面向深度学习评估过程的边界自适应插值策略研究

刘存思[1]，王秋月[2]
1. 飞腾信息技术有限公司，湖南 长沙 410000
2. 中电信数智科技有限公司，湖南 长沙 410011
DOI: 10.61369/TACS.2026070002

摘　　要： 在深度学习模型训练和生成过程中，通常每一轮训练后都会进行模型评估，人脸识别模型训练的早期轮次为例，模型评估过程有概率会因插值边界处理不当而训练程序崩溃，导致训练鲁棒性不够。本文深入剖析了该类错误的产生机理，指出其根源在于训练初期模型性能不足，导致预设的评估操作点超出插值函数的定义域。并且针对此问题，本文提出了一种边界自适应插值策略（Boundary-Adaptive Interpolation Strategy, BAIS）。该策略通过动态感知模型性能边界与评估目标的相对位置，采用四阶段决策机制，将传统的被动错误捕获转变为主动的边界预判与处理。实验表明，该策略在模型性能极差或极优等边界场景下均能保持零错误率，较传统 try-except 方法在评估准确性和代码健壮性上均有显著提升。本研究为构建高可靠、全自动的深度学习训练与评估流程提供了重要的理论依据和技术支撑。

关 键 词： 插值错误；模型评估；训练稳定性；边界自适应；深度学习；人脸识别

Research on Boundary-Adaptive Interpolation Strategy for Deep Learning Evaluation Process

Liu Cunsi[1], Wang Qiuyue [2]
1.Feiteng Information Technology Co., Ltd., Changsha, Hunan 410000
2.China Telecom Digital Intelligence Technology Co., Ltd., Changsha, Hunan 410011

Abstract: In the process of deep learning model training and generation, model evaluation is typically performed after each round of training. Taking the early epoch of face recognition model training as an example, the model evaluation process has a probability of causing the training program to crash due to improper handling of interpolation boundaries, resulting in insufficient training robustness. This paper deeply analyzes the mechanism behind such errors, identifying the root cause as insufficient model performance during the early training stage, which leads to preset evaluation operation points falling outside the domain of the interpolation function. To address this issue, this paper proposes a Boundary-Adaptive Interpolation Strategy (BAIS). By dynamically sensing the relative position between model performance boundaries and evaluation targets, this strategy employs a four-stage decision-making mechanism, replacing traditional passive error capture with active boundary prediction and handling. Experiments demonstrate that the proposed strategy maintains a zero error rate in boundary scenarios where model performance is extremely poor or excellent. Compared with the traditional try-except method, it shows significant improvements in both evaluation accuracy and code robustness. This study provides an important theoretical basis and technical support for constructing highly reliable and fully automated deep learning training and evaluation pipelines.

Keywords: **interpolation error; model evaluation; training stability; boundary-adaptive; deep learning; face recognition**

引言

深度学习模型的性能评估是衡量其泛化能力、指导模型优化的核心环节。在人脸识别、身份验证等应用中，通常采用在特定假阳性率下的真阳性率作为关键性能指标[1]。例如，在 LFW（Labeled Faces in the Wild）等标准评测中，研究者普遍关注当 FAR 被严格限制在0.1%、0.01% 等“操作点”时，模型所能达到的 TPR[2]。这种“固定 FAR，求 TPR”的评估范式，确保了模型性能在统一、有实

作者简介：
刘存思（1982.12—），男，汉族，湖南怀化人，单位：飞腾信息技术有限公司，职称：高级工程师（计算机），学历：硕士，研究方向：云计算、人工智能、国产芯片；
王秋月（1997.08—），女，汉族，河南邓州人，单位：中电信数智科技有限公司，职称：信息系统项目管理师（高级），学历：2021 级学士，研究方向：云计算、人工智能。

Copyright © This Work is Licensed under A Commons Attribution-Non Commercial 4.0 International License.

际业务意义的标准下进行公平比较。

然而，在实际工程实践中，尤其是在模型训练的初始阶段，一个普遍存在但未被充分重视的问题是：评估代码会因插值错误而崩溃。错误信息通常表现为 ValueError: A value ... is below the interpolation range's minimum value。这不仅中断了训练流程，更严重的是，它阻碍了长周期、无人值守训练的实现。国内相关学者在研究大规模人脸识别训练时也指出，训练流程的鲁棒性，特别是评估环节的稳定性，是模型能否成功收敛的关键因素之一[3]。

本文的贡献在于：

1. 系统化分析：从模型性能分布、评估指标计算和插值工具特性三个层面，系统揭示了插值错误的产生原理。

2. 提出 BAIS 策略：设计并实现了一种边界自适应插值策略，从根本上解决了该问题，相比被动容错方法具有更高的评估准确性和代码健壮性。

3. 提供完整解决方案：结合 PyTorch 分布式训练中的其他常见问题（如 autocast 版本兼容性、DDP 参数警告），提供了一套完整的、经过实践检验的训练稳定性优化方案。

一、相关工作

（一）深度学习模型评估指标

模型评估是机器学习流程的核心。对于二分类问题，ROC 曲线和 AUC 是常用指标。但在人脸验证等任务中，FAR 和 TPR 构成的“操作点”更具实际意义[4]。TPR（真阳性率）衡量模型正确识别正样本的能力，而 FAR（假阳性率）则衡量模型将负样本误判为正样本的概率，直接关联系统的安全性[5]。部分研究强调，例如在安防领域，FAR 是首要约束条件，所有性能优化都必须在满足 FAR 阈值的前提下进行[6]。

（二）插值技术在评估中的应用

为了精确获得特定 FAR 下的 TPR，通常需要使用插值技术。Scipy 库的 interp1d 函数是常用工具，它能在已知数据点之间进行线性或样条插值[7]。然而，interp1d 默认不支持外推，即当查询点超出已知数据范围时会抛出异常。相关分析在研究时间序列预测模型时也提到，评估工具对边界条件的处理能力是衡量其鲁棒性的重要标准，但现有研究较少关注训练初期模型性能不足导致的评估失效问题[8]。

（三）深度学习训练稳定性

训练稳定性研究多集中于梯度消失 / 爆炸、学习率调度等方面[9]。对于评估阶段的稳定性，尤其是由模型性能动态变化引发的评估逻辑错误，相关研究和在实际代码中考虑的尚不充分，因此本文聚焦于此。

二、插值错误产生原理分析

（一）评估流程与插值逻辑

标准的评估流程（如 calculate_val 函数）通常包含以下步骤：

1. 生成候选阈值：在所有样本对距离的范围内，生成一组密集的阈值 $T = \{t_1, t_2, \ldots, t_n\}$。

2. 计算 FAR/TPR 序列：对每个阈值 t_i，在验证集的一个子集上计算其对应的 FAR_i 和 TPR_i，得到两组序列 $FAR = \{far_1, far_2, \ldots, far_n\}$ 和 $TPR = \{tpr_1, tpr_2, \ldots, tpr_n\}$（TPR 序列后续其他地方使用，比如在 ROC 曲线绘制时，此处插值讨论暂时不相关）。

3. 插值求解：建立从 FAR 到 T 的插值函数 f，然后通过 threshold = f(far_target) 求解目标 far（如 0.001）对应的精确阈值，再用此阈值计算最终的 tpr_{far_target}。

（二）错误根源：模型能力与评估目标的失配

模型此时在评估的时候，往往体现插值报错，其根源在于模型当前性能无法满足预设的评估目标，具体原因分析如下：

1. 训练初期模型特征分布：在训练初期，模型尚未学习到有效的区分性特征。对于“同一个人”（正样本）和“不同人”（负样本）的样本对，其计算出的特征距离高度重叠，没有明显的分界线。

2. 可实现的 FAR 下限：由于距离分布重叠，无论使用何种阈值（比如：人脸欧式距离），仍会有一些负样本对的距离小于该阈值，被错误识别。（即：不同的两个人的人脸在模型评估后距离过小，被认为是同一个人）例如模型训练 epoch 初期，FAR 最小值无论采用怎样的阈值，最小只能达到 0.002。

3. 评估目标的不可达性：当预设的评估目标 far_target（如 0.001）小于 FAR_min（0.002）时，从数学上讲，模型无法实现这个目标。这好比要求一个只能举起 50 公斤的人去举起 100 公斤，是“能力之外”的任务。

4. 插值函数的失效：scipy.interpolate.interp1d 创建的函数 f，其定义域为 [FAR_min, FAR_max]。当查询 f(0.001) 时，0.001 位于定义域之外，scipy 为防止返回不可靠的外推值，选择抛出 ValueError 异常。

结论：该错误并非代码逻辑缺陷，而是评估代码对“模型能力不足”这一边界情况处理不当的体现。

三、对策分析与边界自适应插值策略

（一）现有对策及其局限性

1.try-except 容错法：在插值调用外包裹 try-except 块，捕获异常后返回一个默认值（如 TPR=0）。

・优点：实现简单，能防止训练崩溃。

 Copyright © This Work is Licensed under A Commons Attribution-Non Commercial 4.0 International License.

・缺点：被动处理，丢失了模型在当前能力下的真实性能信息。返回的默认值（如 TPR=0）可能与模型实际表现（如 TPR=0.1）存在偏差，导致评估不准确。

2. 修改评估目标法：临时调高 far_target（如从0.001改为0.01）。

・优点：能快速绕过错误。

・缺点：治标不治本，改变了评估标准，使得结果与学术界的通用报告的方式不可比，失去了评估的意义。

（二）边界自适应插值策略

为克服上述方法的局限性，我们提出边界自适应插值策略，其核心思想是主动预判并智能处理所有边界情况。

其算法流程如下：

输入：阈值列表 thresholds，距离 distances，标签 labels，目标 FAR_{far_target}

输出：校准后的阈值 threshold

1. 数据预处理：

・计算每个候选阈值对应的 FAR 序列 far_train。

? 对 far_train 和 thresholds 进行排序，并去除 far_train 中的重复值，保留对应的最大阈值，得到 unique_far 和 unique_thresholds。

2. 单调性校验：

? 检查 unique_far 是否为严格单调递增序列。若不满足，则无法进行有效插值，执行退化策略（如返回最严格的阈值 np.max(thresholds)）。

3. 边界判定：

・ 计 算 min_far = np.min(unique_far) 和 max_far = np.max(unique_far)。

? 比较 far_target 与 [min_far, max_far] 的相对位置。

4. 自适应选择：

? 情况1 (far_target < min_far)：目标 FAR 低于插值序列可达点下限。此时，选择产生最低 FAR 的阈值，即 threshold = unique_thresholds[0]。

? 情况2 (far_target > max_far)：目标 FAR 高于插值序列可达点。此时，选择产生最高 FAR 的阈值，即 threshold = unique_thresholds[-1]。

? 情况3 (min_far <= far_target <= max_far)：目标 FAR 在插值范围内。执行安全插值，f = interp1d(unique_far, unique_thresholds)，threshold = f(far_target)。

该策略的伪代码如下：

```
# 核心逻辑
if len(unique_far) > 1 and is_monotonic(unique_far):
min_far, max_far = np.min(unique_far), np.max(unique_far)
#min_far 为模型当前所能达到的最低 FAR
if far_target < min_far:
    threshold = unique_thresholds[0]  # 采用左端点值
#max_far 为模型当前所能达到的最高 FAR
elif far_target > max_far:
    threshold = unique_thresholds[-1] # 采用右端点值
else:
    threshold = f(far_target)        # 安全插值区
else:
threshold = np.max(thresholds)       # 退化策略
```

（三）策略优势分析

与 try-except 法相比，BAIS 策略的优势在于：

・主动性与准确性：在插值前预判边界，避免了异常的发生。在边界情况下，返回的是模型当前能力下的“最佳边界阈值”，而非一个固定的、可能失真的默认值，从而保留了评估的有效性。

? 健壮性与完备性：覆盖了模型性能过优、过差和正常三种情况，逻辑完备，适用于训练的任何阶段。

・代码稳定性：从根本上消除了因插值崩溃导致的训练中断，是实现7x24小时无人值守训练的关键保障。

四、实验与结果分析

为验证 BAIS 策略的有效性，我们在 LFW 数据集上，使用 ArcFace 模型进行了对比实验。

（一）实验设置

? 数据集：LFW

・模型：ArcFace (ResNet-50 backbone)

? 对比方案：(1) 原始插值代码；(2)try-except 容错法；(3) BAIS 策略（本文方法）。

? 评估指标：(1) 训练崩溃次数；(2)TPR 估计偏差（与真实值的差距）；(3) 评估耗时。

（二）结果分析

评估指标	原始插值代码	try-except 法	BAIS 策略（本文）
训练崩溃次数	17 (Epoch 1-3)	0	0
TPR 估计偏差	-	0.15	0.02
评估耗时	8.2 ms	9.1 ms	8.5 ms

分析：

1. 稳定性：BAIS 策略与 try-except 法均成功避免了训练崩溃，而原始代码在模型性能不足时频繁失败。

2. 准确性：BAIS 策略的 TPR 估计偏差（0.02）远低于 try-except 法（0.15），证明了其在边界情况下能更准确地反映模型性能。

3. 效率：BAIS 策略的计算开销仅比原始代码增加约3.7%，远低于 try-except 法因异常处理带来的开销。

实验结果表明，BAIS 策略在保证训练稳定性的同时，最大限度地维持了评估的准确性和效率，是综合性能最优的解决方案。

五、结论

本文针对深度学习训练初期模型评估中常见的插值崩溃问

Copyright © This Work is Licensed under A Commons Attribution-Non Commercial 4.0 International License.

题，进行了深入的原理分析，并提出了一种创新的边界自适应插值策略。该策略通过主动感知模型性能边界，智能选择插值或边界值，从根本上解决了因模型能力不足与评估目标失配导致的训练中断问题。实验证明，该策略在稳定性、准确性和效率上均优于传统方法，为构建鲁棒、高效的深度学习训练评估系统提供了有效的技术路径。

未来的工作将探索该策略在其他评估指标（如 ROC-AUC 计算）和多模态任务中的泛化能力，并研究如何将其与自动化超参数调优框架更紧密地结合。

参考文献

[1] Schroff, F., Kalenichenko, D., & Philbin, J.. FaceNet: A Unified Embedding for Face Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[2] Taigman, Y., Yang, M., Ranzato, M., & Wolf, L.. DeepFace: Closing the gap to human-level performance on face verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

[3] 王伟，李明，张华 . 大规模分布式人脸识别训练中的稳定性优化研究 . 自动化学报，46(8), 1603-1612, 2020.

[4] Deng, J., Guo, J., & Zafeiriou, S.. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[5] 刘洋，陈云 . 面向安防应用的人脸识别系统性能评估方法 . 计算机研究与发展，56(4), 745-756, 2019.

[6] 李静，孙强 . 基于操作点约束的深度学习模型评估与选择 . 软件学报，32(5), 1321-1335, 2021.

[7] Virtanen, P., Gommers, R., Oliphant, T. E., et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods, 17(3), 261-272, 2020.

[8] 张涛，赵磊 . 时间序列预测模型评估中的鲁棒性分析 . 计算机学报，41(11), 2421-2433, 2018.

[9] Goodfellow, I., Bengio, Y., & Courville, A.. Deep Learning. MIT Press, 2016.

 Copyright © This Work is Licensed under A Commons Attribution-Non Commercial 4.0 International License.