计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
63(4): 1010−1020, 2026

锚定边界: 融合双边对比学习和 BMIU 的深度不平衡回归

刘文杰　陈松灿
(南京航空航天大学计算机科学与技术学院　南京　211106)
(模式分析与机器智能工业和信息化部重点实验室(南京航空航天大学)　南京　211106)
(liu_wenjie@nuaa.edu.cn)

Anchors as Boundaries: Integrating Bilateral Contrastive Learning and BMIU for Deep Imbalanced Regression

Liu Wenjie and Chen Songcan
(*College of Computer Science and Technology*, *Nanjing University of Aeronautics and Astronautics*, *Nanjing* 211106)
(*Key Laboratory of Pattern Analysis and Machine Intelligence* (*Nanjing University of Aeronautics and Astronautics*), *Ministry of Industry and Information Technology*, *Nanjing* 211106)

Abstract　Imbalanced regression tackles learning tasks where the labels are continuous but the sample distribution across equidistant label-bins is skewed, with its deep learning extension forming the emerging field of deep imbalanced regression (DIR). Existing algorithms implicitly assume a consistent similarity pattern among samples within equidistant label-bins. For example, this presumption implies uniform facial aging rates across age groups and identical depth perception cues across spatial intervals. Furthermore, DIR also confronts challenges such as significant intra-bin sample heterogeneity and overlapping decision boundaries between adjacent bins. To address the aforementioned challenges, we propose BCLU, a novel strategy integrating bilateral contrastive learning (BCL) and bilateral mixup-induced Universum (BMIU) augmentation, along with the corresponding RBLU algorithm. On one hand, BCL dynamically partitions data into left and right subsets anchored at adaptively determined boundaries, enabling unidirectional matching of strongly similar positive pairs within each subset. This mechanism overcomes the dilemma of constructing negative pairs with weak label similarity but strong semantic affinity, thereby relaxing the homogeneous assumption. Experimental results demonstrate BCL's effectiveness in mitigating the imbalance of samples. On the other hand, BMIU generates hybrid features by mixing anchors with randomly selected samples from bilateral subsets, constructing wedge-shaped features that belong to no specific label-bin. This strategy enforces the discriminant boundary between label-bins to be maximally separated while ensuring that the features within each label-bins are as compact as possible, thereby suppressing inter-bin overlaps. Notably, BMIU inherently preserves label-ordering consistency without explicit any ordinal constraints. Extensive experiments on standard DIR benchmarks thoroughly validate the superior efficacy of the proposed RBLU algorithm.

Key words　machine learning; deep imbalanced regression; contrastive learning; imbalanced data; Universum

摘　要　不平衡回归专指标签连续但其等长标签段(label-bin)内样本数不等的学习任务,而对应的深度延伸形成了当前流行的深度不平衡回归(deep imbalanced regression, DIR)。这些算法均隐含假设了等间隔标签段组成的样本对相似性一致。以年龄和图像景深(depth)估计为例,该假设认为不同年龄段人脸

收稿日期: 2024-06-05; **修回日期**: 2025-10-21
基金项目: 国家自然科学基金项目(62106102)
This work was supported by the National Natural Science Foundation of China (62106102).
通信作者: 陈松灿(s.chen@nuaa.edu.cn)

老化速度和不同空间距离深度视觉线索均相同；与此同时，DIR 还面临标签段内样本差异大以及相邻标签段判别界限重叠的问题。为应对上述挑战，提出一种融合双边对比学习（bilateral contrastive learning, BCL）和双边 mixup 诱导的 Universum（bilateral mixup-induced Universum, BMIU）增广的 BCLU 策略及其对应的 RBLU 算法。一方面，BCL 以锚点为边界将数据划分成左右子集，锚点在各子集内能单向匹配到强相似的正对，克服了仅依靠标签距离构造出标签弱相似但语义强相似负对的困境，达到放宽均匀假设的目的，实验也证明 BCL 显著缓和了样本不平衡；另一方面，BMIU 将锚点与两侧子集内随机选择的样本特征混合，构造出不属于任何标签段且能抑制相邻标签段重叠的楔形（wedge）特征，强制标签段间判别边界尽可能分离以及标签段内特征尽可能紧致以降低类间重叠。值得提及的是，BMIU 能作为副产品在无需引入任何有序约束的前提下便能自然地保持有序特性。最后在 DIR 标准数据集上充分验证了所提 RBLU 算法的有效性。

关键词 机器学习；深度不平衡回归；对比学习；不平衡数据；Universum

中图法分类号 TP391

DOI: 10.7544/issn1000-1239.202550468 **CSTR:** 32373.14.issn1000-1239.202550468

回归作为机器学习的一项基本任务，旨在建立输入到连续输出值之间的映射关系。有时为了便于统计会将某些特定场景的输出离散化成多个等长的标签段[1]（label-bin）。然而，受现实因素制约，这类回归任务中不同标签段所含样本密度差异明显，其中部分标签段包含大量样本，而有的则样本稀少甚至为空[2]，导致了类似分类任务中的不平衡长尾分布[3]（long-tail distribution）。例如连续年龄值对应标签段长度通常为 1 岁，并且相较于样本丰富的年轻群体，100 岁以上的高龄样本显得尤为稀疏[2]。尽管长尾分类已引起学界的广泛关注，但对回归场景的不平衡现象探索有限。不同于分类，回归中相邻标签段的样本因缺乏足够的判别特征[2,4]导致重叠严重，如数值 3.9 和 4.1 传统上对应的年龄边界 3 岁与 4 岁难以区分。为了有效处理这种具有连续标签值的不平衡数据，不平衡回归技术[5-6]逐渐受到关注，鉴于深度化是目前众多领域的主要发展趋势，自然延伸出深度不平衡回归[7]（deep imbalanced regression, DIR）。现有的 DIR 方法大致可分为 2 类：1）回归型方法[7-10]，包括利用邻近样本和有序信息缓解不平衡的传统回归，或结合先进对比策略[10]的方案。2）混合型方法[11-13]，包括在回归损失后添加约束不平衡的分类损失，或将回归视为分类的层次调整方案[13]。本文关注属于回归型的对比策略，但其本身也可用于混合型。

上述 2 类现有方法均隐含假设了等间隔标签段组成的样本对相似性一致，以年龄和图像景深（depth）估计为例，这些方法认为不同年龄段的人脸老化速度和不同空间距离的深度视觉线索均恒定。而 Geng 等人[14]指出人脸在童年、中年、老年呈现非均匀演化规律，即在上述假设下的对比策略[10]尽管已经达到 SOTA，但其构造正负样本对时仅仅利用标签距离而忽略了关键的语义信息，面临着掺杂标签强相似而语义弱相似正对以及标签弱相似而语义强相似负对的困境。

尽管放弃均匀假设更贴近现实，但遗憾的是相似性变化规律，如衰老趋势受基因、环境、生活习惯等因素影响[2,14-15]，难以显式建模。为了克服上述困境，本文提出一种总是以锚点为边界将数据动态地划分为左右互补子集的双边对比学习（bilateral contrastive learning, BCL）策略。BCL 无需知晓样本变化规律，便可在其左右子集内分别构造出符合预期的正负对。例如给定年龄样本集合 D ={17, 18, 19, 20, 21, 22, 23}，当锚点为 20 岁时，BCL 以此为界将 D 划分为左子集 D_{l} ={17, 18, 19}和右子集 D_{r} ={21, 22, 23}，如图 1（b）所示。若正例阈值设置为 1，在 D_{l} 内锚点向左匹配到在标签和语义层面均强相似正例{19}和相对弱相似负例{17, 18}，无需考虑正样本之前{17, 18, 19}的变化规律，在 D_{r} 内同理。而现有方法直接在 D 内构造正例{19, 21}和负例{17, 18, 22, 23}，如图 1（a）所示，但实际可能的语义相似性：$sim(20, 19)<sim(20, 22)\approx sim(20, 23)$，即在特征空间中将语义本不相似的样本视作正例{19}，同时将语义本身相似的样本视作负例{22, 23}会导致矛盾。BCL 放弃同时约束跨越锚点的正负对，以减轻不均匀变化对模型的误导。实验也证明 BCL 能显著缓解不平衡对模型的影响。

针对 DIR 标签段内样本差异显著以及相邻标签段间界限重叠问题，本文尝试构建一种楔形特征，以实现特征空间中标签段内特征紧致、标签段间特征分离。幸运的是，已有研究表明 Universum 作为对立

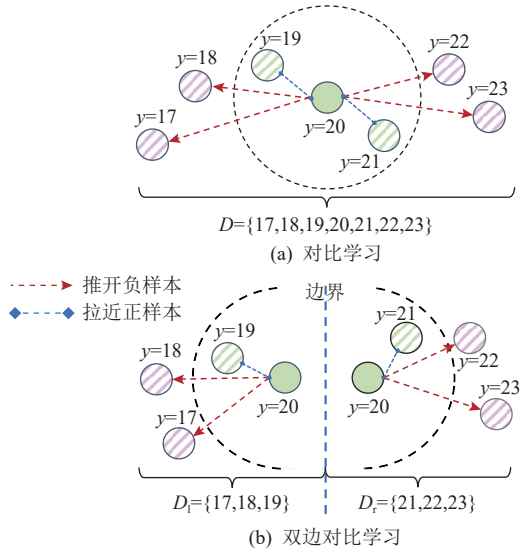


Fig. 1 Contrastive learning and bilateral contrastive learning

图 1 对比学习与 BCL

于给定类别概念的样本(即给定类外的样本)能够有效提升类间的判别性和泛化性[16], 这种特性与本文所追求的楔形特征直觉上高度契合。Universum 最初由文献 [17−18] 提出作为负类用于支持向量分类机(U-SVC), 研究显示只有选择合适的 Universum 方能提升模型性能。此后 Dhar 等人[19] 采用添加噪声等策略构造 Universum 用于支持向量回归机(U-SVR), 将其拓展至类概念不明确的回归任务。但上述通过样本外数据构造的 Universum 并非总是有效, 值得注意的是, 近年来 Han 等人[16] 在分类任务中利用 2 类样本的平衡混合, 直接由数据本身的矛盾类概念构造 MIU(mixup-induced Universum)既简单又有效。受此启发, 本文尝试借鉴 MIU 的构造方式设计楔形特征。但在 DIR 上直接应用 MIU 的生成方式并不有效, 因为会导致特征朝任意方向扩展, 难以保持有序。为迎合 DIR 的独特性, 本文专门在特征空间里设计了既简单又能良好适配 DIR 的双边 mixup 诱导的 Universum(bilateral mixup-induced Universum, BMIU)作为楔形特征, 在标签段两边同时迫使其内部特征向原型靠拢以降低类间重叠。消融实验进一步表明 BMIU 无需如现有方法要额外引入有序约束(如有序编码[1,20] 或排序方法[9,21] 等), 便能自然地保持特征的有序性, 简化了优化过程。

总之, 本文工作的贡献归纳有 3 点:

1)提出一种以锚点为边界, 将数据划分成双侧的 BCL 模块, 以适应 DIR 标签段整体的非均匀性。

2)设计了构造楔形特征的 BMIU 模块, 能紧化标签段内和分离标签段间的特征。有趣的是, BMIU 能作为副产品自然获得有序信息的编码, 免去了显式的有序约束。据我们所知, 这属于首次尝试。

3)融合 BCL 模块和 BMIU 模块形成 BCLU, 最后结合传统回归(regression)得到深度不平衡回归的 RBLU 算法, 并在 DIR 标准数据集上通过丰富的对比实验验证其有效性。

1 相关工作

1.1 不平衡回归

不平衡回归[5-6](IR)专指标签连续但其等长标签段内样本数不等的学习任务。早期的 IR 研究借鉴长尾分类方法, 如合成数据[5,22]、集成模型[23] 或调整损失[24] 等。随着深度化在各个领域的普遍流行, Yang 等人[7] 提出 DIR 概念, 而后逐渐发展出回归型和混合型 2 类技术。对于回归型 DIR, 文献 [7] 借用较相似的邻近样本得到平滑的标签分布和特征分布。RankSim[9] 则鼓励样本在特征空间与标签空间的排序保持一致, 以维持特征的有序性, 但排序算法复杂度高且难以适用高维数据。Ren 等人[8] 提出可用于高维数据的平衡均方误差损失来应对不平衡的视觉回归。此外, 回归型 DIR 也已拓展至分子和聚合物性质预测[25]、太阳高能粒子预测[26] 和心血管年龄预测[27] 等任务。与此同时, 结合分类策略的混合型 DIR 的研究也在不断跟进, 其中 Pintea 等人[11] 证明了回归损失后添加约束不平衡的分类损失能获得性能提升; Zhang 等人[12] 指出与分类损失相比回归损失均方误差(MSE)缺乏高熵的特征表示, 据此提出有序熵弥补。二者从不同角度探讨了回归与分类的联系, 但实际对 DIR 的改进有限。最近 Xiong 等人[13] 提出层次分类调整法, 将原始的等长标签段再次映射到完全平衡但非等长的新标签段, 并结合蒸馏处理 DIR。以上方法均默认遵循均匀假设, 相比之下, RBLU 中的 BCL 模块迎合了 DIR 非均匀性的事实, 并同时利用了邻近样本的相似信息(正例)和标签索引的距离信息(负例)。

1.2 对比学习

对比学习通过拉近相似样本和推开不同样本的方式实现表示学习, 成为当前的主流方法之一[28-31], 并广泛应用于视觉[32]、文本[33]、多视图[34] 及多模态[29] 等领域。它既可用于无监督学习, 如 Chen 等人[29] 提出 SimCLR 框架, 将个体的增广作为正例通过约束语义的一致性进行学习; 也可用于监督学习, 如 Khosla 等人[30] 提出的 SupCon 框架不仅利用了个体与自身增广视图的一致性, 还进一步约束同标签样本语义

的一致性；甚至用于更具挑战的长尾学习以缓和不平衡的影响[35-37]。此外，对比学习近年也拓展到输出值连续的一般回归场景，尽管取得了进展，但仍鲜有学习有序特征的回归方法。就我们所知，仅有专门设计的 Rank-N-Contrast[21] 方法被提出，但其涉及排序且未顾及其中的不平衡。另一方面，作为对比回归的 SOTA 方法 ConR[10] 虽考虑了不平衡，但遗憾地未顾及有序性。最近 SRL[38] 将几何原理整合到数据表示，确保均匀的特征分布。但它们均立足于均匀假设，既不十分合理，也会引起前述的困境。本文提出 BCLU 中的 BCL 模块刚好能克服此困境，并进一步融合专为减轻重叠设计的 BMIU 模块和传统回归以得到 RBLU 算法。

2　融合 BCL 和 BMIU 增广的 BCLU

本节介绍 RBLU 算法的实现细节，其框架如图 2 所示。问题描述如下，给定训练集 $D_{\text{train}}=\{(\boldsymbol{x}_i,y_i)\}_{i=1}^{N}$，$\boldsymbol{x}_i\in\mathbb{R}^d$ 是输入数据，$y_i\in\mathbb{R}$ 是对应的标签，在预处理阶段按具体任务将连续的目标值 y 划分成 B 等长的标签段，$[y_0,y_1),[y_1,y_2),\cdots,[y_{B-1},y_B)$，其中 D_{train} 中各标签段样本数不等，D_{test} 服从均匀分布。输入 $\boldsymbol{x}$ 经过编码器 $f_\theta:\boldsymbol{x}\to\boldsymbol{z}$ 和预测器 $g_\varphi:\boldsymbol{z}\to\hat{y}$，期望输出的 $\hat{y}$ 与实际标签段索引 Y_b 接近。具体步骤为：

1）输入 $\boldsymbol{x}$ 经过编码器 $f_\theta(\cdot)$ 得到特征 $\boldsymbol{z}$；

2）利用 BCL 模块划分左右子集以应对均匀假设；

3）利用 BMIU 构造楔形特征以减轻特征重叠；

4）计算包含 BCLU 和传统回归损失的 RBLU。

通过迭代上述步骤，将无序、分散、边界重叠严重的特征转为相对有序、紧致、边界尽量分离的特征。

2.1　BCL 模块

近年来对比学习在回归场景也渐受关注[10,25]，现有方法通过考虑标签距离来改进传统对比损失，但均立足于均匀假设，同时在锚点两侧构造正负例，会产生一侧标签强相似而语义弱相似的正例而另一侧标签弱相似而语义强相似负例的困境。BCL 模块专为此设计，沿用对比回归范式[10] 并引入动态划分机制（锚定边界）逐个选取样本作锚点，据此将数据划分为左子集 D_{l} 和右子集 D_{r}，确保在 D_{l} 内的正例只对比其负例，放弃对比 D_{r} 的负例，在 D_{r} 内同理。

BCL 具体构造正负例的方式为 $\|y_i-y_j\|\leqslant w$ 为正例，$\|y_i-y_j\|>w$ 且 $\|\hat{y}_i-y_j\|\leqslant w$ 为负例，w 为设定的阈值。若锚点 j 在 D_{l} 中 $K_j^{\text{l}+}$ 与 $K_j^{\text{l}-}$ 非空，其损失为

$$\mathcal{L}_j^{\text{l}}=-\log\frac{1}{N_j^{\text{l}+}}\frac{\ell_{\text{pos}}^{\text{l}}(j)}{\ell_{\text{neg}}^{\text{l}}(j)+\ell_{\text{pos}}^{\text{l}}(j)},\tag{1}$$

否则为 0；若 j 在 D_{r} 中 $K_j^{\text{r}+}$ 与 $K_j^{\text{r}-}$ 非空，其损失为

$$\mathcal{L}_j^{\text{r}}=-\log\frac{1}{N_j^{\text{r}+}}\frac{\ell_{\text{pos}}^{\text{r}}(j)}{\ell_{\text{neg}}^{\text{r}}(j)+\ell_{\text{pos}}^{\text{r}}(j)},\tag{2}$$

否则为 0。

$$\begin{cases}\ell_{\text{pos}}^{*}(j)=\displaystyle\sum_{\boldsymbol{z}_i^{*}\in K_j^{*+}}\exp(\boldsymbol{z}_j\cdot\boldsymbol{z}_i^{*}/\tau),\\ \ell_{\text{neg}}^{*}(j)=\displaystyle\sum_{\boldsymbol{z}_q^{*}\in K_j^{*-}}S_{j,q}\exp(\boldsymbol{z}_j\cdot\boldsymbol{z}_q^{*}/\tau),\end{cases}\tag{3}$$

其中 τ 是温度系数；N_j^{*+} 为正样本数；K_j^{*+}，K_j^{*-} 是正、负样本集合；“*”为左（l）、右（r）标识。

$S_{j,q}$ 为推开负样本的权重，定义为

$$S_{j,q}=\eta_j\mathcal{L}_{\text{reg}}(y_j,y_q)。\tag{4}$$

结合 j 的左右损失得到整体损失函数为

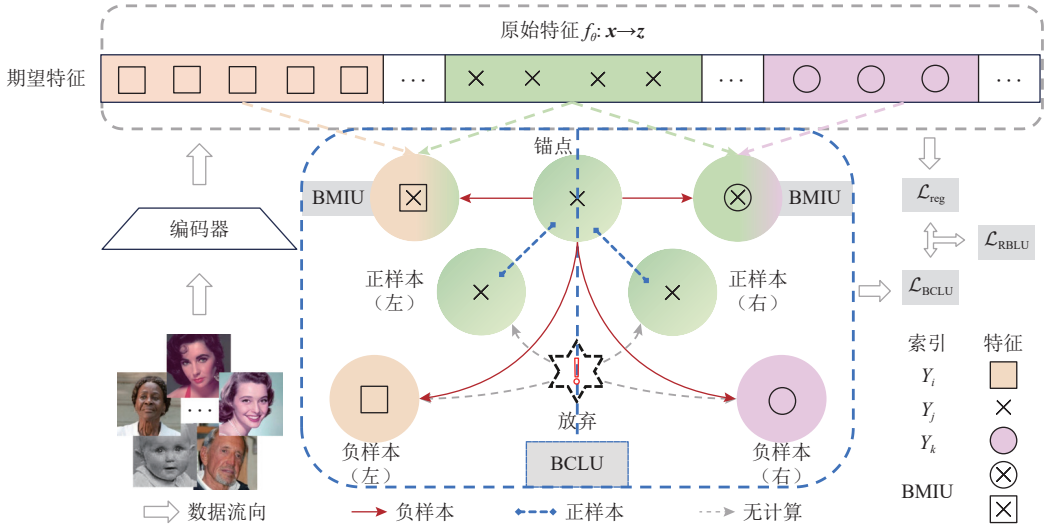


Fig. 2　Model framework of RBLU

图 2　RBLU 模型框架

$$\mathcal{L}_j = \alpha_j \mathcal{L}_j^{\mathrm{l}} + \beta_j \mathcal{L}_j^{\mathrm{r}}, \tag{5}$$

其中α_j, β_j是调节因子, 定义如下:

$$\alpha_j = \frac{n_j^{\mathrm{l}} + n_j^{\mathrm{r}}}{n_j^{\mathrm{l}}}, \beta_j = \frac{n_j^{\mathrm{l}} + n_j^{\mathrm{r}}}{n_j^{\mathrm{r}}}。 \tag{6}$$

BCL 定义的负样本是真实值在锚点阈值 w 外但预测值在 w 内的样本。对全体锚点而言, 调节因子与当前锚点负样本总数(分子)正相关, 即错分样本越多, 对损失惩罚越大; 对各子集而言, 锚点位置决定了左右子集的长度, 定义调节因子与该子集内负样本数(分母)负相关, 旨在平衡锚点位置对左右损失的影响。遍历所有样本, 得到 BCL 总体损失为

$$\mathcal{L}_{\mathrm{BCL}} = \frac{1}{2N} \sum_{j=1}^{2N} \mathcal{L}_j。 \tag{7}$$

BCL 在单侧子集内约束特征空间中正样本靠近锚点, 负样本远离锚点, 而跨越锚点两侧的正负样本不同时作对比, 克服了仅依靠标签距离构造的弱相似正例而强相似负例的困境, 从而放宽均匀假设。

2.2 BMIU 模块

Mixup 利用线性插值增广样本在分类领域获得成功[39], 但对标签值连续样本的 Mixup 会产生任意不确定的标签[40], 将其引入 DIR, 可能会导致本就难区分的特征重叠加剧。而 Han 等人[16]采用 MIU 构造 Universum 的方式能避免此问题。MIU 实现如下:

$$\boldsymbol{u}_j = \lambda \tilde{\boldsymbol{x}}_j + (1-\lambda)\tilde{\boldsymbol{x}}_{q_j}, j = 1,2,\cdots,2N, \tag{8}$$

其中$\tilde{\boldsymbol{x}}_j$是锚点, λ是混合系数, 取值为λ=1/2, q_j由 $\bigcup_{k\neq j} D_k$ 随机采样得到。

遗憾的是将 MIU 直接用于 DIR 并不有效, 会导致特征任意扩展而无法保持有序。为此本文专门面向特征空间设计了 BMIU 模块以应对 DIR 标签段内特征分散和标签段间特征重叠的挑战。具体来说, BMIU 同时在锚点两侧构造出不属于任何标签段的 BMIU 流形作为楔形特征。如图 3 所示, 当锚点位于标签段 Y_b 内, 锚点与 Y_{b-1} 的特征构造左侧楔形特征迫使 Y_b 的左边界向右收缩、Y_{b-1} 的右边界向左收缩; 与此同时, 锚点与 Y_{b+1} 的特征构造右侧楔形特征迫使 Y_b 的右边界向左收缩、Y_{b+1} 的左边界向右收缩。两侧楔形特征同时朝 Y_b 原型方向施压, 强迫其内部特征尽可能紧致, 避免了单侧施压导致标签段原型不确定地偏移。故 BMIU 虽并未给楔形特征赋予新标签, 也无需关心是否保留了实际有序语义[41], 但能自然地保持有序, 这种无需添加有序约束实现特征有序的方式是在 DIR 的首次探索。

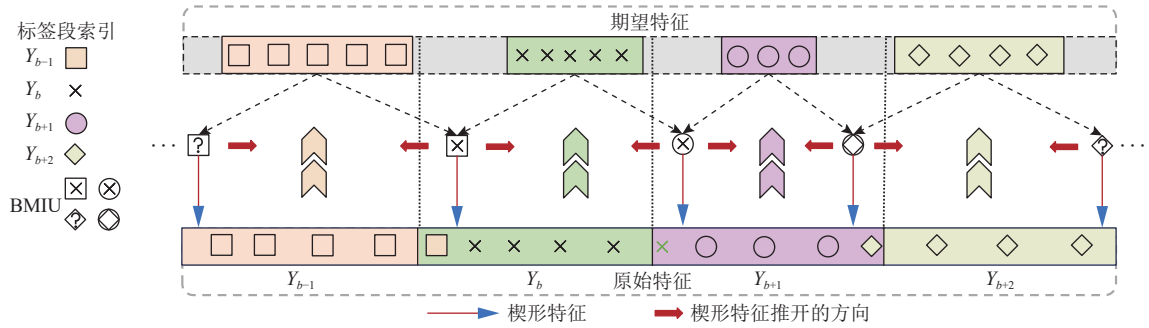


Fig. 3 Specification of BMIU module

图 3 BMIU 模块说明

BMIU 具体实现如下:

$$\begin{cases} z_{\mathrm{u}}^{\mathrm{l}} = \lambda \tilde{z}_j + (1-\lambda)\tilde{z}_j^{\mathrm{l}}, \ \tilde{z}_j^{\mathrm{l}} \in D_{\mathrm{l}}, \\ z_{\mathrm{u}}^{\mathrm{r}} = \lambda \tilde{z}_j + (1-\lambda)\tilde{z}_j^{\mathrm{r}}, \ \tilde{z}_j^{\mathrm{r}} \in D_{\mathrm{r}}, \end{cases} \tag{9}$$

其中$z_{\mathrm{u}}^{\mathrm{l}}$和$z_{\mathrm{u}}^{\mathrm{r}}$为构造的左右楔形特征, λ是混合系数, 取值为λ=1/2, $\tilde{z}_j^{\mathrm{l}}$和$\tilde{z}_j^{\mathrm{r}}$是左右子集内标准化后特征。

2.3 融合 BCL 和 BMIU 的 BCLU

将 2.2 节 BMIU 模块构造的楔形特征作为补充的对立面特征添加至式(5)(6)得到 j 在左右子集最终的损失函数为

$$\mathcal{L}_j^{\mathrm{l,u}} = -\log \frac{1}{N_j^{\mathrm{l+}}} \frac{\ell_{\mathrm{pos}}^{\mathrm{l}}(j)}{\ell_{\mathrm{neg}}^{\mathrm{l}}(j) + \ell_{\mathrm{pos}}^{\mathrm{l}}(j) + \ell_{\mathrm{u}}^{\mathrm{l}}(j)}, \tag{10}$$

$$\mathcal{L}_j^{\mathrm{r,u}} = -\log \frac{1}{N_j^{\mathrm{r+}}} \frac{\ell_{\mathrm{pos}}^{\mathrm{r}}(j)}{\ell_{\mathrm{neg}}^{\mathrm{r}}(j) + \ell_{\mathrm{pos}}^{\mathrm{r}}(j) + \ell_{\mathrm{u}}^{\mathrm{r}}(j)}, \tag{11}$$

其中$\ell_{\mathrm{pos}}^{*}(j)$, $\ell_{\mathrm{neg}}^{*}(j)$与前述类似, $\ell_{\mathrm{u}}^{*}(j)$定义为

$$\begin{cases} \ell_{\mathrm{u}}^{\mathrm{l}}(j) = \sum_{k\in K_{\mathrm{l}}, k\neq j} \exp(z_j \cdot z_{\mathrm{u}}^{k}/\tau), \\ \ell_{\mathrm{u}}^{\mathrm{r}}(j) = \sum_{k\in K_{\mathrm{r}}, k\neq j} \exp(z_j \cdot z_{\mathrm{u}}^{k}/\tau)。 \end{cases} \tag{12}$$

定义 BCLU 损失函数为

$$\mathcal{L}_{j,\mathrm{u}} = \alpha_j \mathcal{L}_j^{\mathrm{l,u}} + \beta_j \mathcal{L}_j^{\mathrm{r,u}}, \tag{13}$$

$$\mathcal{L}_{\mathrm{BCLU}} = \frac{1}{2N} \sum_{j=1}^{2N} \mathcal{L}_{j,\mathrm{u}}。 \tag{14}$$

最小化式(14)会减少总体负样本预测值落在 w

(锚点正样本阈值内, 即错分)的情况。有研究[7]表明不平衡会导致少数样本严重坍塌至头部样本, 即 DIR 中大部分少数标签段的样本被确定为负样本, 优化 BCLU 能有效关注到少数样本, 提升区分准度。

结合传统回归损失 $\mathcal{L}_{\text{reg}}$ 得到 RBLU 损失函数:

$$\mathcal{L}_{\text{RBLU}} = \mathcal{L}_{\text{reg}} + \gamma \mathcal{L}_{\text{BCLU}}, \tag{15}$$

其中 $\mathcal{L}_{\text{reg}}$ 一般为 *MSE*, 平均绝对误差(*MAE*)等, γ 是平衡因子。

RBLU 的流程如算法 1 所示。

算法 1. 深度不平衡回归的 RBLU 算法。

输入: 训练集 D, 等长标签段数 B, 训练轮数 E, 温度系数 τ, 阈值 w, 超参数 λ 和 γ;

输出: 预测结果 $\hat{y}$。

① 初始化 $f_\theta(\cdot)$, $g_\varphi(\cdot)$, $\mathcal{L}_{\text{reg}}=\mathcal{L}_{\text{BCLU}}=\mathcal{L}_{\text{RBLU}}=0$;

② for *epoch* = 1 to E do

③ $\{z_j\}_{j=1}^{2N} \leftarrow f_\theta(\boldsymbol{x}_j)$, $\{\hat{y}_j\}_{j=1}^{B} \leftarrow g_\varphi(z_j)$;

④ $\mathcal{L}_{\text{reg}} \leftarrow \mathcal{L}_{\text{reg}}(\hat{y}, y)$; /*回归损失*/

⑤ for j = 1 to $2N$ do

⑥ $z_{\text{l}} \in D_{\text{l}}$, $z_{\text{r}} \in D_{\text{r}}$; /*锚定边界, 左右子集*/

⑦ if $K_j^{\text{l}+} \neq \varnothing \wedge K_j^{\text{l}-} \neq \varnothing$ then

⑧ $z_{\text{u}}^{\text{l}} \leftarrow \lambda \tilde{z}_j + (1-\lambda)\tilde{z}_j^{\text{l}}$; /*见式(9)*/

⑨ end if

⑩ if $K_j^{\text{r}+} \neq \varnothing \wedge K_j^{\text{r}-} \neq \varnothing$ then

⑪ $z_{\text{u}}^{\text{r}} \leftarrow \lambda \tilde{z}_j + (1-\lambda)\tilde{z}_j^{\text{r}}$; /*见式(9)*/

⑫ end if

⑬ $\mathcal{L}_{j,\text{u}} \leftarrow \alpha_j \mathcal{L}_j^{\text{l,u}} + \beta_j \mathcal{L}_j^{\text{r,u}}$; /*见式(13)*/

⑭ $\mathcal{L}_{\text{BCLU}} \leftarrow \dfrac{1}{2N}\sum_{j=1}^{2N} \mathcal{L}_{j,\text{u}}$; /*见式(14)*/

⑮ end for

⑯ $\mathcal{L}_{\text{RBLU}} \leftarrow \mathcal{L}_{\text{reg}} + \gamma \mathcal{L}_{\text{BCLU}}$; /*见式(15)*/

⑰ end for

3 实验与分析

3.1 实验方案

本文在 DIR 标准数据集上验证 RBLU 的有效性, 包括年龄估计 AgeDB-DIR, IMDB-WIKI-DIR 和深度(depth)估计 NYUD2-DIR, 详细信息如表 1 所示。

Table 1 Detailed Information of Datasets

表 1 数据集详细信息

数据集	标签段长度	标签段样本数	标签范围	训练集样本数	验证集样本数	测试集样本数
AgeDB-DIR	1 岁	1~353	0~101	12 208①	2 140①	2 140①
IMDB-WIKI-DIR	1 岁	1~7 149	0~186	191 509①	11 022①	11 022①
NYUD2-DIR	0.1 m	1.13×10^6~1.46×10^8	0.7~10	3.51×10^{9}②		8.70×10^{5}②

注: ①表示图片数量; ②表示像素点数量。

为全面评估 RBLU 的性能, 选取多种算法进行对比, 包括回归型方法: 传统回归(vanilla)、逆频率加权(SQINV)、标签(LDS)[7]和特征平滑(FDS)[7]、排序(RankSim)[9]、SOTA 方法(ConR)[10]和最新的 SRL[38]。混合型方法: 添加分类损失(CLS)[11]、有序熵(ORDINAL)[12]、层次分类(HCA-d)[13]等。

本文评估方式与文献 [7] 相同, 按各标签段内样本数将评估方式划分为 few-shot、medium-shot、many-shot 区域及 all, 其中年龄估计任务各区域样本数分别为 [0, 20], [20, 100], [100, +∞), 选取平均绝对误差(*MAE*)与几何均值(*GM*)作为评估指标; 深度估计任务各区域像素数为 [0, 1.0×10^7], [1.0×10^7, 2.6×10^7], [2.6×10^7, +∞), 选择均方根误差(*RMSE*)与阈值准确率(δ_1)作为评价指标, 定义如下:

1)平均绝对误差 $MAE = \dfrac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$;

2)几何均值 $GM = \left(\prod_{i=1}^{N}|y_i - \hat{y}_i|\right)^{\frac{1}{N}}$;

3)均方根误差 $RMSE = \sqrt{\dfrac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$;

4)准确率 $\delta_i = \max\left(\dfrac{d_i}{g_i}, \dfrac{g_i}{d_i}\right) < 1.25^i (i = 1,2,3)$。

实验基本配置与文献 [10] 保持一致, 对于年龄估计任务采用 ResNet-50 作为编码器 $f_\theta(\cdot)$ 以及 1 层全连接层作为预测器 $g_\varphi(\cdot)$; 训练 90 轮, 批次大小为 64; 初始学习率为 2.5×10^{-4}, 按照固定系数 0.1 在第 60 轮和 80 轮阶梯式衰减[3]; 阈值 w 为 1 岁, 超参数 γ=8.0, τ=0.2。对于深度估计任务, 输出维度是 114×152, 训练 20 轮, 批次大小为 32; 初始学习率 1.0×10^{-4}, 按照固定系数 0.1 每 5 轮阶梯式衰减; 相似性度量方式与文献 [10] 相同, w=0.1 m, 超参数 γ=0.2, τ=0.7。2 类任务均使用 Adam 优化器, *momentum*=0.9, 权重衰减率是 1.0×10^{-4}。

3.2 实验结果

本节给出各算法在不同数据集的评估结果，由表 2 得出，RBLU 在所有区域的 *MAE* 与 *GM* 指标均最优，相较于 RankSim, few-shot 的 *GM* 降低 15.82%；与 SOTA 方法 ConR 相比，*MAE* 降低 5.42%，*GM* 降低 8.93%，few-shot 的 *MAE* 降低 9.23%。有序熵方法未考虑到不平衡，效果表现一般。

表 3 展示 RBLU 在 IMDB-WIKI-DIR 数据集上的评估结果，除 few-shot 的 *MAE* 与 SOTA 相差 0.01，RBLU 在其他指标最优，其中 medium-shot 和 few-shot 的 *GM* 降低 5% 和 13%，平均降低 2.85%。如表 4 所示，在 NYUD2-DIR 数据集上的实验保持与文献 [11] 设置一致，选择文献 [8] 中最好的结果（GAI 模型），vanilla 在 many-shot 的 *RMSE* 始终占优，但其完全忽略了 medium-shot 和 few-shot。RBLU 在其他区域达到理想结果，较 BMSE 的 few-shot，*RMSE* 降低 8.8%，相较于 ConR 平均降低 2.8%。以往表现较好的 RankSim 无法应对高维数据，RBLU 并没有此种限制。

以上结果表明均匀假设下的 DIR 方法并非最优，RBLU 凭锚定边界的划分方式可有效缓解该困境。层次分类调整方法 HCA-d 强制新的标签段样本数平衡，会破坏有序信息并加剧特征重叠，如图 4 所示，RBLU 利用楔形特征缓解特征重叠的同时能保持一定的有序性。

3.3 RBLU 减少特征重叠

利用训练完成的模型提取 AgeDB-DIR 平衡测试集的特征，其中 LDS, RankSim, SRL 模型在官方代码库下载，由于 ConR 官方未提供模型，此处及后文使

Table 2 Performance Comparison of Each Algorithm on AgeDB-DIR Dataset

表 2 AgeDB-DIR 数据集上各算法性能比较

算法	*MAE*(↓)				*GM*(↓)			
	all	many-shot	medium-shot	few-shot	all	many-shot	medium-shot	few-shot
vanilla (ICML2021)	7.77	6.62	9.55	13.67	5.05	4.23	7.01	10.75
SQINV (ICML2021)	7.81	7.16	8.80	11.20	4.99	4.57	5.73	7.77
LDS (ICML2021)	7.42	6.83	8.21	10.79	4.85	4.39	5.80	7.03
FDS (ICML2021)	7.55	6.99	8.4	10.48	4.82	4.49	5.47	6.58
RankSim (ICML2022)	6.91	6.34	7.79	9.89	4.28	3.92	4.88	6.89
ORDINAL (ICLR2023)	7.46	6.73	8.18	12.38	4.72	4.21	5.36	9.70
HCA-d (CVPR2024)	7.45	6.86	8.22	10.90				
ConR (ICLR2024)	7.20	6.50	8.04	9.73	4.59	3.94	4.83	6.39
SRL (CVPR2025)	7.22	6.64	8.28	9.81	4.50	4.12	5.37	6.29
RBLU (本文)	**6.81**	**6.26**	**7.69**	**9.57**	**4.18**	**3.88**	**4.82**	**5.80**

注：每列中最优的结果以黑体数值突出标注，次优结果以下划线数值突出标注。“↓”表示指标值越小，性能越优。

Table 3 Performance Comparison of Each Algorithm on IMDB-WIKI-DIR Dataset

表 3 IMDB-WIKI-DIR 数据集上各算法性能比较

算法	*MAE*(↓)				*GM*(↓)			
	all	many-shot	medium-shot	few-shot	all	many-shot	medium-shot	few-shot
vanilla (ICML2021)	8.06	7.23	15.12	26.33	4.57	4.17	10.59	20.46
SQINV (ICML2021)	7.87	7.24	12.44	22.76	4.47	4.22	7.25	15.10
LDS (ICML2021)	7.83	7.31	12.43	22.51	4.42	4.19	7.00	13.94
FDS (ICML2021)	7.83	7.23	12.60	22.37	4.42	4.20	6.93	13.48
RankSim (ICML2022)	7.42	6.84	12.12	**22.13**	4.10	3.87	6.74	12.78
BMSE (CVPR2022)	8.08	7.52	12.47	23.29				
HCA-d (CVPR2024)	7.54	6.91	12.69	22.96				
ConR (ICLR2024)	7.33	6.75	11.99	22.22	4.02	3.79	6.98	12.95
SRL (CVPR2025)	7.69	7.08	12.65	22.78	4.28	4.03	7.28	15.25
RBLU (本文)	**7.28**	**6.71**	**11.85**	22.23	**3.97**	**3.75**	**6.63**	**11.26**

注：每列中最优的结果以黑体数值突出标注，次优的结果以下划线数值突出标注。“↓”表示指标值越小，性能越优。

Table 4 Performance Comparison of Each Algorithm on NYUD2-DIR Dataset

表 4 NYUD2-DIR 数据集各算法性能比较

算法	*RMSE*(↓)				δ_1(↑)			
	all	many-shot	medium-shot	few-shot	all	many-shot	medium-shot	few-shot
vanilla (ICML2021)	1.477	**0.591**	0.952	2.123	0.677	**0.777**	0.693	0.570
LDS (ICML2021)	1.387	0.671	0.913	1.954	0.672	0.701	0.706	0.630
FDS (ICML2021)	1.442	0.615	0.940	2.059	0.681	0.760	0.695	0.596
RankSim (ICML2022)								
BMSE(GAI)(CVPR2022)	1.251	0.692	0.959	1.703	0.702	0.676	0.734	0.715
CLS (ICCV2023)	1.454	0.607	0.965	2.077				
HCA-d (CVPR2024)	1.475				0.689			
ConR (ICLR2024)	1.304	0.682	0.889	1.885	0.675	0.699	0.753	0.648
RBLU (本文)	**1.245**	0.764	**0.828**	**1.688**	**0.706**	0.676	**0.757**	**0.716**

注：每列中最优的结果以黑体数值突出标注，次优的结果以下划线数值突出标注。“↓”表示指标值越小，性能越优；“↑”表示指标越大，性能越优。

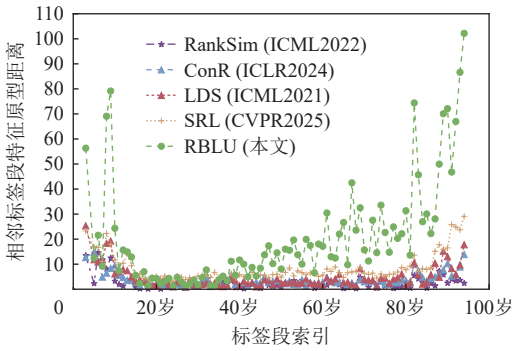


Fig. 4 Prototype distance of encoded features in balanced test sets on AgeDB-DIR

图 4 在 AgeDB-DIR 上的平衡测试集编码特征原型距离

用的 ConR 模型和结果由我们按照官方提供的代码结合论文的参数复现得到，实验环境与 RBLU 一致。

如图 4 所示，将相邻标签段样本特征的原型采用欧式距离度量，在区间 [20, 35] 内 RBLU 优势不显著，其他区域均可将邻近标签段的特征有效区分，而在区间 [20, 35] 的实验结果符合人类面部变化的客观事实，即大部分人在 20~35 岁样貌本就变化细微。以往方法未专门针对回归任务的特征重叠设计，并立足均匀假设导致重叠严重。RBLU 无需对标签段的数据变化显式建模就可适应数据真实的变化规律以及减轻特征重叠。

3.4 消融实验

本节首先在 AgeDB-DIR 数据集上进行 RBLU 中 BCL 模块与 BMIU 模块的消融实验。选取最先 ConR 作为基线(baseline)。bilateral 表示应用 BCL 划分左右子集，BMIU 表示直接(不区分左右)使用 BMIU 增广的楔形特征。表 5 展示消融实验结果，BCL 在 few-shot 的 *MAE* 和 *GM* 获得显著提升，表明 BCL 在应对均匀假设的同时更能关注到 few-shot 样本；将 BMIU 直接用于 baseline 在 few-shot 上的性能提升稍逊，原因是模型受均匀假设误导，many-shot 上的性能提升显著的原因是该区域含有大量样本且实际变化平缓导致重叠严重，在楔形特征引导后效果提升显著。最后融合 BCL 和 BMIU 的 RBLU 达到了最优性能。

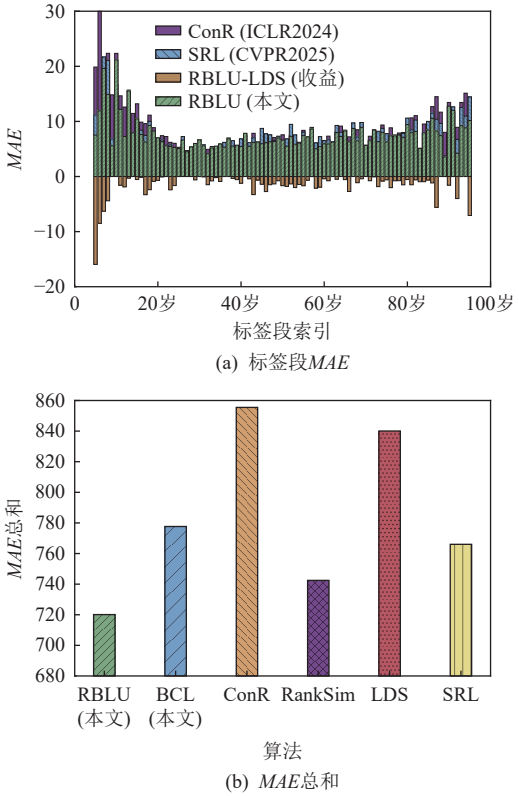
图 5(a)展示 AgeDB-DIR 数据集中各标签段内样本的预测值与真实值之间的 *MAE*，与 ConR 和 SRL 相比，RBLU 在 few-shot 区域优势明显；与 LDS 相比，RBLU 在其他区域表现出不同程度的增益，表明 RBLU 并未忽略任何标签段样本。图 5(b)展示各标签段 *MAE* 求和结果，体现各标签段对总体误差贡献

Table 5 Ablation Study on AgeDB-DIR Dataset

表 5 AgeDB-DIR 数据集上的消融实验

算法			*MAE*(↓)				*GM*(↓)			
baseline	bilateral	BMIU	all	many-shot	medium-shot	few-shot	all	many-shot	medium-shot	few-shot
√			7.59	6.72	8.56	13.07	4.97	4.41	5.71	10.29
√		√	7.18	6.29	8.33	12.41	4.60	4.00	5.69	9.48
√	√		7.10	6.58	8.01	**9.54**	4.54	4.2	5.29	6.17
√	√	√	**6.81**	**6.26**	**7.69**	9.57	**4.18**	**3.88**	**4.82**	**5.80**

注：每列中最优结果以黑体数值突出标注，次优结果以下划线数值突出标注。“↓”表示指标值越小，性能越优，“√”表示使用了该模块。



(a) 标签段*MAE*

(b) *MAE*总和

Fig. 5 Comparison of errors among different algorithms in balanced test sets on AgeDB-DIR

图 5 AgeDB-DIR 平衡测试集上不同算法误差对比

均等, RBLU 的总体误差最小。

与文献 [8] 保持一致, 对 AgeDB-DIR 分布平衡的测试集(年龄从 3~95 岁)特征使用 t-SNE 可视化展示, 如图 6 所示。BCL(图 6(b))相对于 baseline(图 6(a))显著扩展了中间样本段的特征空间, 该区域在均匀假设下难以区分。图 6(c)表明 RBLU(在 BCL 基础上融合 BMIU 模块)编码的特征更加紧致, 表现出有序性, 验证了 BMIU 可作为副产品, 无需增加额外的排序计算开销就能得到有序特征。



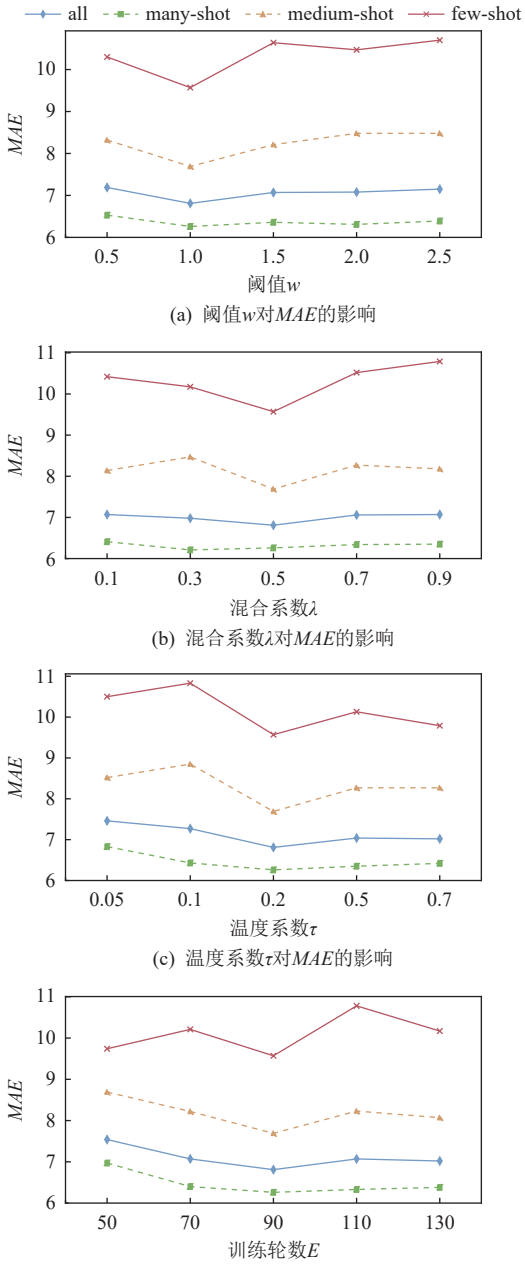
(a) baseline (b) BCL(w.t.BMIU) (c) RBLU(w.BMIU)

Fig. 6 t-SNE visualization of features in test sets on AgeDB-DIR

图 6 在 AgeDB-DIR 上的测试集特征 t-SNE 可视化

3.5 参数敏感性分析

本节以 AgeDB-DIR 数据集为例, 对文中的阈值、混合系数、温度系数以及训练轮数进行参数敏感性分析。如图 7(a), 实验表明当阈值为 1 岁时模型取得最优性能, 随阈值增加, 各区域的误差呈现不同程



(a) 阈值w对*MAE*的影响

(b) 混合系数λ对*MAE*的影响

(c) 温度系数τ对*MAE*的影响

(d) 训练轮数E对*MAE*的影响

Fig. 7 Parameter sensitivity analysis on AgeDB-DIR

图 7 在 AgeDB-DIR 上的参数敏感性分析

度的上升趋势。较大的阈值会扩大正样本的匹配空间, 迫使模型维持多个邻近标签段的特征紧致, 加剧特征重叠。BMIU 中的混合系数 $\lambda \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, 当均匀混合构造 Universum 时, 楔形特征与原始特征的距离最远, 模型达到最佳性能。这与 Han 等人[16] 理论分析得出的结论一致, 即当 λ=0.5 时, BMIU 诱导的 Universum 可较好地描述总体负例。温度系数 $\tau \in \{0.05, 0.1, 0.2, 0.5, 0.7\}$, 如图 7(c)所示, 当 τ=0.2 时模型效果最好。对于训练轮数 E, 随着 E 从 50 增加到 90, *MAE* 呈现下降趋势, 说明模型在不断学习,

性能逐步提升，当 E 超过 90 后，*MAE* 开始上升，表明模型出现了过拟合。总体上，few-shot 指标对参数敏感度高于 all 和 many-shot。

4 结　论

本文首先提出了 BCL 模块，无需对样本语义变化规律显式建模即可克服对比学习构造弱相似正对而强相似负对的困境。此外，本文针对标签值连续的回归任务设计了 BMIU 模块，BMIU 不仅能区分相邻标签段的特征原型和减轻特征重叠，而且可作为副产品避免施加有序约束就能自然地保持有序。最后结合传统回归发展出 RBLU 算法，并在 DIR 标准数据集上验证了 RBLU 可以应对连续数据不平衡以及不均匀变化的挑战。未来的工作考虑将 RBLU 应用到更具挑战性的场景，如：1)面向部分标签缺失的半监督回归任务；2)处理连续与离散标签共存的多输出任务；3)应对连续标签段大规模数据缺失的开集回归场景；4)适应标签分布漂移的在线对比学习框架；5)结合大模型实现跨域下游任务的知识迁移与范式创新等。

作者贡献声明：刘文杰负责提出算法模型，制定实验方案并完成实验，撰写和修改论文；陈松灿提出研究题目，讨论和修改论文。

参　考　文　献

[1] Shah D, Xue Ziyu, Aamodt T M. Label encoding for regression networks[C/OL]//Proc of the 10th Int Conf on Learning Representations. 2022[2024-02-02]. https://openreview.net/forum?id=8WawVDdKqlL

[2] Cao Meng, Tian Qing, Ma Tinghuai, et al. Human facial attributes estimation: A survey[J]. Journal of Software, 2019, 30(7): 2188−2207 (in Chinese)
(曹猛, 田青, 马廷淮, 等. 人类面部属性估计研究: 综述[J]. 软件学报, 2019, 30(7): 2188−2207)

[3] Jing Lei, Max G, Alessandro R, et al. Distribution-free prediction bands for non-parametric regression[J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2014, 76(1): 71−96

[4] Wen Xin, Li Biying, Guo Haiyun, et al. Adaptive variance based label distribution learning for facial age estimation[C]//Proc of the 16th European Conf on Computer Vision. Berlin: Springer, 12368: 379–395

[5] Branco P, Torgo L, Ribeiro R P. SMOGN: A pre-processing approach for imbalanced regression[C]//Proc of the 1st Int Workshop on Learning with Imbalanced Domains: Theory and Applications. New York: PMLR, 2017: 36−50

[6] Song Xinyue, Dao N, Branco P. DistSMOGN: Distributed SMOGN for imbalanced regression problems[C]//Proc of the 4th Int Workshop on Learning with Imbalanced Domains: Theory and Applications. New York: PMLR, 2022: 38−52

[7] Yang Yuzhe, Zha Kaiwen, Chen Yingcong, et al. Delving into deep imbalanced regression[C]//Proc of the 38th Int Conf on Machine Learning. New York: ACM, 2021: 11842−11851

[8] Ren Jiawei, Zhang Mingyuan, Yu Cunjun, et al. Balanced MSE for imbalanced visual regression[C]//Proc of the 40th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 7916–7925

[9] Gong Yu, Mori G, Tung F. RankSim: Ranking similarity regularization for deep imbalanced regressionn[C]//Proc of the 39th Int Conf on Machine Learning. New York: ACM, 2022: 7634−7649

[10] Keramati M, Meng Lili, Evans R D. ConR: Contrastive regularizer for deep imbalanced regression[C/OL]//Proc of the 12th Int Conf on Learning Representations. 2024[2024-09-09]. https://openreview.net/forum?id=RIuevDSK5V

[11] Pintea S L, Lin Yancong, Dijkstra J, et al. A step towards understanding why classification helps regression[C]//Proc of the 19th IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2023: 19972−19981

[12] Zhang Shihao, Yang Linlin, Mi M B, et al. Improving deep regression with ordinal entropy[C/OL]//Proc of the 11th Int Conf on Learning Representations. 2023[2024-02-02]. https://openreview.net/forum?id=raU07GpP0P

[13] Xiong Haipeng, Yao A. Deep imbalanced regression via hierarchical classification adjustment[C]//Proc of the 41st IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2024: 23721−23730

[14] Geng Xin, Wang Qin, Xia Yu. Facial age estimation by adaptive label distribution learning[C]//Proc of the 22nd Int Conf on Pattern Recognition. Piscataway, NJ: IEEE, 2014: 4465–4470

[15] Tian Qing, Cao Meng, Chen Songcan, et al. Relationships self-learning based gender-aware age estimation[J]. Neural Processing Letters, 2019, 50(3): 2141−2160

[16] Han Aiyang, Geng Chuanxing, Chen Songcan. Universum-inspired supervised contrastive learning[J]. IEEE Transactions on Image Processing, 2023, 32: 4275−4286

[17] Vapnik V. Estimation of Dependences Based on Empirical Data[M]. Berlin: Springer, 2006: 1–26

[18] Weston J, Collobert R, Sinz F, et al. Inference with the universum[C]//Proc of the 23rd Int Conf on Machine Learning. New York: ACM, 2006: 1009–1016

[19] Dhar S, Cherkassky V. Universum learning for SVM regression [C]//Proc of the 30th Int Joint Conf on Neural Networks. Piscataway, NJ: IEEE, 2017: 2161−4407

[20] Chen Ke, Gong Shaogang, Xiang Tao, et al. Cumulative attribute space for age and crowd density estimation[C]//Proc of the 26th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2013: 2467–2474

[21] Zha Kaiwen, Cao Peng, Son J, et al. Rank-N-Contrast: Learning continuous representations for regression[C]//Proc of the 37th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2023: 17882–17903

[22] Torgo L, Ribeiro R P, Pfahringer B, et al. SMOTE for Regression[M]. Berlin: Springer, 2013: 378–389

[23] Branco P, Torgo L, Ribeiro R P. REBAGG: Resampled bagging for imbalanced regression[C]//Proc of the 2nd Int Workshop on Learning with Imbalanced Domains: Theory and Applications. New York: PMLR, 2018: 67–81

[24] Steininger M, Kobs K, Davidson P, et al. Density-based weighting for imbalanced regression[J]. Machine Learning, 2021, 110(8): 2187–2211

[25] Liu Gang, Zhao Tong, Inae E, et al. Semi-supervised graph imbalanced regression[C]//Proc of the 29th ACM SIGKDD Conf Knowledge Discovery and Data Mining. New York: ACM, 2023: 1453–1465

[26] Moukpe J. Survey on imbalanced data, representation learning and sep forecasting[J]. arXiv preprint, arXiv: 2310.07598, 2023

[27] Nie G, Zhao Qinghao, Tang Gongzheng, et al. Deep imbalanced regression to estimate vascular age from PPG data: A novel digital biomarker for cardiovascular health[J]. arXiv preprint, arXiv: 2406.14953, 2024

[28] He Kaiming, Fan Haoqi, Wu Yuxin, et al. Momentum contrast for unsupervised visual representation learning[C]//Proc of the 37th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 9726–9735

[29] Chen Ting, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations[C]//Proc of the 37th Int Conf on Machine Learning. New York: ACM, 2020: 1597–1607

[30] Khosla P, Teterwak P, Wang Chen, et al. Supervised contrastive learning[C]//Proc of the 34th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2020: 18661–18673

[31] Kang Bingyi, Li Yu, Xie Sa, et al. Exploring balanced feature spaces for representation learning[C/OL]//Proc of the 9th Int Conf on Learning Representations. 2021[2024-12-09]. https://openreview.net/forum?id=OqtLIabPTit

[32] Misra I, Van Der Maaten L. Self-supervised learning of pretext-invariant representations[C]//Proc of the 37th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 6706–6716

[33] Logeswaran L, Lee H. An efficient framework for learning sentence representations[C/OL]//Proc of the 6th Int Conf on Learning Representations. 2018[2025-01-09]. https://openreview.net/forum?id=rJvJXZb0W

[34] Zhao Yuhan, Chen Songcan. Multi-view clustering algorithm via eliminating the complementarity controversy[J]. Journal of Computer Research and Development, 2025, 62(5): 1216−1225 (in Chinese)
(赵玉涵, 陈松灿. 消除互补性争议的多视图聚类算法[J]. 计算机研究与发展, 2025, 62(5): 1216−1225)

[35] Cui Jiequan, Zhong Zhisheng, Liu Shu, et al. Parametric contrastive learning[C]//Proc of the 17th IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2021: 695–704

[36] Zhu Jianggang, Wang Zheng, Chen Jingjing, et al. Balanced contrastive learning for long-tailed visual recognition[C]//Proc of the 39th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 6898–6907

[37] Du Chaoqun, Wang Yulin, Song Shiji, et al. Probabilistic contrastive learning for long-tailed visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(9): 5890−5904

[38] Dong Zijian, Wu Yilei, Chen Chongyao, et al. Improve representation for imbalanced regression through geometric constraints[J]. arXiv preprint, arXiv: 2503.00876, 2025

[39] Zhang Hongyi, Cisse M, Dauphin Y N, et al. Mixup: Beyond empirical risk minimization[C/OL]//Proc of the 6th Int Conf on Learning Representations. 2018[2024-07-16]. https://openreview.net/forum?id=r1Ddp1-Rb

[40] Yao Huaxiu, Wang Yiping, Zhang Linjun, et al. C-mixup: Improving generalization in regression[C]//Proc of the 36th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT, 2022: 3361–3376

[41] Zheng Jiyang, Yao Yu, Han Bo, et al. Enhancing contrastive learning for ordinal regression via ordinal content preserved data augmentation[C/OL]//Proc of the 12th Int Conf on Learning Representations. 2024[2024-09-04]. https://openreview.net/forum?id=kx2XZlmgB1



Liu Wenjie, born in 2001. Master candidate. Student member of CCF. His main research interests include machine learning and pattern recognition.

刘文杰, 2001 年生。硕士研究生。CCF 学生会员。主要研究方向为机器学习、模式识别。



Chen Songcan, born in 1962. PhD, professor, PhD supervisor. Senior member of CCF. His main research interests include pattern recognition, machine learning, and neural computing.

陈松灿, 1962 年生。博士, 教授, 博士生导师。CCF 高级会员。主要研究方向为模式识别、机器学习、神经计算。