Artificial Intelligence and Robotics Research 人工智能与机器人研究, 2025, 14(2), 304-312
Published Online March 2025 in Hans. https://www.hanspub.org/journal/airr
https://doi.org/10.12677/airr.2025.142030

Hans汉斯

基于弱点增强的LLM知识蒸馏算法

李　崭

中国电子科技集团公司第十研究所，四川 成都

收稿日期：2025年1月21日；录用日期：2025年3月4日；发布日期：2025年3月12日

摘　要

得益于大语言模型的技术发展，自然语言处理领域的知识蒸馏范式发生了颠覆式的改变。大模型提示知识获取方式，使得知识蒸馏的方式向着更通用的知识获取以及数据增强的方式发展。面向大模型时代知识提炼模式转变以及小模型小样本学习挑战，本文提出了一种基于弱点增强的LLM知识蒸馏算法，结合LLM的语义理解以及文本生成能力，实现小样本情况下的学生模型弱点分析，通过LLM教师模型针对学生模型弱点进行增强样本构建迭代训练强化，增强学生模型的能力。通过在多种自然语言处理实验结果表明，本文提出的方法在少样本标注需求下，通过知识蒸馏，可以大幅度提升模型的训练效果，充分证明了方法的有效性。

关键词

知识蒸馏，小样本弱点分析，大模型数据增强

A Knowledge Distillation Algorithm for LLM Based on Weakness Enhancement

Zhan Li

The 10th Research Institute of China Electronic Technology Group Corporation, Chengdu Sichuan

Received: Jan. 21st, 2025; accepted: Mar. 4th, 2025; published: Mar. 12th, 2025

Abstract

Thanks to the technological advancements in large language models, the knowledge distillation paradigm in the field of natural language processing has undergone a revolutionary change. The knowledge acquisition method prompted by LLM has led to a shift in knowledge distillation towards more universal knowledge acquisition and data augmentation approaches. In response to the transformation of knowledge extraction patterns in the era of LLM and the challenges of small-model, few-sample learning, this paper proposes a knowledge distillation algorithm for LLM based on weakness

文章引用: 李崭. 基于弱点增强的 LLM 知识蒸馏算法[J]. 人工智能与机器人研究, 2025, 14(2): 304-312.
DOI: 10.12677/airr.2025.142030

enhancement. By leveraging the semantic understanding and text generation capabilities of LLM, this algorithm enables the analysis of student model weaknesses under few-sample conditions. The LLM teacher model enhanced samples to train and strengthen the student model to enhance student model's capabilities. Experimental results in various NLP tasks demonstrate that the proposed method, under the requirement of few labeled samples, can significantly improve the training effectiveness of models through knowledge distillation, fully proving the effectiveness of the method.

Keywords

Knowledge Distillation, Weakness Analysis in Few Shot Learning, LLM Data Augmentation

1. 引言

知识蒸馏[1] (Knowledge Distillation, KD)是人工智能发展过程中逐渐兴起的新型技术，其核心是将参数量较大、结构更复杂的教师模型知识转移到参数量较小、结构更简单的学生模型，以达到在精度损失较小的情况下更高效的推理效率。尤其在当前人工智能发展，一方面向着更大参数量更复杂结构的多任务、多模态智能模型进行演进；另一方面向着资源受限的环境下进行智能赋能，如移动设备、边缘计算平台等。这两方面发展的冲突，加速智能模型在知识蒸馏方向上的发展。

随着大语言模型[2] [3] (Large Language Model, LLM)时代的到来，自然语言处理(Natural Language Process)领域的知识蒸馏范式发生了颠覆式的改变。传统知识蒸馏，主要是同任务类的智能模型，基于大参数模型向小参数模型上进行知识蒸馏。大语言模型得益于其强大的上下文学习[4]能力，通过零样本[5]、小样本学习[6]就可实现多任务上的知识蒸馏。与此同时，大语言模型的提示学习能力为知识提取的手段提供了一种更加灵活和动态的方法，打破了模型知识蒸馏的壁垒，其作为教师模型，可以实现向不同架构、不同任务的自然语言处理模型进行知识传递，极大提升了知识蒸馏范式的通用性以及泛化性。

2. 技术现状

当前知识蒸馏技术在自然语言领域的发展，以大语言模型的介入为区分，主要分为大语言模型知识蒸馏框架以及专项任务模型知识蒸馏框架。

专项任务知识蒸馏主要是用小参数量的学生模型模仿较大参数教师模型的输出，构建知识蒸馏的过程。受限于教师模型大部分不支持多任务的数据，此类知识蒸馏模式教师学生模型基本均为相同的推理任务。周孝青[7]等提出了一种半知识蒸馏方法和递进式半知识蒸馏方法，在机器翻译领域将复杂且参数量大的 NMT 模型压缩为精简参数量小的 NMT 模型；俞亮[8]等提出了一种基于知识蒸馏的隐式篇章关系识别方法；黄友文[9]等提出在文本分类任务上，以 MPNetGCN 模型通过知识蒸馏得到学生模型 DistillBIGRU；廖胜兰[10]等提出一种知识蒸馏意图分类方法，通过 BERT 向 Text-CNN 小规模模型进行知识蒸馏；顾佼佼[11]等提出一种预训练 BERT 模型知识蒸馏到 BiGRU-CRF 进行信息抽取任务的知识迁移。受限于专项任务教师模型的知识瓶颈，在此框架下的知识蒸馏主要集中在同构模型以及同任务模型上。

相较于专项知识蒸馏框架，大语言模型(LLM)的出现彻底改变了知识蒸馏的面貌。LLM 基于生成式

任务一统自然语言多任务处理能力，基于 prompt 工程[12]以及 chain-of-thought 构建[13]等方式，使得知识的提取效能大幅度提升。知识蒸馏发展，逐步转向显性的数据增强模式[14]，其技术核心转向基于 LLM 的 prompt 以及 chain-of-thought 等技术强化领域任务小样本下的推理效果，提取推理结果作为领域知识指导学生模型进行学习。由此，知识蒸馏与数据增强[15] [16] (Data Augmentation)之间的关系变得越来越紧密，不同于传统的数据增强技术是以某种机械方式扩展训练数据集，在 LLM 能力的加持下，数据增强转变为面向特定领域和技能的生产新颖、丰富训练数据[15]。这种方法的转变，可以实现大量多任务、高质量的数据集，这些数据集不仅数量庞大，而且具有多样性和特定性[16]。基于此种模式使蒸馏过程更有效，确保提取的模型不仅复制了教师模型的输出行为，还进一步强化其模型的泛化推理能力。

在此基础上，本文提出了一种基于弱点增强的 LLM 知识蒸馏算法，结合 LLM 的语义理解以及文本生成能力，实现小样本情况下的学生模型弱点分析，通过 LLM 教师模型针对学生模型弱点进行增强样本构建迭代训练强化，增强学生模型的能力。实验结果表明，本文提出的方法在三类自然语言处理任务中，在少样本标注需求下，通过知识蒸馏，可以大幅度提升了模型的训练效果，充分证明了方法的有效性，下面将详细介绍该技术。

3. 基于弱点增强的 LLM 知识蒸馏算法

本文提出了基于弱点增强的 LLM 知识蒸馏算法，其算法主要适用于自然语言处理领域，算法具体流程如下。首先，通过学生模型的小样本学习以及训练样本集的评估，计算构建样本数据的错误价值；其次，基于错误价值对样本数据进行采样，并通过 LLM 教师模型的提升推理能力以及多样性样本评估函数构建训练增强样本集；最后，通过不断训练评估学习模型能力，迭代生成训练增强样本集，直至训练出测试集上最优解，完成模型的知识蒸馏过程。

下面将详细介绍基于弱点增强的 LLM 知识蒸馏算法的核心步骤以及实现方案。首先，将介绍如何对学生模型在数据集上的学习弱点分析；其次介绍基于 LLM 如何针对弱点分析结果，进行多样性的增强训练样本构建；最后模型知识蒸馏过程进行详细介绍。

3.1. 模型弱点分析

在模型弱点分析阶段，本文提出算法采用小样本进行特征学习，对学生模型进行训练。设构建标注训练样本集 C_{label_train} 样本数 K 个，构建标注测试样本集 C_{label_test} 样本数 M 个，构建标注样本集 C_{label} 样本数为 N，其中 $N=M+K$。通过小样本模型训练，完成对学生模型的初始化构建，表示为 $model_{studentk}$。基于 $model_{studentk}$ 对标注样本集 C_{label} 进行预测，并对每一条样本数据的错误价值 Err_{value} 进行计算评估，其中 Err_{value} 的计算公式如下：

$$Err_{value}=\frac{1}{e^{(2-h_x)}},h_x\in[0,1]$$

其中，e 为自然对数底数，模型推理计算这条数据 h_x 为模型推理这条数据可用价值函数，其值越高表示模型对这条数据的推理准确性越高，通过对每条数据计算，完成对标注样本集数据的弱点分析。

3.2. 样本数据增强

在样本数据增强阶段，基于样本数据计算错误价值 Err_{value} 的分值比例进行样本采样，获取样本数据种子。基于样本数据种子，通过 LLM 思维链提示推理，构建训练增强样本，其中思维链提示 Prompt 构建包括：语义角色(role)、任务要求(mission)、样本标签(tag)、样本文本(text)。语义角色设定为：文本样本生成器；任务要求设定为：根据样本标签，仿照样本文本重生生成一段全新文本以及文本标签，生成

的文本中需要包括样本标签数据；样本标签设置为文本中包含的所有信息抽取标签以及标签值；样本文本设置为样本文本信息。生成完成样本后，基于字符比对技术对生成样本以及样本标签进行二次校验，对生成样本与生成的信息抽取标签进行校验，校验成功的数据进行存储，实现训练增强样本构建。

同时，在训练增强样本构建同时，为减少大量高度雷同的数据生成，导致模型知识蒸馏过程中泛化能力丢失，在训练增强样本生成完成后，对其多样性进行计算。其多样性计算公式如下所示：

$$score_{multi}^{a}=\frac{\sum dist<dist_{\min},i\in C_{label}\left(\frac{1}{1+dist_{a,i}}\right)}{1+\sum dist<dist_{\min},i\in C_{label}\left(\frac{1}{1+dist_{a,i}}\right)}$$

其中，$dist_{\min}$ 为样本之间近邻最小距离阈值，$dist_{a,i}$ 为生成样本 a 与标注样本 i 之间的语义距离，其距离计算公式为：

$$dixt=1-\frac{vec_a\cdot vec_b}{|vec_a|*|vec_b|}$$

通过多样性计算，筛选出生成样本与标注样本有更远语义距离的数据，增强样本的多样性。在语义距离计算中，vec 为对应文本的特征向量值，采用的文本特征化模型是基于非对称 encoder-decoder 结构 RetroMAE 模型架构的 BGE 文本 Embedding 模型，进行文本向量化建模。

设训练增强样本集 F_{label_train}，构建样本集时，首先 N_F 倍数的训练增强样本生成，基于生成的样本，通过多样性计算评判，对得分高的数据进行高分优先采样，完成 F_{label_train} 的构建。

3.3. 模型知识蒸馏

模型知识蒸馏总共包含三个阶段，分别为：模型弱点分析、样本数据增强以及模型迭代训练。在模型弱点分析阶段，通过模型弱点分析，实现标注样本集 C_{label} 的构建以及标注样本数据计算错误价值 Err_{value}。在样本数据增强阶段，基于第一阶段的分析结果，完成训练增强样本集 F_{label_train} 的构建。

在模型迭代训练阶段，设 h_{total}^{n} 为模型在测试样本集 C_{label_test} 上的评估效果。首次训练，基于训练增强样本集 $F_{label\_train}^{0}$ 与训练标注样本集 C_{label_train} 进行混合构建训练样本集 T_0，基于训练样本集对模型进行训练，通过 C_{label_test} 进行评估获得首次数据增强评估效果 h_{total}^{n}。

重复样本数据增强以及模型迭代训练步骤，直到模型评估效果 h_{total}^{n} 值不再增长，停止模型迭代训练，其中，迭代中混合构建训练样本集构建方式为 $T_n=F_{label\_train}^{n}+T_{n-1}$，完成模型知识蒸馏。

4. 实验

4.1. 数据集

为了验证本文算法在自然语言领域多任务上的能力，本文算法采用多种任务的公开数据集进行验证，分别在信息抽取、文本摘要生成、文本分类三项任务上进行验证。其中，实体抽取任务采用人民日报命名实体识别数据集，数据集中的数据是在句子级进行了实体信息的标注；文本摘要生成任务采用自建数据集，其数据集是通过收集新浪微博、网易新闻等社交媒体以及新闻门户网站，通过数据爬虫、数据清洗、形成测试数据集，数据集包括文本详细信息以及文本简要介绍；文本分类采用通义实验室开源的中文文本领域分类测试集，数据集主要针对文本构建不同领域的分类标签。

基于上述三个测试集，分布在信息抽取任务、文本生成任务以及文本分类任务上对算法进行测试，以更加全面的评测知识蒸馏算法的效果。

4.2. 评估函数

本文所提出的知识蒸馏框架，选择在信息抽取、文本摘要生成、文本分类三项任务上进行验证，所以算法中的针对单条数据的 h_x 评估函数以及针对测试集的 h_{total}^n 评估函数针对不同任务需要进行不同定义。

信息抽取任务中，本文算法中的评估函数采用 $F1$ 值进行评估，$F1$ 值计算公式如下所示：

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

其中，$Precision$、$Recall$ 分别为模型在单条数据或测试数据集上的准确率与召回率。

文本摘要生成任务中，本文算法中的评估函数采用 $F_{ROUGE-L}$ 值进行评估，$F_{ROUGE-L}$ 值计算公式如下所示：

$$F_{\mathrm{ROUGE-L}} = \frac{2 * lcs^2}{lcs * len_t + lcs * len_p}$$

其中，lcs 为生成摘要与标注摘要的最长公共子序列长度，len_t 为标注摘要长度，len_p 为生成摘要长度。

文本分类任务中，本文算法中的评估函数采用识别准确率 acc 值进行评估，acc 值计算公式如下所示：

$$acc = \frac{N_{Precision}}{N_{Predict}}$$

其中，预测标签的数量为 $N_{Predict}$，预测正确的结果数为 $N_{Precision}$。

5. 实验结果与分析

针对本文提出的基于弱点增强的 LLM 知识蒸馏算法，在教师模型的选择上，采用 Qwen2.0 系列的大语言模型，模型参数量选择 72 B 参数量进行实验验证。当前，Qwen 系列模型作为中文系开源模型中效果最好的模型之一，在语义理解以及知识数据生成上都有非常好的表现，作为教师模型可以更好的评测 LLM 知识蒸馏框架的知识迁移能力。基于以上教师模型选型，将从文本分类、信息抽取、摘要生成三类自然语言处理任务上，对知识蒸馏框架的能力进行实验评估。

5.1. 文本分类实验结果

文本分类实验将针对算法在文本分类任务上的知识蒸馏性能进行数据实验以及分析。在数据的验证准备上，采用通义实验室开源的中文文本领域分类测试集，数据集主要包括不同领域的文本信息以及对应的领域标签，鉴于数据中存在超短文本信息，其文本信息与分类标签的语义关联度过低，通过数据清洗与人工判决的方式筛选部分数据作为实验的训练测试数据。其中，对语料的 18 类领域标签均匀采样，构建 360 条数据作为训练数据，进行学生模型初始小样本训练，另对各标签均匀采样 1000 条数据作为测试数据进行学生模型能力验证。

学生模型选择 ERNIE3.0 模型[17] [18]，作为文本分类模型，ERNIE 模型基于 Transformer 架构，采用多层 Transformer-XL 作为骨干网络构建模型，通过输出层多标签的分类概率输出，实现模型的多分类任务。知识蒸馏具体实验结果如下图 1 所示，其中，X 轴为训练数据规模，训练数据包括 360 条标注数据以及生成的仿真样本数据，Y 轴为训练模型在测试数据集上的分类准确率。通过实验可以看出，在小样本情况下，模型的标签分类能力明显不足，通过知识蒸馏对训练样本进行增强，其分类准确在初期大幅度提升，后续逐渐趋于稳定。验证了框架在小样本的情况下，对分类任务具有较好效果。

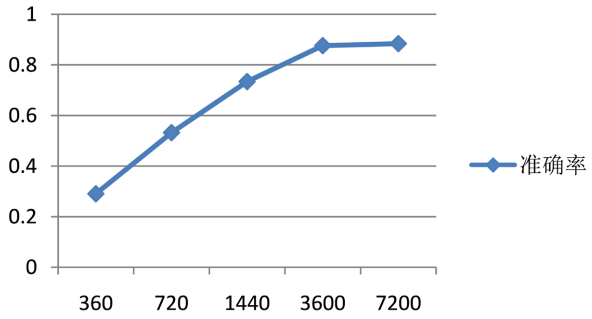

Figure 1. Experimental results of text classification knowledge distillation
图 1. 文本分类知识蒸馏实验结果

5.2. 信息抽取实验结果

信息抽取实验将针对算法在实体抽取任务上的知识蒸馏性能进行数据实验以及分析。在数据的验证准备上，采用人民日报命名实体识别数据集，数据集中的数据是在句子级进行了实体信息的标注，鉴于实验中选择的学生模型具有非常强的小样本学习能力，随机采样 3 条数据作为实验的训练数据，随机采样 200 条数据作为实验的测试数据。

学生模型选择 UIE [19]信息抽取模型，UIE 抽取模型在 2022 年由百度和中科院软件所共同提出，其作为一种通用信息抽取[20]技术，通过多任务统一的标签架构设计以及学习实现通用信息抽取能力，同时，其在小样本学习以及零样本推理中具有较强能力。知识蒸馏具体实验结果如下图 2 所示，其中，X 轴为训练数据规模，训练数据包括 3 条标注数据以及生成的仿真样本数据，Y 轴为训练模型在测试数据集上的实体抽取准确率、召回率以及 *F*1 值。通过实验可以看出，鉴于 UIE 模型在信息抽取任务上具备较高的零样本以及小样本学习推理能力，在知识蒸馏过程中，其预测准确率小幅度下降、召回率能够大幅度提升，其 *F*1 值能够实现小幅度提升，验证了知识蒸馏框架在实体抽取任务上对学生模型具有一定的知识传递能力。

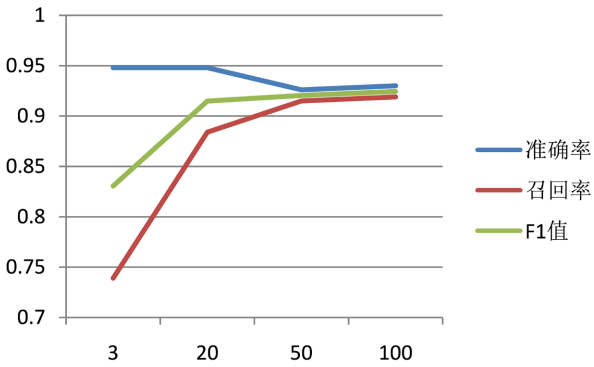

Figure 2. Experimental results of info extraction knowledge distillation
图 2. 信息抽取知识蒸馏实验结果

5.3. 摘要生成实验结果

摘要生成实验将针对算法在摘要生成任务上的知识蒸馏性能进行数据实验以及分析。在数据的验证准备上，采用自建数据集，其数据集是通过收集新浪微博、网易新闻等社交媒体以及新闻门户网站，通过数据爬虫、数据清洗、形成测试数据集，数据集包括文本详细信息以及文本简要介绍。分别随机采样

200 条数据作为实验的训练数据以及测试数据。

学生模型选择 Pegasus 模型[21] [22]，Pegasus 模型是一种生成式文本摘要模型，模型采用 Transformer-based 的编码器-解码器架构，通过 GSG 等训练任务构建，提升其摘要生成的效果。知识蒸馏具体实验结果如下图 3 所示，其中，X 轴为训练数据规模，训练数据包括 200 条标注数据以及生成的仿真样本数据，Y 轴为训练模型在测试数据集上的 rouge-1、rouge-2、rouge-L、BLEU-4 性能评估值。通过实验可以看出，通过本文算法的知识蒸馏，在摘要生成的任务上，模型的能力在各个指标上持续增长，验证了知识蒸馏框架在文本生成任务上，对学生模型的知识传递上具有很好的效果。

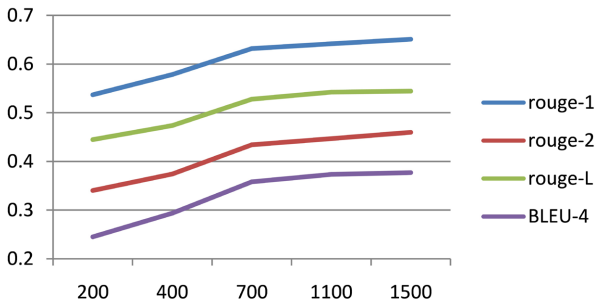

Figure 3. Experimental results of abstract generation knowledge distillation
图 3. 摘要生成知识蒸馏实验结果

5.4. 对比实验结果

对比实验将对同等规模下基于蒸馏训练与基于标注数据训练的文本分类模型、摘要生成模型、实体抽取模型进行对比实验测试，对比其数据蒸馏与标注数据的训练模型推理能力差距，实验结果如图 4 所示。在信息抽取任务上，Y 轴为训练模型在测试数据集上的实体抽取 *F*1 值，训练集规模为 100 条，可以看出知识蒸馏训练与标注样本训练在针对此项任务上的能力增强上几乎没有太大差距，标注样本训练略有提高。在文本分类任务上，Y 轴为训练模型在测试数据集上的准确率值，训练集规模为 7200 条，可以看出标注样本训练较知识蒸馏训练有更好效果。在摘要生成任务上，Y 轴为训练模型在测试数据集上摘要 Rouge-L 值，训练集规模为 1500 条，可以看出知识蒸馏训练与标注样本训练在摘要任务上，知识蒸馏相较于标注样本训练有更好的体现。

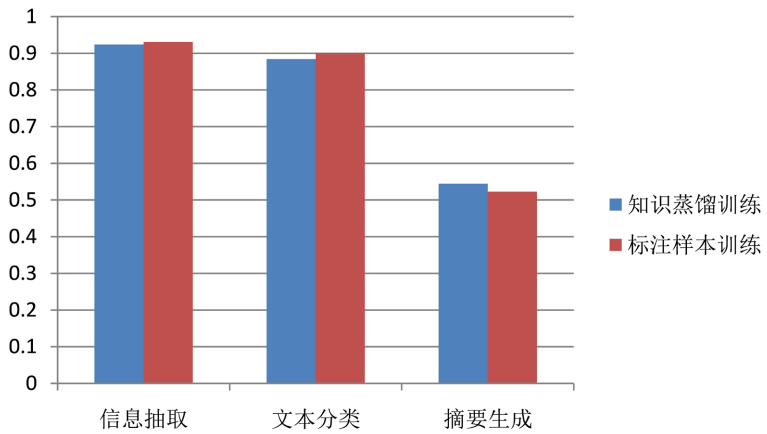

Figure 4. Comparison results of distillation training and label training
图 4. 蒸馏训练与标注训练对比结果

针对以上结果，通过训练过程观察可知，针对文本生成式任务，基于弱点增强的 LLM 知识蒸馏算法，相较于普通人类有更统一的样本生成标准以及样本生成质量，在摘要总结等生成任务上的样本构造能力有超越普通人的技能水平。在信息提取、文本分类等标签类的特征提取任务上，相较于其他任务其标签定义的误差更小，其生成样本可以达到接近人类标注的能力。

6. 结束语

本文提出了基于弱点增强的 LLM 知识蒸馏算法，结合的 LLM 的语义理解以及文本生成能力，实现小样本情况下的学生模型弱点分析，通过 LLM 教师模型针对学生模型弱点进行增强样本构建迭代训练强化，增强学生模型的能力，通过自然语言处理的多任务实验，其知识蒸馏下的学生模型在对应任务推理上可以达到接近升级超越基于人类标注训练的学生模型能力。后续，将对此知识蒸馏研究拓展到多模态领域，验证框架多模态下的知识传递能力，为进一步降低边缘推理、私有化专项领域智能推理任务的落地成本，提升边缘模型推理性能给出参考研究意见。

基金项目

国家自然科学基金资助项目(项目编号：62406263)。

参考文献

[1] 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.

[2] 朱炫鹏, 姚海东, 刘隽, 等. 大语言模型算法演进综述[J]. 中兴通讯技术, 2024, 30(2): 9-20.

[3] 张钦彤, 王昱超, 王鹤羲, 等. 大语言模型微调技术的研究综述[J]. 计算机工程与应用, 2024, 60(17): 17-33.

[4] Liu, P.F., Yuan, W.Z., Fu, J.L., *et al*. (2021) Pre-Train Prompt and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing.

[5] Chen, S., Chen, S., Xie, G., Shu, X., You, X. and Li, X. (2024) Rethinking Attribute Localization for Zero-Shot Learning. *Science China Information Sciences*, **67**, 184-196. https://doi.org/10.1007/s11432-023-4051-9

[6] 赵凯琳, 靳小龙, 王元卓. 小样本学习研究综述[J]. 软件学报, 2021, 32(2): 349-369.

[7] 周孝青, 段湘煜, 俞鸿飞, 等. 基于递进式半知识蒸馏的神经机器翻译[J]. 中文信息学报, 2021, 35(2): 52-60.

[8] 俞亮, 魏永丰, 罗国亮, 等. 基于知识蒸馏的隐式篇章关系识别[J]. 计算机科学, 2021, 48(11): 319-326.

[9] 黄友文, 魏国庆, 胡燕芳. DistillBIGRU: 基于知识蒸馏的文本分类模型[J]. 中文信息学报, 2022, 36(4): 81-89.

[10] 廖胜兰, 吉建民, 俞畅, 等. 基于 BERT 模型与知识蒸馏的意图分类方法[J]. 计算机工程, 2021, 47(5): 73-79.

[11] 顾佼佼, 翟一琛, 姬嗣愚, 等. 基于 BERT 和知识蒸馏的航空维修领域命名实体识别[J]. 电子测量技术, 2023, 46(3): 19-24.

[12] Shi, C., Su, J., Chu, C., Wang, B. and Feng, D. (2024) Balancing Privacy and Robustness in Prompt Learning for Large Language Models. *Mathematics*, **12**, Article No. 3359. https://doi.org/10.3390/math12213359

[13] Feng, K., Luo, L., Xia, Y., Luo, B., He, X., Li, K., *et al*. (2024) Optimizing Microservice Deployment in Edge Computing with Large Language Models: Integrating Retrieval Augmented Generation and Chain of Thought Techniques. *Symmetry*, **16**, Article No. 1470. https://doi.org/10.3390/sym16111470

[14] Do, D., Nguyen, M. and Nguyen, L. (2025) Enhancing Zero-Shot Multilingual Semantic Parsing: A Framework Leveraging Large Language Models for Data Augmentation and Advanced Prompting Techniques. *Neurocomputing*, **618**, Article ID: 129108. https://doi.org/10.1016/j.neucom.2024.129108

[15] Ullah, F., Gelbukh, A., Zamir, M.T., Riverón, E.M.F. and Sidorov, G. (2024) Enhancement of Named Entity Recognition in Low-Resource Languages with Data Augmentation and BERT Models: A Case Study on Urdu. *Computers*, **13**, Article No. 258. https://doi.org/10.3390/computers13100258

[16] Feng, S.J.H., Lai, E.M. and Li, W. (2024) Geometry of Textual Data Augmentation: Insights from Large Language Models. *Electronics*, **13**, Article No. 3781. https://doi.org/10.3390/electronics13183781

[17] 温浩, 杨洋. 融合 ERNIE 与知识增强的临床短文本分类研究[J/OL]. 计算机工程与应用, 2025: 1-10. http://kns.cnki.net/kcms/detail/11.2127.TP.20240527.1040.004.html, 2025-01-15.

[18] 杨笑笑, 陆奎. 融合 ERNIE 和深度学习的文本分类方法[J]. 湖北民族大学学报(自然科学版), 2023, 41(4): 506-512.

[19] Lu, Y.J., Liu, Q., Dai, D., *et al*. (2022) Unified Structure Generation for Universal Information Extraction.

[20] Ye, Z., Qi, D., Liu, H., Yan, Y., Chen, Q. and Liu, X. (2024) Rouie: A Method for Constructing Knowledge Graph of Power Equipment Based on Improved Universal Information Extraction. *Energies*, **17**, Article No. 2249. https://doi.org/10.3390/en17102249

[21] Zhang, J., Zhao, Y., Saleh, M., *et al*. (2020) Pegasus: Pre-Training with Extracted Gap-Sentences for Abstractive Summarization. *International Conference on Machine Learning*. *PMLR*, 13-18 July 2020, 11328-11339.

[22] Wang, J., Zhang, Y., Zhang, L., *et al*. (2022) Fengshenbang 1.0: Being the Foundation of Chinese Cognitive Intelligence.