

面向电子商务评论的多标签长尾分类样本感知损失函数

谢周坚

贵州大学省部共建公共大数据国家重点实验室, 贵州 贵阳

收稿日期: 2026年3月9日; 录用日期: 2026年3月20日; 发布日期: 2026年5月15日

摘要

在电子商务平台中, 商品评论的标签分布往往呈现长尾特性: 少数热门标签拥有大量训练样本, 而大量细粒度标签仅有极少样本支持。这种严重的数据不平衡导致传统分类方法过度拟合头部类别, 严重损害尾部标签的分类性能, 限制了电商平台对用户评论深度理解的准确性。针对这一问题, 本文提出一种面向电商评论的样本感知损失函数, 通过三个核心组件的协同设计系统地缓解长尾分布带来的性能瓶颈。首先, 引入温度感知的标签权重机制, 利用可调节的温度参数灵活控制各类别权重的分布锐度, 实现对电商评论中低频标签的自适应强化。其次, 提出多标签适配的LDAM损失, 仅对正标签施加类别边际约束, 在保留电商评论多标签语义结构的同时融入基于边际的优化策略。第三, 设计协同混合损失函数, 通过边际调整重构分类决策边界, 再在其基础上应用Focal损失调制机制, 使二者在训练过程中产生协同效应, 提升模型对电商评论中难分样本的感知能力。在EUR-Lex和Wiki10-31K两个公开数据集上的实验结果表明, SAL方法在P@k和N@k等主流评价指标上均显著优于现有的基线方法。同时, 通过PSP@k指标的评估证实, SAL方法在保持优秀整体性能的同时, 有效兼顾了电商评论中尾部标签的识别能力, 验证了各模块设计的有效性与整体方法的优越性。

关键词

电子商务评论分类, 多标签文本分类, 长尾学习, 混合损失, 样本感知损失

Multi-Label Long-Tail Classification Sample-Aware Loss Function for E-Commerce Reviews

Zhoujian Xie

Provincial-Ministry Joint Construction of State Key Laboratory of Public Big Data, Guizhou University, Guiyang
Guizhou

Abstract

In e-commerce platforms, the label distribution of product reviews often exhibits long-tail characteristics: a small number of popular labels have a large number of training samples, while numerous fine-grained labels are supported by only a few samples. This severe data imbalance leads to overfitting of traditional classification methods on head categories, severely impairing the classification performance of tail labels and limiting the accuracy of e-commerce platforms' deep understanding of user reviews. To address this issue, this paper proposes a Sample-Aware Loss function for e-commerce reviews, which systematically alleviates the performance bottleneck caused by long-tail distributions through the collaborative design of three core components. Firstly, a temperature-aware label weight mechanism is introduced, utilizing adjustable temperature parameters to flexibly control the sharpness of the distribution of weights for various categories, achieving adaptive reinforcement for low-frequency labels in e-commerce reviews. Secondly, a multi-label adaptive LDAM loss is proposed, which only imposes category margin constraints on positive labels, integrating margin-based optimization strategies while preserving the multi-label semantic structure of e-commerce reviews. Thirdly, a collaborative hybrid loss function is designed, which reconstructs the classification decision boundary through margin adjustment, and then applies the Focal loss modulation mechanism on top of it, enabling the two to produce a synergistic effect during training and enhancing the model's perception ability for difficult-to-classify samples in e-commerce reviews. Experimental results on two public datasets, EUR-Lex and Wiki10-31K, demonstrate that the SAL method significantly outperforms existing baseline methods in mainstream evaluation metrics such as P@k and N@k. Meanwhile, evaluation through the PSP@k metric confirms that the SAL method effectively balances the recognition ability of tail labels in e-commerce reviews while maintaining excellent overall performance, verifying the effectiveness of each module design and the superiority of the overall method.

Keywords

E-Commerce Review Classification, Multi-Label Text Classification, Long-Tail Learning, Hybrid Loss, Sample-Aware Loss

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着电子商务的迅猛发展，网络购物已成为人们日常生活中不可或缺的一部分。据中国互联网络信息中心数据显示，截至 2025 年底，我国网络购物用户规模已超过 9 亿，电子商务交易额持续增长。在这一背景下，商品评论作为连接消费者与商家的关键纽带，承载着海量的用户反馈信息。对这些评论进行深度挖掘和理解，不仅能够帮助消费者作出更明智的购买决策，也能为商家提供产品改进的方向，同时为电商平台的推荐系统、舆情监控和售后服务体系提供重要支撑。

多标签文本分类作为自然语言处理领域的重要研究方向，在电商评论分析中具有广泛的应用前景。一条电商评论往往同时包含多个维度的信息，如对产品质量的评价、对物流速度的反馈、对客服态度的描述等[1]。通过对这些多维标签的准确识别，电商平台可以构建更加精细化的用户画像，优化商品推荐

算法,提升用户体验。然而,在实际应用中,电商评论的标签分布往往呈现天然长尾特性:少数“头部”标签如“好评”“差评”等拥有充足的训练样本,出现频率极高;而大量具有细粒度语义的“尾部”标签,如“包装精美”“物流迅速”“售后贴心”“性价比高”等,则仅有极少的样本支持,出现次数非常有限[2]。

这一现象在极端多标签分类任务中尤为突出。以电商领域常用的公开数据集 Wiki10-31K 为例,该数据集包含超过三万个候选标签,最频繁的头部标签可出现数千次,而数以千计的尾部标签在整个数据集中仅出现一次。这种严重倾斜的长尾分布给电商评论的多标签学习带来了严峻挑战。具体而言,以二元交叉熵(BCE)为代表的传统损失函数在优化过程中容易被头部标签主导,模型因而倾向于拟合高频类别,导致对电商评论中具有重要商业价值的细粒度尾部标签的分类性能严重下降。这一问题直接影响了电商平台对用户评论深度理解的准确性,进而制约了基于评论分析的下游应用效果,如精准推荐、舆情预警、售后服务优化等。

基于上述背景,本文提出一种基于预训练语言模型的电商评论多标签长尾分类方法——样本感知损失函数(SAL, Sample-Aware Loss)。该方法由三个核心组件构成,旨在系统性地缓解电商评论数据中长尾分布带来的性能瓶颈。首先,我们引入温度感知的标签权重机制,通过可调节的温度参数 τ 灵活控制各类别权重的分布锐度,从而实现对电商评论中低频标签的自适应强化。其次,针对电商评论多标签场景的特点,我们提出多标签适配的 LDAM(标签分布感知边际)损失[3],仅对正标签施加类别边际约束,在保留电商评论多标签语义结构的同时,有效融入基于边际的优化策略。第三,我们设计了一种协同混合损失函数,该模块首先通过边际调整重构分类决策边界,再在其基础上应用 Focal loss(焦点损失)[4]调制机制,使二者在训练过程中产生协同效应,提升模型对电商评论中难分样本的感知能力。通过在多个公开数据集上的实验验证,本文所提出的 SAL 方法在长尾多标签分类任务中表现出显著的性能提升,验证了各模块设计的有效性与整体方法的优越性,为电子商务评论分析提供了新的技术解决方案。

2. 相关研究

近年来,长尾分布问题在机器学习领域引起了广泛关注,众多研究者从不同角度提出了针对性解决方案。从方法论角度来看,现有研究大致可归纳为三类主流范式:一是重采样方法,通过对训练数据进行过采样或欠采样,调整各类别的样本分布,以缓解数据不平衡带来的偏差;二是重加权方法,根据类别频率为不同类别的损失项赋予差异化权重,从而提升尾部类别在优化过程中的影响力;三是基于边际的方法,通过为稀有类别设置更大的分类边际,引导模型形成更具区分性的决策边界。在上述方法中, Focal loss 凭借其聚焦于难分样本的调制机制,在长尾分类任务中展现出显著效果; LDAM 则通过引入与类别频率相关的边际约束,有效提升了尾部类别的泛化能力。

在电子商务领域,多标签文本分类研究也取得了系列进展。针对电商商品分类中的长尾问题, Wu 等[5]提出了分布平衡损失函数,有效缓解了多标签分类中的数据不平衡问题。Wei 等[6]设计了面向尾部标签的鲁棒预测方法 ReRank,通过标签倾向分数重排序提升尾部标签的分类性能,该方法可应用于不同电商平台的评论分析系统。Qaraci 等[7]则提出了保持凸性的 PW 损失函数,可适配于线性分类器、深度学习架构和基于 Transformer 的方法,为电商场景下的多标签分类提供了新的解决方案。

值得注意的是,上述代表性方法主要面向单标签分类任务设计,其直接迁移至多标签场景仍存在明显局限。电商评论数据中样本常同时关联多个标签,标签之间的共现关系复杂,且不同标签可能分属头尾两端,使得传统的类别均衡策略难以直接适配。LightXML [8]作为当前极端多标签文本分类任务的 SOTA 模型,通过动态负采样策略有效降低了计算复杂度,但未专门针对长尾分布进行优化。本研究在 LightXML 的基础上,引入样本感知损失函数,旨在在进一步提升电商评论多标签分类在长尾场景下的性能表现。

3. 方法

给定训练集 $\{(x_i, y_i)\}_{i=1}^N$ ，其中 x_i 是原始文本， $y_i \in \{0, 1\}^L$ 是由 L 维多热向量表示的 x_i 的标签。我们的目标是提高模型在多标签文本分类中的长尾表现。

本节主要介绍 SAL 方法的实现方式。模型框架如图 1 所示。

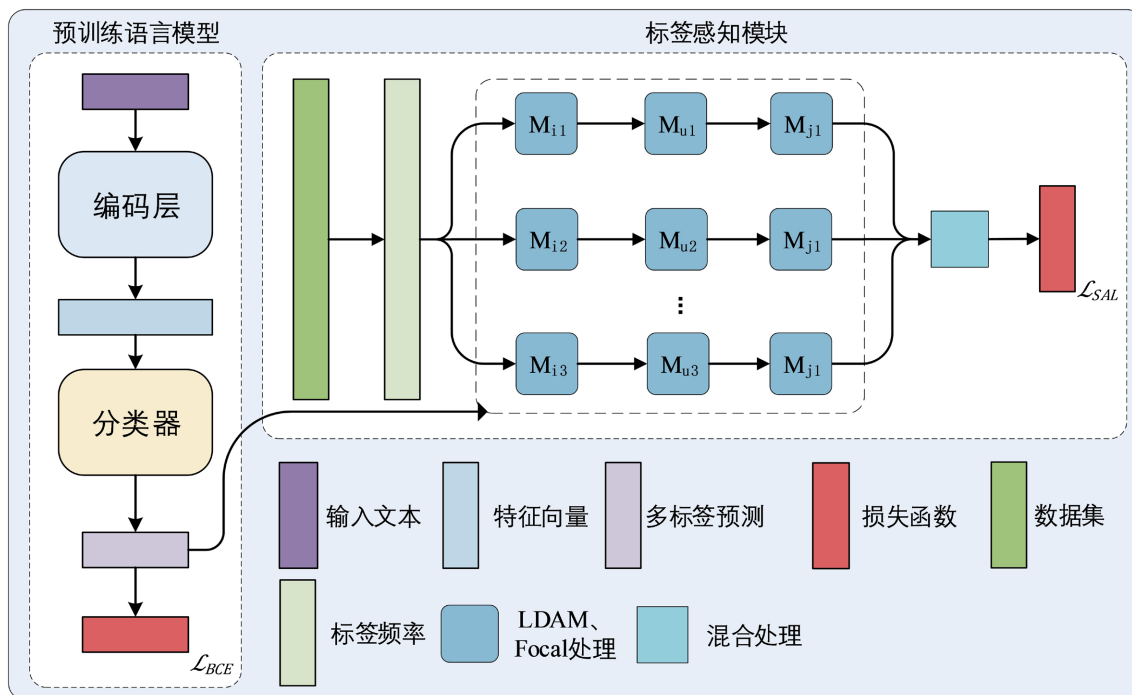


Figure 1. SAL model framework diagram
图 1. SAL 模型框架图

本文中使用了当前最强大的极端多标签文本分类的预训练语言模型 LightXML。LightXML 是针对极端多标签文本分类任务的高效深度学习模型，其工作流程分为两个核心阶段：首先在标签检索阶段，模型通过预训练 Transformer 编码器将输入文本转换为语义向量，同时维护一个可学习的标签嵌入矩阵，通过计算文本与所有标签嵌入的点积相似度，快速筛选出 Top-K 个候选标签，将计算复杂度从 $O(N)$ 降低到 $O(\log N)$ ；随后在精细排序阶段，模型对选中的候选标签进行更深入的交互计算，通过注意力机制捕捉文本与每个候选标签之间的细粒度语义关系，最终输出精确的标签预测结果。整个流程采用端到端的联合学习方式，动态负采样策略使得模型在训练过程中能够自适应地选择最具信息量的负样本，而分组标签嵌入和参数共享等轻量级设计则确保了即使在数万到数百万级别的超大规模标签空间中，LightXML 依然能够保持高效的计算性能和较低的内存占用。

具体做法如下，首先给定训练集中每个类别 c 的样本频率 n_c （即包含该类别的样本数），首先计算基础权重：

$$w_c^{\text{base}} = \frac{1}{\sqrt{n_c + \epsilon}}, \epsilon = 10^{-8} \quad (1)$$

引入温度参数 τ 对权重进行锐化或平滑：

$$w_c^{\text{temp}} = \left(w_c^{\text{base}}\right)^{\tau} \quad (2)$$

最后归一化使得权重的平均值为 1:

$$w_c = \frac{w_c^{\text{temp}}}{\frac{1}{C} \sum_{j=1}^C w_j^{\text{temp}}} \quad (3)$$

其中, C 是类别总数。得到的 w_c 将用于后续损失函数中的样本加权和正样本加权。

接下来是 LDAM 损失中每个类别 c 对应的 margin 基于其频率计算:

$$m_c = \beta \cdot \frac{n_c^{-1/4}}{\max_j n_j^{-1/4}} \quad (4)$$

其中, β 是用于缩放 margin 大小的参数。该设计使得稀有类别获得更大的 margin, 从而在训练中受到更强的约束。

对于样本 i 的 logits $x_{i,c}$ 和真实标签 $y_{i,c} \in \{0, 1\}$, 加权 BCE 损失定义为:

$$\mathcal{L}_{w\text{-BCE}} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C \left[w_c y_{i,c} \log \sigma(x_{i,c}) + (1 - y_{i,c}) \log(1 - \sigma(x_{i,c})) \right] \quad (5)$$

其中, $\sigma(\cdot)$ 是 sigmoid 函数。该损失通过参数在正样本上施加类别权重 w_c 。

再计算标准 BCE 损失项(不带权重):

$$l_{i,c}^{\text{BCE}} = -\left[y_{i,c} \log \sigma(x_{i,c}) + (1 - y_{i,c}) \log(1 - \sigma(x_{i,c})) \right] \quad (6)$$

令 $p_{i,c} = \sigma(x_{i,c})$, 定义:

$$p_{i,c}^{(t)} = y_{i,c} \sigma(x_{i,c}) + (1 - y_{i,c}) (1 - \sigma(x_{i,c})) \quad (7)$$

Focal loss 引入调制因子:

$$\text{focal}_{i,c} = \left(1 - p_{i,c}^{(t)}\right)^{\gamma} \quad (8)$$

以及类别平衡因子(α 为正样本权重):

$$\alpha_{i,c} = y_{i,c} \alpha + (1 - y_{i,c}) (1 - \alpha) \quad (9)$$

则样本 i 上类别 c 的损失为:

$$l_{i,c}^{\text{Focal}} = \alpha_{i,c} \cdot \text{focal}_{i,c} \cdot l_{i,c}^{\text{BCE}} \quad (10)$$

最后, 为每个样本引入标签权重平均因子:

$$\bar{w}_i = \frac{\sum_{c=1}^C y_{i,c} w_c}{\sum_{c=1}^C y_{i,c} + \varepsilon} \quad (11)$$

将样本权重广播到所有类别, 得到总损失:

$$\mathcal{L}_{\text{Focal}} = \frac{1}{N} \sum_{i=1}^N \bar{w}_i \sum_{c=1}^C l_{i,c}^{\text{Focal}} \quad (12)$$

接着计算多标签的 LDAM loss, 对 logits 施加基于标签的 margin 偏移:

$$\widetilde{x}_{i,c} = x_{i,c} - m_c \cdot y_{i,c} \quad (13)$$

然后在此调整后的 logits 上计算 BCE 损失(不带权重):

$$l_{i,c}^{\text{LDAM}} = -\left[y_{i,c} \log \sigma(\widetilde{x}_{i,c}) + (1 - y_{i,c}) \log(1 - \sigma(\widetilde{x}_{i,c})) \right] \quad (14)$$

同样应用样本权重平均因子, 得到:

$$\mathcal{L}_{\text{LA}} = \frac{1}{N} \sum_{i=1}^N \overline{w}_i \sum_{c=1}^C l_{i,c}^{\text{LDAM}} \quad (15)$$

最后混合损失结合了 LDAM 的 margin 调整和 Focal loss 的调制机制。首先计算调整后的 logits:

$$\widetilde{x}_{i,c} = x_{i,c} - m_c \cdot y_{i,c} \quad (16)$$

然后计算调整后的概率 $\widetilde{p}_{i,c} = \sigma(\widetilde{x}_{i,c})$ 和对应的正确类别概率:

$$\widetilde{p}_{i,c}^{(t)} = y_{i,c} \widetilde{p}_{i,c} + (1 - y_{i,c})(1 - \widetilde{p}_{i,c}) \quad (17)$$

Focal 调制因子为 $(1 - \widetilde{p}_{i,c}^{(t)})^\gamma$, 平衡因子 $\alpha_{i,c}$ 同上。则混合损失项为:

$$l_{i,c}^{\text{Hybrid}} = \alpha_{i,c} \cdot (1 - \widetilde{p}_{i,c}^{(t)})^\gamma \cdot \left[-\left(y_{i,c} \log \widetilde{p}_{i,c} + (1 - y_{i,c}) \log(1 - \widetilde{p}_{i,c}) \right) \right] \quad (18)$$

最终同样应用样本权重:

$$\mathcal{L}_{\text{Hybrid}} = \frac{1}{N} \sum_{i=1}^N \overline{w}_i \sum_{c=1}^C l_{i,c}^{\text{Hybrid}} \quad (19)$$

总的损失函数为:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{BCE}} + \lambda_{\text{longtail}} \cdot \mathcal{L}_{\text{Hybrid}} \quad (20)$$

4. 实验

4.1. 电商场景下的数据集与实验设置

我们利用两个多标签文本分类数据集来验证我们模型的效果, 它们是 EUR-Lex 和 Wiki10-31K, EUR-Lex 包含欧洲联盟法律文件的多标签分类任务, 其标签分布呈现显著的长尾特性, 类似于电商平台中商品评论的属性标签分布——少数高频标签(如“好评”“差评”)样本丰富, 而大量细粒度标签(如“包装精美”“物流迅速”“售后贴心”)样本稀缺。Wiki10-31K 源自维基百科文档分类任务, 包含超过 3 万个候选标签, 最频繁的头部分类出现数千次, 而数千个尾部标签仅出现一次。这一分布特性与大型电商平台(如淘宝、京东、亚马逊)的商品评论标签分布高度吻合, 能够有效检验模型在真实电商场景下的泛化能力。这两个数据集是标签大小从 4 k 到 31 k 不等的大规模数据集。统计数据详见表 1。

Table 1. Dataset parameters

表 1. 数据集参数

数据集	训练样本数量	测试样本数量	标签总数	样本的平均标签数	标签的平均样本数	训练集的平均单词数	测试集的平均单词数
EUR-Lex	15,449	3865	3956	5.3	20.79	1248.58	1230.4
Wiki10-31K	14,146	6616	30,938	18.64	8.52	2484.3	2425.45

对于所有的实验，我们使用 PyTorch-1.13.1 实现我们的模型，并在具有 48 GB 显存的单个 NVIDIA A40 GPU 上训练它们。我们使用 Adam 训练我们的模型，初始学习率设置为 $1e-4$ ，权重衰减为 0.01。我们在文本表示层之后执行 dropout，通常设置为 0.5。为了加快实验速度和减小模型的尺寸，我们使用了 NVIDIA 公司的 AMP (自动混合精度)。此外，我们利用 SWA (随机加权平均) 来避免过度拟合，这通常用于训练深度学习模型。

4.2. 电商场景下的评估方法

为全面评估模型在电商评论分类任务中的性能，我们选取三类主流评价指标：

$P@k$ 衡量模型预测的前 k 个标签中正确标签的比例。在电商场景中，该指标直接反映评论属性识别的准确率，例如当 $k = 3$ 时， $P@3$ 表示模型预测的前 3 个评论属性中实际正确的比例，对电商平台的标签推荐和属性提取具有直接指导意义。

$N@k$ 考虑标签排序位置的权重，赋予排名靠前的正确标签更高的分值。在电商评论分析中，排序靠前的标签往往代表用户最关注的评论维度，如“产品质量”“物流速度”等，因此该指标更能反映模型在实际应用中的用户体验。

$PSP@k$ 针对尾部标签预测性能的专门评估指标，通过倾向性评分对长尾标签赋予更高权重。在电商场景中，该指标能够有效衡量模型对低频但具有商业价值的细粒度属性(如“赠品丰富”“退换货便捷”)的识别能力，避免模型过度追求整体准确率而忽视尾部标签。

上述三个指标构成完整的评估体系： $P@k$ 评估基础准确率， $nDCG@k$ 考量排序质量， $PSP@k$ 聚焦尾部性能，全面反映模型在电商评论分类任务中的实用价值。

4.3. 电商场景下的实验结果及分析

首先，为了说明我们的方法在多标签文本分类任务上的有效性，我们选择了六个基线来比较我们实验的有效性，详细信息如下：

-**Base Model**: 当前极端多标签文本分类任务的 SOTA 模型 LightXML，采用动态负采样策略降低计算复杂度，作为本研究的基准框架。

-**Focal loss** 是解决类别失衡的最具代表性的模型，但主要针对多类别而不是多标签任务。

-**CB [9]** 考虑到影响模型训练的样本数与实际样本数不同，定义了一个公式来计算有效样本数，并用新定义的有效样本数重新加权损失函数。

-**NTR** 是分配平衡损失的第二部分。第一部分是过度抽样的改进，我们没有采用。该方法改进了 BCE 损失函数，缓解了负样本的梯度抑制。

-**ReRank** 是为消除标签先验造成的影响而设计的模块。它利用每个标签的不同倾向分数来重新排序其预测结果。

-**PW** 是一个凸代换 0~1 损失的损失函数，避免了失去凸性和下界等重要性质，并结合基于标签频率的再平衡，进一步适应了标签不平衡。

所有基线方法均在 LightXML 框架下实现，确保对比的公平性。SAL 方法集成 BERT、XLNet 和 RoBERTa 三个预训练语言模型进行表示学习，通过集成策略进一步提升预测效果。所有结果列于表 2 中，其中最佳结果用黑体标记。

实验结果表明，对于 $P@k$ 和 $N@k$ 度量值，我们的方法在 EUR-Lex 和 Wiki10-31K 数据集上的表现优于所有其他基线模型。在 $P@1$ 度量上，我们的模型分别比这两个数据集上的次优结果提高了 0.69% 和 0.3%；在 $P@3$ 度量上，我们的方法比次优结果分别提高了 1.49% 和 0.48%；在 $P@5$ 度量上，我们的方法

比次优结果分别提高了 1.57%和 1.11%。在 N@3 度量上, 我们的模型分别比这两个数据集上的次优结果提高了 1.41%和 0.44%; 在 N@5 度量上, 我们的方法比次优结果分别提高了 1.5%和 1.21%。

Table 2. Experimental results
表 2. 实验结果

EUR-Lex								
Method	P@1	P@3	P@5	N@3	N@5	PSP@1	PSP@3	PSP@5
Base Model	86.21	74.40	62.12	77.62	71.64	39.85	49.90	54.47
Focal	85.17	73.01	60.76	76.32	70.29	40.21	49.38	53.55
CB	85.98	73.00	60.26	76.49	70.09	38.52	47.37	51.43
NTR	86.91	74.33	61.30	77.72	71.75	39.63	49.77	53.63
PW	86.13	73.76	61.46	77.14	71.11	43.40	51.50	54.65
ReRank	68.67	66.30	59.90	67.30	65.73	49.31	53.66	56.05
SAL	87.66	75.89	63.69	79.13	73.25	42.49	51.85	56.18
Wiki10-31K								
Method	P@1	P@3	P@5	N@3	N@5	PSP@1	PSP@3	PSP@5
Base Model	87.96	77.75	69.19	80.09	73.58	14.84	15.81	16.66
Focal	86.90	75.33	66.99	77.93	71.50	15.55	16.24	16.98
CB	88.35	77.35	67.16	79.89	72.21	12.94	16.42	13.87
NTR	89.16	78.37	68.70	80.88	73.56	12.98	14.43	15.66
PW	88.48	77.56	68.21	80.03	72.95	13.77	14.56	15.41
ReRank	60.71	57.26	54.26	58.10	55.83	19.68	18.96	19.08
SAL	89.46	78.85	70.30	81.32	74.79	15.00	15.75	16.53

上述提升在电商评论分类场景中具有重要意义: P@1 提升 0.3%~0.7%意味着对用户评论核心属性的识别准确率显著增强; P@3 和 P@5 提升 0.5%~1.6%表明模型对多维评论属性的综合识别能力大幅优化; N@k 指标的同步提升证实了 SAL 方法在标签排序质量上的优势, 这对电商平台的标签推荐系统至关重要。

对于 PSP@k 度量值, 我们的方法在这两个数据集仅次于 ReRank 方法。但 ReRank 方法牺牲了许多 P@k 和 N@k 的性能, 对于头部标签限制过大。相比之下, SAL 方法在保持顶尖整体性能的同时(P@k 和 N@k 均排名第一), 实现了优秀的尾部标签识别能力(PSP@k 排名第二)。这一平衡特性完全符合电商平台的实际需求——既要确保热门评论属性(如“好评”“差评”)的准确识别, 也要提升细粒度属性(如“包装精美”“物流迅速”“售后贴心”)的挖掘能力。

4.4. 电商场景应用商业价值

从电商平台实际应用角度, SAL 方法的性能提升具有以下商业价值:

P@k 指标的全面提升意味着平台能够更准确地从海量评论中提取关键属性信息, 包括产品质量、物流服务、售后体验等维度, 为商家改进产品和优化服务提供数据支撑。N@k 指标的改善表明模型对评论标签的排序质量更高, 能够将用户最关注的属性排在前面, 有助于构建更精细化的用户画像, 提升商品推荐的相关性和转化率。PSP@k 指标的保持证明模型对低频但具有商业价值的细粒度属性(如“赠品丰

富”“退换货便捷”“客服态度好”)具备良好的识别能力,帮助平台发现潜在的改进点和商机。高精度的自动化评论属性提取可大幅减少人工标注和审核的工作量,提升电商平台的运营效率和响应速度。

5. 结论

本文针对电子商务评论多标签文本分类中的长尾分布问题,提出了一种面向电商场景的样本感知损失函数(SAL)。该方法通过温度感知的标签权重机制实现对低频标签的自适应强化,通过多标签适配的LDAM损失在保留评论语义结构的同时引入边际约束,通过协同混合损失函数增强模型对难分样本的感知能力,三个核心组件的协同设计系统性地解决了电商评论数据中头部标签主导、尾部标签识别困难的挑战。在EUR-Lex和Wiki10-31K两个具有典型长尾分布特征的数据集上的实验结果表明,SAL方法在P@k和N@k指标上均显著优于现有基线方法,同时通过PSP@k指标证实其在保持顶尖整体性能的前提下有效兼顾了尾部标签的识别能力。本研究对电子商务领域具有重要的应用价值:可集成至电商平台的智能评论分析系统,实现对海量用户评论的自动化多维属性提取;可为精准推荐引擎提供更丰富的用户偏好信息,构建多维度用户画像;可帮助商家及时发现产品问题和改进机会,优化售后服务体系;可大幅减少人工标注和审核的人力投入,提升平台运营效率。未来研究将引入真实电商评论数据集进行验证,探索小样本与零样本学习以解决冷启动问题,优化模型轻量化设计以满足实时性要求,并拓展多模态评论分析和跨领域迁移学习等方向,进一步提升方法在电商场景中的实用性和泛化能力。

参考文献

- [1] 李可悦,陈轶,牛少彰.基于BERT的社交电商文本分类算法[J].计算机科学,2021,48(2):87-92.
- [2] 蒋佳,周玉婷.电商平台内“评论劫持”不正当竞争行为规制研究[J].中国价格监管与反垄断,2026(2):35-37.
- [3] Kaidi, C., Colin, W., Adrien, G., Nikos, A. and Tengyu, M. (2019) Learning Imbalanced Datasets with Label-Distribution-Aware Margin Loss. *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, Vancouver, 8-14 December 2019, 32.
- [4] Lin, T., Goyal, P., Girshick, R., He, K. and Dollar, P. (2020) Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **42**, 318-327. <https://doi.org/10.1109/tpami.2018.2858826>
- [5] Wu, T., Huang, Q., Liu, Z., Wang, Y. and Lin, D. (2020) Distribution-Balanced Loss for Multi-Label Classification in Long-Tailed Datasets. In: Vedaldi, A., et al., Eds., *Computer Vision—ECCV 2020*, Springer International Publishing, 162-178. https://doi.org/10.1007/978-3-030-58548-8_10
- [6] Wei, T., Tu, W., Li, Y. and Yang, G. (2021) Towards Robust Prediction on Tail Labels. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, Singapore, 14-18 August 2021, 1812-1820. <https://doi.org/10.1145/3447548.3467223>
- [7] Qaraei, M., Schultheis, E., Gupta, P. and Babbar, R. (2021) Convex Surrogates for Unbiased Loss Functions in Extreme Classification with Missing Labels. *Proceedings of the Web Conference 2021*, Ljubljana, 19-23 April 2021, 3711-3720. <https://doi.org/10.1145/3442381.3450139>
- [8] Jiang, T., Wang, D., Sun, L., Yang, H., Zhao, Z. and Zhuang, F. (2021) LightXML: Transformer with Dynamic Negative Sampling for High-Performance Extreme Multi-Label Text Classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, **35**, 7987-7994. <https://doi.org/10.1609/aaai.v35i9.16974>
- [9] Cui, Y., Jia, M., Lin, T., Song, Y. and Belongie, S. (2019) Class-Balanced Loss Based on Effective Number of Samples. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 16-21 June 2019, 9260-9269. <https://doi.org/10.1109/cvpr.2019.00949>