本文工作的贡献可总结为以下 3 个方面:

1)提出一种全新的知识增强自适应原型学习方法,专注于将领域特定知识融入小样本关系分类任务,特别是领域适应任务。

2)在 KEAP 中引入交互注意力模块,以选择具有更大信息量的支持实例,同时采用自适应融合模块将实例与关系信息相结合,生成更紧凑的原型表示。

3)本文进行了广泛的比较实验和消融实验,实验结果表明,所提方法的性能取得了显著提升,跨领域小样本分类准确率明显优于基线模型。

1 相关工作

1.1 小样本学习

当前,元学习^[9]已成为小样本学习任务的主流框架,有助于处理小样本关系分类任务,元学习模型

的运行机制与直接通过训练神经网络以执行特定任务的传统模型存在本质区别。元学习在各种相似但不同的任务上进行训练,以获得快速适应新任务的能力。现有基于元学习的小样本学习^[13]方法可以分为 3 个子类:基于度量的方法、基于优化的方法和基于参数的方法。基于度量的方法旨在学习如何将样本嵌入到 1 个统一联合空间中,使得相似样本更具有区分性^[14]。原型网络是最通用的一种方法,假设每个关系类别都有 1 个原型,并在嵌入空间中学习原型,最后使用最近邻规则对查询实例进行分类。基于优化的方法^[15]侧重于参数更新策略,该方法认为单个任务产生的梯度不一定是全局最优的。这类方法的框架通常包括 1 个元学习器和 1 个基础学习器。外部循环逐渐训练元学习器,跨多个不同任务,并学习共享参数的全局初始化。内部循环则仅使用几个梯度下降步骤来快速微调全局初始化,以完成特定任务。基于参数的方法^[16]对从源类别训练所得的模型进行微调,使其快速适应目标类别。这类方法旨在学习跨任务的可迁移先验知识,并使用训练样本为类似的未知任务生成网络参数。

1.2 小样本关系分类

近些年,应用于小样本分类的架构不断被提出,针对小样本图像分类的方法已经得到深入研究,文本领域缺乏相关公共数据集一直未得到发展。之后,HAN 等^[6]结合小样本学习和关系分类模型,创建了大规模的 FSRC 基准 FewRel,用于评估小样本关系分类的基线方法,以探究小样本关系分类,并将经典方法如元网络^[7]、原型网络^[8]应用于关系分类。

其中,GAO 等^[17]提出的 HATT-Proto 是一种基于混合注意力的原型网络,分别关注关键实例和特征,以减小噪声数据和稀疏特征的影响。YE 等^[18]提出了多级匹配和基于聚合原型的模型,以充分学习句子实例之间的匹配信息。他们研究了训练实例和测试实例之间的潜在关联,并提出了多级匹配聚合网络,以提高小样本关系分类的准确性。然而,当标注数据为非常规,且不能反映关系的一般模式时,这些方法仍然是有限的。为获得更好的性能,研究人员利用额外信息来补充标注数据。许多工作将关系信息集成到模型中,以辅助原型表示的学习。QU 等^[19]提出的 REGRAB 是通过贝叶斯元学习建模不同关系之间的后验分布,其训练速度较慢。HAN 等^[20]提出的 HCRP 是通过对比学习方法来支持模型训练,以获得更好的原型。除关系描述以外,YANG 等^[21]提出的 ConceptFCRE 是一种实体概念增强的方法,引入实体的固有概念,为关系预测

提供线索。YANG 等^[22]提出的 TD-Proto 是通过关系和实体描述来增强原型网络。这些研究通过不同的方式引入了更多的外部信息,以显著提升性能。本文所提的方法就属于这一类。DONG 等^[23]提出的 MapRE 同时考虑了标签无关和标签感知的语义映射信息,用于低资源关系的提取。此外,有研究侧重于进一步训练预训练语言模型(PLM)来处理 RC 任务。基于相同实体对的句子可能表达相同关系的假设^[24],构建了大规模预训练数据集,并提出了“匹配空白”的预训练范式。这些研究表明良好的嵌入对小样本学习非常重要。DaFeC^[25]是一种基于归纳聚类的框架,利用聚类生成的伪标签来训练小样本分类器,以解决 FSRC 中的领域适应问题。BERT-PAIR^[12]对抗抗性训练应用于 FSRC 任务,通过将每个查询实例与所有支持实例进行配对,然后将这些连接的序列输入到 BERT^[26]序列分类模型,以获得 2 个实例表示相同关系的得分。

不同于上述现有方法,为构建更具区分性和泛化性的原型网络,本文提出一种知识增强的自适应原型网络。该网络一方面使用交互注意力机制对支持实例进行评分,以突出关键实例;另一方面在交互实例和关系信息的引导下逐步完善原型表示。

2 问题定义

对于小样本关系分类任务,其目标是训练 1 个分类模型,能够在给定 1 个支持集 S 的情况下预测

查询实例 q 的标签。形式上,1 个 N -way K -shot 任务由 1 个包含 N 个关系类的标注实例的支持集和查询集组成,任务目标是在给定支持集中的 K 个支持实例的情况下,对查询集中的关系进行预测。每个实例由 1 组样本 (x, e, r) 组成,其中 x 是 1 个句子, $e = (e1, e2)$ 分别表示头实体和尾实体, r 是实例 (x, e) 的关系标签。

此外,为了训练 1 个良好的模型,需要 1 个基类数据集 D_{base} ,以提高小样本学习的性能。基于元学习的方法通过模仿小样本学习的范式,使用基类数据集来训练模型。在训练迭代过程中,从基类数据集中随机选择 N 个关系,然后从每个关系的标注数据中随机选择 K 个实例,以构建训练支持集 $S = \{s_{i,k}; 1 \leq i \leq N, 1 \leq k \leq K\}$,其中 i 表示类, k 代表序号, $s_{i,k}$ 表示第 i 个类别的第 k 个实例。此外,还会从 N 个关系的剩余实例中随机抽取 M 个实例,构建查询集 $Q = \{(q_j, y_j); 1 \leq j \leq M\}$,其中 q_j 表示第 j 个查询实例, $y_j \in \{1, \dots, N\}$ 表示实例的标签。模型通过从基类数据集中抽取的少量任务集合进行训练优化,能够快速适应新任务。

在小样本关系分类任务中,2-way 1-shot 场景例如表 1 所示,训练阶段数据与测试阶段 1 的数据均采集自 Wikipedia 数据集。然而,在领域适应场景中,数据存在显著差异。这一差异主要体现在:训练阶段的数据源于 Wikipedia,测试阶段 2 的数据则来自 Biomedicine,两者的数据来源分属不同领域。

表 1 2-way 1-shot 场景示例

Table 1 Examples of 2-way 1-shot scenarios

阶段	关系	实体	标注实例
训练阶段	支持集	role_of	Actress Sally Hawkins was asked to play [Persuasion]e2 protagonist [Anne Elliot]e1
		place_of birth	[Elvis Presley]e1 was born in [Memphis, Tennessee]e2
	查询	role_of 或	Ulrich Matthes is a German actor, best known for his role as [Joseph Goebbels]e1 in the film
		place_of birth	[Downfall]e2
测试阶段 1	支持集	religion	He founded a literary prize which in 1901 was won by [Ndoc Nikaj]e1, a [Catholic]e2 priest
		mother	In 1845, [Cemile Sultan]e1 's mother, Empress [Dzdidil Kadn]e2, died
	查询	religion 或 mother	He was married to [Eva Funck]e2 and they have a son [Gustav]e1
		disposition_of	whether [ritonavir]e1 boosted [protease inhibitor]e2(pi/r) monotherapy...
测试阶段 2	支持集	classified_as	Several outbreaks of [trichinellosis]e1 associated with the consumption ... pecies of [trichinella infection]e2 in the ...
	查询	disposition_of 或	A prevalence study was designed to evaluate host and seasonal risk factors associated with three
		classified_as	systemic [mycoses]e1 in dogs: [blastomycosis]e2, coccidioidomycosis, and histoplasmosis

FSRC 的目标是通过优化以下目标函数来实现:

$$L = \frac{1}{M} \sum_{(q,y) \in Q} P_{\phi}(y | S, q) \quad (1)$$

式中: P_{ϕ} 表示计算查询实例属于该支持集的概率。

3 本文方法

本节将详细介绍知识增强自适应原型学习的方法,其整体结构如图 1 所示,整体结构中 3 个相同的形状代表不同的关系类别,同时将 N 个关系的名称

和描述作为输入。所有输入共享同一个句子编码器。在模型中,由支持向量表示计算出的基础原型 c_i ,再由交互注意力机制得到交互向量表示 s_i ,最后,通过自适应混合机制,由交互向量表示与关系向

量表示得到交互实例原型 z_i ,最后通过自适应系数将基础原型 c_i 与交互实例原型 z_i 计算出混合原型 P_i 。模型会计算混合原型与每个查询之间的距离,用于训练和预测。

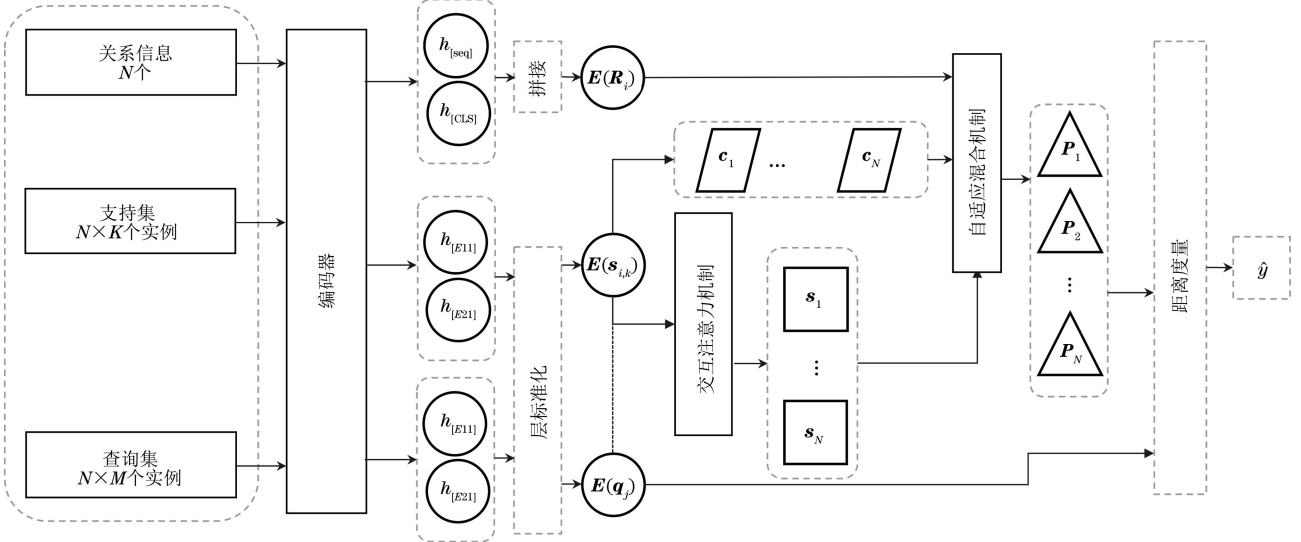


图 1 KEAP 模型的整体结构

Fig.1 Overall structure of KEAP model

3.1 实例编码器

KEAP 模型使用 BERT^[26] 作为编码器,以获取实例的上下文嵌入,遵循 SOARES 等^[24] 所提的编码方式,通过引入额外的序列信息来指定命名实体的位置,包括 4 个保留词片段 [E11], [E12], [E21], [E22] 增加到输入实例中,以标记句子中提到每个实体的开始和结束,其中头实体与尾实体的隐藏状态表示为 $h_{[E11]}$ 与 $h_{[E21]}$, 并进行连接以及层标准化(LN)得到实例向量表示。对于每个关系,将关系名称与关系描述输入到编码器中,语义融合后得到 [CLS] 所对应标记的隐藏状态 $h_{[CLS]}$ 和所有标记的隐藏状态的平均值 $h_{[seq]}$, 将其拼接起来以获得关系向量表示。为了使得公式表达更简便,将编码过程记为 E 。其中,支持实例向量表示为 $\{E(s_{i,k}) \in \mathbb{R}^{2d}; 1 \leq i \leq N, 1 \leq k \leq K\}$, 查询实例向量表示为 $\{E(q_j) \in \mathbb{R}^{2d}; 1 \leq j \leq M\}$, 关系向量表示为 $\{E(R_i) \in \mathbb{R}^{2d}; 1 \leq i \leq N\}$, 其中 d 表示 BERT 输出的隐藏尺寸。

3.2 自适应混合原型学习

原型网络^[8] 通过编码器来获取支持集中每个关系类中的支持表示,然后对每个关系类中的 K 个支持表示取平均值,从而得到基础原型表示为 $\{c_i \in \mathbb{R}^{2d}; 1 \leq i \leq N\}$:

$$c_i = \frac{1}{K} \sum_{j=1}^K E(s_{i,j}) \quad (2)$$

然而,在小样本场景中,由于缺乏充分的支持数据,当 1 个实例特征与其他实例差异很大时,相应的原型可能会出现较大的偏差,因此模型提取的特征与查询实例无关。为了更好地将注意力集中在与查询实例相关的特征上,需要改进原型网络的结构。

针对上述问题,本文采用交互注意力机制来学习实例之间的特征信息,确保模型更加关注查询实例的相关特征。由于查询实例与支持集中所有实例之间的相关性不是一致的,因此需要计算每个支持实例与查询实例之间的注意力分数。对于每个支持实例与查询实例,都会有 1 个相关性权重 $a_j \in [0,1]$, 定义为 $a = [a_1, a_2, \dots, a_K]^T$ 。通过交互注意力计算得到了交互向量表示 s_i :

$$s_i = \frac{1}{K} \sum_{j=1}^K a_{i,j} E(s_{i,j}) \quad (3)$$

$$a_{i,j} = \text{Softmax}\left(-\frac{1}{|M|} \text{LN}(E(s_{i,j}) \circ E(q))\right) \quad (4)$$

式中: $\text{LN}(\cdot)$ 为线性层; $\circ$ 为对应位置元素相乘的运算; $|M|$ 为查询集中的实例总数。

当 2 个类别特别相似时,网络模型学习到的样本分布可能无法有效分离类别特征,这可能会影响分类性能。然而,关系信息可以帮助模型更好地分离在嵌入空间中具有相似类别的特征,关系信息包含了有关类别整体分布的最基本信息。对于每个类

别,本文将关系信息与交互表示进行整合,将关系信息作为锚点来动态融合原型表示。通过引入关系信息,可得到包含更多有效信息和更具稳健性的混合原型。混合原型的计算如式(5)所示:

$$\mathbf{P}_i = \lambda_i \mathbf{c}_i + (1 - \lambda_i) \mathbf{z}_i \quad (5)$$

基础原型表示 $\mathbf{c}_i$ 与交互实例原型 $\mathbf{z}_i$ 通过自适应系数 λ_i 融合得到混合原型 $\mathbf{P}_i$ 。其中自适应系数取决于关系信息的表示,以确保计算出的类别原型是兼容的。交互实例原型的计算如式(6)所示:

$$\mathbf{z}_i = \text{Concat}[\mathbf{E}(\mathbf{R}_i); \mathbf{s}_i] \quad (6)$$

为得到交互实例原型,将关系向量表示 $\mathbf{E}(\mathbf{R}_i)$ 与交互实例表示的平均值 $\mathbf{s}_i$,经拼接 $\text{Concat}[\cdot]$ 操作,得到交互实例原型 $\mathbf{z}_i$ 。自适应系数的计算如式(7)所示:

$$\lambda_i = \frac{1}{1 + \exp(-\text{MLP}(\mathbf{E}(\mathbf{R}_i)))} \quad (7)$$

式中:MLP 为两层全连接网络;自适应系数 $\lambda_i \in (0, 1)$ 。

3.3 训练预测

该模型使用向量点积方式来计算查询实例与每个混合原型之间的距离,然后将这些距离作为输入,形成交叉熵损失,训练损失的定义如下:

$$L = - \sum_{j=1}^M \log_a \frac{\exp(\mathbf{P}_i \cdot \mathbf{E}(\mathbf{q}_j))}{\exp(\mathbf{P}_k \cdot \mathbf{E}(\mathbf{q}_j))} \quad (8)$$

在预测阶段,选择查询向量表示与原型向量表示空间距离最短的关系原型作为预测结果,也就是模型分配给查询实例属于关系 r_i 概率最高的标签。

$$\hat{y} = \underset{i}{\operatorname{argmax}} \mathbf{P}(y = i | \mathbf{E}(\mathbf{q}_j)) \quad (9)$$

4 实验

4.1 数据集与训练细节

在实验过程中,本文使用 2 个基准数据集进行评估,分别是 FewRel 1.0 和 FewRel 2.0-DA。FewRel 1.0 数据集由人工标注维基百科语料库构建,包含 100 个关系类,每个关系有 700 个标注实例。实验遵循官方基准测试中使用的分割标准,64 个基类用于训练,16 个关系类用于验证,20 个新类用于测试。此外,还给出了每个关系的名称和描述,为每个关系提供了额外的可解释性。FewRel 2.0-DA 数据集使用 FewRel 1.0 的训练集,测试集来自大量生物医学文献的数据库 PubMed 与生物医学领域的大型知识库 UMLS 进行对齐,只给出了关系标签的名称,但没有提供它们的描述,这使得任务更具挑战性。测试集包含 25 个关系和每个关系

的 100 个实例。值得注意的是,FewRel 只发布了训练集和验证集,测试集的标签并没有公开。因此,在 FewRel 的验证集上进行了大部分性能分析,对于测试集,将最佳模型的预测文件提交到 CodaLab 官方排行榜,以获得测试集的最终结果。

模型运行的实验环境中,所采用的操作系统是 Ubuntu 18.04.6,CPU 型号为 Intel Xeon E5-2678 v3 @ 2.50 GHz,GPU 型号为 NVIDIA GeForce RTX 1080 Ti,内存大小为 31 GB,CUDA 版本为 11.4。

在 5-way 1-shot、5-way 5-shot、10-way 1-shot 和 10-way 5-shot 4 种场景下进行测试,通过分类准确率来评估模型性能。与官方评估结果基本一致,通过从验证数据中随机采样 10 000 个任务来评估模型性能,从而选择最佳模型进行测试。

模型在实验中的超参数如表 2 所示。设置类别数 $N = 5$ 来构造训练关系类别。在句子预处理中,将最大句子长度设置为 128。使用 PyTorch 和 Huggingface 的 Transformer 库来实现。采用 BERTbase-uncased 模型作为模型中的编码器。BERTbase 模型有 12 个变换器层,隐藏尺寸大小为 768。训练模型应用 AdamW 优化器来最小化损失,初始学习率设置为 1×10^{-5} 。该模型经过 30 000 次迭代训练,每 1 000 次迭代进行 1 次评估。Dropout rate 设为 0.25,批量大小设置为 1。

表 2 模型在实验中的超参数

Table 2 Hyperparameters of the model in the experiment

实验设置	参数	值
训练设置	训练时关系类别数	5
	支持集中的关系类别数	5
	查询数中的关系类别数	1
	训练批次	1
编码器设置	隐藏尺寸大小	768
	隐藏层数量	12
	注意力头的数量	12
优化	优化器	AdamW
	学习率	1×10^{-5}
	Dropout rate	0.25
	批量大小	1

对于 FSRC 任务,采用典型的 N -way K -shot 设置,将标记的训练数据根据关系类别分为 2 个部分: Dtrain 和 Dtest。这 2 个数据集有不同的标签空间。将 Dtrain 进一步拆分为 2 个部分: Dtrain-support 和 Dtrain-query。具体来说,在每次训练迭代中,从 Dtrain 中随机抽出 N 个类,然后,对 N 个类中的每个类抽出 K 个实例,构成 Dtrain-support。

此外,从这 N 个类的剩余样本中随机抽取 M 个实例来构建 Dtrain-query。同样,也可以用类似的方法获得 Dtest-support 和 Dtest-query。

4.2 与现有方法的比较

为了验证所提方法的有效性,本文需要与多个基线模型进行比较。本文选择 2 个基于 CNN 的模型和 8 个基于 BERT 的模型。基于 CNN 的模型包括:1)Proto^[8] 使用原型网络;2)MLMAN^[18] 是一种多级匹配和聚合的原型网络。基于 BERT 编码器的模型包括:

1)Proto-BERT^[6] 测量查询实例与每个关系原型之间的距离;2)BERT-PAIR^[12] 是一种基于 BERT 中的序列分类模型来衡量句子对相似性的方法,预测 2 个实例是否表达相同的关系。

以上 4 种方法没有使用额外的信息。另外,本文还选择了以下 6 种方法:

1)ConceptFERE^[21] 基于实体概念增强方案,首先为每个实体选取合适的概念,再整合概念和句子的语义来提高关系分类性能。

2)TD-Proto^[22] 通过关系和实体的文本描述来增强原型网络;

3)REGRAB^[19] 是一种带有外部全局关系图的贝叶斯元学习方法;

4)HCRP^[20] 基于对比学习,利用关系标签信息来学习更好的表示;

5)PRM^[27] 是一种原型校正方法,促进小样本关系分类,通过引入关系信息来校正原始原型网络。

6)HR-DProto^[28] 是一种融合混合表征的方法,用于解决原型偏差问题。

针对跨域小样本关系分类任务的研究很少,因此 FewRel 2.0-DA 的基线数量会降低。当前,主流方法主要涵盖基于 CNN 的方法和基于 BERT 的方法。Proto-BERT 表示本文基准方法,这意味着模型在未引入所提任何改进的情况下得出的结果。这一结果也将在消融研究中进行分析和显示。

在 FewRel 1.0 测试集上的小样本关系分类准确率如表 3 所示,加粗表示最优数据(下同)。表 3 中的数据来自 FewRel 1.0 公共排行榜或者其他论文,* 表示在训练过程中无引入外部信息。以 CNN 作为编码器的模型分类准确率较低,说明他们对文本数据的建模结果较差。同样的原型网络(Proto)模型将编码器换成 BERT,性能就得到了有效提高。当不使用外部信息时,KEAP 模型较 BERT-PAIR 模型性能有较大的提高,可以更好地对实例进行关系分类。与添加外部信息的模型相比,模型将交互实例与关系信息进行自适应融合。KEAP 模型在 5-shot 设置下与其他基准模型相比取得了最佳的结果。其原因为当每个关系类给出多个实例时,能提供更多的交互信息。

表 3 在 FewRel 1.0 测试集上的小样本关系分类准确率
Table 3 Few-shot relation classification accuracy on the FewRel 1.0 testset

编码器	模型	5-way 1-shot	5-way 5-shot	10-way 1-shot	10-way 5-shot	%
CNN	Proto-CNN *	74.52	88.40	62.38	80.45	
	MLMAN *	82.98	92.66	73.59	87.29	
BERT	Proto-BERT *	80.68	89.60	71.48	82.89	
	BERT-PAIR *	88.32	93.22	80.63	87.29	
	ConceptFERE	89.21	90.34	75.72	81.82	
	TD-Proto	84.76	92.38	74.32	85.92	
	REGRAB	90.30	94.25	84.09	89.93	
	KEAP *	91.67	96.32	86.14	93.21	
	HCRP	93.76	95.66	89.95	92.10	
	PRM	94.22	96.51	91.42	92.79	
	HR-DProto	90.16	96.26	85.24	94.18	
	KEAP	93.94	96.68	90.35	94.58	

在 FewRel 2.0 领域适应性测试集上的小样本关系分类准确率如表 4 所示。从表 4 可以看出,KEAP 模型总体性能优于强基线,分类准确率取得了显著的提高。与 PRM 相比,在 5-way 1-shot 和 10-way 1-shot 场景下分别提高了 7.15 和 6.28 百分点。实验

结果表明,该模型具有一定的稳定性和有效性,主要原因可能是交互实例能提供更具有泛化性的表示特征,自适应混合机制增强实例间的语义相关性,提高了模型的鲁棒性。在 1-shot 场景下也就是样本更少的情况下,KEAP 模型相比现有模型分类准确率更高。

表 4 在 FewRel 2.0 领域适应性测试集上的小样本关系分类准确率

Table 4 Few-shot relation classification accuracy on the FewRel 2.0 domain-adaptive testset

编码器	模型	5-way 1-shot	5-way 5-shot	10-way 1-shot	10-way 5-shot
CNN	Proto-CNN	35.09	51.50	22.98	35.22
	Proto-BERT	40.12	51.50	26.45	36.93
	Proto-ADV	42.21	58.71	28.91	44.35
BERT	BERT-PAIR	67.41	78.57	54.89	66.85
	HCRP	76.34	83.03	63.77	72.94
	PRM	73.98	88.38	62.72	79.43
	KEAP	81.13	91.84	69.00	83.15

4.3 消融实验

本文通过消融实验评估 KEAP 各模块的作用。由于 FewRel 数据集没有标注测试集,因此为了实验方便,对验证集进行消融实验。在 FewRel 2.0-DA 验证集上的消融实验结果如表 5 所示。首先,删除了交互注意力机制,在表中记为-IAM,以验证实例内交互信息的有效性,然后,移除自适应混合机制,记为-AIM,并将式(5)前的系数设为 $\lambda=0.5$ 。其中,移除自适应系数的模块记为-AIM#1。移除关系信息的模块记为-AIM#2,在删除式(6)与式(7)的同时,式(5)中的混合原型 P_i 由基础原型 c_i 与交互向量表示 s_i 直接计算得到。最后,同时删除交互注意力机制与自适应混合机制,在表中记为-IAM & -AIM 以验证整个 KEAP 网络的有效性。

表 5 在 FewRel 2.0-DA 验证集上的消融实验结果

Table 5 Ablation experiment results on the FewRel 2.0-DA validation set

模型	5-way-1-shot	10-way-1-shot
KEAP	80.80	69.75
-IAM	79.08	68.80
-AIM#1	79.52	68.48
-AIM#2	77.36	66.23
-IAM & -AIM	69.44	60.71

5 结束语

本文提出一种适用于领域适应任务的 KEAP 模型,用于解决关系分类中缺乏特定领域训练数据的问题。首先,通过交互注意力机制捕获实例间的交互信息,选择具有更大信息量的支持实例;然后,使用自适应混合机制权衡原型与关系信息的重要性,学习更具有区分度的原型表征。在 FewRel 1.0 和 FewRel 2.0 数据集上的实验结果表明,与基线模型相比,KEAP 模型的分类准确率取得了显著提升。下一步将知识增强自适应原型网络扩展到更需要特定领域背景知识的任务中。

参考文献

- [1] TRISEDDYA B D, WEIKUM G, QI J Z, et al. Neural relation extraction for knowledge base enrichment [C]// Proceedings of Annual Meeting of the Association for Computational Linguistics. [S. l.]: Association for Computational Linguistics, 2019: 229-240.
- [2] LI Z F, LIU H, ZHANG Z L, et al. Learning knowledge graph embedding with heterogeneous relation attention networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(8): 3961-3973.
- [3] 胡晗,刘鹏远.小样本关系分类研究综述[J].中文信息学报, 2022, 36(2):1-11.
- [4] HU H, LIU P Y. Few-shot relation classification: a survey [J]. Journal of Chinese Information Processing, 2022, 36(2):1-11. (in Chinese)
- [5] YUAN J B, GUO H, JIN Z W, et al. One-shot learning for fine-grained relation extraction via convolutional siamese neural network[C]//Proceedings of 2017 IEEE International Conference on Big Data. Washington D. C., USA: IEEE Press, 2017: 21-35.
- [6] SUNG F, YANG Y X, ZHANG L, et al. Learning to compare: relation network for few-shot learning [C]// Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 1199-1208.
- [7] HAN X, ZHU H, YU P F, et al. FewRel: a large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation[EB/OL]. [2023-10-12]. <https://arxiv.org/pdf/1810.10147>.
- [8] MUNKHDALAI T, YU H. Meta networks [C]// Proceedings of International Conference on Machine Learning. New York, USA: ACM Press, 2017: 102-109.
- [9] SNELL J, SWERSKY K, ZEMEL R S. Prototypical networks for few-shot learning [EB/OL]. [2023-10-12]. <https://arxiv.org/pdf/1703.05175>.
- [10] HOSPEDALES T, ANTONIOU A, MICAELLI P, et al. Meta-learning in neural networks: a survey [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44: 5149-5169.
- [11] ZHANG S C, LI X L, ZONG M, et al. Efficient kNN classification with different numbers of nearest neighbors[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(5): 1774-1785.
- [12] ZHAO A, DING M Y, LU Z W, et al. Domain-adaptive few-shot learning[C]//Proceedings of 2021 IEEE Winter Conference on Applications of Computer Vision. Washington D. C., USA: IEEE Press, 2021: 1389-1398.
- [13] GAO T Y, HAN X, ZHU H, et al. FewRel 2.0: towards more challenging few-shot relation classification[EB/OL]. [2023-10-12]. <https://arxiv.org/abs/1910.07124v1>.

- [13] TIAN Y J, ZHAO X X, HUANG W. Meta-learning approaches for learning-to-learn in deep learning: a survey [J]. *Neurocomputing*, 2022, 44: 203-223.
- [14] JIANG W, HUANG K, GENG J, et al. Multi-scale metric learning for few-shot learning [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 31(3): 1091-1102.
- [15] RAVI S, LAROCHELLE H. Optimization as a model for few-shot learning [C] // *Proceedings of International Conference on Learning Representations*. New York, USA: ACM Press, 2017: 135-139.
- [16] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks [EB/OL]. [2023-10-12]. <https://arxiv.org/pdf/1703.03400>.
- [17] GAO T Y, HAN X, LIU Z Y, et al. Hybrid attention-based prototypical networks for noisy few-shot relation classification [C] // *Proceedings of AAAI Conference on Artificial Intelligence*. [S. l.]: AAAI Press, 2019: 6187-6195.
- [18] YE Z X, LING Z H. Multi-level matching and aggregation network for few-shot relation classification [EB/OL]. [2023-10-12]. <https://arxiv.org/abs/1906.06678>.
- [19] QU M, GAO T, KHONNEUX L P, et al. Few-shot relation extraction via bayesian meta-learning on relation graphs [C] // *Proceedings of International Conference on Machine Learning*. New York, USA: ACM Press, 2020: 7867-7876.
- [20] HAN J L, CHENG B, LU W. Exploring task difficulty for few-shot relation extraction [EB/OL]. [2023-10-12]. <https://arxiv.org/abs/2109.05473?context=cs.CL>.
- [21] YANG S, ZHANG Y F, NIU G L, et al. Entity concept-enhanced few-shot relation extraction [EB/OL]. [2023-10-12]. <https://arxiv.org/abs/2106.02401?context=cs.AI>.
- [22] YANG K, ZHENG N, DAI X, et al. Enhance prototypical network with text descriptions for few-shot relation classification [C] // *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. New York, USA: ACM Press, 2020: 2273-2276.
- [23] DONG M P, PAN C G, LUO Z P. MapRE: an effective semantic mapping approach for low-resource relation extraction [EB/OL]. [2023-10-12]. <https://arxiv.org/abs/2109.04108?context=cs>.
- [24] SOARES L B, FITZGERALD N, LING J, et al. Matching the blanks: distributional similarity for relation learning [EB/OL]. [2023-10-12]. <https://arxiv.org/abs/1906.03158>.
- [25] CONG X, YU B W, LIU T W, et al. Inductive unsupervised domain adaptation for few-shot classification via clustering [C] // *Proceedings of Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Berlin, Germany: Springer, 2021: 624-639.
- [26] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional Transformers for language understanding [C] // *Proceedings of Conference on the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. San Diego, USA: Association for Computational Linguistic, 2019: 4171-4186.
- [27] LIU Y, HU J, WAN X, et al. Learn from relation information: towards prototype representation rectification for few-shot relation extraction [C] // *Proceedings of the Association for Computational Linguistics*. San Diego, USA: Association for Computational Linguistic, 2022: 1822-1831.
- [28] 刘昊鑫, 董超, 勾智楠, 等. 融合混合表征的小样本关系抽取方法 [J]. *计算机工程*, 2023, 49(8): 63-68.
- LIU H X, DONG C, GOU Z N, et al. Few-shot relation extraction method fusing with hybrid representation [J]. *Computing Engineering*, 2023, 49(8): 63-68. (in Chinese)

编辑 薛晋栋