计算机科学
COMPUTER SCIENCE

# 基于提示学习与自适应损失加权的汉越产业文本分类

陈霖，马龙轩，张勇丙，黄于欣，高盛祥，余正涛

引用本文

陈霖，马龙轩，张勇丙，黄于欣，高盛祥，余正涛. 基于提示学习与自适应损失加权的汉越产业文本分类[J]. 计算机科学，2026，53(2)：312-321.

CHEN Lin, MA Longxuan, ZHANG Yongbing, HUANG Yuxin, GAO Shengxiang, YU Zhengtao. Industrial Text Classification for Chinese and Vietnamese Based on Prompt Learning and AdaptiveLoss Weighting [J]. Computer Science, 2026, 53(2): 312-321.

---

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.250300038

# 基于提示学习与自适应损失加权的汉越产业文本分类

陈 霖 马龙轩 张勇丙 黄于欣 高盛祥 余正涛
昆明理工大学信息工程与自动化学院 昆明 650500
昆明理工大学云南省人工智能重点实验室 昆明 650500
(chenlin0@stu.kust.edu.cn)

**摘 要** 跨境产业文本分类是支撑跨境产业大数据分析的基础任务。随着东南亚地区跨境产业数据的快速增长,对产业数据的分析和处理,特别是对产业文本分类的需求也在日益增加。然而,当前跨境产业文本面临不同语种间的语言差异、语种间数据不均衡以及标注数据稀缺等问题,尤其在低资源语言中更加突出,导致跨境产业数据分类难度加大。针对这一问题,提出了一种基于提示学习和自适应损失加权策略的少样本跨境产业文本分类方法,显著提升了模型在跨境场景中的分类性能。具体而言,该模型基于提示学习框架缓解数据稀缺问题,利用预训练模型的先验知识增强少样本的学习能力;其次,通过构建跨语言文本对,实现语义空间的知识迁移和语义对齐;同时创新性地设计动态混合损失函数,将交叉熵损失、焦点损失和标签平滑损失进行多目标优化,并基于不确定性加权机制动态调整各损失项权重:交叉熵损失保障基础分类能力,焦点损失强化对难分类样本的关注,标签平滑则有效抑制过拟合风险。实验结果表明,所提方法在中文和越南语产业文本分类任务中显著优于现有主流方法,特别是在数据稀缺和语种不平衡的少样本学习场景下,提供了高效的解决方案,为低资源语言的处理提供了新的研究思路。

**关键词**:跨境产业文本分类;少样本学习;提示学习;自适应损失加权
**中图分类号** TP391

# Industrial Text Classification for Chinese and Vietnamese Based on Prompt Learning and Adaptive Loss Weighting

CHEN Lin,MA Longxuan,ZHANG Yongbing,HUANG Yuxin,GAO Shengxiang and YU Zhengtao
Faculty of Information Engineering and Automation,Kunming University of Science and Technology,Kunming 650500,China
Yunnan Key Laboratory of Artificial Intelligence,Kunming University of Science and Technology,Kunming 650500,China

**Abstract** Cross-border industrial text classification is a fundamental task that supports big data analysis in cross-border industries. With the rapid growth of cross-border industrial data in Southeast Asia,there is an increasing demand for the analysis and processing of industrial data,particularly with respect to industrial text classification. However,cross-border industrial text classification faces several challenges,including linguistic differences across languages,data imbalance among languages,and the scarcity of annotated data. These issues are particularly pronounced in low-resource languages,making cross-border industrial data classification more difficult. To address this issue,this paper proposes a few-shot cross-border industrial text classification method based on prompt learning,combined with an adaptive loss weighting strategy,which significantly enhances the model's classification performance in cross-border scenarios. Specifically,the proposed model mitigates the issue of data scarcity within the prompt-learning framework by leveraging the prior knowledge of pre-trained models to enhance few-shot learning capabilities. Furthermore,cross-lingual text pairs are constructed to facilitate knowledge transfer and semantic alignment in semantic space. Additionally,an innovative dynamic hybrid loss function is designed,integrating cross-entropy loss,focal loss,and label smoothing loss in a multi-objective optimization framework. The loss terms are dynamically weighted based on an uncertainty-based weighting mechanism:cross-entropy loss ensures fundamental classification capability,focal loss enhances the focus on hard-to-classify samples,and label smoothing effectively mitigates the risk of overfitting. Experimental results demonstrate that the proposed method significantly outperforms existing mainstream approaches in cross-border Chinese and Vietnamese industrial text classification tasks,particularly in few-shot learning scenarios with data scarcity and language imbalance. This approach provides an efficient solution and offers new research perspectives for processing low-resource languages.

到稿日期:2025-03-10 返修日期:2025-05-27
基金项目:国家自然科学基金(U23A20388,U21B2027);云南省重点研发计划(202303AP140008,202402AG050007,202302AD080003);云南省基础研究项目(202301AT070393);昆明理工大学"双一流"科技重大专项(202402AG050007)
This work was supported by the National Natural Science Foundation of China(U23A20388,U21B2027),Yunnan Provincial Key Research and Development Program(202303AP140008,202402AG050007,202302AD080003),Yunnan Provincial Basic Research Project(202301AT070393) and Double First-Class Science and Technology Major Project of Kunming University of Science and Technology(202402AG050007).
通信作者:高盛祥(gaoshengxiang.yn@foxmail.com)

**Keywords**　Cross-border industrial text classification，Few-shot learning，Prompt learning，Adaptive loss weighting

## 1　引言

文本分类是自然语言处理（Natural Language Processing，NLP）中的一项关键任务，广泛应用于情感分析、垃圾邮件检测、观点分析等领域[1-2]。在“一带一路”倡议的推动下，中国与东南亚国家的跨境产业合作日益密切，来自新闻站点、社交媒体、政府公告等渠道的非结构化文本数据呈现快速增长态势。以中方规范对跨境产业文本数据进行分类，可以为以中方为主体的跨境产业信息挖掘和决策制定提供重要的支撑。

跨境产业文本分类面临的主要挑战如下。1）不同语种间的语言差异和语料资源不均衡。在将境外产业文本信息对齐到中方规范的过程中，需要进行跨语言的知识迁移和语义对齐。2）标签数据的缺乏。在相对低资源语言一方，产业文本标注数据资源稀缺且标注成本高昂。

传统的产业文本分类研究主要基于单语言场景下的模型优化[3-4]。例如：Ji 等[5]通过引入知识图谱的实体关系信息增强文本表征，提升文本分类效果；Meng 等[6]提出融合标签语义相关性信息的多粒度特征提取方法，提升了模型在标签混淆问题上的表现。上述方法试图引入外部知识或标签信息提升模型对样本的理解能力，但这些方法不仅难以应用于标签数据缺乏的场景，还缺少跨语言的知识迁移，因此无法适用于跨境产业文本分类。

近年来，针对低资源语言场景，少样本学习模型受到越来越多的关注[3]。这类方法旨在利用少量训练样本，实现与全量数据训练后的模型可比的分类性能[4]，同时解决跨境产业文本分类的标注数据稀缺问题。Zhang 等[7]提出结合主题建模与提示微调的增强方法（TPT），通过扩展标签词空间（Verbalizer），增强模型理解单语环境下语义对应关系的能力，但其 Verbalizer 设计无法有效对齐不同语言间的共性特征，难以解决知识迁移问题。Winata 等[8]的研究表明，使用英语数据训练的模型可以在多语言少样本分类任务中取得良好表现，但在特定领域对齐不同语种的语义时，仍存在困难。

为了解决跨境产业文本分类面临的挑战，本文提出了一种融合提示学习与自适应损失加权策略的少样本跨境产业文本分类方法。首先，构建了跨语言提示模板，借助预训练语言模型（Pre-trained Language Model，PLM）的隐式对齐能力捕捉语言间的共性特征，并创新性地将多语言分类任务重构为跨语义空间的文本对相关性估计任务，有效解决了跨境产业文本分类中的知识迁移和语义对齐问题。其次，为了更好地解决语种不平衡问题，设计了一种动态混合损失函数，通过可学习的权重参数动态调节损失贡献度，提升模型对语种不平衡样本分类的鲁棒性，缓解过拟合问题。实验结果表明，所提出的方法在中文和越南语产业文本分类任务中显著优于现有主流方法，也优于用全量数据训练的更大规模的模型。

本文的主要贡献包括以下 3 个方面：

1）针对跨境产业少样本文本分类场景，提出基于语义相关度度量的提示学习方法。该方法将提示学习模板作为跨语言语义对齐的工具，通过扩展文本对构建策略（支持同语言及跨语言输入）并扩大 Verbalizer 的语义覆盖范围，有效适应跨境产业场景。

2）针对跨境产业文本分类语种不平衡问题，设计了一种动态混合损失函数，将交叉熵、焦点损失和标签平滑损失相结合，并通过动态加权缓解语种不平衡问题，增强了模型的泛化能力与鲁棒性。

3）收集了汉越产业文本分类数据集，并通过大量实验验证了所提出方法的有效性。对实验结果进行的详细分析，说明了方法有效的原因。本文使用的代码和数据将在 Github 上公开。

## 2　相关工作

跨境产业文本分类和低资源自然语言处理是近年来的研究热点。由于低资源语言越南语的标注数据稀缺和获取成本高昂，因此针对少样本场景的跨境产业文本分类研究具有重要意义。为应对跨境产业文本分类中有限语言资源的挑战，已有研究从双语词典、机器翻译、跨语言嵌入等角度出发，通过知识迁移、微调和提示学习等技术，提升模型在跨语言文本分类任务中的表现[9-10]。近年来，提示学习在低资源场景中展现出显著优势，进一步推动了相关研究的发展[11]。

### 2.1　提示学习

随着深度学习的发展，训练大量标注数据的需求不断增加，但在很多实际应用中，标注数据稀缺性仍然是制约深度学习广泛应用的主要障碍。为解决这一问题，少样本学习（Few-shot Learning）应运而生，它能够在仅有少量训练样本的条件下实现有效学习并进行高效的分类[12]。

提示学习（Prompt Learning）作为少样本学习的典型方法之一，其核心思想是通过重新构造下游任务，将分类任务转化为掩码语言模型（Masked Language Model，MLM）任务，利用上下文信息推断 MASK 标签对应的词汇，从而充分挖掘预训练模型的潜力[13]。该方法通过定义提示函数将输入文本填充到提示模板中，使 PLM 更自然地适应新任务；模型根据提示后的文本生成词汇表上的概率分布，最终通过词汇映射器（Verbalizer）将输出概率分布映射为目标类别。自 Brown 等[14]提出以来，提示学习已在文本分类[15]、文本生成[16]、命名实体识别[17-18]和知识探测[19-20]等文本挖掘任务中得到广泛应用。随着 GPT-3 等强大语言模型的出现，提示学习在多语言文本分类任务中展现出显著优势，尤其在数据稀缺场景下表现出良好的适应性。

在模板构建方面，早期跨语言自然语言处理任务的研究主要关注离散和软提示的训练策略[21]。Schick 等[22]探讨了手动定义模板在文本分类中的有效性。Shin 等[23]提出了一种基于梯度搜索的方法用于优化模板设计。Liu 等[24]提出了一种基于知识增强的提示学习方法（SKPT），通过引入外部知识（如开放三元组）来优化提示模板，从而提升了少样本文本分类的效果。近年来，Zhou 等[25]研究了通用提示的构建方法，在少样本多语言迁移设置中取得了显著成果，缓解了源语言训练与目标语言推理之间的差异的问题。Dementieva 等[26]提出了一种无监督的跨语言知识迁移方法，避免了手动

标注数据，并利用大规模多语言编码器和翻译系统来提升跨语言文本分类的效果。与现有研究相比，本文方法在提示模板的设计上进行了改进，使得模板能够不依赖于具体任务类别，从而实现更加通用的设计。

### 2.2 词汇映射器

传统微调方法通常通过池化预训练语言模型(PLM)最后一层的隐藏状态生成句子表示，并借助任务特定的分类头进行预测。这种方法在训练数据充足时表现良好，但在少样本场景下容易面临过拟合及迁移适配困难问题。这主要源于任务分类头在少样本下容易过拟合，以及下游任务形式与预训练目标之间的差距影响了模型的迁移能力。为应对上述挑战，提示学习方法被提出。在少样本场景下，提示模型显著缓解了过拟合问题[27]。然而，提示模型的性能在很大程度上依赖于提示模板和词汇映射器的设计[11]。

词汇映射器是提示学习中至关重要的部分，其质量直接决定了模型的性能。Schick 等[22]设计了基于领域知识的手工词汇映射器。然而，手工词汇器的主要缺点在于，在缺乏领域知识的情况下，设计高质量词汇器的成本较高。为避免这一问题，Gao 等[28]和 Wang 等[29]借助预训练语言模型自动生成映射器，Shin 等[23]采用语言模型和梯度引导搜索生成额外的词汇器，Hu 等[30]利用知识图谱和预训练的嵌入丰富提示词汇器，从而提升其生成有效提示的能力。此外，软词汇映射器(Soft Verbalizer)作为一种基于连续空间的方法，近年来受到更多关注。Hambardzumyan 等[31]将映射器视为可优化的连续嵌入，与模型参数一同训练。Cui 等[32]结合原型学习和提示学习，利用原型向量作为软映射器，进一步提升模型性能。还有的研究通过拓展外部知识来提升性能。Ji 等[5]提出了一种基于知识图谱信息约束和知识融合的文化产业文本分类方法，借助知识图谱提取实体和关系信息，从而优化了模型在复杂文本分类中的表现。此外，Sun 等[33]提出了一种结合对比学习和对抗网络的框架，利用外部知识增强少样本分类任务中的元学习过程。该方法通过增强类原型的嵌入质量，提升了模型对新任务的适应能力。这些方法在提升提示学习性能方面展示了显著的效果。

然而，当前方法存在候选词构造优化难度大的问题，限制了在其他任务中的迁移性和通用性。本文在现有方法 MetricPrompt[34]的基础上，对当前的 Verbalizer 进行了拓展，引入更多的相关词汇，以更全面地捕捉上下文信息和语义特征，从而提高模型在分类任务中的性能和适应性。

## 3 本文方法

本文方法框架包括跨境产业文本对构建阶段、训练学习阶段和推理阶段。模型框架如图 1 所示。

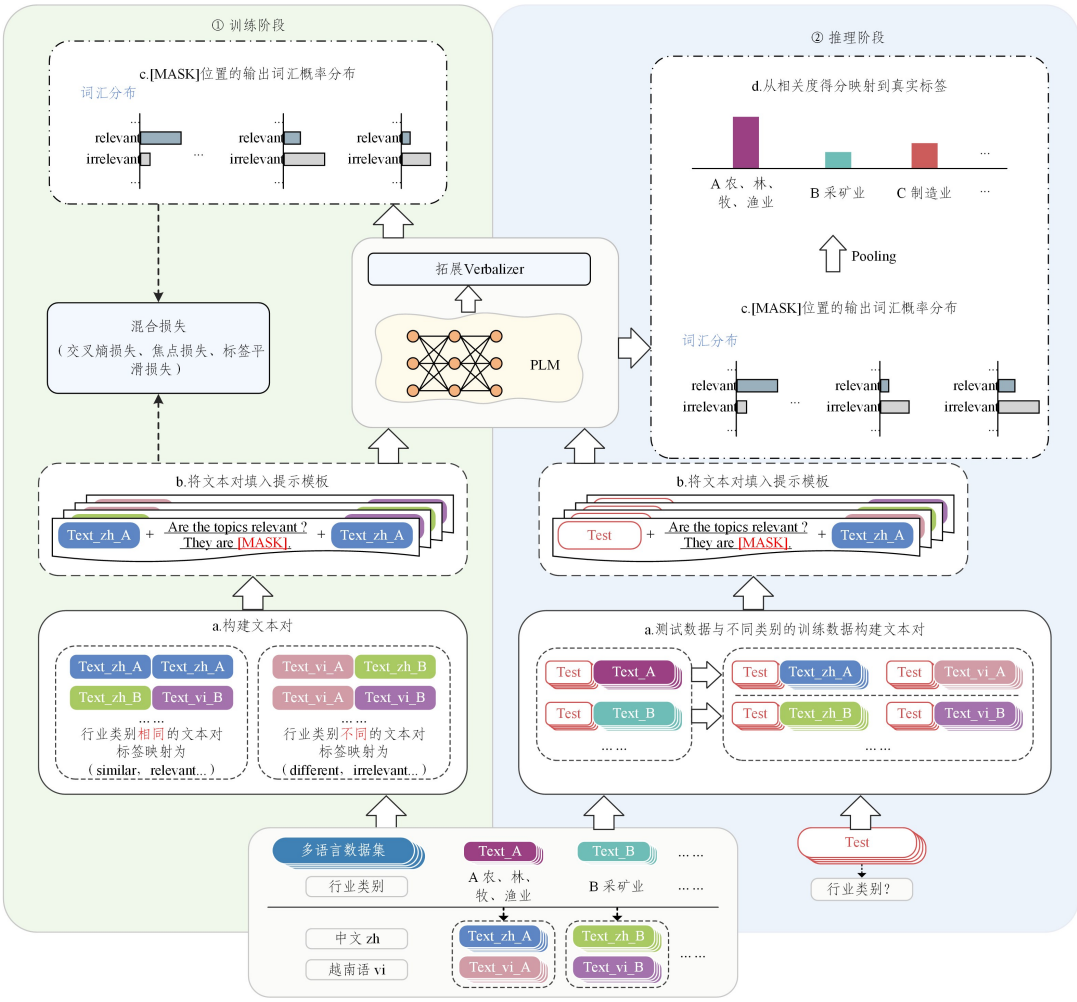


图 1 整体模型框架(电子版为彩图)

Fig. 1 Overall model framework

### 3.1　提示模板构建

为适应提示学习机制，设计并构建了一个通用化提示模板，用于构造产业文本分类任务的输入示例。具体的模板形式可以表示为：

$$x_p=[text1][\mathrm{SEP}]\underline{Are\ the\ topics\ relevant?\ They\ are}[\mathrm{MASK}].[text2] \tag{1}$$

其中，下划线部分表示模板，MASK 标记用于预训练语言模型预测掩码位置，$text1$ 和 $text2$ 分别代表文本对中的两个样本文本。

在实现细节上，分别对[$text1$]和[$text2$]进行了 Segment 标记，用于区分不同的文本段，每个文本段的最大长度是 250。提示模板不进行 Segment 标记。该模板的目标是引导模型识别文本对之间的语义关系，其核心设计在于通过填充[MASK]词元，使模型能够预测出文本对是否属于相同类别。例如，输出“relevant，consistent，similar”表示同一类别，输出“irrelevant，inconsistent，different”表示不同类别。借助这一结构，模板实现了对不同语言文本输入格式的统一，同时有效增强了模型在不同语言语义对齐方面的能力，从而使其在学习过程中能够更加深入地理解并匹配跨语言语义的相似性。

### 3.2　文本对构建

为了充分利用提示学习的优势，对原始数据集$\mathcal{D}$进行了重组，即将原始单样本转换为成对样本，从而将原有的多类别分类问题转换为基于相关性估计的分类任务。这种数据重组策略能够显著提升模型在少样本场景中的表现，尤其是在跨境产业文本分类任务中，有助于模型在不同语言环境下有效捕捉类别间的细微差异。给定一个包含中文和越南语的多语言少样本产业文本分类数据集$\mathcal{D}$，用$\mathcal{D}_t$表示训练数据，用$\mathcal{D}_q$表示查询样本集（即测试集）。具体步骤如下：

#### 3.2.1　构建训练数据

给定训练数据$\mathcal{D}_t$，其中每个样本表示为 $d=(x_d,y_d)$，$x_d$ 表示样本文本，$y_d$ 表示其对应样本的标签。将样本两两配对，生成成对的样本 $T=(d_i,d_j)\mid d_i,d_j\in\mathcal{D}_t$。每对样本包括两个文本段$(x_{d_i},x_{d_j})$，并且以相同类和不同类的配对形式标记相关性标签，使模型能够从多语言输入中学习类别间的相关性特征。比如当构建的是相同类的文本对时，对于相同的类别，可以是两个文本都是中文，或两个文本都是越南语，还可以一个是中文一个是越南语。训练数据构建过程如式(2)所示：

$$\mathcal{D}_t^{\mathcal{M}}=\bigcup_{(d_i,d_j)\in\mathcal{D}_t\times\mathcal{D}_t}\{(p(x_{d_i},x_{d_j}),y_{ij})\} \tag{2}$$

其中，$p(\cdot,\cdot)$是提示函数，将文本对$(x_{d_i},x_{d_j})$填入提示模板，生成模型输入；标签 $y_{ij}=I(y_{d_i}=y_{d_j})$表示该文本对是否属于同一类别，若属于同一类别，则 $y_{ij}=1$，否则，$y_{ij}=0$。这种成对数据生成方法能够有效引导模型学习类别间的相似性和差异性。

在少样本设置中，为了确保不同语言间数据的均衡性，从每个类别中选择 $k$ 个样本，其中$\lceil k/2\rceil$个样本为中文文本，剩余的 $k-\lceil k/2\rceil$个样本为越南语文本。这种样本划分策略能够确保模型在多语言环境下均衡地学习类别特征，减少语言偏差对模型的影响。如果仅随机选取少样本，可能会导致数据偏重于某一特定语言，进而影响模型对类别的普适性判断。因此，均衡划分中文和越南语样本，使得模型能够更充分地学习两种语言间的类别关系，从而减少单一语言对模型训练的干扰。在同类别的中越文本对中，模型可以逐渐识别出类别不变的语义特征；而在不同类别的文本对中，模型则能够学习到类别之间的差异性特征。这种方法不仅能够缓解语言特定特征对模型训练的干扰，还能够促使模型更深入地挖掘类别的本质特征。

#### 3.2.2　构建查询数据

在推理阶段，查询数据集$\mathcal{D}_q$的每个样本也采用成对样本的方式进行重组。具体地，将查询样本与训练数据中的样本进行配对，生成测试样本对 $Q=(d_i,d_j)\mid d_i\in\mathcal{D}_q,d_j\in\mathcal{D}_t$，以便在推理时，模型能够基于文本间的相似性判断来输出分类结果。通过这种方式构建查询数据，模型能够在推理阶段利用对文本间相关性的估计来进行类别判定。查询数据构建过程如式(3)所示：

$$\mathcal{D}_q^{\mathcal{M}}=\bigcup_{(d_i,d_j)\in\mathcal{D}_q\times\mathcal{D}_t}(p(x_{d_i},x_{d_j}),\hat{y}_{ij}) \tag{3}$$

其中，$\hat{y}_{ij}$表示待学习的相关性标签。

整体数据构建过程如图 1 所示，包括训练阶段和推理阶段。通过上述数据构建过程，原本的多类别分类任务被转换为基于成对样本的相关性估计任务。这一方法特别适用于多语言产业数据的实际需求，能够有效支持模型在少样本、跨语言的情境下学习文本对之间的相似性和一致性，从而实现跨语言语义对齐，提高分类任务的准确性。尤其在训练数据有限的情况下，成对样本的设计可以显著增强模型的泛化能力，充分利用提示学习在少样本和多语言环境下的优势，帮助模型在复杂的多语言少样本场景中更有效地捕捉类别关联，从而在跨境产业文本分类任务中表现得更加稳健。

### 3.3　Verbalizer 构建

在提示学习中，Verbalizer 是将模型预测的词汇映射到类别标签的关键组件，其设计质量直接影响分类性能。然而，在少样本多语言场景下，现有方法（如 MetricPrompt[34]）中的 Verbalizer 候选词汇数量相对有限，难以全面覆盖跨语言场景中的多样化语义表达，进而限制了模型的分类能力和泛化性能。

为了解决上述问题，受 KPT[30] 的启发，认识到扩展 Verbalizer 的重要性，特别是在少样本和多语言场景下，构建与特定类别相关的标签词集合对于提升分类性能至关重要。本文采用 Related Words[1)] 作为外部知识资源，从中检索与映射标签最相关的 15 个词汇。通过引入更多同义词和关联词汇，对 Verbalizer 进行了扩展，从而扩大其语义覆盖范围。扩展后的 Verbalizer（记作$V_{\mathrm{Ext}}$）能更精准地捕捉语义差异，有效提升了模型在多语言文本分类任务中的性能。

具体而言，Verbalizer 的核心功能是定义词汇表中标签词

1) https://relatedwords.org

集合 $V$ 与标签空间 $Y$ 之间的映射关系，通常表示为一个映射函数 $f(\cdot):V\to Y$。在预训练语言模型(PLM)$M$ 的基础上，Verbalizer 的作用是通过预测[MASK]位置上的词汇，间接推断出文本所属的类别。具体而言，对于经过模板处理后的输入文本 $x_p$，模型对标签词集合的概率分布可表示为：

$$P(y\in Y\mid x_p)=P_M([\text{MASK}]=v\in V_y\mid x_p) \tag{4}$$

其中，$V_y$ 表示被映射到标签 $y$ 的词汇子集，映射关系由 $f(\cdot)$ 给出；$P_M([\text{MASK}]=v\in V_y\mid x_p)$ 表示(PLM)$M$ 在给定输入文本 $x_p$ 的条件下，预测[MASK]位置填充为单词 $v$ 的概率。通过这一机制，文本分类任务可以转换为标签词概率预测问题。例如，将原始的 $V=\{\text{relevant}\}$ 扩展为 $V=\{\text{consistent, similar, aligned, pertinent, related, …}\}$，使得 Verbalizer 能更全面地表达类别语义。

### 3.4 优化方法

在跨境产业数据少样本分类任务中，数据通常来源于真实的互联网资讯，难免受到标签噪声、跨语言及样本稀疏性的影响。单一的交叉熵损失在此类任务中表现出一定的局限性，主要体现在其对主导类别优化的倾向性，这导致在少数类别或难分类样本上的表现较差。此外，互联网数据中的标签噪声源于标注主观性和语言表达差异，进一步加大了模型的学习难度。为了解决这些问题，本文采用了一种动态混合损失函数(Dynamic Hybrid Loss，DHL)，并将其应用于本文模型，以优化少样本任务。

#### 3.4.1 优化目标

将少样本文本分类任务重新构建为文本对的相关性估计任务，相比传统的软标签设计方法，其减少了对特定任务词表的需求。通过这种任务重构，我们的优化目标与预训练语言模型(PLM)的预训练目标更加一致。

设 $P(\cdot;\theta)$ 为一个由参数 $\theta$ 参数化的掩码语言模型(MLM)，$P_{\text{vocab}}(\cdot;\theta)$ 为该模型在[MASK]位置上的输出词概率。本文将优化目标定义为：

$$\begin{aligned}\hat{\theta}&=\arg\min_{\theta}\sum_{d^M\in\mathcal{D}_t^M}\mathcal{L}(P_{\text{vocab}}(x_{d^M};\theta),y_{d^M})\\&=\arg\min_{\theta}\sum_{d^M\in\mathcal{D}_t^M}\mathcal{L}_{\text{DHL}}(f(P_{\text{vocab}}(x_{d^M};\theta)),\phi(y_{d^M}))\\&=\arg\min_{\theta}\sum_{d^M\in\mathcal{D}_t^M}\mathcal{L}_{\text{DHL}}(f_{\text{cls}}(x_{d^M};\theta),\phi(y_{d^M}))\end{aligned} \tag{5}$$

其中，$\mathcal{L}_{\text{DHL}}$ 为该模型的损失函数，其形式为包含交叉熵损失(Cross-Entropy)、标签平滑损失(Label Smoothing)和焦点损失(Focal Loss)的加权组合；$\phi(\cdot)$ 表示标签类别上的概率分布，输入样本的正确标签位置设为 1，其余位置设为 0。$f(\cdot)$ 表示一个预定义的、任务通用的标签映射词表(Verbalizer)，用于将输出词概率分布 $P_{\text{vocab}}(\cdot;\theta)$ 映射到二分类概率分布 $f_{\text{cls}}(\cdot;\theta)$ 上。具体而言，该词表将{relevant, similar, consistent}对应的 logits 归为标签 1 的预测得分，同时将{irrelevant, inconsistent, different}的 logits 归为标签 0 的预测得分。

#### 3.4.2 动态混合损失

在本文的损失设计中，采用 3 种损失的加权组合，以兼顾分类性能、泛化能力和对难样本的关注度。设 $\mathcal{L}_{\text{ce}}$ 为交叉熵损失，$\mathcal{L}_{\text{ls}}$ 为标签平滑损失，$\mathcal{L}_{\text{fl}}$ 为焦点损失。

交叉熵损失(Cross-Entropy)虽然在常规分类任务中有效，但在少样本、多语言场景下的泛化性较弱，且实验结果不稳定，具有较大的波动性。交叉熵损失定义为：

$$\mathcal{L}_{\text{ce}}=-\sum_{k=1}^{N}y_k\log P_{\text{vocab}}(x_{d^M};\theta)_k \tag{6}$$

其中，$y_k$ 为第 $k$ 个类别的真实标签分布(独热编码)，$P_{\text{vocab}}(\cdot;\theta)$ 为该模型在[MASK]位置上的输出词概率，$N$ 是类别数量。

针对数据噪声及跨语种的问题，本文引入焦点损失(Focal Loss)以对难分类样本赋予更高权重，增强模型对少数类别的关注。Lin 等[35]验证了焦点损失在不平衡数据上的有效性，这为焦点损失在少样本分类任务中的应用提供了理论基础。焦点损失通过对难分类的样本赋予更高的权重，来解决衡语种不平衡问题。其定义为：

$$\mathcal{L}_{\text{fl}}=-\alpha(1-p_t)^{\gamma}\log(p_t) \tag{7}$$

其中，$p_t$ 表示模型对目标类别的预测概率；$\alpha$ 为样本平衡因子；$\gamma$ 为聚焦因子，用于控制对难样本的关注程度。

为缓解标签噪声对模型的干扰，本文同时采用标签平滑(Label Smoothing)策略，将“独热”标签分布平滑化，以弱化模型对单一类别的过度拟合。Szegedy 等[36]指出，标签平滑不仅能够提升模型的鲁棒性，还能减少对错误标签的敏感性。标签平滑损失通过将真实标签分布转换为平滑分布，以减少模型对噪声标签的过度拟合，定义如式(8)所示：

$$\mathcal{L}_{\text{ls}}=-\sum_{k=1}^{N}\tilde{y}_k\log P_{\text{vocab}}(x_{d^M};\theta)_k \tag{8}$$

其中，$\tilde{y}_k$ 是经过平滑处理的标签分布，满足 $\tilde{y}_k=(1-\beta)y_k+\beta/N$，$\beta$ 为平滑系数。

在传统方式下，多项损失的加权组合需要人为调整这些损失的权重 $\lambda_1$，$\lambda_2$，$\lambda_3$(即超参数)，如式(9)所示：

$$\mathcal{L}_{\text{total}}=\lambda_1\mathcal{L}_{\text{ce}}+\lambda_2\mathcal{L}_{\text{fl}}+\lambda_3\mathcal{L}_{\text{ls}} \tag{9}$$

然而，在实际应用中，手动调整超参数的方式不仅增加了寻找最优超参数组合的成本，还降低了训练过程的灵活性，难以适应不同任务对损失权重的动态需求。例如，在优化过程中，某一损失项可能主导训练，导致模型性能受限。因此，本文借鉴了 Cipolla 等[37]提出的自适应损失加权(Adaptive Loss Weighting)方法，引入可学习的不确定性参数 $\sigma_j$，通过权重项 $1/2\exp(\sigma_j)$ 动态调整 3 种损失函数(Cross-Entropy, Label Smoothing, Focal Loss)的权重。本文定义的动态混合损失如式(10)所示：

$$\begin{aligned}\mathcal{L}_{\text{DHL}}&=\sum_{j\in\{ce,fl,ls\}}\frac{\mathcal{L}_j}{2\exp(\sigma_j)}+\sigma_j\\&=\frac{\mathcal{L}_{\text{ce}}}{2\exp(\sigma_{\text{ce}})}+\sigma_{\text{ce}}+\frac{\mathcal{L}_{\text{fl}}}{2\exp(\sigma_{\text{fl}})}+\sigma_{fl}+\\&\quad\frac{\mathcal{L}_{\text{ls}}}{2\exp(\sigma_{\text{ls}})}+\sigma_{\text{ls}}\end{aligned} \tag{10}$$

其中，$\mathcal{L}_{\text{ce}}$ 为交叉熵损失，提供了基本的分类能力；$\mathcal{L}_{\text{fl}}$ 为焦点损失，对难分类样本的优化尤为重要；$\mathcal{L}_{\text{ls}}$ 为标签平滑损失，可以缓解过拟合；$\sigma_{\text{ce}}$，$\sigma_{\text{fl}}$，$\sigma_{\text{ls}}$ 是可学习参数，表示对应损失项的不确定性。训练过程中，模型通过权重项 $1/2\exp(\sigma_j)$ 动态调整损失影响：当 $\sigma_j$ 增大时，对应损失权重指数衰减，表明模型认为该任务不确定性较高(如噪声干扰或优化困难)，从而降低其干扰；而 $\sigma_j$ 自身作为正则化项，通过线性惩罚约束参数增长，迫使模型在降低损失权重与抑制 $\sigma_j$ 值之间进行平衡，避免单一损失项因梯度贡献过大而主导优化方向。通过联合优化

$\mathcal{L}_j$ 与 $\sigma_j$，模型可自主实现“高不确定性损失降权、低不确定性损失聚焦”的动态平衡，无需人工干预即可适应任务复杂性与数据分布变化。

### 3.5 推理阶段

在完成模型优化后，提示模型在推理阶段被用作一种相关性度量方法。如图 1 所示，将查询样本 $d_q$（以红色表示）与所有的训练样本（以不同的颜色表示）进行配对，生成推理阶段的样本对。对于每个原始的训练样本 $d_i$，通过式(11)计算训练样本 $d_i$ 与查询样本 $d_q$ 的相关性得分 $s_{d_i}$：

$$s_{d_i}=\Delta(f_{\text{cls}}(p(x_{d_q},x_{d_i});\hat{\theta})) \tag{11}$$

其中，$p(x_{d_q},x_{d_i})$ 表示输入的查询样本与训练样本的拼接形式；$f_{\text{cls}}(\cdot;\hat{\theta})$ 表示分类模型的输出函数，参数 $\hat{\theta}$ 为已优化的模型参数；$\Delta(\cdot)$ 用于计算二项分布在标签为 1 和 0 上的概率差。

在少样本分类任务中，用 $l$ 表示类别，定义 $\mathcal{D}_l=\{d_i\mid d_i\in\mathcal{D}_t,y_{d_i}=l_i\}$，即类别 $l$ 对应的所有训练样本集合。其中，$\mathcal{D}_t$ 表示训练样本的全集；$y_{d_i}$ 表示训练样本 $d_i$ 的真实标签。通过汇总查询样本 $d_q$ 与 $\mathcal{D}_l$ 中样本的相关性得分，采用式(12)计算类别 $l$ 的分类得分 $s_l$：

$$s_l=\frac{\sum_{d_i\in\mathcal{D}_l}s_{d_i}}{|D_l|} \tag{12}$$

其中，$|\mathcal{D}_l|$ 表示类别 $l$ 中样本的数量。

最终，根据所有分类得分选择最相关的类别作为预测结果：

$$\hat{l}=\arg\max_l s_l \tag{13}$$

通过上述方法，本文将提示模型转换为一种相关性度量工具。由于提示模型能够一次输入两段文本样本，因此可以利用跨样本跨语言的交互信息来更精确地估计相关性。在少样本任务中，本文模型能够基于训练样本的相关性对查询样本进行准确分类，从而显著提升推断阶段的性能。

## 4 实验

### 4.1 数据集及评价指标

#### 4.1.1 实验数据集

本文的数据来源涵盖金融业、采矿业、制造业等九大行业领域，具体数据来自中文和越南语的产业新闻网站，均为公开的互联网数据（如环球网、越通社等平台）。

在数据收集过程中，共获取了 130 539 条产业新闻数据，其中中文数据 113 781 条，越南语数据 16 758 条，具体数据统计如表 1 所列。这些数据涵盖多行业和多语言特性，来源广泛且具有代表性，为行业分类研究提供了基础和数据支持。由于部分中文和越南语类别数据缺乏现成完整标签，因此本文采用细粒度数据采集策略。以“农、林、牧、渔业”为例，分别收集农业、林业、畜牧业、渔业等相关数据，构建完整的类别数据集。在数据划分方面，按 8∶2 的比例将数据集划分为训练集和测试集。此外，测试集经过人工校验与补充标注，确保标注质量。

表 1　数据集统计情况

Table 1　Statistical overview of the datasets

| 行业名称 | 标签 | 总数据量 | 中文数据量 | 越南语数据量 |
|---|---|---|---|---|
| 农、林、牧、渔业 | 0 | 23 563 | 20 243 | 3 320 |
| 采矿业 | 1 | 4 133 | 3 880 | 253 |
| 制造业 | 2 | 8 462 | 8 252 | 210 |
| 交通运输、仓储和邮政业 | 3 | 19 433 | 18 854 | 579 |
| 金融业 | 4 | 10 340 | 7 425 | 2 915 |
| 房地产业 | 5 | 10 695 | 9 958 | 737 |
| 教育 | 6 | 12 925 | 9 995 | 2 930 |
| 卫生和社会工作 | 7 | 18 556 | 15 590 | 2 966 |
| 文化、体育和娱乐业 | 8 | 22 432 | 19 584 | 2 848 |

#### 4.1.2 评价指标

本文采用精准率（Precision）、召回率（Recall）和 F1 分数（F1）作为评估指标，用于衡量产业文本分类的性能。上述指标在后续章节中均基于加权平均进行计算，以确保评价结果能准确反映模型在各类别上的整体表现。

### 4.2 实验参数设置

本文在 2-shot，4-shot 和 8-shot 设置下开展少样本分类实验，实验实现基于 PyTorch[38] 和 Huggingface[39] 框架完成。为确保对比的公平性，实验模型与基线模型均采用 BERT-base-multilingual-uncased[40] 作为骨干模型，所有实验均在 NVIDIA RTX 3090 上运行。模型参数通过 AdamW 优化器[41] 进行优化，学习率设定为 $1\times10^{-5}$，批大小设置为 8。模型微调的训练轮数根据训练集的 shot 数量动态调整，2-shot，4-shot，8-shot 设置下的训练轮数分别为 35，25，15。

### 4.3 基线方法

本文选择以下方法作为基线模型。

1）mBERT[40]：利用 mBERT 得到的 768 维句子嵌入作为文本表示，并通过线性层将其映射到具体类别进行分类。

2）PET[22]：使用手动模板来构造提示，通过在输入文本前添加前缀或后缀，并屏蔽掉某些标记，将输入实例转换为填空形式的短语，以帮助语言模型理解之前给定的任务。

3）PBML[42]：结合提示机制与元学习框架，通过分配标签词学习和模板学习，提升少样本文本分类的准确性，同时减少元学习对大规模数据的依赖。

4）MetricPrompt[34]：将少样本文本分类转换为相关性估计任务，利用提示模型作为相关性度量，并通过交叉熵损失对模型进行监督训练。

5）KPT[30]：引入外部知识来扩展标签词的搜索空间，并在预测前使用预训练语言模型细化扩展的标签词空间，旨在提升和稳定 prompt-tuning 的效果。

### 4.4 主实验

为了评估本文方法在中文和越南语文本分类的低资源场景下的性能，针对不同少样本（few-shot）条件下的产业文本分类任务进行了实验。具体而言，在 2-shot，4-shot 和 8-shot 的设置下，采用 mBERT，PET 和 MetricPrompt 等基线模型进行性能对比。详细实验结果如表 2 所列。

从表 2 的实验结果可以看出，在少样本场景下，本文方法在不同样本规模和评估指标上均优于基线模型，展现了强大的性能优势。具体而言，在 2-shot，4-shot 和 8-shot 3 种设置下，本文方法的平均 F1 值分别达到了 75.11%，80.35% 和

83.75%，相较于性能较优的基线模型 MetricPrompt 分别提升了 11.65 个百分点、6.11 个百分点和 4.01 个百分点。值得注意的是，本文方法在 4-shot 条件下的性能已超过所有基线模型在 8-shot 设置下的最高水平。这一结果表明，本文方法在极低样本场景下能够显著提升文本分类性能，实现更优的分类效果。

表 2 不同少样本设置下的实验结果

Table 2 Experimental results under different few-shot settings

(%)

| *k*-shot | Model | *Precision* | *Recall* | *F*1 |
|---|---|---|---|---|
| 2-shot | mBERT | 60.20 | 56.48 | 55.73 |
| | PET | 32.61 | 27.51 | 26.06 |
| | PBML | 63.23 | 63.23 | 63.23 |
| | MetricPrompt | 66.69 | 64.33 | 63.46 |
| | KPT | 73.27 | 72.61 | 71.72 |
| | Ours | **76.63** | **75.40** | **75.11** |
| 4-shot | mBERT | 64.52 | 65.61 | 64.64 |
| | PET | 40.09 | 38.23 | 36.61 |
| | PBML | 69.10 | 69.10 | 69.09 |
| | MetricPrompt | 76.80 | 74.37 | 74.24 |
| | KPT | 78.18 | 77.51 | 77.21 |
| | Ours | **81.85** | **80.37** | **80.35** |
| 8-shot | mBERT | 78.32 | 78.70 | 77.26 |
| | PET | 61.58 | 56.22 | 56.15 |
| | PBML | 72.77 | 72.77 | 72.77 |
| | MetricPrompt | 80.85 | 79.67 | 79.74 |
| | KPT | 80.41 | 78.83 | 77.79 |
| | Ours | **84.45** | **83.72** | **83.75** |

在实验设计中，本文对训练数据进行了语言均匀采样，将样本划分为等量的中文和越南语，使模型能够更加均衡地学习两种语言的类别特征。这种设计有效减少了单一语言对模型学习产生的偏移问题，进一步提升了分类效果。尤其是在 2-shot 设置下，本文方法的 F1 值达到了 75.11%，显著高于其他基线模型，相较于基线模型 mBERT 提升了 19.38 个百分点，且显著优于最优的基线模型 KPT(+3.39 个百分点)。这充分说明了语言均衡采样策略在极低资源场景下的关键作用。

针对越南语任务，本文采用参数量更大的预训练模型 mT5-base 作为基线对比模型，在完整训练集上进行有监督训练，并与本文方法在 8-shot 场景下的训练效果进行对比分析，结果如表 3 所列。

表 3 越南语任务中 mT5 模型与本文模型在 8-shot 设置下的性能对比

Table 3 Performance comparison between the mT5 model and the proposed method under the 8-shot setting on the Vietnamese task

(%)

| Model | *Precision* | *Recall* | *F*1 |
|---|---|---|---|
| mT5 | 80.53 | 80.40 | 80.35 |
| Ours_8-shot | **84.84** | **84.52** | **84.55** |

从表 3 的实验结果可以看出，在每个类别仅使用 8 个标注样本的极端低资源场景下，本文方法在精确率、召回率和 F1 值 3 项核心指标上均实现了显著提升，全面超越基于完整训练集训练的大规模预训练模型 mT5。这有力印证了本文框架在跨语言知识迁移方面的有效性，其通过构建深度语义关联，成功实现了多语言表征空间的协同优化。值得注意的是，该方法不仅突破了传统预训练模型对海量标注数据的依赖，更通过参数共享机制，在保持模型轻量化的同时，显著提升了单一语言的语义捕获能力。

### 4.5 消融实验

为评估 Verbalizer 拓展对性能的影响，本文在不同的少样本设置下进行了消融实验，实验结果如表 4 所列。

表 4 Verbalizer 的消融实验

Table 4 Ablation study of the Verbalizer

(%)

| k-shot | Model | *Precision* | *Recall* | *F*1 |
|---|---|---|---|---|
| 2-shot | Ours | **76.63** | **75.40** | **75.11** |
| | Ours w/o $V_{Ext}$ | 75.65 | 73.98 | 73.39 |
| 4-shot | Ours | **81.85** | 80.37 | 80.35 |
| | Ours w/o $V_{Ext}$ | 81.81 | **80.59** | **80.47** |
| 8-shot | Ours | **84.45** | **83.72** | **83.75** |
| | Ours w/o $V_{Ext}$ | 83.77 | 83.30 | 83.28 |

从表 4 可以看出，拓展后的 Verbalizer 对模型性能的提升尤为重要。在极低样本量的 2-shot 场景下，使用 Verbalizer 扩展(Ours)的平均 F1 值比未使用 Verbalizer 扩展(Ours w/o $V_{Ext}$)提升了 2.34%，说明通过 Verbalizer 扩展提供的更多相关语义信息，能够帮助模型更全面地捕捉特征，进一步提高分类性能。而在 4-shot 及以上场景，模型对扩展信息的依赖度降低，说明样本量增加时，基础语义特征已能支撑模型决策。

与此同时，为了进一步验证本文提出的自适应损失加权方法($\mathcal{L}_{DHL}$)的有效性，在 2-shot 设置下设计了消融实验。通过对数据进行随机采样(每次采样 80 条文本，允许出现单一语种主导情况，如全为越南语样本)，验证该方法在极少样本和语种不平衡条件下的性能表现。为降低数据选择过程中随机性的影响，分别进行了 5 组实验，并以模型在相同测试集上的平均性能作为最终结果。实验结果如表 5 所列。

表 5 在 2-shot 少样本场景下的消融实验

Table 5 Ablation study under the 2-shot setting

(%)

| Model | *Precision* | *Recall* | *F*1 |
|---|---|---|---|
| Ours | **68.97** | **66.11** | **65.83** |
| Ours w/o $\mathcal{L}_{ce}$ | 64.94 | 62.06 | 61.98 |
| Ours w/o $\mathcal{L}_{fl}$ | 67.75 | 65.10 | 64.52 |
| Ours w/o $\mathcal{L}_{ls}$ | 67.31 | 64.73 | 63.79 |

从表 5 的实验结果可以看出，在 2-shot 场景下，采用 $\mathcal{L}_{DHL}$ 完整损失组合(Ours)的方法，其 F1 值达到 65.83，显著优于各个损失单独移除的情况。从 F1 值来看，与只移除交叉熵损失(Ours w/o $\mathcal{L}_{ce}$)相比，完整损失组合的 F1 值提升了约 3.85 个百分点；与移除焦点损失(Ours w/o $\mathcal{L}_{fl}$)的64.52%和移除标签平滑损失(Ours w/o $\mathcal{L}_{ls}$)的 63.79%相比，分别提升了 1.31 个百分点和 2.04 个百分点。在 Precision 和 Recall 指标上也呈现出相同趋势，完整损失组合分别以 68.97%和 66.11%领先于其他变体。这些结果验证了多损失协同优化策略的有效性。

值得注意的是，移除任何单一损失项都会导致模型性能下降。其中，交叉熵损失的缺失导致 F1 值下降幅度最大(相对降幅 6.21%)，这表明交叉熵损失在提升模型基础分类性

能方面具有关键作用。相比之下，移除焦点损失和标签平滑损失分别导致 F1 值相对完整损失组合下降 2.03% 和 3.20%，这表明二者在缓解样本不均衡和防止过拟合方面具有互补作用。

上述结果表明，自适应混合损失方法能够有效适应多语言少样本场景的特点。通过在训练过程中动态调整不同损失项的权重，$\mathcal{L}_{\mathrm{DHL}}$ 显著降低了多语言间语种分布不平衡对模型性能的负面影响。这种动态加权策略使模型在优化过程中更为平衡，能够更充分地利用语言间的共享信息，同时避免语种分布不均导致的偏倚，从而提高模型在产业文本分类任务中的整体表现，实现更高的准确性和鲁棒性。

### 4.6　案例分析

为验证模型在不同语言场景下对细粒度语义特征的捕捉能力，对中文和越南语语料各随机选取一个示例样本进行分析。选取的具体示例如表 6 所列。通过对比本文方法与 MetricPrompt 基线模型在相同样本上的不同类别关联度得分分布，揭示不同模型在细粒度分类任务中的决策差异，并使用热力图进行可视化分析。

表 6　选取的示例样本

Table 6　Selected example samples

| 语言 | 标签(行业类别) | 示例 |
|---|---|---|
| 中文 | 1(采矿业) | “俄罗斯亿万富翁 AlisherUsmanov 控股的贝加尔湖矿业公司周二表示，该公司已开始在西伯利亚东部偏远地区乌多坎铜矿兴建大型采矿和冶金工厂。该矿床的铜最早于 20 世纪 40 年代被发现，由于矿石的特性，以及该矿床地处偏远地区，基础设施落后、地震活动频繁、冻土条件恶劣，因此很难提取。乌多坎铜矿的总储量约为 2670 万吨，是俄罗斯最大的未开发铜矿，也是世界上最大的铜矿之一。” |
| 越南语 | 4(金融业) | “Theo đó, ông Bùi Thành Lâm (địa chỉ tại 88 Sương Nguyệt Ánh, phường Bến Thành, quận 1, TP Hồ Chí Minh) bị xử phạt vi phạm hành chính theo Quyết định số 1006/QĐ-XPHC, mức tiền phạt 75 triệu đồng, do ông Lâm đã giao dịch ngoài khoảng thời gian đăng ký. Trước đó, ông Bùi Thành Lâm là thành viên Hội đồng quản trị Công ty Cổ phần Tập đoàn Bamboo Capital (mã chứng khoán: BCG), đăng ký bán 1.000.000 cổ phiếu BCG vào ngày 3/10/2022. Tuy nhiên, ông Lâm đã bán 1.000.000 cổ phiếu BCG (tương ứng với10.000.000.000đồng tính theo mệnh giá cổ phiếu BCG) vào ngày 18/10/2022. Cùng việc bị phạt tiền, ông Bùi Thành Lâm còn bị đình chỉ hoạt động giao dịch chứng khoán 2 tháng, quy định tại điểm a khoản 7 Điều 33 Nghị định số 156/2020/NĐ-CP được sửa đổi, bổ sung theo quy định tại khoản 27 Điều 1 Nghị định số 128/2021/NĐ-CP.” |

通过训练完成的模型参数，分别计算两个模型在九大行业类别上的标准化概率分布。图 2、图 3 分别展示了中文和越南语样本的热力图可视化结果，其中横轴表示行业类别，热力值表示经 Min-Max 归一化处理的预测置信度得分。

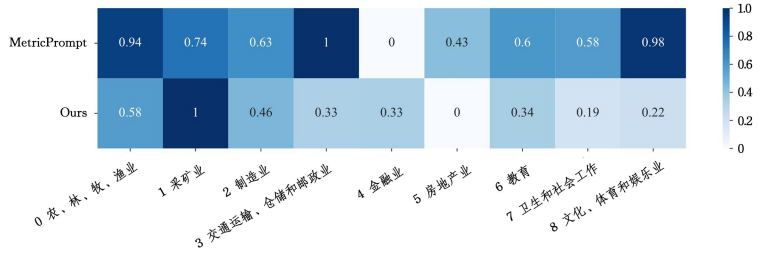


图 2　中文示例中各类别得分可视化

Fig. 2　Visualization of category scores in Chinese examples

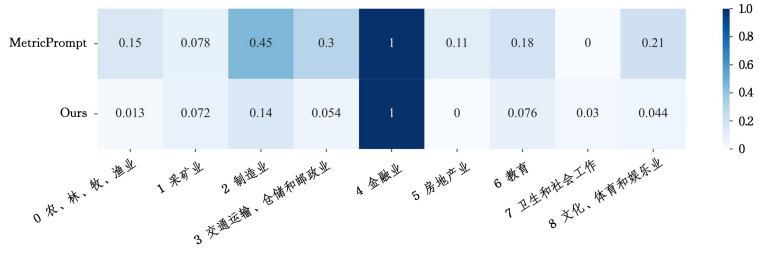


图 3　越南语示例中各类别得分可视化

Fig. 3　Visualization of category scores in Vietnamese examples

从热力图可以看出，本文方法展现出更优的类别区分能力。在中文样本分析中(见图 2)，MetricPrompt 模型在正确类别“采矿业”(标签 1)的置信度仅为 0.74，而对于标签 0、标签 3 和标签 8 都有着较高的干扰置信度，反映出基线模型对“矿业公司”等行业术语的误判倾向。相比之下，本文方法在正确类别中获得显著区分度(1.00 vs. 次高 0.58)，且非相关类别的最大置信度降幅达 77.6%(标签 8：0.98→0.22)，有效验证了本文方法在抑制语义漂移方面的优势。

对于越南语样本(见图 3)，本文方法展现出更加优异且更清晰的决策边界。虽然两个模型均在正确类别“金融业”(标签 4)取得峰值置信度，但本文方法的次高置信度类别得分仅为 0.14(标签 2：制造业)，显著低于 MetricPrompt 的次高得分 0.45(标签 2)，降幅为 68.9%。值得注意的是，MetricPrompt 在语义关联性较低的“教育”(标签 6)和“文化、

体育和娱乐业”(标签8)中分别产生了0.18和0.21的干扰得分,而本文方法将这两类干扰项的置信度压制在0.076以下,显著降低了跨行业误判风险。这一现象表明,本文的训练策略能够有效捕捉金融领域中的判别性特征,如“证券交易”和“违规处罚”等,而非单纯依赖关键词匹配。

进一步分析发现,在中文样本中,本文方法对于6个非相关类别的置信度严格收敛于0.4以下;在越南语样本中,5个非相关类别的置信度均低于0.06。这一突出的概率分布特征得益于本文设计的动态混合损失函数,该函数在增强类间差异性和类内紧密度方面起到了关键作用,从而显著提升了模型对行业特征边界的学习能力。

**结束语** 本文针对跨境产业文本分类任务,提出了一种融合提示学习与动态混合损失函数的分类方法。该方法扩展了提示学习中文本对的构建方式,支持同语言(如中文-中文、越南语-越南语)和跨语言(如中文-越南语)输入,从而为跨境产业文本分类任务提供了更灵活的解决方案。在提示学习基础上,结合交叉熵损失、焦点损失和标签平滑损失,并引入基于不确定性加权的动态调节机制,使模型能够在语种分布不平衡的多语言输入条件下实现高效适配和优化。与此同时,本文通过对Verbalizer的拓展,进一步扩大了其语义覆盖范围,使模型在多语言提示学习任务中能够更准确地理解和生成语义信息,从而提升分类的鲁棒性和泛化能力。在汉越产业数据集上的实验结果表明,本文方法显著提升了少样本场景下模型的分类性能,有效缓解了语言间差异对分类任务的影响。

在未来的研究工作中,计划进一步探索适用于多语言场景的连续提示模板生成策略,通过引入更加灵活的模板生成机制,更有效地捕捉文本间的语法和语义关系。此外,将尝试结合无监督对比学习方法,增强模型的表示能力,以进一步提升其在低资源和少样本场景中的性能。

## 参 考 文 献


[1] BRAUWERS G, FRASINCAR F. A survey on aspect-based sentiment classification[J]. ACM Computing Surveys, 2022, 55(4): 1-37.

[2] MINAEE S, KALCHBRENNER N, CAMBRIA E, et al. Deep learning—based text classification: a comprehensive review[J]. ACM Computing Surveys, 2021, 54(3): 1-40.

[3] GU Y, HAN X, LIU Z, et al. PPT: Pre-trained Prompt Tuning for Few-shot Learning[C]// Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. 2022: 8410-8423.

[4] LU Y, 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. 2022: 5755-5772.

[5] JI X. A cultural industry text classification method based on knowledge graph information constraints and knowledge fusion [J]. International Journal of Web Engineering and Technology, 2024, 19(2): 127-147.

[6] MENG C, TODO Y, TANG C, et al. MFLSCI: Multi-granularity fusion and label semantic correlation information for multi-label legal text classification[J]. Engineering Applications of Artificial Intelligence, 2025, 139: 109604.

[7] ZHANG Y, XU Y, DONG F. An enhanced few-shot text classification approach by integrating topic modeling and prompt-tuning[J]. Neurocomputing, 2025, 617: 129082.

[8] WINATA G I, MADOTTO A, LIN Z, et al. Language Models are Few-shot Multilingual Learners[C]// Proceedings of the 1st Workshop on Multilingual Representation Learning. 2021: 1-15.

[9] QI K, WAN H, DU J, et al. Enhancing cross-lingual natural language inference by prompt-learning from cross-lingual templates [C]// Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. 2022: 1910-1923.

[10] CHEN Y, HARBECKE D, HENNIG L. Multilingual Relation Classification via Efficient and Effective Prompting[C]// Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. 2022: 1059-1075.

[11] LIU P, YUAN W, FU J, et al. prompt, and predict: A systematic survey of prompting methods in natural language processing [J]. ACM Computing Surveys, 2023, 55(9): 1-35.

[12] WANG Y, YAO Q, KWOK J T, et al. Generalizing from a few examples: A survey on few-shot learning[J]. ACM Computing Surveys, 2020, 53(3): 1-34.

[13] SALAZAR J, LIANG D, NGUYEN T Q, et al. Masked Language Model Scoring[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020: 2699-2712.

[14] BROWN T, MANN B, RYDER N, et al. Language models are few-shot learners[J]. Advances in Neural Information Processing Systems, 2020, 33: 1877-1901.

[15] LIU X, ZHENG Y, DU Z, et al. GPT understands, too[J]. AI Open, 2024, 5: 208-215.

[16] LI X L, LIANG P. Prefix-Tuning: Optimizing Continuous Prompts for Generation[C]// Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. 2021: 4582-4597.

[17] CUI L, WU Y, LIU J, et al. Template-Based Named Entity Recognition Using BART[C]// Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. 2021: 1835-1845.

[18] HOU Y, CHEN C, LUO X, et al. Inverse is Better! Fast and Accurate Prompt for Few-shot Slot Tagging[C]// Findings of the Association for Computational Linguistics: ACL 2022. 2022: 637-647.

[19] PETRONI F, ROCKTÄSCHEL T, RIEDEL S, et al. Language Models as Knowledge Bases? [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). 2019: 2463-2473.

[20] TALMOR A, ELAZAR Y, GOLDBERG Y, et al. oLMpics-on what language model pre-training captures[J]. Transactions of the Association for Computational Linguistics, 2020, 8: 743-758.

[21] ZHAO M, SCHÜTZE H. Discrete and Soft Prompting for Mul-

tilingual Models[C]// Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021：8547-8555.

[22] SCHICK T，SCHÜTZE H. Exploiting Cloze-Questions for Few-Shot Text Classification and Natural Language Inference[C]// Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics. 2021：255-269.

[23] SHIN T，RAZEGHI Y，LOGAN IV R L，et al. AutoPrompt：Eliciting Knowledge from Language Models with Automatically Generated Prompts[C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing（EMNLP）. 2020：4222-4235.

[24] LIU J，YANG L. Knowledge-enhanced prompt learning for few-shot text classification[J]. Big Data and Cognitive Computing，2024，8(4)：43.

[25] ZHOU M，LI X，JIANG Y，et al. Enhancing Cross-lingual Prompting with Dual Prompt Augmentation[C]// Findings of the Association for Computational Linguistics：ACL 2023. 2023：11008-11020.

[26] DEMENTIEVA D，KHYLENKO V，GROH G. Cross-lingual Text Classification Transfer：The Case of Ukrainian[C]// Proceedings of the 31st International Conference on Computational Linguistics. 2025：1451-1464.

[27] SCHICK T，SCHÜTZE H. It's Not Just Size That Matters：Small Language Models Are Also Few-Shot Learners[C]// Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics：Human Language Technologies. 2021：2339-2352.

[28] GAO T，FISCH A，CHEN D. Making Pre-trained Language Models Better Few-shot Learners[C]// Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. 2021：3816-3830.

[29] WANG H，XU C，MCAULEY J. Automatic Multi-Label Prompting：Simple and Interpretable Few-Shot Classification[C]// Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics：Human Language Technologies. 2022：5483-5492.

[30] HU S，DING N，WANG H，et al. Knowledgeable Prompt-tuning：Incorporating Knowledge into Prompt Verbalizer for Text Classification[C]// Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. 2022：2225-2240.

[31] HAMBARDZUMYAN K，KHACHATRIAN H，MAY J. WARP：Word-level adversarial reprogramming[C]// ACL-IJCNLP 2021—59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing，Proceedings of the Conference. 2021：4921-4933.

[32] CUI G，HU S，DING N，et al. Prototypical Verbalizer for Prompt-based Few-shot Tuning[C]// Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. 2022：7014-7024.

[33] SUN X，YANG Y，LIU Y. External Knowledge Enhancing Meta-learning Framework for Few-Shot Text Classification via Contrastive Learning and Adversarial Network[C]// Asia-Pacific Web（APWeb）and Web-Age Information Management（WAIM）Joint International Conference on Web and Big Data. Singapore：Springer，2024：46-58.

[34] DONG H，ZHANG W，CHE W. Metricprompt：Prompting model as a relevance metric for few-shot text classification[C]// Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023：426-436.

[35] LIN T Y，GOYAL P，GIRSHICK R，et al. Focal Loss for Dense Object Detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2018，42(2)：318-327.

[36] SZEGEDY C，VANHOUCKE V，IOFFE S，et al. Rethinking the inception architecture for computer vision[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016：2818-2826.

[37] CIPOLLA R，GAL Y，KENDALL A. Multi-task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR）. IEEE Computer Society，2018：7482-7491.

[38] PASZKE A，GROSS S，MASSA F，et al. PyTorch：an imperative style，high-performance deep learning library[C]// Proceedings of the 33rd International Conference on Neural Information Processing Systems. 2019：8026-8037.

[39] WOLF T，DEBUT L，SANH V，et al. Transformers：State-of-the-art natural language processing[C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing：System Demonstrations. 2020：38-45.

[40] 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. 2019：4171-4186.

[41] LOSHCHILOV I，HUTTER F. Decoupled weight decay regularization[J]. arXiv：1711.05101，2017.

[42] ZHANG H，ZHANG X，HUANG H，et al. Prompt-based meta-learning for few-shot text classification[C]// Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. 2022：1342-1357.

**CHEN Lin**，born in 2000，postgraduate. His main research interests include natural language processing and cross-border industrial big data analysis.

**GAO Shengxiang**，born in 1977，Ph. D，professor，is a member of CCF（No. 38040M）. Her main research interests include natural language processing，information retrieval and machine translation.

（责任编辑：柯颖）