软件学报 ISSN 1000-9825, CODEN RUXUEW
2026,37(4):1449−1471 [doi: 10.13328/j.cnki.jos.007521] [CSTR: 32375.14.jos.007521]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

基于高质量样本选择的跨领域方面级情感分析*

赵传君[1,2], 孙绪壮[1], 康 璐[1], 李 旸[3], 王素格[4,5], 李德玉[4,5]

[1](山西财经大学 信息学院, 山西 太原 030006)

[2](数据要素创新与经济决策分析山西省重点实验室 (山西财经大学), 山西 太原 030006)

[3](山西财经大学 金融学院, 山西 太原 030006)

[4](山西大学 计算机与信息技术学院, 山西 太原 030006)

[5](计算智能与中文信息处理教育部重点实验室 (山西大学), 山西 太原 030006)

通信作者: 王素格, E-mail: wsg@sxu.edu.cn; 李德玉, E-mail: lidy@sxu.edu.cn

摘　要: 跨领域方面级情感分析利用源领域的已标注样本来帮助训练目标领域上的方面级情感分析任务, 但并非所有源领域样本均适合进行迁移训练, 部分样本会对迁移模型训练产生负迁移效应, 需要进行样本筛选工作. 现有的跨领域实例迁移方法所考虑的迁移依据比较片面, 忽略了样本间的协同作用, 影响跨领域泛化性能. 为了解决方面级情感分析任务中的特定领域训练样本匮乏与跨领域迁移中的样本筛选问题, 以多领域情感分析为开放环境, 结合高可信机器学习理论及建模中的领域适应方法, 提出一种基于高质量样本选择的跨领域方面级情感分析方法. 首先, 该方法分别设计了域间及域内高质量样本选择指标, 依次对源领域数据进行领域层面和样本层面的筛选, 兼顾了两种样本选择粒度的优势. 其次, 全面地设计了源领域与目标领域间相似性的衡量指标, 并通过图神经网络进行高效计算. 最后, 将多源领域迁移的场景纳入跨领域 ABSA (aspect-based sentiment analysis) 的讨论范围中, 设计了域间联合适应性分数, 通过平衡领域特征的重合性与差异性来选择领域间协同性高的多源领域组合. 在涵盖 6 个领域的基准数据集上设计了跨领域迁移任务, 并在方面级情感分析的 3 种子任务上进行了实验来验证所提方法的有效性.

关键词: 跨领域方面级情感分析; 高质量样本选择; 领域适应; 迁移学习

中图法分类号: TP18

中文引用格式: 赵传君, 孙绪壮, 康璐, 李旸, 王素格, 李德玉. 基于高质量样本选择的跨领域方面级情感分析. 软件学报, 2026, 37(4): 1449–1471. http://www.jos.org.cn/1000-9825/7521.htm

英文引用格式: Zhao CJ, Sun XZ, Kang L, Li Y, Wang SG, Li DY. Cross-domain Aspect-based Sentiment Analysis Based on High-quality Sample Selection. Ruan Jian Xue Bao/Journal of Software, 2026, 37(4): 1449–1471 (in Chinese). http://www.jos.org.cn/1000-9825/7521.htm

Cross-domain Aspect-based Sentiment Analysis Based on High-quality Sample Selection

ZHAO Chuan-Jun[1,2], SUN Xu-Zhuang[1], KANG Lu[1], LI Yang[3], WANG Su-Ge[4,5], LI De-Yu[4,5]

[1](School of Information, Shanxi University of Finance and Economics, Taiyuan 030006, China)

[2](Shanxi Key Laboratory of Data Element Innovation and Economic Decision Analysis (Shanxi University of Finance and Economics), Taiyuan 030006, China)

[3](School of Finance, Shanxi University of Finance and Economics, Taiyuan 030006, China)

[4](School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China)

* 基金项目: 国家自然科学基金 (62376143, 62473241, 61906110); 山西省高等学校青年学术带头人项目 (2024Q018); 山西省基础研究计划 (202303021211139)
本文由“高可信自适应机器学习”专题特约编辑王熙照教授、张敏灵教授、曹付元教授推荐.
收稿时间: 2025-04-09; 修改时间: 2025-06-30, 2025-08-15; 采用时间: 2025-08-20; jos 在线出版时间: 2025-09-02
CNKI 网络首发时间: 2025-11-21

[5](Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education (Shanxi University), Taiyuan 030006, China)

Abstract: Cross-domain aspect-based sentiment analysis (ABSA) uses annotated samples from the source domain to help train ABSA tasks on the target domain. However, not all samples from the source domain are suitable for transfer training, and some samples may have negative transfer effects on the training of the transfer model, which requires sample screening. Existing cross-domain instance transfer methods consider a one-sided transfer basis, ignoring the synergistic effect between samples and affecting cross-domain generalisation performance. In order to solve the problems of insufficient domain-specific training samples and sample screening in cross-domain transfer for ABSA tasks, this study proposes a cross-domain ABSA method based on high-quality sample selection by combining high-reliability machine learning theories and domain adaptation methods in modelling with the open environment of multi-domain sentiment analysis. First, inter-domain and intra-domain high-quality sample selection metrics are designed to filter the source domain data at the domain level and the sample level in turn, which takes into account the advantages of the two sample selection granularities. Second, similarity metrics between source and target domains are comprehensively designed and efficiently calculated through a graph neural network. Finally, the scenarios of multi-source domain transfer are included in the discussion of cross-domain ABSA, and inter-domain joint adaptability scores are designed to select the multi-source domain combinations with high inter-domain synergies by balancing the overlap and difference of domain features. A cross-domain transfer task is designed on a benchmark dataset covering six domains, and experiments are conducted on three sub-tasks of ABSA to validate the effectiveness of the proposed method.

Key words: cross-domain aspect-based sentiment analysis; high-quality sample selection; domain adaptation; transfer learning

随着互联网和 Facebook、Twitter、微博等社交媒体的迅猛发展, 用户可以方便地在这些平台上分享对各种产品、服务、事件等的意见和情感. 用户生成内容 (user-generated content, UGC) 呈现爆发式增长. 这些内容具有实时性、海量性和多样性, 包含着丰富的用户情感信息. 对这些内容进行情感分析对于了解用户需求、提升用户体验、改进产品和服务等方面具有重要意义[1]. 然而在许多领域, 方面级情感分析 (aspect-based sentiment analysis, ABSA) 的训练支持样本仍然十分匮乏, 因此产生了跨领域 ABSA 任务[2]. 为解决特定领域标注样本匮乏的问题, 需要将训练样本丰富的源领域情感知识迁移到目标领域. 然而, 针对不同领域, 用户对于情感的表达方式存在较大差异[3−5]. 例如, 在电商领域, 用户的观点可能包含针对更多产品特点方面的情感表达; 在服务领域, 用户观点则可能包含更多针对服务态度等方面的情感. 这导致简单的模型套用无法达到目标领域上的性能预期, 研究人员开始探索如何借助源领域的带标签样本集中的潜在信息来帮助训练目标领域上的 ABSA 任务.

近年来, 随着深度学习技术的快速发展, 各类深度学习模型在 ABSA 领域表现优异[6]. 深度学习模型的迁移应用及预训练微调模型的出现也使得跨领域 ABSA 任务的性能水平快速提升. 无论是深度学习模型的训练还是预训练模型的微调都依赖大量的样本来学习特征与模式, 以实现精准的预测和分类[7−9]. 而跨领域 ABSA 任务涉及不同领域的方面级数据, 在词汇、语义等诸多方面存在显著差异. 若不加甄别地使用所有训练样本, 模型容易受到领域差异带来的干扰, 无法有效捕捉关键情感信息. 因此, 科学合理的实例迁移方法至关重要. 一方面, 它能够筛选出具有代表性、通用性的样本, 减少无关信息的干扰, 提升训练效率; 另一方面, 针对目标领域的特性筛选与之适配度高的源领域样本, 能够有效缩小领域差异, 优化迁移学习的效果. 因此, 基于高质量样本选择的领域适应方法在跨领域 ABSA 领域有很大的优势. 通过这种面向开放环境的高可信机器学习理论及建模中的领域适应方法, 能够有效提升众多领域上跨领域 ABSA 任务的性能.

因此, 跨领域 ABSA 任务急需一种能够提升泛化性能的高质量样本选择方法, 以筛选出兼具领域通用性及目标领域特征的高质量源领域训练数据, 从而缓解负迁移效应, 提高迁移学习的效果. 然而, 跨领域高质量数据选择面临诸多难点, 主要体现在以下 3 个方面.

(1) 样本选择粒度问题. 在跨领域 ABSA 中, 样本选择的粒度是一个关键问题. 即样本选择应基于领域整体, 或是具体到样本级别. 以领域为单位进行样本选择的优点在于其能够借助领域内的整体模式, 使模型在迁移时能够更好地捕获全局情感趋势. 然而, 这种方法会引入一些不适合目标领域的样本, 从而产生负迁移. 而以样本为单位的精细选择能够避免这种问题, 但单一的样本选择往往难以利用领域整体情感特征, 导致模型在迁移过程中无法识别出领域的全局模式.

(2) 源领域与目标领域相似性的衡量. 要选择适合用于迁移的源领域数据, 首先需要衡量源领域与目标领域的

相似性. 以往的相似度衡量方法通常关注单一维度, 如语义或特征相似性. 然而跨领域的相似性并非单一的标准, 以往衡量方法难以全面反映领域间的差异.

(3) 源领域间的协同性衡量. 协同性指的是不同源领域数据在情感倾向和模式上的重合性与差异性的平衡. 重合性有助于确保迁移模型在目标领域中能够识别到通用的情感特征, 减少因通用情感表达不一致而带来的负迁移风险. 差异性可以增强多源领域数据组合的情感表达模式的多样性, 使模型在目标领域中表现出稳健的适应性. 以往的样本选择方法大多只关注单源域或对多源域协同性考虑不足, 限制了模型泛化性能.

尽管跨领域 ABSA 通过迁移源领域知识为目标领域提供支持成为可行路径, 但由于不同领域间在语言表达、情感倾向及关注重点方面存在显著差异, 直接迁移往往会引发负迁移问题, 影响模型性能. 跨领域间关注焦点和用词习惯的差异, 导致模型在未经过筛选的数据迁移过程中容易学习到与目标任务无关甚至冲突的特征. 因此, 仅依赖传统的迁移方法难以充分发挥源领域数据的价值. 为此, 如何从多源领域中科学筛选出兼具领域通用性与目标领域适配性的高质量训练样本, 成为提升跨领域 ABSA 任务性能的关键. 这不仅能够缓解负迁移带来的性能瓶颈, 还能显著提升模型对目标领域情感特征的捕捉能力, 从而实现更有效的知识迁移与泛化, 正是本文开展研究的主要动机所在.

为了解决样本选择粒度, 跨领域相似度度量以及源领域间的协同性衡量问题, 提出了一种基于高质量样本选择的跨领域 ABSA 方法. 方法兼顾源领域与目标领域间相似性及多源领域间协同性, 设计了域间和域内两个层面的高质量样本选择指标, 依次对多源领域数据进行领域层面和样本层面的评估和筛选. 在域间层面, 综合考虑词汇、语义及特征设计相似度衡量, 同领域协同性共同构成域间高质量样本选择指标; 在域内层面, 从样本的领域通用性与目标领域匹配性两个角度, 综合考虑词汇、语义及特征设计了域内高质量样本选择指标. 为了验证所提方法的性能, 在共涵盖 6 个领域的 ABSA 数据集上, 针对方面级情感分析的 3 个子任务进行了全面实验. 结果表明, 所提出的方法可以有效选择出多源领域中的高质量样本, 降低迁移模型在目标领域的泛化误差界, 提升了多个类别的跨领域 ABSA 模型的性能. 本文的 4 点主要贡献如下.

(1) 针对跨领域 ABSA 任务中样本选择粒度的问题, 设计了域间高质量领域选择指标和域内高质量样本选择指标, 分别从领域层面和样本层面筛选源领域数据, 兼顾领域通用性与目标领域适配性, 缓解负迁移风险.

(2) 为提升源目标领域间相似性计算的准确性, 引入共现图和图卷积网络, 从结构和语义双重视角提取领域特征, 提出跨领域模式共现相似度, 全面衡量领域之间的关联性, 增强了样本选择的可靠性.

(3) 面向多源领域迁移场景, 构建了联合适应性分数指标, 综合考虑领域特征重合性与差异性, 指导多源领域组合策略优化, 提升模型在多源复杂情境下的泛化性能.

(4) 在涵盖多个领域和 ABSA 3 个子任务的数据集上进行了大量实验, 结果表明所提方法在不同模型和任务中均显著优于现有基线方法, 具备良好的适应性与推广价值.

本文第 1 节介绍跨领域 ABSA 及跨领域实例迁移方法的相关工作和研究现状. 第 2 节介绍本文的研究任务. 第 3 节介绍本文提出的基于高质量样本选择的跨领域 ABSA 方法. 第 4 节通过与基线系统的对比实验验证了所提方法的有效性, 通过消融实验, 统计性分析实验以及案例分析对所提方法进行深入分析. 最后第 5 节总结全文并对未来研究方向进行了展望.

1 相关工作

1.1 跨领域方面级情感分析

近年来, 跨领域 ABSA 已成为自然语言处理领域的重要研究方向. 该任务旨在利用源领域丰富的情感信息帮助解决训练样本匮乏的目标领域上的 ABSA 任务, ABSA 任务主要包含 3 种子任务: 方面术语提取 (aspect term extraction, ATE)、方面类别检测 (aspect category detection, ACD) 与方面级情感分类 (aspect-based sentiment classification, ABSC)[10]. 现有的跨领域 ABSA 方法可以大致分为基于规则、基于传统机器学习、基于深度学习以及基于预训练微调的跨领域 ABSA 方法[11].

基于规则的跨领域 ABSA 方法主要通过人工总结和制定的情感分析规则与模式来对不同领域文本中的特定方面进行情感分析. 早期研究者的研究聚焦于各个特定领域, 分别设计一套完备的规则集合, 并借助领域词典精准定位方面词并判别情感类别[12]. 例如, Jakob 等人[13]提出了一种基于条件随机场的方面术语提取方法 (CRF-based approach for opinion target extraction), 首先设计了句法依赖关系链接规则, 并据此规则构建特征作为输入, 对句子中的方面术语进行提取. Li 等人[14]提出了一种关系自适应引导处理 (relation adaptive guidance processing, RAP) 方法, 用于源领域中方面与观点间句法规则的构建.

基于传统机器学习的跨领域 ABSA 方法利用带标注方面级文本数据进行模型训练, 并通过领域适应技术进行模型迁移[15]. 早期研究者们运用支持向量机等模型实施跨领域 ABSA 研究, 通过特征选择筛选关键特征对模型进行训练, 并采用领域适应技术将其迁移至目标领域[16]. 在此之后, Wang 等人[17]提出了一种基于渐进式机器学习范式的 ABSA 方法, 从简单样本开始, 通过迭代因子图推理逐步标注更具挑战性的样本, 在无需人工标注的情况下实现准确的机器标注. Pathan 等人[18]使用隐含狄利克雷分布主题建模方法和潜在语义分析主题建模方法提取方面主题, 并使用多项式朴素贝叶斯方法和支持向量机模型处理 ABSC 任务.

基于深度学习的跨领域 ABSA 方法运用神经网络架构自动从大量源领域文本数据中学习文本的深层次语义特征表示及跨领域的领域无关特征表示. 深度学习的兴起为跨领域 ABSA 提供了新思路, 特别是循环神经网络 (recurrent neural network, RNN) 及其变体成为研究的焦点. Wang 等人[19]提出了一种用于跨领域方面术语提取的 RNN, 通过句法关系来有效地减少词语层面的领域偏移, 并设计辅助任务来预测句法依赖树中任意两个相邻单词之间的关系. 注意力机制的出现进一步增强了语义理解能力. Liu 等人[20]提出了一种基于注意力机制的情感推理器, 为句子中的不同单词分配重要度, 设计了两种注意力机制: 内部注意力和全局注意力. Yang 等人[21]在利用上下文依赖的基础上, 应用词汇嵌入来合并额外的词汇线索. 并提出了依存注意力, 来在注意力推断中捕捉单词之间的句法依存线索. 生成对抗训练方式的出现也为跨领域 ABSA 领域提供了新的思路. Knoester 等人[22]通过领域对抗训练方法扩展了用于 ABSA 的 LCR-Rot-hop++模型来实现迁移学习策略, 同时彻底摆脱了对目标领域标注样本的依赖. Yin 等人[23]提出了一种基于对抗训练的生成技术, 从源领域中获取带标签样本, 并利用其生成具有方面级标签的目标领域样本.

基于预训练微调的跨领域 ABSA 方法在海量数据上进行无监督预训练, 获取通用知识和语义表示, 并在特定领域的少量标注数据上微调以实现模型迁移. Liu 等人[24]提出了一种基于 BERT 的跨领域 ABSA 算法, 该算法使用 BERT 结构提取句子级和方面级表示向量, 通过改进的卷积神经网络提取局部特征. 使用领域对抗神经网络使从源领域和目标领域提取的特征具有更高的相似性, 以期望分类器在源领域和目标领域都能取得良好的情感分类效果. Zhao 等人[25] 提出了一种基于预训练和微调策略的低资源跨领域基于方面的情感分类方法 (CDABSC). Zhou 等人[26]对 ABSA 领域的大语言模型进行了全面评估, 涉及 13 个数据集、8 个 ABSA 子任务和 6 个大语言模型, 在评估中设计了一种统一的任务表述方式, 统一了多种范式下多个 ABSA 子任务的多种大语言模型.

综上所述, 基于规则的跨领域 ABSA 方法的规则制定过程非常依赖人工, 同时所制定规则难以有效适配多个领域, 泛化性较差; 基于传统机器学习的跨领域 ABSA 方法需要深厚的领域知识及大量实验探索来进行特征工程的构建, 训练过程较为复杂; 无论是基于深度学习还是基于预训练微调的跨领域 ABSA 方法, 虽提升了预测准确性, 但在模型的训练或微调过程中都在一定程度上忽略了样本选择的影响, 从而限制目标模型的性能和可解释性.

1.2 跨领域实例迁移方法

跨领域实例迁移特别是高质量样本选择是处理低资源情境下模型训练难题的有效方法之一. 目的是通过对带标注源领域样本的情感知识进行迁移, 对标注样本匮乏的目标领域进行补充[27]. 根据是否有新样本生成, 跨领域实例迁移方法可以分为两种类型: 基于样本选择的跨领域实例迁移与基于样本生成的跨领域实例迁移.

基于样本选择的跨领域实例迁移方法旨在通过选择最有价值的样本进行标注和学习, 减少对大量标注数据的依赖, 以提高学习效率和模型性能. 此类方法的主流思路是基于领域差异度量来计算源领域样本对目标领域的重要性, 从而根据设定的度量阈值对源领域样本进行选择, 并与目标领域样本共同训练目标模型. 例如, Wu 等人[28]提出了一种主动样本选择方法, 通过 MMD 距离来度量源领域与目标领域之间的领域差异, 选择最具信息量的源

领域样本, 使用主动学习技术对源领域样本进行标注并迁移至目标领域. Wu 等人[29]提出了一种知识保存与分布对齐模型, 通过联合最小化信息损失和最大化领域分布对齐来实现异构领域适应. 另一种思路是通过聚类算法, 将不同领域数据样本根据特征向量相似性进行聚类后进行样本选择. 平瑞等人[30]通过对多数类样本聚类, 并采用弱平衡准则对集群进行降采样, 将选择的多数类样本与原训练集的少数类样本融合, 使少数类样本得到充分学习. Li 等人[31]提出了一种跨领域自适应聚类方法来解决有标签和无标签目标样本间及无标签目标样本与源领域间的不一致问题, 实现领域间和领域内适应.

基于样本生成的跨领域实例迁移方法旨在学习源领域和目标领域样本特征并构建统一特征空间, 从而生成尽可能与目标领域相似的伪目标领域样本. 此类方法通常利用源领域样本训练生成模型以生成伪目标领域新样本来训练目标领域模型, 从而利用生成的新样本实现跨领域知识迁移, 提升目标领域模型性能. Saito 等人[32]提出了一种非对称的三训练方法来给未标记的样本提供伪标签, 进行无监督领域迁移. Rotman 等人[33]提出了一种深度上下文自训练算法, 利用在序列标记任务上训练的表示模型, 将其应用于未标记数据来获取伪标签. Yu 等人[34]提出了一种基于领域自适应语言模型的跨领域数据增强方法 (cross-domain data augmentation approach based on domain-adaptive language modeling, DA^{2}LM), 使用领域自适应语言模型学习跨领域的共享上下文和注释, 为未标记的目标领域数据分配伪标签. 同时生成对抗训练形式的诞生使得伪标签生成模型的训练具备更强的自主性. Li 等人[35]提出了一种基于选择性对抗学习 (selective adversarial learning, SAL) 的跨领域 ABSA 方法, 可以动态学习每个词的权重, 使得更重要的词可以具有更高的权重, 从而实现细粒度词语级的领域自适应. Zhou 等人[36]在选择性对抗学习方法基础上提出了一种自适应混合框架 (adaptive hybrid framework, AHF), 利用在目标数据上生成的伪标签来训练任务分类器, 进一步提升了模型性能.

此外, 随着迁移学习向多源领域情景的扩展, 跨领域实例迁移的研究对多个源领域样本之间的协同性愈发重视. 例如, He 等人[37]提出了一种多源领域自适应的协同样本选择方法, 通过衡量不同源领域样本与目标领域的相关性和互补性, 构建了一个协同样本选择框架, 旨在解决多源领域数据存在差异时的样本选择问题. Dai 等人[38]提出了一个两阶段的领域适应框架, 在第 1 阶段采用多任务架构为有标签源领域建模领域通用特征和领域特定特征, 在第 2 阶段采用选择性领域适应方法及协同集成方法从最接近的源领域处进行知识迁移.

综上所述, 现有的基于样本选择的跨领域实例迁移方法在处理新任务或新领域时需要重新设计样本选择机制或样本聚类算法, 对不同任务及领域的泛化性能比较差; 而基于样本生成的跨领域实例迁移方法通常会选择模型对无标签数据的高置信度预测作为伪标签, 这会导致模型更偏向于容易处理的样本, 而对于困难样本的处理不足. 同时, 跨领域实例迁移方法大多未考虑样本间的协同性, 且均单独面向实例迁移过程, 并不能同时进行目标领域模型的训练, 一定程度上影响了训练效率.

2 任务描述

对于 ABSA 任务的 3 种子任务: ATE、ACD 及 ABSC 任务, 本文均将其视为文本序列标记问题. 针对 ATE 任务, 对于给定的一条输入文本, 目标是预测该文本中的方面术语标签序列, 其中每个方面术语标签来自标签集{O, ASP}. 标签 O 表示该单词不是方面术语, 标签 ASP 表示该单词为方面术语. 图 1 中给出了一条示例文本及 ABSA 各子任务的标签序列. 在示例文本中, “screen”和“battery”是两个方面术语, 分别被标注为 ASP 标签.

针对 ACD 任务, 对于给定的一条输入文本, 目标是预测该文本对应的方面类别标签序列, 其中每个方面类别标签来自预先定义的标签集{O, ASP-C-1, ASP-C-2, ASP-C-3,…}. 标签 O 表示该单词未表达方面类别信息, 而 ASP-C-1、ASP-C-2、ASP-C-3 等则分别代表单词表达的不同方面类别信息. 在图 1 示例文本中, “screen”和“battery”这两个方面术语被进一步识别出属于不同的方面类别, 分别为显示质量 (display quality) 和电池 (battery).

针对 ABSC 任务, 对于给定的一条输入文本 $X=\{w_1, w_2,\ldots, w_n\}$, 任务目标是预测该文本对应的方面级情感标签序列 $Y=\{l_1, l_2,\ldots, l_n\}$, 其中每个方面级情感标签 l 来自标签集{O, T-POS, T-NEG, T-NEU}, 标签 O 表示这个单词未表达出方面级情感, T-POS、T-NEG、T-NEU 分别表示该单词表达积极、消极和中性情感. 在图 1 示例文本中, “screen”和“battery”是两个方面术语, 它们在句子中表达的情感极性分别是积极和消极.

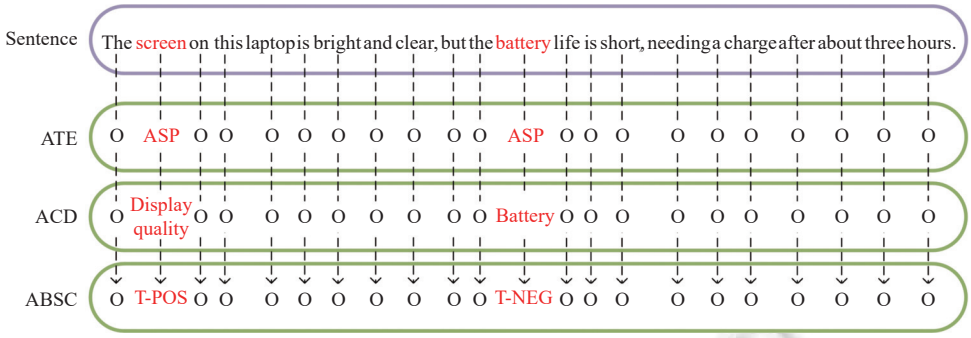


图 1 ABSA 各子任务示例句子及相应标签

在跨领域 ABSA 任务中, 源领域带标签样本记为 $D_S=\{X_S^k,Y_S^k\}_{k=1}^{N_S}$, 目标领域无标签样本记为 $D_T=\{X_T^k\}_{k=1}^{N_T}$. 其中 X 表示文本, Y 表示方面级情感标签序列. 目标为在源领域带标签数据集 D_S 上训练一个 ABSA 模型 $M_S: X_S \to Y_S$, 然后将模型迁移至目标领域 D_T 得到目标模型 $M_T: X_T \to Y_T$, 预测目标领域中无标签样本 X_T 对应的情感标签序列 Y_T.

为提升跨领域 ABSA 任务的效果, 须关注目标模型在目标领域的性能, 本文通过目标领域分类错误率来对此进行量化. 目标模型可表示为:

$$M_T^*=\underset{M_T}{\operatorname{argmin}} Prob\left(M_T(X_T)\neq Y_T \mid (X_T,Y_T)\sim D_T\right) \tag{1}$$

在跨领域环境下训练迁移模型时, 通常只有目标领域的未标记数据, 目标领域样本的标签对于模型是不可见的. 为了解决上述问题, 将目标转化为通过降低目标模型的泛化误差来优化目标模型的性能. 目标领域的泛化误差可以表示为:

$$\varepsilon_T=\varepsilon_{\text{base}}+\lambda_1\cdot Complexity+\lambda_2\cdot\delta_{\text{base}}(D_S,D_T) \tag{2}$$

其中, $\varepsilon_{\text{base}}$ 为目标模型在源领域上的经验误差, $Complexity$ 为模型复杂性度量, $\delta_{\text{base}}(D_S,D_T)$ 为源领域与目标领域的领域适应性损失, λ_1, λ_2 为对应项的系数.

显而易见的是, 模型的复杂性对于模型的泛化误差有重要影响. 过于复杂的模型可能会导致过拟合, 即在源领域训练集上表现良好但在目标领域上性能不佳. 因此, 本文讨论固定模型的情况下源领域数据选择对模型泛化误差界的优化效果. 对于同一个模型, $Complexity$ 相同, 源领域训练数据的选择对于优化源领域经验误差 $\varepsilon_{\text{base}}$ 和领域适应性损失 $\delta_{\text{base}}(D_S,D_T)$ 至关重要. 为了探究源领域训练数据的选择对于模型在目标领域泛化误差界的影响, 本文引入了样本的数据质量因素 (Q), 目标模型泛化误差表示为:

$$\varepsilon_T=(\varepsilon_{\text{base}}-k_1\cdot Q)+\lambda_1\cdot Complexity+\lambda_2\cdot(\delta_{\text{base}}(D_S,D_T)-k_2\cdot Q) \tag{3}$$

其中, $\varepsilon_{\text{base}}-k_1\cdot Q$ 为考虑了源领域训练样本的数据质量要素 Q 后目标模型在源领域上的经验误差, 记作 ε_S^Q; $\delta_{\text{base}}(D_S,D_T)-k_2\cdot Q$ 为考虑了源领域训练样本的数据质量要素 Q 后源领域训练数据与目标领域数据间的领域适应性误差, 记作 $\delta^Q(D_S^-,D_T)$; k_1, k_2 分别为数据质量 Q 对经验误差和领域适应性误差的影响系数.

因此, 模型结构固定的情况下, 为了降低目标模型的泛化误差并提升模型迁移效果, 训练目标转变为寻找能够使模型经验误差 ε_S^Q 及领域适应性误差 $\delta^Q(D_S^-,D_T)$ 最小化的源领域训练数据集 D_S^*:

$$D_S^*=\underset{D_S^-}{\operatorname{argmin}}\left(\varepsilon_S^Q+\lambda_1\cdot Complexity+\lambda_2\cdot\delta^Q(D_S^-,D_T)\right),\ D_S^-\subseteq D_S \tag{4}$$

3 基于高质量样本选择的跨领域方面级情感分析方法

为了高效筛选源领域训练数据, 提出了一种基于高质量样本选择的跨领域 ABSA 方法 (HQSS-CDABSA). 方法基本框架如图 2 所示. HQSS-CDABSA 方法包括 3 个部分: 一是领域间选择, 即在多个领域中选择特定领域作为迁移训练的源领域; 二是领域内选择, 即在源领域内选择其子集作为迁移训练的源领域样本; 三是模型训练, 即在被选择源领域样本集合上训练下游任务迁移模型.

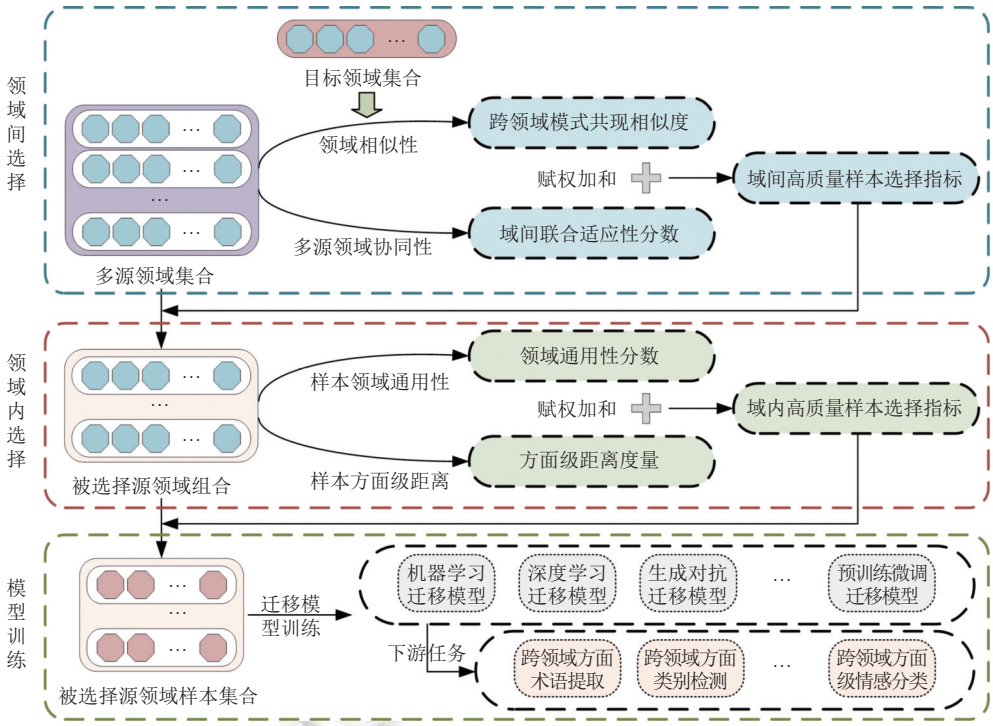


图 2 基于高质量样本选择的跨领域 ABSA 方法框架图

3.1 领域间选择

为了从多个领域中选择出最合适进行迁移模型训练的一个或多个源领域, 综合考虑了源领域与目标领域间的相似性和多源领域之间的协同效应, 提出了一种领域选择指标——域间高质量样本选择指标 Q_{inter}. 对于包含 k 个源领域的集合 $\{D_S^1, D_S^2, \ldots, D_S^k\}$, Q_{inter} 定义为:

$$Q_{\text{inter}} = \frac{\alpha}{k} \cdot \sum_{i=1}^{k} A\left(D_S^i, D_T\right) + \frac{2 \cdot \beta}{k^2 - k} \cdot \sum_{i=1}^{k} \sum_{j=i+1}^{k} C\left(D_S^i, D_S^j\right) \tag{5}$$

其中, $A(D_S^i, D_T)$ 代表第 i 个源领域与目标领域之间的跨领域模式共现相似度; $C\left(D_S^i, D_S^j\right)$ 代表第 i 个源领域与第 j 个源领域之间的联合适应性分数; α 和 β 分别代表两部分的权重系数. 具体计算流程如图 3 所示.

(1) 跨领域模式共现相似度

为了衡量源领域与目标领域之间的关系, 提出了一种跨领域模式共现相似度. 具体来说, 对于源领域 D_S 及目标领域 D_T 的句子样本集合, 分别记为 $Text_S = \{X_S^k\}_{k=1}^{N_S}$ 和 $Text_T = \{X_T^k\}_{k=1}^{N_T}$, 其中 X 代表句子样本.

对每个领域的句子样本集合进行以下操作.

• 特征提取: 将每条句子文本进行分词操作得到词汇列表, 并对词汇进行停用词去除预处理操作. 通过 BERT 预训练模型将词汇转化为高维向量表示, 以捕捉词汇间的语义关系.

• 共现图构建: 对源领域和目标领域分别构建其词共现矩阵 U_S 和 U_T, 矩阵的元素 u_{ij} 表示词汇 v_i 和 v_j 在句子中的共现次数. 分别将两个领域的共现矩阵 U_S 和 U_T 转换为领域共现图 $G_S = (V_S, E_S)$ 和 $G_T = (V_T, E_T)$, 图中每个节点代表一个词汇, 边代表两词汇存在共现情况, 边的权重代表共现强度.

• 图神经网络构建: 以目标领域为例, 将领域共现图 $G_T = (V_T, E_T)$、共现矩阵 U_T、词特征矩阵 $\mathbf{E} \in \mathbb{R}^{|V_T| \times d}$ 作为图神经网络的输入. 网络包含多个图卷积层和一个池化层, 设第 l 个图卷积层的节点表示为 $\mathbf{H}^{(l)}$, 则节点更新公式为:

$$\mathbf{H}^{(l+1)} = ReLU\left(\tilde{\mathbf{A}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right) \tag{6}$$

其中, $\tilde{\mathbf{A}} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$ 为归一化的邻接矩阵, $\mathbf{A}$ 为共现矩阵 U_T 单位化得到的邻接矩阵, $\mathbf{D}$ 为度矩阵.

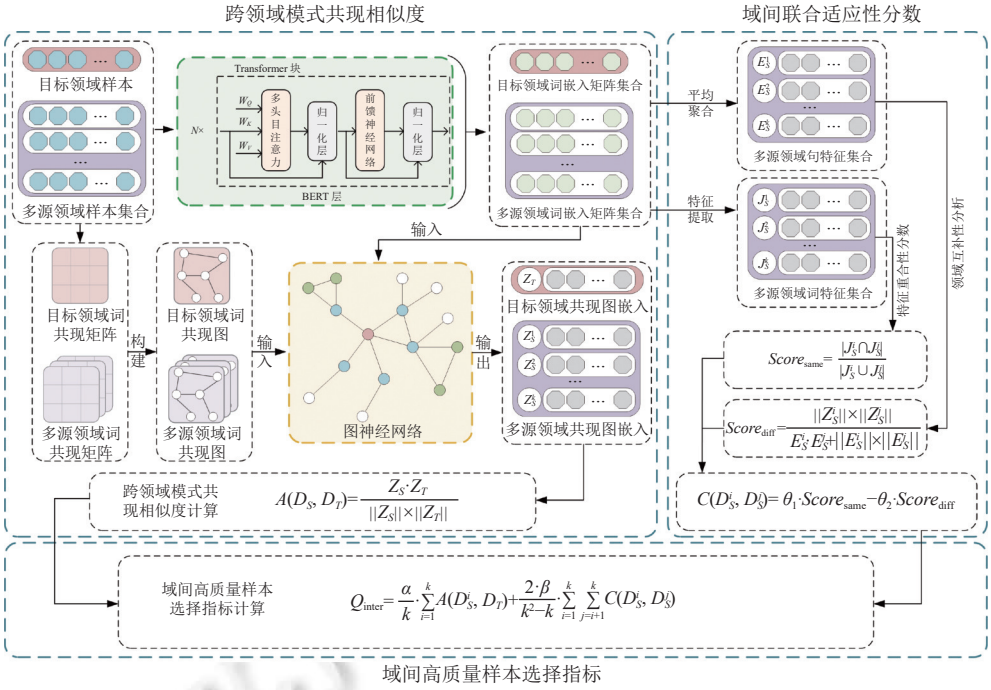


图 3 领域间高质量样本选择流程图

经过最后的图卷积层处理得到的节点嵌入表示 $\mathbf{H}^{(\text{last})}$, 由池化层进行处理, 得到目标领域的领域共现图的整体嵌入表示:

$$Z_T = \frac{1}{|V_T|}\sum_{i\in V_T}\mathbf{H}_i^{(\text{last})} \tag{7}$$

同理可得到源领域的领域共现图的整体嵌入表示 Z_S.

• 模式共现相似度计算: 计算源领域与目标领域的领域共现图嵌入表示 Z_S 和 Z_T 之间的余弦相似度作为两领域间的跨领域模式共现相似度:

$$A(D_S, D_T) = \frac{Z_S \cdot Z_T}{\|Z_S\| \times \|Z_T\|} \tag{8}$$

(2) 域间联合适应性分数

为了衡量多领域之间的领域互补性及领域特征重合性, 提出了一种域间联合适应性分数. 具体来说, 对于源领域 D_S^i 及 D_S^j 的句子样本集合, 分别记为 $Text_i = \left\{X_i^k\right\}_{k=1}^{N_S^i}$ 和 $Text_j = \left\{X_j^k\right\}_{k=1}^{N_S^j}$, 其中 X 代表句子样本. 对两个源领域样本集合进行以下操作.

• 特征提取: 将每条句子文本进行分词操作得到词汇列表, 并对词汇进行停用词去除预处理操作. 通过 BERT 预训练模型将词汇转化为高维向量表示, 以捕捉词汇间的语义关系. 对每条句子的词向量矩阵采用平均聚合方式得到该条句子的句特征向量, 将两个领域内所有句子的句特征向量采用平均聚合方式分别得到两个领域的特征向量表示 E_S^i 和 E_S^j.

• 领域互补性分数计算: 根据以下公式计算两个源领域的领域互补性分数:

$$Score_{\text{diff}} = \frac{\left\|E_S^i\right\| \times \left\|E_S^j\right\|}{E_S^i \cdot E_S^j + \left\|E_S^i\right\| \times \left\|E_S^j\right\|} \tag{9}$$

• 领域特征重合性分数计算: 对于两个源领域内的每个句子, 提取其包含的词汇特征, 分别得到两个源领域的词特征集合 J_S^i 和 J_S^j. 根据公式 (10) 计算两个源领域的领域特征重合性分数:

$$Score_{\text{same}} = \frac{\left|J_S^i \cap J_S^j\right|}{\left|J_S^i \cup J_S^j\right|} \tag{10}$$

• 域间联合适应性分数计算: 根据公式 (11) 计算两个源领域的域间联合适应性分数:

$$C\left(D_S^i, D_S^j\right) = \theta_1 \cdot Score_{\text{same}} - \theta_2 \cdot Score_{\text{diff}} \tag{11}$$

其中, θ_1 和 θ_2 代表领域特征重合性分数与领域互补性分数的权重.

域间联合适应性分数代表了选择多个源领域时在领域特征重合性与领域互补性之间的权衡. 选择源领域组合时, 尽量确保特征相似的领域能够互相补充, 而差异较大的领域则可以提供更多样化的信息. 这种选择能够确保模型在目标领域上具有良好的泛化能力.

3.2 领域内选择

为了从源领域中选择出最合适进行迁移模型训练的源领域样本, 综合考虑了源领域样本的领域通用性、源领域样本与目标领域之间的方面级距离, 提出了一种样本选择指标——域内高质量样本选择指标 Q_{intra}. 对于选择出的一个源领域样本数据 X_S, Q_{intra} 定义为:

$$Q_{\text{intra}} = \mu \cdot Gen(X_S) - \eta \cdot Dis_{\text{asp}}(X_S, D_T) \tag{12}$$

其中, $Gen(X_S)$ 为样本 X_S 的领域通用性分数, $Dis_{\text{asp}}(X_S, D_T)$ 为样本 X_S 与目标领域的方面级距离度量, μ 和 η 分别为两种分数的权重系数. 具体计算流程如图 4 所示.

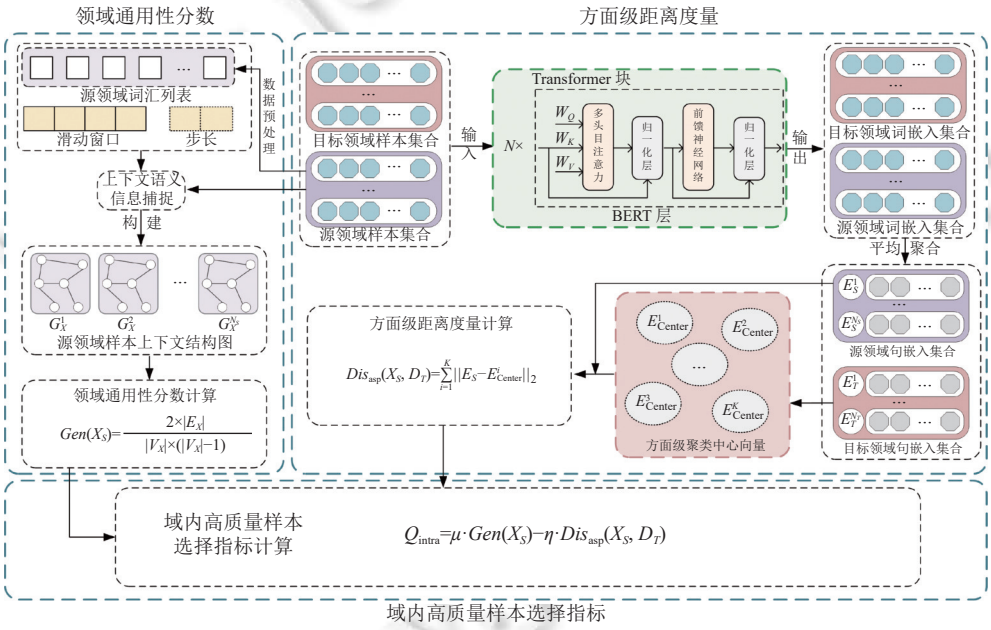


图 4 领域内高质量样本选择流程图

(1) 领域通用性分数

为了筛选出在迁移训练过程中泛化性能更强的源领域数据, 提出了一种领域通用性分数来指导数据的选择. 具体来说, 对于一条源领域文本数据 X_S, 进行以下操作.

• 数据预处理: 将文本数据 X_S 进行分词操作得到其词汇列表, 并对词汇进行停用词去除、词汇去重等预处理操作.

• 上下文信息捕捉: 对文本数据 X_S 采用滑动窗口的方式捕捉文本数据中的上下文语义信息. 为了捕捉更细微的上下文语义信息, 将滑动窗口大小设置为 4, 步长为 2.

• 上下文结构图构建: 将文本 X_S 对应的词汇列表中词汇作为上下文结构图的节点, 对于每个滑动窗口中共现的词汇列表元素, 在结构图中添加表示共现关系的边. 构建完成后的文本 X_S 对应的上下文结构图记作 $G_X=(V_X,E_X)$.

• 领域通用性分数计算: 根据公式 (13) 计算领域通用性分数:

$$Gen(X_S)=\frac{2\times|E_X|}{|V_X|\times(|V_X|-1)} \tag{13}$$

从公式 (13) 可以发现, 领域通用性分数更高的文本, 其上下文结构图连通性更高, 说明文本中词汇之间的联系更强. 这导致这些词汇在更多的上下文环境中适用, 从而使文本在多个方面更具领域通用性, 因为它能更全面地反映出领域内的共同特征和情感信息. 这种具备高连通性结构图的源领域数据为迁移学习提供了更丰富的基础.

(2) 方面级距离度量

由于领域通用性分数更关注的是样本在本领域内的词汇联系程度, 缺乏与目标领域的联系, 提出了一种方面级距离度量来对域内高质量样本的选择进行补充. 具体来说, 对于一条源领域样本 X_S 和目标领域 D_T, 进行以下操作.

• 特征提取: 将目标领域 D_T 中每条句子文本进行分词操作得到词汇列表, 并对词汇进行停用词去除预处理操作. 通过 BERT 预训练模型将词汇转化为高维向量表示, 以捕捉词汇间的语义关系. 对每条句子的词向量矩阵采用平均聚合方式得到该条句子的句特征向量, 最终得到源领域样本 X_S 的句向量 E_S 与目标领域 D_T 的句向量集合 $\{E_T^1,E_T^2,\dots,E_T^{N_T}\}$.

• 目标领域样本聚类: 将根据目标领域的常见 K 个方面类别, 分别为每个方面类别构造 n 条伪目标领域样本, 采用半监督聚类及监督聚类方式对目标领域句向量集合进行方面级聚类操作. 具体来说, 从构造的 $K\times n$ 条伪目标领域样本中对各个类别分别选取一条样本作为初始聚类中心样本, 同时将剩余的伪目标领域样本用于聚类算法的监督训练. 最终得到目标领域 K 个方面类别的类中心向量 $\{E_{\text{Center}}^1,E_{\text{Center}}^2,\dots,E_{\text{Center}}^K\}$.

• 方面级距离度量计算: 根据公式 (14) 计算源领域样本 X_S 与目标领域 D_T 间的方面级距离度量:

$$Dis_{\text{asp}}(X_S,D_T)=\sum_{i=1}^{K}\left\|E_S-E_{\text{Center}}^i\right\|_2 \tag{14}$$

方面级距离度量通过量化源领域样本与目标领域方面级类中心之间的距离, 补充了领域通用性分数的不足, 确保所选样本不仅在源领域内具备良好的词汇联系, 还能在目标领域中保持相关性.

3.3 模型训练

经过领域间和领域内两个层面的筛选工作后得到高质量训练样本集合, 记为 $Data_{\text{Select}}$, 并在 $Data_{\text{Select}}$ 基础上进行下游任务迁移模型的训练.

为了全面探究所提方法对不同模型架构在下游任务迁移中的影响, 本文设置了不同类型的模型进行实验: 基于深度学习的模型、基于生成对抗训练的模型以及基于预训练的模型. 这 3 种类型的模型具有不同的优势: 基于深度学习的模型能够快速从数据中学习复杂的特征表示; 基于生成对抗训练的模型可以通过生成器和判别器的对抗学习来挖掘数据的潜在分布; 基于预训练的模型可以借助大规模通用数据上预训练的知识, 快速适应新的任务.

针对基于深度学习的模型, 本文选择了更擅长处理长文本序列数据的循环神经网络作为基本网络架构, 并将其变体模型——双向循环神经网络模型及基于注意力机制的循环神经网络模型引入实验的讨论范围. 针对基于生成对抗训练的模型, 构建了生成器和判别器. 生成器生成与源领域样本相似的样本, 判别器负责区分生成的样本和真实的源领域样本. 通过交替训练生成器和判别器, 使生成器生成更加逼真的样本. 针对基于预训练的模型, 利用在大规模通用数据集上的预训练 BERT 模型, 后接多层感知机模型, 保留预训练模型的通用知识, 同时让模型能够适应下游任务.

在训练过程中, 通过调整超参数, 如学习率、批次大小和训练轮数, 以确保模型能够稳定收敛并达到最佳性

能. 采用反向传播算法来更新模型的参数, 通过计算损失函数来衡量模型的预测结果与真实标签之间的差异. 在 $Data_{Select}$ 上完成模型训练后, 在目标领域数据集上对各模型的效果进行验证. 通过计算 $F1$ 值来评估模型在目标领域的表现, 反映模型在目标领域的分类任务上的能力.

4 主要实验结果

4.1 实验数据

为了评估所提方法的性能, 在共 6 个领域的 ABSA 数据集上进行了实验. 针对 ATE 及 ABSC 任务, 所使用数据集来自 4 个不同的领域: 餐厅 (R)、笔记本电脑 (L)、设备 (D) 和服务 (S), 其统计数据如表 1 所示. 餐厅数据集由 SemEval-2014[39]、SemEval-2015[40]和 SemEval-2016[41]的餐厅评论组成. 笔记本电脑数据集包含了 SemEval-2014 的笔记本电脑评论. 设备数据集由 Hu 等人[42]创建, 包含了来自 5 种不同数字产品的评论数据. 服务数据集包含来自 Web 服务的评论, 并由 Toprak 等人[43]标注.

针对 ACD 任务, 所使用数据集来自 3 个不同的领域: 餐厅 (R_{ACD})、旅馆 (H_{ACD}) 和街道 (S_{ACD}), 其统计数据如表 2 所示. 餐厅数据集由 SemEval-2016 的餐厅评论组成. 旅馆数据集包含了 SemEval-2015 的旅馆领域评论. 街道数据集由 Saeidi 等人[44]创建, 包含了针对街区领域的评论数据.

表 1 ATE 及 ABSC 任务数据集统计数据

Dataset	Domain	#Sentences	#Train	#Test
R	Restaurant	6035	3877	2158
L	Laptop	3845	3045	800
D	Device	3836	2557	1279
S	Service	2239	1492	747

表 2 ACD 任务数据集统计数据

Dataset	Domain	#Sentences	#Train	#Test
R_{ACD}	Restaurant	2676	2000	676
S_{ACD}	Sentihood	4468	2977	1491
H_{ACD}	Hotel	1330	1064	266

本文在实验中使用 $F1$ 分数作为评价指标, $F1$ 分数的计算公式如下:

$$F1 = \frac{2\times Precision\times Recall}{Precision + Recall} \tag{15}$$

其中, $Precision$ 为准确率, $Recall$ 为召回率, 计算公式如下:

$$Precision = \frac{TP}{TP + FP} \tag{16}$$

$$Recall = \frac{TP}{TP + FN} \tag{17}$$

其中, TP 表示真正例, 即模型正确预测为正例的数量; FP 表示假正例, 即模型错误地将负例预测为正例的数量; FN 表示假负例, 即模型错误地将正例预测为负例的数量.

4.2 实验设置及基准模型

本文使用 PyTorch 的 Embedding 模块及 BERT 预训练模型来获取词嵌入表示, 其中向量维数设定为 100. 输入数据的批次处理大小设置为 32. 在模型结构方面, 实验中各基线模型分类器的隐藏层大小为 128, 为避免模型过拟合, 在隐藏层两侧设置 dropout 层, dropout 率为 0.2. 针对 ATE、ACD 及 ABSC 任务, 分类器的输出维度分别为 2、7、4. 在评估标准方面, 当对某条文本的所有分词标签预测均准确时, 判定为对该文本标签序列的预测正确.

模型参数优化采用 Adam 算法, 学习率设定为 0.001. 为确保实验结果的可靠性与稳定性, 对每个迁移任务均进行了 5 次独立实验, 并取 5 次实验结果的平均值作为最终结果. 实验运行环境为 Windows 系统, 硬件配置包括 NVIDIA RTX A4000 GPU (16 GB×2) 和 Intel (R) Xeon (R) W-2175 CPU (@ 2.50 GHz), 软件采用 PyTorch 2.1.2 框架.

为了全面评估所提方法对不同类型深度学习模型性能的增强效果, 本文针对 ABSA 任务在以下 5 种基线模型上进行了比较.

(1) RNN[45]: 一种处理序列数据的神经网络模型, 通过隐藏状态传递历史信息, 适合时间序列任务, 但存在梯度消失问题, 难以捕捉长距离依赖关系. 本实验中, RNN 的隐藏层维度设置为 128, 层数为 1, 激活函数采用 tanh.

(2) Bi-LSTM[46]: LSTM 的扩展模型, 通过双向结构同时捕捉前后文信息, 有效解决长序列依赖问题, 广泛应用于文本分类和序列标注任务. 实验中 Bi-LSTM 的隐藏层维度为 128, 层数为 1, 采用 CrossEntropyLoss 计算模型损失.

(3) Att-LSTM[47]: 在 LSTM 模型基础上加入注意力机制, 动态分配权重以聚焦关键信息, 提升模型对重要特征的捕捉能力, 适用于复杂序列建模. 该模型中 LSTM 部分的参数与 Bi-LSTM 模型保持一致, 注意力机制采用加性注意力, 其隐藏层维度为 64.

(4) GAN[48]: 由生成器和判别器组成, 通过生成对抗训练生成目标领域的伪情感数据, 缓解领域间数据分布差异, 提升模型在目标领域的泛化能力. 本实验中, 生成器和判别器均采用 RNN 结构, 隐藏层维度设置为 128. GAN 训练中采用 BCEWithLogitsLoss 计算损失, 用于判别器对数据真伪的判断和生成器对判别器的欺骗训练. 情感分类器采用 CrossEntropyLoss 进行损失计算.

(5) BERT[49]: 基于 Transformer 的预训练语言模型, 通过双向上下文理解文本语义. BERT 通过大规模预训练捕捉通用语言特征, 能够有效缓解领域差异问题, 显著提升数据稀缺的目标领域中的分类性能. 实验采用 BERT-base-uncased 作为预训练模型, 将输出向量维度设置为 128, 用于对 BERT 输出特征进行降维.

为了验证所提方法相比其他迁移学习方法在性能上的优越性, 使用以下模型作为对比模型.

(1) FMIM[50]: 一种采用互信息最大化技术的基于特征的领域自适应方法.

(2) CDRG[51]: 一种跨领域数据增强方法, 通过利用带标注的源领域样本, 基于掩码语言模型来生成带标注的目标领域样本.

(3) GCDDA[52]: 一种生成式跨领域数据增强框架, 利用预先训练的 BART 模型来生成具有方面级注释的目标领域数据.

(4) DA^{2}LM[34]: 一种基于领域自适应语言模型的跨领域数据增强方法, 通过领域自适应语言模型为未标记的目标领域数据分配伪标签.

4.3 预实验

为了确定 HQSS-CDABSA 方法最佳的样本选择比例, 针对跨领域 ABSC 任务在 Att-LSTM 及 BERT 模型上进行了预实验, 预实验的实验结果如表 3 所示, 其中粗体数据为不同任务最高的数值.

表 3 不同模型在不同样本选择比例下处理跨领域 ABSC 任务的 $F1$ 值比较结果 (%)

模型	Data	D→R	D→S	L→R	L→S	R→D	R→L	R→S	S→D	S→L	S→R	Avg
Att-LSTM	All	42.49	16.30	38.46	11.97	42.59	40.07	19.18	47.98	39.01	32.66	33.07
	90%	42.86	16.99	39.21	12.44	42.77	40.51	19.73	48.36	39.72	32.94	33.55
	70%	**44.43**	**21.95**	**41.75**	**15.41**	**44.65**	**43.32**	**21.12**	**53.51**	**41.78**	**34.66**	**36.26**
	50%	38.27	12.03	33.24	7.35	37.31	36.22	14.03	40.09	33.28	26.49	27.83
	30%	21.11	4.65	18.67	4.02	22.49	18.87	10.14	23.88	19.96	14.28	15.81
BERT	All	60.53	40.96	61.36	36.53	67.71	61.97	38.62	64.40	55.62	61.83	54.95
	90%	60.82	42.86	61.44	36.54	67.99	62.29	39.47	66.21	56.88	61.97	55.65
	70%	**62.69**	**48.49**	**63.12**	**36.99**	**69.70**	**64.37**	**42.05**	**75.53**	**58.10**	**62.64**	**58.37**
	50%	39.27	28.44	39.96	24.33	41.74	37.26	25.11	40.05	36.38	38.97	35.15
	30%	28.95	24.37	31.22	20.08	25.16	25.48	20.19	30.58	24.46	24.87	25.54

表 3 中的实验结果表明, HQSS-CDABSA 方法在不同样本选择比例下对模型性能的影响存在显著差异. Att-LSTM 和 BERT 模型均在 70% 的样本选择比例下取得最佳性能, 平均 $F1$ 值分别为 36.26% 和 58.37%, 显著高于其他比例下的性能. 这源于在 70% 的样本选择比例下, HQSS-CDABSA 方法能够有效筛选出高质量样本, 同时保留足够的训练数据量, 避免了因数据量过少导致的模型欠拟合问题. 高质量样本的筛选减少了噪声数据对模型训练的干扰, 从而提升了模型的泛化能力和性能. 在全样本训练和 90% 的样本选择比例下, 模型性能虽然较高, 但未达到最佳. 这源于训练数据中的噪声数据过多, 尽管数据量充足, 但噪声数据的存在限制了模型的性能提升. 而

在 50% 和 30% 的样本选择比例下, 由于过少的训练数据量导致模型出现欠拟合现象, 模型性能显著下降.

综上所述, 在上述 5 种样本选择比例中, HQSS-CDABSA 方法在 70% 的样本选择比例下表现最优. 这表明, 70% 的样本选择比例能够在保证筛选出的样本高质量的同时, 提供足够的训练数据量, 从而提升模型的性能. 相比之下, 更高的样本选择比例由于保留的源领域样本过多, 无法有效筛除其中的噪声数据. 而更低的样本选择比例尽管筛选出的源领域样本质量更高, 但其训练数据量的不足仍然会限制模型的性能.

4.4 对比实验

为了验证方法的优越性, 本节以跨领域 ATE 任务为下游任务, 将 HQSS-CDABSA 方法与 4 种对比模型进行了性能对比. 实验结果如表 4 所示, 其中 HQSS-CDABSA 方法的样本选择比例为 70%, 以 BERT 模型为下游模型.

表 4 不同模型处理跨领域 ATE 任务的 $F1$ 值比较结果 (%)

模型	D→R	D→S	L→R	L→S	R→D	R→L	R→S	S→D	S→L	S→R	Avg
FMIM	61.64	**49.53**	68.67	**51.68**	36.11	50.57	47.60	35.26	39.14	57.43	49.76
CDRG	57.51	43.19	68.63	51.07	34.89	55.50	**49.97**	38.59	39.49	60.20	49.90
GCDDA	53.70	30.74	58.00	30.31	44.25	64.06	35.69	39.16	43.95	63.53	46.34
DA^{2}LM	**63.86**	38.20	**68.72**	41.06	44.29	54.55	43.41	43.24	44.96	**65.78**	50.80
HQSS-CDABSA	62.69	48.49	63.12	36.99	**69.70**	**64.37**	42.05	**75.53**	**58.10**	62.64	**58.37**

表 4 中的实验结果表明, 所提出 HQSS-CDABSA 方法在 10 组跨领域 ATE 任务中整体表现优越, 平均 $F1$ 值达到 58.37%, 分别较 FMIM、CDRG、GCDDA 和 DA^{2}LM 方法提升了 8.61%、8.47%、12.03% 和 7.57%, 充分验证了 HQSS-CDABSA 方法在跨领域任务中缓解源领域样本负迁移效应的有效性与优越性.

具体来看, HQSS-CDABSA 方法在 S→D、R→D、S→L 和 S→R 迁移任务中取得显著优势, 其中在 S→D 任务中 $F1$ 值达到 75.53%, 远超其他模型. 这一现象表明, HQSS-CDABSA 方法能够在源领域和目标领域存在较大语义差异的情境下, 有效筛选出对目标任务真正有利的源领域样本, 提升模型在目标领域的泛化性能. 这是由于 HQSS-CDABSA 方法引入了高质量样本选择机制, 能够从多个角度评估源领域样本与目标任务的匹配度, 有效过滤掉可能带来负迁移的无关或噪声样本.

传统迁移方法中, 通常默认全部源领域样本都具有迁移价值, 而在实际场景中, 不同领域间存在语义不一致、风格差异或标签分布偏移等问题, 使得部分源领域样本对目标任务无益, 反而产生负迁移效应, 降低目标领域模型的性能. HQSS-CDABSA 方法通过融合多角度指标, 构建领域间及领域内的高质量样本选择指标, 识别并迁移对目标任务有益的源领域样本, 从根本上提升了迁移效果.

而在 L→S、D→S、R→S 等任务中, 由于领域差异较大, 源领域中的高质量样本占比本身较低, HQSS-CDABSA 方法在样本总量受限的条件下未能充分捕获领域共性. 但 HQSS-CDABSA 方法在大多数任务上依旧展现出良好的迁移稳定性, 表明其所使用的高质量样本选择策略具有良好的通用性和鲁棒性.

综上所述, 所提出的 HQSS-CDABSA 方法通过高质量样本选择, 有效缓解了跨领域迁移中普遍存在的负迁移问题, 在保证训练数据数量的同时, 较大程度保留了对目标领域有利的信息, 从而显著提升模型的跨领域泛化能力.

4.5 单源领域场景实验结果

针对跨领域 ATE、跨领域 ACD 及跨领域 ABSC 任务, 在单源领域场景下的具体实验结果分别如表 5–表 7 所示. 表 5–表 7 中展示了在两种不同的训练数据选择策略下不同模型在各子任务上的效果对比. 两种训练数据选择策略分别为: (1) 使用全样本训练集训练迁移模型; (2) 通过所提出的 HQSS-CDABSA 方法选择 70% 数据量的训练集样本训练迁移模型.

表 5 中的实验结果表明, 所提出的 HQSS-CDABSA 方法能够有效筛选出适合迁移训练的高质量样本, 从而提升 5 种模型的性能表现. 5 种模型在 10 组跨领域 ATE 任务上的平均 $F1$ 值均有所提升, 具体表现为: RNN 从 20.08% 提升至 22.57%, Bi-LSTM 从 26.59% 提升至 28.62%, Att-LSTM 从 33.07% 提升至 36.26%, GAN 从

29.60% 提升至 33.85%, BERT 从 54.95% 提升至 58.37%. 这一结果充分证明了 HQSS-CDABSA 方法的有效性, 能够有效筛选训练数据中的高质量样本, 显著减少源领域与目标领域之间的分布差异.

表 5 不同模型在不同数据选择策略下处理跨领域 ATE 任务的 $F1$ 值比较结果 (%)

模型	Data	D→R	D→S	L→R	L→S	R→D	R→L	R→S	S→D	S→L	S→R	Avg
RNN	All	32.87	4.95	24.50	2.90	26.06	27.62	2.90	31.85	24.58	22.59	20.08
	70%	35.45	5.20	28.30	3.41	28.99	29.72	4.69	34.22	29.20	26.53	22.57
Bi-LSTM	All	37.06	5.70	34.28	4.44	42.45	36.39	2.64	44.10	32.49	26.31	26.59
	70%	38.06	6.21	35.81	4.69	43.69	37.87	4.95	49.09	37.10	28.71	28.62
Att-LSTM	All	42.49	16.30	38.46	11.97	42.59	40.07	19.18	47.98	39.01	32.66	33.07
	70%	44.43	21.95	41.75	15.41	44.65	43.32	21.12	53.51	41.78	34.66	36.26
GAN	All	38.32	5.45	34.47	4.69	39.63	34.14	3.67	56.29	41.71	37.60	29.60
	70%	39.69	5.70	38.51	5.45	53.44	45.91	5.19	58.30	47.60	38.75	33.85
BERT	All	60.53	40.96	61.36	36.53	67.71	61.97	38.62	64.40	55.62	61.83	54.95
	70%	**62.69**	**48.49**	**63.12**	**36.99**	**69.70**	**64.37**	**42.05**	**75.53**	**58.10**	**62.64**	**58.37**

表 6 不同模型在不同数据选择策略下处理跨领域 ACD 任务的 $F1$ 值比较结果 (%)

模型	Data	S_{ACD}→R_{ACD}	S_{ACD}→H_{ACD}	H_{ACD}→R_{ACD}	H_{ACD}→S_{ACD}	R_{ACD}→S_{ACD}	R_{ACD}→H_{ACD}	Avg
RNN	All	70.17	57.73	37.63	42.76	52.62	48.60	51.58
	70%	71.83	63.85	50.31	51.48	61.51	49.63	58.10
Bi-LSTM	All	70.83	60.19	70.14	67.55	61.79	52.38	63.81
	70%	71.25	63.87	71.66	73.92	73.37	58.00	68.68
Att-LSTM	All	70.60	61.40	69.81	67.62	68.68	57.85	65.99
	70%	71.66	64.93	70.58	74.51	73.02	62.71	69.57
GAN	All	56.95	52.74	40.16	49.27	56.15	50.44	50.95
	70%	61.19	56.28	50.98	55.42	58.66	53.51	56.01
BERT	All	72.05	62.67	75.15	73.90	72.10	73.38	71.54
	70%	**72.26**	**64.37**	**76.33**	**74.51**	**73.65**	**75.28**	**72.73**

表 7 不同模型在不同数据选择策略下处理跨领域 ABSC 任务的 $F1$ 值比较结果 (%)

模型	Data	D→R	D→S	L→R	L→S	R→D	R→L	R→S	S→D	S→L	S→R	Avg
RNN	All	29.95	3.67	23.54	1.59	25.84	24.74	2.41	23.27	22.62	19.34	17.69
	70%	32.13	4.69	26.19	3.41	27.99	27.10	3.08	24.65	26.56	21.69	19.75
Bi-LSTM	All	29.84	5.45	29.10	3.93	39.22	32.34	2.38	34.59	26.36	26.85	23.00
	70%	36.61	5.96	33.05	5.20	41.97	34.32	4.18	39.61	29.03	27.38	25.73
Att-LSTM	All	40.62	8.43	35.61	7.19	39.64	36.59	13.84	40.59	29.77	32.66	28.49
	70%	41.55	12.68	39.66	10.55	42.44	41.16	17.39	39.30	42.19	31.76	31.87
GAN	All	35.34	4.94	34.71	3.15	33.42	30.47	2.89	44.89	32.32	34.97	25.71
	70%	38.29	5.70	42.13	4.94	45.64	44.40	4.44	63.17	44.66	38.23	33.16
BERT	All	60.52	34.00	60.09	31.76	73.27	61.85	33.63	61.13	54.82	58.18	52.93
	70%	**63.10**	**35.69**	**60.88**	**32.47**	**75.25**	**63.45**	**35.40**	**66.26**	**62.49**	**59.07**	**55.41**

表 6 中的实验结果表明, 针对跨领域 ACD 任务, 所提方法同样能够有效提升 5 种模型的性能表现. 5 种模型在 6 组跨领域 ACD 任务上的平均 $F1$ 值均有所提升, 具体表现为: RNN 从 51.58% 提升至 58.10%, Bi-LSTM 从 63.81% 提升至 68.68%, Att-LSTM 从 65.99% 提升至 69.57%, GAN 从 50.95% 提升至 56.01%, BERT 从 71.54% 提升至 72.73%. 这一结果进一步证明了所提方法在筛选训练数据中的高质量样本方面的有效性.

表 7 中的实验结果表明, 针对跨领域 ABSC 任务, 所提方法在提升 5 种模型的性能表现方面仍具备有效性. 5 种模型在 10 组跨领域 ABSC 任务上的平均 $F1$ 值均有所提升, 具体表现为: RNN 从 17.69% 提升至 19.75%, Bi-LSTM 从 23.00% 提升至 25.73%, Att-LSTM 从 28.49% 提升至 31.87%, GAN 从 25.71% 提升至 33.16%, BERT 从

52.93% 提升至 55.41%. 综上所述, 针对跨领域 ABSA 的 3 种子任务, 所提出的 HQSS-CDABSA 方法通过筛选训练数据中的高质量样本, 能够显著减少源领域与目标领域之间的分布差异, 从而提升不同类型模型在跨领域任务上的性能.

从提升幅度来看, 5 种模型在 ATE 任务上的提升分别为 2.49%、2.03%、3.19%、4.25% 和 3.42%; 在 ACD 任务上的提升分别为 6.52%、4.87%、3.58%、5.06% 和 1.19%; 在 ABSC 任务上的提升分别为 2.06%、2.73%、3.38%、7.45% 和 2.48%; 5 种模型在 3 种任务上平均提升分别为 3.69%、3.21%、3.38%、5.59% 和 2.36%. 其中 GAN 的提升最大, 其次是 RNN、Att-LSTM 和 Bi-LSTM, 而 BERT 的提升相对较小, 这种差异表明了模型的结构和能力的差异. 5 种模型中, GAN 通过生成对抗机制来利用筛选后的高质量样本进行生成器的训练, 从而生成目标领域的伪样本, 对训练数据的依赖性最强. RNN、Bi-LSTM 和 Att-LSTM 的结构相对简单, 捕捉领域间复杂关系的能力有限, 因此这些模型对数据质量的依赖较高, 筛选高质量样本同样能够显著提升其性能. 同时 Bi-LSTM 和 Att-LSTM 由于其双向结构和注意力机制的使用, 使其在未筛选数据上能够部分缓解领域差异, 因此两者在筛选后数据上训练的增益略低于 RNN. 而 BERT 通过大规模预训练已经具备了强大的领域适应能力, 能够捕捉丰富的语义信息. 在未筛选数据上, BERT 的表现已经维持在较高水平, 因此筛选高质量样本后的增益在 5 种模型中最小. 尽管 BERT 的提升幅度最小, 但其在 3 种子任务中的绝对性能始终最优. 这表明 BERT 通过预训练获得的语言表示能力在跨领域任务中具有显著优势.

以上实验结果为提升跨领域 ABSA 任务的效果水平提供了一条有效途径, 即通过结合强大的模型与高质量数据筛选方法来提升跨领域任务的性能.

4.6 多源领域场景实验结果

为研究多源领域场景对跨领域 ABSA 任务的影响, 本文以跨领域 ABSC 任务为例, 对不同模型在多源领域场景下的跨领域迁移任务进行了实验. 表 8 中展示了针对不同的目标领域, 各源领域组合通过 HQSS-CDABSA 方法计算得到的域间高质量样本选择指标 Q_{inter} 的值.

表 8 各目标领域条件下各源领域组合的 Q_{inter} 值

Target	L+R+S	D+R+S	D+L+S	D+L+R	D+R	D+L	D+S	L+R	L+S	R+S
D	11.46	—	—	—	—	—	—	**11.98**	11.07	10.28
L	—	11.13	—	—	**11.71**	—	10.87	—	—	10.09
R	—	—	10.21	—	—	**10.58**	10.03	—	9.27	—
S	—	—	—	10.37	**10.86**	9.81	—	10.44	—	—

针对单源领域场景下以 S 领域为目标领域的跨领域 ABSC 任务性能较低的问题, 本文在多源领域场景下对该任务进行实验, 以探究多源领域场景对跨领域迁移任务的增强作用. 表 9 中展示了 5 种跨领域 ABSC 模型在不同源领域组合场景下向 S 领域迁移的 $F1$ 值. 其中名为 Avg 的行中首先记录了不同的数据选择策略下 5 种模型的平均性能, 其次名为 Avg 的列中记录了表中每行数据的平均值.

表 8 的实验结果表明, 以 S 领域为目标领域时, 源领域组合 D+R、D+L、L+R 及 D+L+R 的 Q_{inter} 值分别为 10.86、9.81、10.44 及 10.37. 这种差异源于不同领域与 S 领域在语义、情感分布及主题上的相似性的差异. 从领域特性来看, D 领域包含多种电子产品评论, 主题广泛, 涵盖硬件、软件、用户体验等多个方面; R 领域围绕餐饮服务、食物质量、环境等主题; L 领域包含笔记本电脑领域的相关评论; S 领域包含网络服务领域的相关评论, 主题广泛, 涉及软件服务、用户体验等多个方面. D 领域与 S 领域的主题有较高的语义重叠, R 领域主题与 S 领域差异较大, 但其多样性为模型提供了额外的泛化能力. D+R 源领域组合通过结合多种电子设备以及餐厅领域的多样性, 能够更好地覆盖 S 领域的特征, 因此 Q_{inter} 值最高, 达到 10.86. 而 L 领域主题集中在笔记本电脑领域, 相较于 D 领域与 S 领域的关联性较弱, 因此 L+R 源领域组合的 Q_{inter} 值略低于 D+R 源领域组合, 为 10.44. D+L+R 源领域组合的 Q_{inter} 值为 10.37, 尽管其包含了所有源领域, 但其 Q_{inter} 值并未显著高于 D+R 源领域组合. 增加 L 领域并未显著提升向 S 领域的迁移潜力, 这表明了 L 领域单一主题带来的局限性. 而 D+L 源领域组合由于主题相互包含,

无法为迁移到 S 领域提供足够的多样性支持, 因此其 Q_{inter} 值最低, 仅 9.81. 综上所述, 在以上源领域组合中, D+R 源领域组合最适合向 S 领域进行迁移任务的训练.

表 9　不同模型在不同源领域组合下向 S 领域迁移的 ABSC 任务的 $F1$ 值比较结果 (%)

模型	Data	D	L	R	D+R	D+L	L+R	D+L+R	Avg
RNN	All	3.67	1.59	2.41	4.98	4.03	4.41	4.32	3.63
	70%	4.69	3.41	3.08	6.37	4.94	6.07	5.33	4.84
Bi-LSTM	All	5.45	3.93	2.38	6.71	5.55	6.14	5.87	5.15
	70%	5.96	5.20	4.18	8.16	6.18	7.84	6.34	6.27
Att-LSTM	All	8.43	7.19	13.84	18.72	10.41	15.31	14.26	12.59
	70%	12.68	10.55	17.39	23.64	15.77	19.46	17.88	16.77
GAN	All	4.94	3.15	2.89	6.58	5.22	6.03	5.47	4.90
	70%	5.70	4.94	4.44	8.04	6.19	7.63	6.81	6.25
BERT	All	34.00	31.76	33.63	39.37	35.76	37.14	36.03	35.38
	70%	**35.69**	**32.47**	**35.40**	**41.38**	**36.57**	**38.35**	**37.21**	**36.72**
Avg	All	9.50	7.97	9.36	13.62	10.45	12.23	11.37	10.64
	70%	13.06	11.4	13.03	17.66	14.06	16.00	15.02	14.32

表 9 的实验结果表明, 多源领域组合在跨领域迁移任务中能够有效提升模型性能. 以全样本训练为例, 5 种模型在单一源领域 D、L、R 上的平均 $F1$ 值分别为 9.50%、7.97%、9.36%, 而在多源领域组合 D+R、D+L、L+R、D+L+R 上的平均 $F1$ 值分别为 13.62%、10.45%、12.23%、11.37%, 多源领域组合场景平均比单一源领域场景的性能高出 2.98%. 这表明多源领域组合能够通过领域间的互补性提供更丰富的语义信息, 从而提升迁移效果. 同时表中展示了在不同的数据选择策略下 5 种模型在所有源领域组合上的平均性能, 使用 HQSS-CDABSA 方法筛选的 70% 数据进行训练的平均 $F1$ 值为 14.32%, 高出使用全样本训练策略 3.68%. 这表明 HQSS-CDABSA 方法的领域内高质量样本选择方法在单源领域场景和多源领域场景均能有效适用.

通过对表 9 中的实验结果进一步分析, D+R 组合在全样本训练策略下的平均 $F1$ 值为 13.62%, 分别比 D+L、L+R、D+L+R 组合高出 3.17%、1.39%、2.25%, 这表明 D+R 源领域组合在迁移到 S 领域时相比其他源领域组合的泛化性能更优. 在通过 HQSS-CDABSA 方法筛选 70% 样本训练的策略下, D+R 组合的平均 $F1$ 值进一步提升至 17.66%, 分别比 D+L、L+R、D+L+R 组合高出 3.60%、1.66%、2.64%. D+R 组合与其他源领域组合的性能差距进一步拉开, 这进一步表明了 D+R 源领域组合的优越性以及 HQSS-CDABSA 方法的有效性. D+R 组合在使用 HQSS-CDABSA 方法后平均 $F1$ 值提升了 4.04%, 而 D+L 组合的平均 $F1$ 值提升幅度较小, 仅 3.61%, 表明 L 领域主题单一的局限性限制了 HQSS-CDABSA 方法的作用. L+R 组合的平均 $F1$ 值提升了 3.77%, 虽高于 D+L 组合, 但仍低于 D+R 组合, 表明 R 领域的多样性虽能弥补部分不足, 但无法完全抵消 L 领域的局限性. D+L+R 组合的平均 $F1$ 值提升了 3.65%, 低于 D+R 组合, 表明增加 L 领域并未显著提升迁移效果. 综上所述, 多源领域组合能够通过领域间的互补性提升迁移效果, 而 D+R 的源领域组合由于多样化的主题具备更强的泛化性, 更适合向 S 领域进行迁移模型的训练.

4.7　消融实验

为了研究框架中每个成分的影响, 本文对完整的 HQSS-CDABSA 方法与其变体之间进行了比较. 所提出 HQSS-CDABSA 方法的变体如下.

• w/o C: 在 HQSS-CDABSA 方法的领域间高质量样本选择阶段将联合适应性分数的系数 β 设置为 0, 在源领域组合的筛选中移除了多源领域之间协同效应的影响.

• w/o *Gen*: 在 HQSS-CDABSA 方法的领域内高质量样本选择阶段将 μ 设置为 0, 在源领域内样本的筛选过程中移除样本领域通用性的影响.

• w/o *Dis*: 在 HQSS-CDABSA 方法的领域内高质量样本选择阶段将 η 设置为 0, 在源领域内样本的筛选过程中移除样本与目标领域相似性的影响.

以向 S 领域迁移的跨领域 ABSC 任务为例, 表 10 中展示了 w/o C 方法与完整的 HQSS-CDABSA 方法计算得到的不同源领域组合的 Q_{inter} 值. 图 5 中展示了在表 10 的 Q_{inter} 值条件下选择各源领域组合的概率. 源领域组合的被选择概率定义为在相同目标领域条件下, 该源领域组合的 Q_{inter} 值与所有源领域组合中 Q_{inter} 最大值的比值.

表 10 以 S 领域为目标领域时各源领域组合在不同方法下计算所得 Q_{inter} 值

方法	D+L+R	D+R	D+L	L+R
HQSS-CDABSA	10.37	**10.86**	9.81	10.44
w/o C	8.87	8.97	**9.07**	8.58

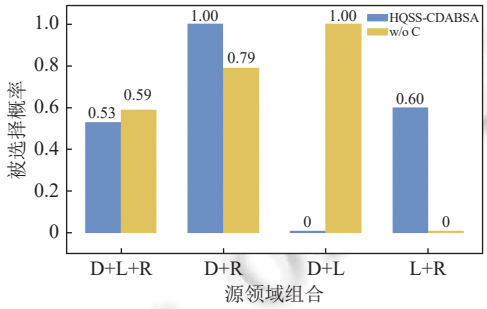


图 5 HQSS-CDABSA 及 w/o C 方法下各源领域组合被选择概率

表 10 的实验结果表明, HQSS-CDABSA 方法与 w/o C 方法在计算源领域组合的域间高质量样本选择指标 Q_{inter} 值时存在显著差异. 具体来说, HQSS-CDABSA 方法下 D+L+R、D+R、D+L、L+R 这 4 个源领域组合的 Q_{inter} 值分别为 10.37、10.86、9.81、10.44, 而 w/o C 方法下对应的 Q_{inter} 值分别为 8.87、8.97、9.07、8.58. 两种方法下 Q_{inter} 值的最高值分别出现在 D+R 组合与 D+L 组合.

结合图 5 所示的源领域组合被选择概率, HQSS-CDABSA 方法下 D+R 组合的被选择概率为 1.00, 分别比 D+L+R、D+L、L+R 的组合高出 0.47、1.0、0.4. 结合第 4.6 节所得“4 种源领域组合中 D+R 组合更适合向 S 领域进行迁移训练”的结论, 验证了 HQSS-CDABSA 方法在领域间高质量样本选择过程中的有效性. 相比之下, w/o C 方法移除了多源领域间的联合适应性分数, 其领域间高质量样本选择过程仅依靠跨领域模式共现相似度来进行筛选. 由于 D 领域与 L 领域在主题上最为接近, 因此 w/o C 方法下的最优源领域组合为 D+L 组合, 被选择概率为 1.0, 而真正的最优源领域组合 D+R 的被选择概率则为 0.79. w/o C 方法对最优源领域组合的误选进一步验证了在 HQSS-CDABSA 方法中, 考虑领域间协同效应而加入联合适应性分数的有效性.

以 Att-LSTM 模型为例, 表 11 中展示了分别在 w/o *Gen* 方法、w/o *Dis* 方法及 HQSS-CDABSA 方法筛选的样本上训练得到的 Att-LSTM 模型在不同的单源跨领域 ABSC 任务上的 *F*1 值.

表 11 HQSS-CDABSA 及其变体下 Att-LSTM 在各任务上的 *F*1 值效果对比 (%)

方法	D→R	D→S	L→R	L→S	R→D	R→L	R→S	S→D	S→L	S→R	Avg
w/o *Gen*	38.24	8.29	31.64	6.08	31.87	35.41	11.37	27.44	33.58	22.75	24.67
w/o *Dis*	36.11	7.12	29.81	7.15	30.68	35.48	10.41	27.01	31.28	23.07	23.81
HQSS-CDABSA	**41.55**	**12.68**	**39.66**	**10.55**	**42.44**	**41.16**	**17.39**	**39.30**	**42.19**	**31.76**	**31.87**

表 11 的实验结果表明, 两种变体与完整的 HQSS-CDABSA 方法相比, 平均 *F*1 值均有不同程度的下降, 证明 HQSS-CDABSA 方法的域内高质量样本选择指标 Q_{intra} 的各部分对模型性能的提升均有积极作用. 具体来说, w/o *Gen*、w/o *Dis* 两种变体方法的平均 *F*1 值与完整的 HQSS-CDABSA 方法相比分别下降了 7.20% 和 8.06%, 在各迁移任务上也显现出了同样的趋势. 其中 w/o *Dis* 方法的性能损失最大, 说明在 HQSS-CDABSA 方法的域内高质量样本选择指标 Q_{intra} 中, 方面级距离度量 Dis_{asp} 对领域内高质量样本选择过程的贡献最大, 起着最重要的指导作用; 同时 w/o *Gen* 方法性能的下降也证明了在域内高质量样本选择指标 Q_{intra} 中引入领域通用性分数 *Gen* 的有效性.

4.8 敏感性实验

为了确定域间高质量样本选择指标 Q_{inter} 及域内高质量样本选择指标 Q_{intra} 的最优超参数设置, 本文以向 S 领域迁移的跨领域任务为例, 针对 Q_{inter} 的超参数 α、β 及 Q_{intra} 的超参数 μ、η 进行了敏感性实验.

针对领域间高质量样本选择过程, 本文在不同的 Q_{inter} 超参数 α、β 组合下进行了实验, 表 12 中展示了在不同的超参数 α、β 组合下各源领域组合的 Q_{inter} 值.

表 12 中的实验结果表明, 当 α 取 0.6、β 取 0.4 时, 各源领域组合 Q_{inter} 值的平均值达到最高, 为 10.37. 说明在该超参数组合下能够有效地兼顾领域特征差异性与语义相似性, 提升源领域样本在迁移到目标领域过程中的整体质量评估效果. 该超参数设置更侧重于语义相似性项的权重, 适度减少了对领域差异项的关注. 这种权重分配使得在选取源领域样本时, 倾向于选择在语义层面与目标领域文本更接近的样本, 提高模型对目标领域语义的适应性, 更准确地识别目标领域中的方面和情感表达. 同时表中结果表明在该超参数组合下, 各源领域组合上 Q_{inter} 值较为均衡, 波动较小, 说明该设置具有更好的鲁棒性, 避免个别源领域组合因特征不匹配而导致的 Q_{inter} 值分布差异过大问题, 增强方法在不同源领域组合下的一致性和泛化能力.

在 (0.6, 0.4) 的超参数 α、β 组合下, 本文选取其中 Q_{inter} 值最大的 D+R 源领域组合作为待迁移源领域组合. 针对领域内高质量样本选择过程, 本文在不同的 Q_{intra} 超参数 μ、η 组合下进行了实验, 表 13 中展示了在不同的超参数 μ、η 组合下不同跨领域任务的实验结果. 表中粗体数据为不同任务最高的数值.

表 12 以 S 领域为目标领域时各源领域组合在不同超参数下计算所得 Q_{inter} 值

(α, β)	D+L+R	D+R	D+L	L+R	Avg
(0.2, 0.8)	10.18	11.36	9.27	10.65	10.36
(0.4, 0.6)	10.25	10.98	9.54	10.51	10.32
(0.6, 0.4)	10.37	10.86	9.81	10.44	**10.37**
(0.8, 0.2)	10.44	10.48	10.08	10.30	10.32

表 13 以 S 领域为目标领域时不同跨领域任务在不同超参数下 $F1$ 值效果对比 (%)

(μ, η)	ATE	ABSC
(0.2, 0.8)	39.84	35.24
(0.4, 0.6)	42.31	36.77
(0.6, 0.4)	44.65	38.26
(0.8, 0.2)	**50.27**	**41.38**

表 13 中实验结果表明, 在领域内高质量样本选择指标 Q_{intra} 的不同超参数组合下, 各跨领域任务的实验结果表现出明显性能差异. 随着超参数中 μ 逐步增大, 方法在两个任务上的 $F1$ 值呈现出上升趋势. 其中, 在 μ 取 0.8、η 取 0.2 的设置下, ATE 任务达到最高 $F1$ 值 50.27%, ABSC 任务达到最高 $F1$ 值 41.38%. 说明在领域内样本选择过程中, 更加重视样本的领域通用性评分, 而相对降低其与目标领域类中心距离的影响, 能够挑选对迁移更具代表性和普适性的样本, 提升下游任务性能. 当 μ 取 0.2、η 取 0.8 时, 更侧重选择与目标领域类中心距离较近的样本, 对样本通用性关注较少, 性能相对较低, 表明单纯考虑与目标领域类中心的相似性不足以保证样本在迁移过程中的有效性, 甚至引入噪声样本, 限制模型的泛化能力.

4.9 案例研究

本文从 S→L 迁移任务的网络服务 (S) 数据集中选择了一些示例句子, 并在表 14 中给出了各句子经过 HQSS-CDABSA 方法计算得到的对应样本的域内高质量样本选择指标 Q_{intra} 及数据质量 Q 的值, 并标注了各句子是否被选中作为训练数据. 其中, 第 1 列为句子编号, 第 2 列显示选定的示例样本, 第 3、4 列展示了 HQSS-CDABSA 方法计算得到的对应样本的域内高质量样本选择指标 Q_{intra} 及数据质量 Q 的值, 第 5 列展示了在 70% 的样本选择比例下句子是否被选中作为训练数据. 从表 14 中可以发现, 在 70% 的样本选择比例下, 前 3 个句子被选中作为训练数据参与模型训练, 其余两句均未入选. 对于第 1 个句子, 该句子描述了用户在使用网站时的体验, 虽然主题与网络服务相关, 但其中涉及的“易用性”概念在 L 领域中也具有普遍性. 同时句中“easy”一词与用户体验相关, 这种情感表达在跨领域任务中具有一定的通用性. 第 2 个句子则涉及账户管理和平台集成, 这种技术性描述在 L 领域中同样存在, 例如不同软件或服务的账户集成. 第 3 个句子描述了在线社区的功能, 这种社交功能在 L 领域中同样存在类似应用, 例如在线论坛或聊天软件. 因此上述句子适合作为向 L 领域迁移的训练样本. 而其余两个句子主题分别围绕电话铃声和书籍内容, 与 L 领域的关联性极弱, 不适合向 L 领域进行迁移训练.

表 14 案例研究表

No.	Sentence	Q_{intra}	Q	Select
1	Finding different parts of the site was just as easy.	−7.25	2.75	√
2	I had to “marry” each of these eGroup accounts with a Yahoo!	−8.66	1.34	√
3	eGroups is a place where people with common interests can join groups that post messages or chat.	−10.13	−0.13	√
4	That telephone just rang and rang and rang.	−12.15	−2.15	×
5	Books were community college level or lower.	−13.75	−3.75	×

4.10 统计性分析实验

(1) 数据质量 Q

为了探究所提出 HQSS-CDABSA 方法对源领域训练数据的数据质量的影响, 针对单源领域场景下的 10 对跨领域迁移任务, 本文对使用 HQSS-CDABSA 方法进行高质量样本选择前后的源领域训练数据的数据质量进行了统计分析. 为了直观展示筛选前后数据质量的差距, 将数据质量 Q 定义为域内高质量样本选择指标 Q_{intra} 与 10 的和. 实验结果如图 6 所示.

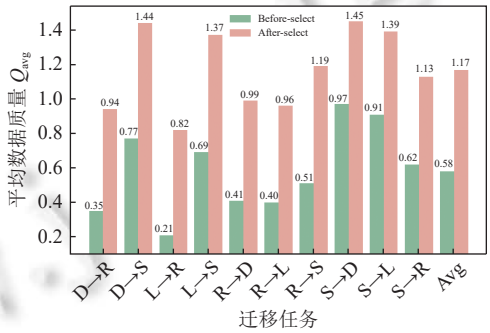


图 6 不同迁移任务中源领域经 HQSS-CDABSA 方法筛选前后的平均数据质量 Q_{avg}

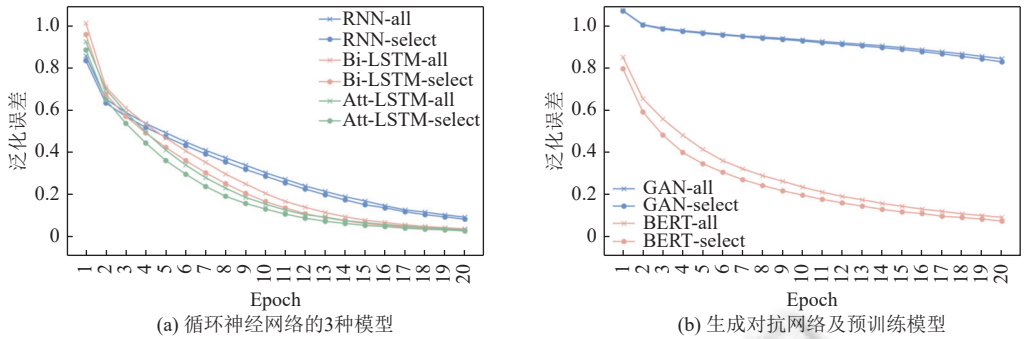
图 6 中展示了每对迁移任务中经 HQSS-CDABSA 方法筛选前后的源领域数据集的平均数据质量 Q_{avg} 以及所有迁移任务的 Q_{avg} 的平均值. 其中, 绿色柱体代表经 HQSS-CDABSA 方法筛选前的源领域数据集的平均数据质量 Q_{avg}, 红色柱体代表经 HQSS-CDABSA 方法筛选后的源领域数据集的平均数据质量 Q_{avg}.

图 6 的实验结果表明, 10 组迁移任务中的源领域在经过 HQSS-CDABSA 方法进行高质量样本选择之后, 其平均数据质量 Q_{avg} 均有明显增长. 综合 10 组迁移任务的平均效果来看, Q_{avg} 由 0.58 增长至 1.17, 提升了 0.59, 表明 HQSS-CDABSA 方法能够有效提升源领域训练数据的数据质量.

从具体任务来看, D→S 任务的 Q_{avg} 值从 0.77 提升至 1.44, 提升幅度最大, 达到 0.67, 这源于 D 领域与 S 领域在语义和主题上具有较高的重叠性, 相比于其他迁移任务, HQSS-CDABSA 方法能够更有效地筛选出与目标领域相关的高质量样本. 类似地, L→S 任务的 Q_{avg} 值从 0.69 提升至 1.37, 提升幅度为 0.68, 进一步验证了 HQSS-CDABSA 方法在筛选高质量样本时的有效性. 相比之下, D→R 任务的 Q_{avg} 值从 0.35 提升至 0.94, 提升幅度为 0.59, 虽然提升显著, 但由于 D 领域与 R 领域的主题差异较大, 筛选后的 Q_{avg} 值仍低于其他任务.

(2) 模型泛化误差 ε_T

为了探究本节所提出 HQSS-CDABSA 方法对不同模型在跨领域任务中泛化误差的影响, 针对单源领域场景下的跨领域 ABSC 任务, 本节对使用 HQSS-CDABSA 方法进行高质量样本选择前后的不同模型的泛化误差进行了统计分析. 为了直观展示 HQSS-CDABSA 方法筛选前后模型性能的差距, 通过折线图对不同模型的泛化误差曲线进行了绘制. 图 7 中展示了每种模型在跨领域 ABSC 任务上的泛化误差曲线, 曲线值为对应模型在 10 组单源领域场景下跨领域 ABSC 任务上的泛化误差的平均值. 其中图 7(a) 中展示了循环神经网络的 3 种模型的性能, 图 7(b) 中展示了生成对抗网络及预训练模型两类模型的性能.



(a) 循环神经网络的3种模型 (b) 生成对抗网络及预训练模型

图 7 跨领域 ABSC 任务上各模型经 HQSS-CDABSA 方法处理前后的泛化误差曲线

图 7 中的实验结果表明, HQSS-CDABSA 方法通过筛选源领域中的高质量样本显著降低了不同模型在跨领域 ABSC 任务中的泛化误差, 从而提升了模型的性能. 具体来说, 图 7(a) 表明 RNN、Bi-LSTM、Att-LSTM 这 3 种循环神经网络类模型在经过样本筛选后模型的泛化误差均有所下降, 且 3 种模型的性能依次递增. 这种下降趋势表明, HQSS-CDABSA 方法能够有效筛选出高质量样本, 减少噪声数据对模型训练的干扰, 从而提升模型的泛化能力. 此外, 3 种循环神经网络模型中 Att-LSTM 的泛化误差最低, 表明其注意力机制能够更好地捕捉跨领域任务中的关键信息, 筛选出的高质量样本的利用效率最高.

图 7(b) 表明 GAN 模型在经过样本筛选后泛化误差有所下降, 但下降幅度不大且总体水平仍然较高. 这源于 ABSC 任务的标签复杂性较高, 导致 GAN 模型在生成伪目标领域数据时数据标注可靠性不足, 影响模型的性能. 尽管如此, HQSS-CDABSA 方法仍然在一定程度上改善了 GAN 的泛化能力. 而 BERT 模型尽管其泛化误差本身在 5 种模型中已处于最低水平, 但在经过样本筛选后其泛化误差同样得到降低. 这表明, 即使对于预训练语言模型如 BERT, HQSS-CDABSA 方法仍然能够通过筛选高质量样本进一步提升其性能. BERT 的优异表现得益于其强大的语义表示能力, 而 HQSS-CDABSA 方法则进一步优化了训练数据的质量, 从而降低了模型的泛化误差.

5 总 结

跨领域 ABSA 任务作为自然语言处理领域的重要研究方向之一, 本文主要考虑其 3 种子任务: 跨领域 ATE、跨领域 ACD 及跨领域 ABSC 任务. 针对 ABSA 领域中标注样本匮乏及跨领域实例迁移问题, 提出了一种基于高质量样本选择的跨领域 ABSA 方法. 方法兼顾源领域与目标领域间相似性及多源领域间协同性, 设计了领域间和领域内两个层面的高质量样本选择指标, 依次对多源领域数据进行领域层面和样本层面的评估和筛选. 经过在 ABSA 数据集上的大量迁移任务实验证明, 所提方法在大部分的迁移任务以及平均性能上显著优于目前的先进方法. 通过所提方法, 实现了源领域样本数量与源领域样本质量上的平衡, 提升了各迁移任务的效果. 本文提出的高质量样本选择策略包含对数据噪声和跨领域分布偏移的处理, 符合高可信机器学习对模型稳定性, 可解释性及动态适应能力的要求, 为解决开放环境下领域迁移问题提供了理论支撑与方法创新.

本文提出的基于高质量样本选择的跨领域 ABSA 方法在提升模型迁移性能方面取得了显著成效, 但仍有进一步优化空间, 尤其是在多源领域协同效应和样本选择机制的进一步强化上. 未来研究可探索更加精细的跨领域相似度度量方法, 以实现更高效的数据筛选和特征融合. 在多模态情感分类工作中, 在样本选择与模型训练阶段引入模态一致性检测机制, 对文本、图像、语音等模态的情感输出进行一致性度量, 并结合冲突敏感融合策略, 在检测到跨模态情感冲突时动态调整模态权重或选择性融合, 以降低冲突样本对模型判别性能的干扰. 随着大规模预训练模型的不断进步, 结合高质量样本选择的迁移学习策略有望在多个应用领域中展现更强的适应能力和更优的性能表现.

References

[1] Zhang WX, Li X, Deng Y, Bing LD, Lam W. A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. IEEE Trans. on Knowledge and Data Engineering, 2023, 35(11): 11019–11038. [doi: 10.1109/TKDE.2022.3230975]

[2] Zhao CJ, Wu ML, Yang XY, Zhang WY, Zhang SX, Wang SG, Li DY. A systematic review of cross-lingual sentiment analysis: Tasks,

strategies, and prospects. ACM Computing Surveys, 2024, 56(7): 177. [doi: 10.1145/3645106]

[3] Wu Z, Dai XY. Separated syntax and semantics modeling for cross-domain aspect-level sentiment classification. Scientia Sinica Informationis, 2023, 53(7): 1299–1313 (in Chinese with English abstract). [doi: 10.1360/SSI-2021-0166]

[4] Zhao CJ, Wang SG, Li DY. Research progress on cross-domain text sentiment classification. Ruan Jian Xue Bao/Journal of Software, 2020, 31(6): 1723–1746 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6029.htm [doi: 10.13328/j.cnki.jos.006029]

[5] Zhuang FZ, Luo P, He Q, Shi ZZ. Survey on transfer learning research. Ruan Jian Xue Bao/Journal of Software, 2015, 26(1): 26–39 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4631.htm [doi: 10.13328/j.cnki.jos.004631]

[6] Chen Z, Qian TY, Li WL, Zhang T, Zhou S, Zhong M, Zhu YY, Liu MC. Low-resource aspect-based sentiment analysis: A survey. Chinese Journal of Computers, 2023, 46(7): 1445–1472 (in Chinese with English abstract). [doi: 10.11897/SP.J.1016.2023.01445]

[7] Sun XZ. Research on cross-domain aspect-based sentiment analysis method based on instance transfer learning [MS. Thesis]. Taiyuan: Shanxi University of Finance and Economics, 2025 (in Chinese).

[8] Wu BC, Deng CL, Guan B, Chen XL, Zan DG, Chang ZJ, Xiao ZY, Qu DC, Wang YJ. Dynamically transfer entity span information for cross-domain Chinese named entity recognition. Ruan Jian Xue Bao/Journal of Software, 2022, 33(10): 3776–3792 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6305.htm [doi: 10.13328/j.cnki.jos.006305]

[9] Li SC, Wang ZQ, Zhou GD. LLM enhanced cross domain aspect-based sentiment analysis. Ruan Jian Xue Bao/Journal of Software, 2025, 36(2): 644–659 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/7156.htm [doi: 10.13328/j.cnki.jos.007156]

[10] Zhao CJ, Kang L, Sun XZ, Xi XX, Shen LH, Gao J, Wang YJ. Aspect-level sentiment classification of consumer reviews utilizing BERT and category-aware multi-head attention. IEEE Trans. on Consumer Electronics, 2025, 71(2): 3329–3339. [doi: 10.1109/TCE.2025.3563150]

[11] Zhao GY, Lv CG, Fu GH, Liu ZL, Liang CF, Liu T. Domain specific sentiment words based attention model for cross-domain attribute-oriented sentiment analysis. Journal of Chinese Information Processing, 2021, 35(6): 93–102 (in Chinese with English abstract). [doi: 10.3969/j.issn.1003-0077.2021.06.010]

[12] Das R, Singh TD. Multimodal sentiment analysis: A survey of methods, trends, and challenges. ACM Computing Surveys, 2023, 55(13s): 270. [doi: 10.1145/3586075]

[13] Jakob N, Gurevych I. Extracting opinion targets in a single and cross-domain setting with conditional random fields. In: Proc. of the 2010 Conf. on Empirical Methods in Natural Language Processing. Cambridge: ACL, 2010. 1035–1045.

[14] Li FT, Pan SJ, Jin O, Yang Q, Zhu XY. Cross-domain co-extraction of sentiment and topic lexicons. In: Proc. of the 50th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). Jeju Island: ACL. 2012. 410–419.

[15] Zhang XG, Chan FTS, Yan C, Bose I. Towards risk-aware artificial intelligence and machine learning systems: An overview. Decision Support Systems, 2022, 159: 113800. [doi: 10.1016/j.dss.2022.113800]

[16] Brauwers G, Frasincar F. A survey on aspect-based sentiment classification. ACM Computing Surveys, 2023, 55(4): 65. [doi: 10.1145/3503044]

[17] Wang YY, Chen Q, Shen JQ, Hou BY, Ahmed M, Li ZH. Aspect-level sentiment analysis based on gradual machine learning. Knowledge-based Systems, 2021, 212: 106509. [doi: 10.1016/j.knosys.2020.106509]

[18] Pathan AF, Prakash C. Cross-domain aspect detection and categorization using machine learning for aspect-based opinion mining. Int'l Journal of Information Management Data Insights, 2022, 2(2): 100099. [doi: 10.1016/j.jjimei.2022.100099]

[19] Wang WY, Pan SJ. Recursive neural structural correspondence network for cross-domain aspect and opinion co-extraction. In: Proc. of the 56th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). Melbourne: ACL, 2018. 2171–2181. [doi: 10.18653/v1/P18-1202]

[20] Liu N, Shen B, Zhang ZJ, Zhang ZY, Mi K. Attention-based sentiment reasoner for aspect-based sentiment analysis. Human-centric Computing and Information Sciences, 2019, 9(1): 35. [doi: 10.1186/s13673-019-0196-3]

[21] Yang T, Yin Q, Yang L, Wu O. Aspect-based sentiment analysis with new target representation and dependency attention. IEEE Trans. on Affective Computing, 2022, 13(2): 640–650. [doi: 10.1109/TAFFC.2019.2945028]

[22] Knoester J, Frasincar F, Truşcă MM. Cross-domain aspect-based sentiment analysis using domain adversarial training. World Wide Web, 2023, 26(6): 4047–4067. [doi: 10.1007/s11280-023-01217-4]

[23] Yin YH, Li X, Yan ZT, Wang XL. Adversarial generative model for cross-domain aspect-based sentiment analysis. In: Proc. of the 6th Int'l Conf. on Data-driven Optimization of Complex Systems. Hangzhou: IEEE, 2024. 806–811. [doi: 10.1109/DOCS63458.2024.10704498]

[24] Liu N, Zhao JH. A BERT-based aspect-level sentiment analysis algorithm for cross-domain text. Computational Intelligence and Neuroscience, 2022, 2022: 8726621. [doi: 10.1155/2022/8726621]

[25] Zhao CJ, Wu ML, Yang XY, Sun XZ, Wang SG, Li DY. Cross-domain aspect-based sentiment classification with a pre-training and fine-

tuning strategy for low-resource domains. ACM Trans. on Asian and Low-resource Language Information Processing, 2024, 23(4): 59. [doi: 10.1145/3653299]

[26] Zhou CZ, Song DD, Tian YH, Wu ZJ, Wang H, Zhang XY, Yang J, Yang ZY, Zhang SH. A comprehensive evaluation of large language models on aspect-based sentiment analysis. arXiv:2412.02279, 2024.

[27] Mao DZ, Li HY, Shao YQ. Semi-supervised domain adaptation for semantic dependency parsing. Journal of Chinese Information Processing, 2022, 36(2): 22–28 (in Chinese with English abstract). [doi: 10.3969/j.issn.1003-0077.2022.02.003]

[28] Wu FZ, Huang YF, Yan J. Active sentiment domain adaptation. In: Proc. of the 55th Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). Vancouver: ACL, 2017. 1701–1711. [doi: 10.18653/v1/P17-1156]

[29] Wu HR, Wu QY, Ng MK. Knowledge preserving and distribution alignment for heterogeneous domain adaptation. ACM Trans. on Information Systems, 2022, 40(1): 16. [doi: 10.1145/3469856]

[30] Ping R, Zhou SS, Li D. Cost sensitive random forest classification algorithm for highly unbalanced data. Pattern Recognition and Artificial Intelligence, 2020, 33(3): 249–257 (in Chinese with English abstract). [doi: 10.16451/j.cnki.issn1003-6059.202003006]

[31] Li JC, Li GB, Shi YM, Yu YZ. Cross-domain adaptive clustering for semi-supervised domain adaptation. In: Proc. of the 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 2505–2514. [doi: 10.1109/CVPR46437.2021.00253]

[32] Saito K, Ushiku Y, Harada T. Asymmetric tri-training for unsupervised domain adaptation. In: Proc. of the 34th Int'l Conf. on Machine Learning. Sydney: PMLR, 2017. 2988–2997.

[33] Rotman G, Reichart R. Deep contextualized self-training for low resource dependency parsing. Trans. of the Association for Computational Linguistics, 2019, 7: 695–713. [doi: 10.1162/tacl_a_00294]

[34] Yu JF, Zhao QK, Xia R. Cross-domain data augmentation with domain-adaptive language modeling for aspect-based sentiment analysis. In: Proc. of the 61st Annual Meeting of the Association for Computational Linguistics (Vol. 1: Long Papers). Toronto: ACL, 2023. 1456–1470. [doi: 10.18653/v1/2023.acl-long.81]

[35] Li Z, Li X, Wei Y, Bing LD, Zhang Y, Yang Q. Transferable end-to-end aspect-based sentiment analysis with selective adversarial learning. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing. Hong Kong: ACL, 2019. 4590–4600. [doi: 10.18653/v1/D19-1466]

[36] Zhou Y, Zhu FQ, Song P, Han JZ, Guo T, Hu SL. An adaptive hybrid framework for cross-domain aspect-based sentiment analysis. In: Proc. of the 35th AAAI Conf. on Artificial Intelligence. AAAI Press, 2021. 14630–14637. [doi: 10.1609/aaai.v35i16.17719]

[37] He JZ, Jia X, Chen SJ, Liu JZ. Multi-source domain adaptation with collaborative learning for semantic segmentation. In: Proc. of the 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 11003–11012. [doi: 10.1109/CVPR46437.2021.01086]

[38] Dai Y, Liu J, Zhang J, Fu HG, Xu ZL. Unsupervised sentiment analysis by transferring multi-source knowledge. Cognitive Computation, 2021, 13(5): 1185–1197. [doi: 10.1007/s12559-020-09792-8]

[39] Pontiki M, Galanis D, Pavlopoulos J, Papageorgiou H, Androutsopoulos I, Manandhar S. SemEval-2014 task 4: Aspect based sentiment analysis. In: Proc. of the 8th Int'l Workshop on Semantic Evaluation. Dublin: ACL, 2014. 27–35. [doi: 10.3115/v1/S14-2004]

[40] Pontiki M, Galanis D, Papageorgiou H, Manandhar S, Androutsopoulos I. SemEval-2015 task 12: Aspect based sentiment analysis. In: Proc. of the 9th Int'l Workshop on Semantic Evaluation. Denver: ACL, 2015. 486–495. [doi: 10.18653/v1/S15-2082]

[41] Pontiki M, Galanis D, Papageorgiou H, Androutsopoulos I, Manandhar S, Al-Smadi M, Al-Ayyoub M, Zhao AA, Qin B, de Clercq O, Hoste V, Apidianaki M, Tannier X, Loukachevitch N, Kotelnikov E, Bel N, Jiménez-Zafra SM, Eryiğit G. SemEval-2016 task 5: Aspect based sentiment analysis. In: Proc. of the 10th Int'l Workshop on Semantic Evaluation. San Diego: ACL, 2016. 19–30. [doi: 10.18653/v1/S16-1002]

[42] Hu MQ, Liu B. Mining and summarizing customer reviews. In: Proc. of the 10th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. Seattle: ACM, 2004. 168–177. [doi: 10.1145/1014052.1014073]

[43] Toprak C, Jakob N, Gurevych I. Sentence and expression level annotation of opinions in user-generated discourse. In: Proc. of the 48th Annual Meeting of the Association for Computational Linguistics. Uppsala: ACL, 2010. 575–584.

[44] Saeidi M, Bouchard G, Liakata M, Riedel S. SentiHood: Targeted aspect based sentiment analysis dataset for urban neighbourhoods. In: Proc. of the 26th Int'l Conf. on Computational Linguistics: Technical Papers. Osaka: ACL, 2016. 1546–1556.

[45] Du CS, Huang L. Text classification research with attention-based recurrent neural networks. Int'l Journal of Computers Communications & Control, 2018, 13(1): 50–61. [doi: 10.15837/ijccc.2018.1.3142]

[46] Bin Y, Yang Y, Shen FM, Xu X, Shen HT. Bidirectional long-short term memory for video description. In: Proc. of the 24th ACM Int'l Conf. on Multimedia. Amsterdam: ACM, 2016. 436–440. [doi: 10.1145/2964284.2967258]

[47] Wang YQ, Huang ML, Zhu XY, Zhao L. Attention-based LSTM for aspect-level sentiment classification. In: Proc. of the 2016 Conf. on

Empirical Methods in Natural Language Processing. Austin: ACL, 2016. 606–615. [doi: 10.18653/v1/D16-1058]

[48] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial networks. Communications of the ACM, 2020, 63(11): 139–144. [doi: 10.1145/3422622]

[49] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional Transformers for language understanding. In: Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1 (Long and Short Papers). Minneapolis: ACL, 2019. 4171–4186. [doi: 10.18653/v1/N19-1423]

[50] Chen X, Wan XJ. A simple information-based approach to unsupervised domain-adaptive aspect-based sentiment analysis. arXiv: 2201.12549, 2022.

[51] Yu JF, Gong CG, Xia R. Cross-domain review generation for aspect-based sentiment analysis. In: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. ACL, 2021. 4767–4777. [doi: 10.18653/v1/2021.findings-acl.421]

[52] Li JJ, Yu JF, Xia R. Generative cross-domain data augmentation for aspect and opinion co-extraction. In: Proc. of the 2022 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Seattle: ACL, 2022. 4219–4229. [doi: 10.18653/v1/2022.naacl-main.312]

附中文参考文献

[3] 吴震, 戴新宇. 基于语法和语义分割的跨领域方面级情感分类. 中国科学: 信息科学, 2023, 53(7): 1299–1313. [doi: 10.1360/SSI-2021-0166]

[4] 赵传君, 王素格, 李德玉. 跨领域文本情感分类研究进展. 软件学报, 2020, 31(6): 1723–1746. http://www.jos.org.cn/1000-9825/6029.htm [doi: 10.13328/j.cnki.jos.006029]

[5] 庄福振, 罗平, 何清, 史忠植. 迁移学习研究进展. 软件学报, 2015, 26(1): 26–39. http://www.jos.org.cn/1000-9825/4631.htm [doi: 10.13328/j.cnki.jos.004631]

[6] 陈壮, 钱铁云, 李万理, 张婷, 周燊, 钟鸣, 祝园园, 刘梦赤. 低资源方面级情感分析研究综述. 计算机学报, 2023, 46(7): 1445–1472. [doi: 10.11897/SP.J.1016.2023.01445]

[7] 孙绪壮. 基于实例迁移学习的跨领域方面级情感分析方法研究 [硕士学位论文]. 太原: 山西财经大学, 2025.

[8] 吴炳潮, 邓成龙, 关贝, 陈晓霖, 昝道广, 常志军, 肖尊严, 曲大成, 王永吉. 动态迁移实体块信息的跨领域中文实体识别模型. 软件学报, 2022, 33(10): 3776–3792. http://www.jos.org.cn/1000-9825/6305.htm [doi: 10.13328/j.cnki.jos.006305]

[9] 李诗晨, 王中卿, 周国栋. 大语言模型驱动的跨领域属性级情感分析. 软件学报, 2025, 36(2): 644–659. http://www.jos.org.cn/1000-9825/7156.htm [doi: 10.13328/j.cnki.jos.007156]

[11] 赵光耀, 吕成国, 付国宏, 刘宗林, 梁春丰, 刘涛. 基于领域特有情感词注意力模型的跨领域属性情感分析. 中文信息学报, 2021, 35(6): 93–102. [doi: 10.3969/j.issn.1003-0077.2021.06.010]

[27] 毛达展, 李华勇, 邵艳秋. 半监督跨领域语义依存分析技术研究. 中文信息学报, 2022, 36(2): 22–28. [doi: 10.3969/j.issn.1003-0077.2022.02.003]

[30] 平瑞, 周水生, 李冬. 高度不平衡数据的代价敏感随机森林分类算法. 模式识别与人工智能, 2020, 33(3): 249–257. [doi: 10.16451/j.cnki.issn1003-6059.202003006]

作者简介

赵传君, 博士, 副教授, CCF 专业会员, 主要研究领域为情感计算, 迁移学习.

孙绪壮, 硕士, 主要研究领域为自然语言处理, 情感计算.

康璐, 硕士生, 主要研究领域为情感计算.

李旸, 博士, 副教授, 主要研究领域为情感分析, 人工智能.

王素格, 博士, 教授, 博士生导师, CCF 专业会员, 主要研究领域为自然语言处理, 情感分析.

李德玉, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为数据挖掘, 粒计算, 机器学习.