基于 EDANN 与 ECDAN 的跨领域弱标注文本分类方法

薛连政[1] 刘海燕[1] 张兵权[1]
XUE Lianzheng LIU Haiyan ZHANG Bingquan

摘 要 针对跨领域弱标注文本分类中标注数据稀缺和领域差异导致的性能瓶颈，文章提出了一种迁移学习方法。采用动态标签清洗的半监督学习，提升数据利用效率，并引入多头注意力机制增强特征建模，设计针对高相似度场景的 EDANN 模型和中、低相似度场景的 ECDAN 模型，结合 SimCSE、SBERT、BERT 预训练模型，通过差异化对抗训练实现特征对齐。实验结果显示，F_1 值在高、中、低相似度场景分别达 95.49%、91.70%、93.43%，相较 BCN、AM-AdpGRU 等基线模型显著提升，验证了方法的有效性和泛化能力。

关键词 弱标注；文本分类；标签清洗；多头注意力

doi：10.3969/j.issn.1672-9528.2025.11.001

0 引言

随着自然语言处理（natural language processing, NLP）技术的迅猛发展，文本分类在信息检索、情感分析等应用中具有重要价值 [1]。然而，标注数据稀缺和领域差异显著导致模型常遭遇性能瓶颈 [2]。传统方法依赖高质量标注数据集，但数据标注成本高昂 [3]。因此，如何在弱标注跨领域背景下提升文本分类性能，已成为 NLP 领域的研究重点。

现有研究在特征提取和分类模型方面取得一定进展，但仍面临挑战 [4]。文献 [5] 和 [6] 虽在数据标注方面降低成本，但前者在 IMDB 数据集 F_1 值仅达 80.08%，后者在 Weibo-STD 数据集 F_1 值仅达 74.18%。文献 [7] 虽提出了 TLMT-BBIC-HS 模型将多任务学习与迁移学习结合，但计算资源需求极大。文献 [8] 虽应用对抗训练，但缺乏领域适应性机制。文献 [9] 虽利用贝叶斯网建模但未充分利用预训练语言模型捕捉深层语义特征。文献 [10] 虽结合双通道特征提取和半监督学习，但其特征提取方法单一。

针对上述提出的问题，通过集成多头注意力机制和动态标签清洗机制，结合 BERT（bidirectional encoder representations from transformers）、SimCSE（simple contrastive sentence embedding）和 SBERT（Sentence-bert），提出 EDANN（enhanced domain-adversarial neural network）模型和 ECDAN（enhanced conditional adversarial domain adaptation）模型，以实现特征对齐和性能提升。实验设计针对高、中、低三个数据相似度实验场景，包括预训练模型对比、消融实验、任务验证及与 BCN 模型 [10]、AM-AdpGRU 模型 [11] 的对照实验，所提方法在分类性能和泛化能力上取得显著改进。

1 模型设计

为充分利用弱标注数据，引入半监督学习机制，通过动态标签清洗增强模型性能。动态特征更新根据验证集表现自适应调整特征集合，增强跨域适应性。LabelCleaner 类根据模型预测的置信度 Softmax 概率筛选可靠样本，清洁掩码由高置信度样本组成，动态扩展有效训练集，提升模型对目标域的适应能力。同时，为增强特征的全局依赖建模能力，ECDAN 模型引入多头注意力机制，计算采用缩放点积形式，拼接多头结果后通过投影矩阵输出，捕捉长距离依赖，弥补线性层不足。

EDANN 模型包含特征提取器、任务分类器和领域分类器。特征提取器将特征映射到隐藏空间，结合批归一化加速训练并提升稳定性。任务分类器基于特征提取器进行文本分类，通过交叉熵损失优化，促使特征提取器学习对分类任务有益的特征。领域分类器通过 GRL 接收反转梯度，区分源域和目标域特征，EDANN 模型架构如图 1 所示。

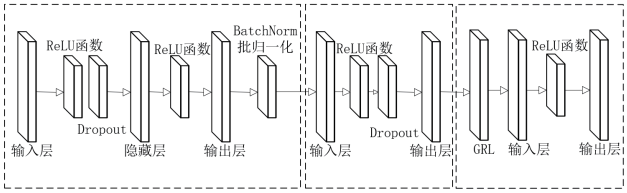

图 1 EDANN 模型架构图

1. 北华航天工业学院 河北廊坊 065000
[基金项目] 河北省教育厅科研资助项目（CXY2023013）；北华航天工业学院硕士研究生创新课题项目（YKY-2025-60）

ECDAN 模型包括特征提取器、源域分类器、目标域分类器和领域分类器。多头注意力机制在中相似度场景下嵌入特征提取器，在低相似度场景下置于特征提取器和任务分类器之间。ECDAN 模型架构如图 2 所示。

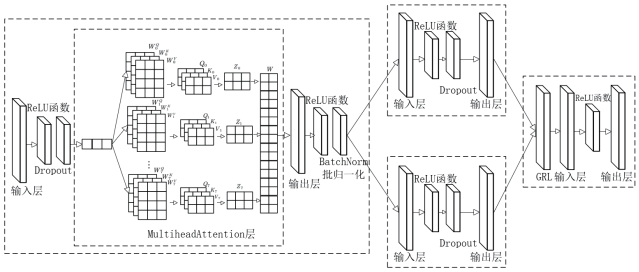

图 2 ECDAN 模型架构图

2 实验设计与结果分析

2.1 实验设置

实验分别在高、中、低三个相似度场景下实施。高相似度场景源领域选用 IMDB 电影评论数据集，目标领域选用 Amazon 商品评论数据集；中相似度场景源领域选用 20 Newsgroups 数据集（comp.* 类别），目标领域选用 20 Newsgroups 数据集（sci.* 类别）；低相似度场景源领域选用 AG News 数据集中的 World News 类别，目标领域选用 Yelp Review 数据集中的餐厅类评论。源域和目标域数据体量之比为 5 000:500，由原始数据集随机抽取获得。其中，源域包括 50% 标注数据，目标域仅包含 10% 标注数据，标注数据 80% 来自原始标注，20% 来自随机产生的错误标签作为噪声标注，模拟弱标注场景。数据预处理过程主要包括文本清洗、分词、去除停用词和词形还原。去除掉原始文本中 HTML 标签、非字母字符等不携带显著语义信息或对文本分类任务产生噪声干扰的因素，将连续的文本拆分为独立的词单元，过滤掉文本中的停用词，最后采用 NLTK 的 WordNetLemmatizer 对分词结果进行词形还原，将词语统一为其基本形式，从而提高特征提取的鲁棒性。模型采用 PyTorch 框架实现，优化器选用 AdamW。

EDANN 模型优化器学习率为 5e-5，权重衰减为 1e-6。源域和目标域的加权交叉熵损失定义为：

$$L_{t/d} = -\frac{1}{N_{t/d}}\sum_{i=1}^{N_{t/d}} \boldsymbol{\omega}_{y_i} \log p_{y_i} \tag{1}$$

式中：$L_{t/d}$表示任务损失和领域损失；$N_{t/d}$表示任务或领域样本数。

ECDAN 模型优化器学习率为 5e-4，权重衰减为 1e-4。源域和目标域的加权交叉熵损失分别定义为：

$$L_{s/t} = -\frac{1}{N_{s/t}}\sum_{i=1}^{N_{s/t}} \boldsymbol{\omega}_{y_i} \log p_{y_i} \tag{2}$$

$$\text{weight} = \frac{1}{\text{class\_counts}+(1e-8)} \tag{3}$$

式中：$\boldsymbol{\omega}_{y_i}$为类别权重；$N_{s/t}$为源域或目标域样本数；$p_{y_i}$为预测概率。领域损失为标准交叉熵损失：

$$L_{s/t} = -\frac{1}{N}\sum_{i=1}^{N} \log p_{d_i} \tag{4}$$

总损失由源域任务损失、目标域任务损失和领域损失共同组成。

实验过程包括数据预处理、特征提取、标签清洗、模型构建、训练优化、实验评估和迭代优化，如图 3 所示。

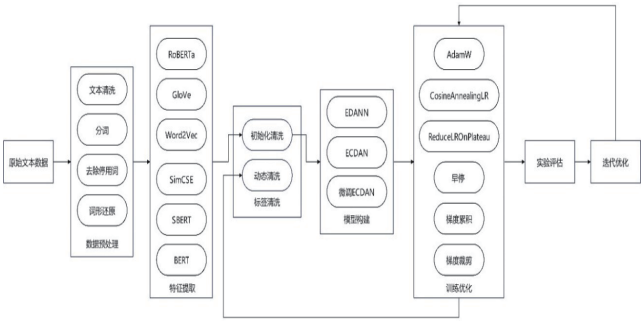

图 3 实验流程图

2.2 预训练模型对比实验

为验证数据选用的合理性，采用 SBERT 计算各场景源领域数据集与目标领域数据集的语义相似度为 0.707 7、0.487 4、0.184 5，如图 4 所示。

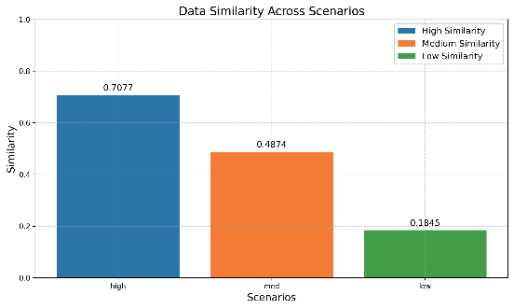

图 4 领域间语义相似度

结果清晰呈现从高到低的相似度梯度，验证了数据集设计的合理性。这一梯度设计为跨领域迁移任务提供了多样化的实验场景。

特征提取方面，采用 6 种预训练模型，系统性地评估特征提取策略，设计支持向量机模型（SVM）测试模型原始性能，结果如图 5 所示。

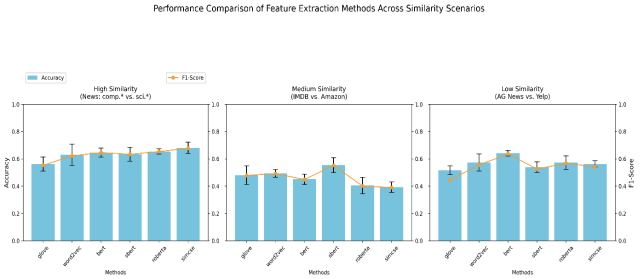

图 5 预训练模型对比实验

通过预训练模型对比实验，预训练模型选取结果如表 1 所示。

表 1 消融实验结果 F_1 值分析

相似度场景	选取策略	F_1 值
高	SimCSE	0.677 4
中	SBERT	0.552 0
低	BERT	0.635 6

2.3 消融实验

设置 3 个实验组：无领域适应的全连接神经网络作为基线模型、无领域适应机制的标准对抗模型以及 EDANN 和 ECDAN 模型。数据采用分层抽样将标注数据划分为 80% 训练集和 20% 验证集。消融实验结果如表 2 所示，验证了标签清洗机制和多头注意力机制的有效性。

表 2 消融实验结果 F_1 值分析

场景	基线模型	标准模型	最终模型
高	0.898 0	0.915 6	0.954 9
中	0.888 4	0.900 4	0.917 0
低	0.899 4	0.916 9	0.934 3

高相似度场景，动态标签清洗机制能更准确地识别目标域中可靠样本，通过较高的置信度阈值筛选出可靠样本，中、低相似度场景分布差异较大，通过逐步提高置信度阈值有效过滤掉噪声样本，减少错误标注的影响，增强模型鲁棒性。在中相似度场景，多头注意力机制通过增强特征提取器的能力，帮助模型更好地捕捉文本的语义特征，而低相似度场景，文本的语义和语言风格差异显著，多头注意力机制置于特征提取器和任务分类器之间，通过建模深层语义依赖，进一步提升了模型对目标域的理解能力。消融实验结果验证了动态标签清洗机制和多头注意力机制的有效性。

2.4 多场景实验

高相似度场景验证集损失函数为：

$$L_{\text{Val}} = -\frac{1}{N_V}\sum_{i=1}^{N_V}\log p_{y_i} \tag{5}$$

式中：N_V为验证集样本数。

对照实验表明 EDANN 模型性能显著提升，结果如表 3 所示。

表 3 高相似度场景最佳性能对比

模型	最佳 F_1	最佳 Val Loss
基线模型	0.898 0	0.340 6
DANN	0.915 6	0.331 3
EDANN	0.954 9	0.178 9

中、低相似度场景验证集损失函数用公式表示为：

$$L_{\text{Val}} = \text{src}_{\text{weight}}\cdot L_s + \text{tgt\_weight}\cdot L_t \tag{6}$$

对照实验表明 ECDAN 模型性能显著提升，结果如表 4 所示。

表 4 中、低相似度场景最佳性能对比

模型	最佳 F_1	最佳 Val Loss
基线模型（中）	0.888 4	0.811 1
基线模型（低）	0.899 4	0.610 3
CDAN（中）	0.900 4	0.453 8
CDAN（低）	0.916 9	0.175 9
ECDAN（中）	0.917 0	0.144 5
ECDAN（低）	0.934 3	0.222 9

模型在各相似度场景性能差异主要源于数据分布相似度、模型组件适配性及特征提取方法的差异。高相似度下数据相似度为 0.707 7，EDANN 利用 SimCSE 提取的高质量特征和动态标签清洗机制有效对齐相近的领域分布，F_1 分数达 0.954 9。中相似度下数据相似度为 0.487 4， ECDAN 通过嵌入多头注意力机制和 SBERT 特征增强语义建模，应对标签类别差异，F_1 分数为 0.917 0。低相似度下数据相似度为 0.184 5，ECDAN 结合 BERT 特征和条件对抗训练，通过多头注意力捕捉深层语义，动态标签清洗提升噪声鲁棒性，F_1 分数达 0.934 3。EDANN 模型实验指标如图 6~7 所示，中相似度下的 ECDAN 模型实验指标如图 8~9 所示，低相似度下的 ECDAN 模型验证指标如图 10~11 所示。

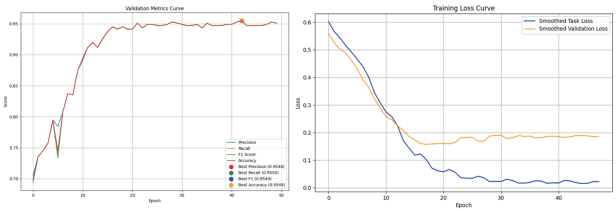

图 6 高相似度场景验证指标图

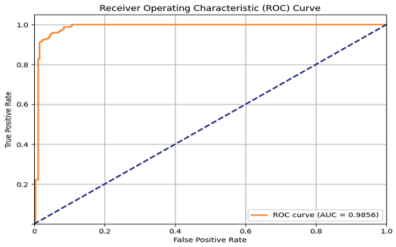

图 7 高相似度场景 ROC 曲线图

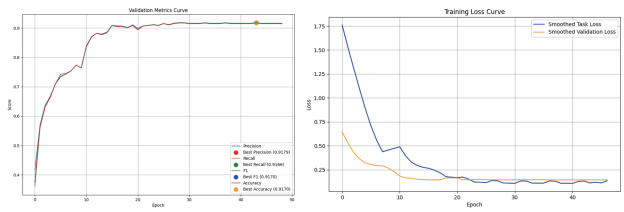

图 8 中相似度场景验证指标图

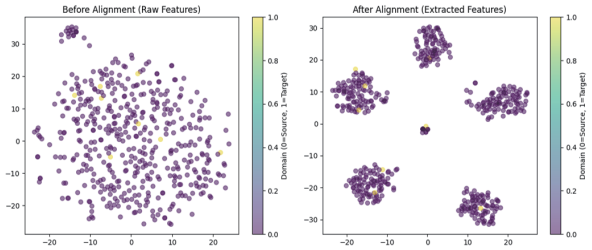

图 9 中相似度场景 t-SNE 可视化结果

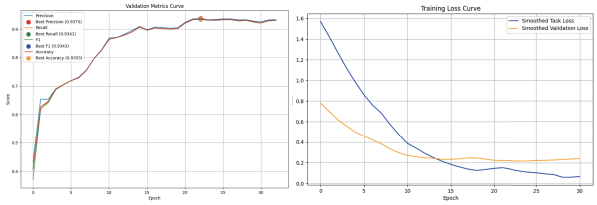

图 10 低相似度场景损失曲线图

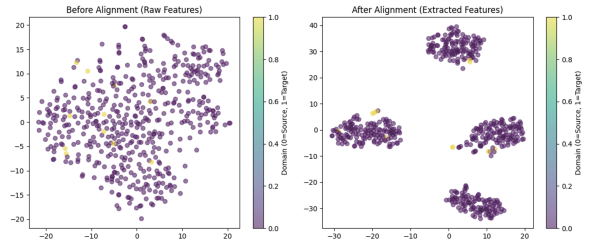

图 11 低相似度场景 t-SNE 可视化结果

2.5 验证实验

为验证所提方法的有效性，将所提方法与现有方法 BCN 模型、AM-AdpGRU 模型进行对比验证实验，实验结果如表 5 所示。

表 5 不同实验组的最佳性能对比

场景	本文模型最佳 F_1	BCN 最佳 F_1	AM-AdpGRU 最佳 F_1
高	0.954 9	0.719 6	0.739 9
中	0.917 0	0.720 0	0.680 0
低	0.934 3	0.560 0	0.616 2

表 5 为本文模型与 BCN 模型、AM-AdpGRU 模型的对比实验结果。结果表明，所提方法在各种数据相似度场景下性能均显著提升，表明本文方法的有效性，同时具有较强的泛化能力。

3 结论

针对弱标注跨领域文本分类中标注数据稀缺和领域差异的挑战，提出的 EDANN 和 ECDAN 模型通过动态标签清洗、多头注意力机制及差异化对抗训练显著提升了性能。实验结果显示，高、中、低相似度场景的 F_1 分数分别达 95.49%、91.70% 和 93.43%，较基线模型、标准模型、现有方法，有显著提升。统计分析表明，SimCSE 在高相似度场景下优化特征对齐，SBERT 和 BERT 分别适配中、低相似度场景，而多头注意力机制在中、低相似度场景下通过捕捉长距离依赖弥补语义差异。动态标签清洗通过高置信度筛选增强噪声鲁棒性，尤其在低相似度场景效果显著。

参考文献：

[1] 石运来，崔运鹏，杜志钢．基于 BERT 和深度主动学习的农业新闻文本分类方法 [J]. 农业图书情报学报，2022, 34(8): 19-29.

[2] 姚洁茹，韩军伟，张鼎文．一种基于点标注的弱监督目标检测方法 [J]. 中国科学 (信息科学), 2022, 52(3): 461-482.

[3] 刘康晟，凌青，闫文君，等．基于弱监督小波 KAN 网络的弱标注辐射源识别算法 [J]. 雷达学报 (中英文), 2025, 14(2): 338-352.

[4] ASHA H J, SCHOLAR M, JOSEPHINE J A. Cross-domain text classification: transfer learning approaches[C/OL]//2024 International Conference on Emerging Smart Computing and Informatics (ESCI). Piscataway:IEEE,2024[2025-10-12]. https://ieeexplore.ieee.org/document/10497318.DOI: 10.1109/ESCI59607.2024.10497318.

[5] 李自强，杨薇，杨先凤，等．基于弱标签争议的半自动分类数据标注方法 [J]. 电子学报，2024, 52(8): 2891-2899.

[6] 方晔玮，王铭涛，陈文亮，等．基于自动弱标注数据的跨领域命名实体识别 [J]. 中文信息学报，2022, 36(3): 73-81.

[7] 韩普，顾亮，叶东宇，等．基于多任务和迁移学习的中文医学文献实体识别研究 [J]. 数据分析与知识发现，2023, 7(9): 136-145.

[8] 蔡国永，林强，任凯琪．基于域对抗网络和 BERT 的跨领域文本情感分析 [J]. 山东大学学报 (工学版), 2020, 50(1): 1-7.

[9] 刘慧清，郭延哺，李维华．基于贝叶斯网的跨领域情感分析方法 [J]. 计算机应用与软件，2020, 37(12): 119-126.

[10] 余本功，汲浩敏．基于 DW-TCI 的半监督文本分类方法研究 [J]. 数据分析与知识发现，2020, 4(10): 58-69.

[11] 吴峰，谢聪，姬少培．基于跨领域迁移的 AM-AdpGRU 金融文本分类 [J]. 应用科学学报，2022, 40(5): 828-837.

【作者简介】

薛连政（2002—），男，河北廊坊人，硕士研究生，研究方向：计算机技术。

刘海燕（1981—），女，河北廊坊人，硕士，副教授，研究方向：计算机技术、软件高可信测试技术。

张兵权（2000—），男，河南周口人，硕士研究生，研究方向：自然语言处理。

（收稿日期：2025-05-19 修回日期：2025-11-11）