

DOI: 10.19650/j.cnki.cjsi.J2107390

基于自适应序列罚权深度神经网络的 膝关节关节炎等级评分

刘伟强^{1,2,3}, 罗林开^{1,2,3}, 彭洪^{1,3}, 章其敏⁴, 黄玮^{5,6}

(1. 厦门大学航空航天学院自动化系 厦门 361005; 2. 厦门大学健康医疗大数据国家研究院 厦门 361005;
3. 厦门市大数据智能分析与决策重点实验室 厦门 361005; 4. 武汉大学附属同仁医院 武汉 430000;
5. 华中科技大学同济医学院附属协和医院骨科 武汉 430000; 6. 鹤峰县中心医院 恩施 445800)

摘要: 膝关节关节炎(OA)是老年人活动受限和身体残疾的主要原因之一,对膝关节关节炎的早期发现和干预可以帮助病人减缓OA的恶化。目前膝关节关节炎的早期发现通过X光片进行诊断,参照Kellgren-Lawrence(KL)标准进行评分,但医师的评分相对主观,不同医生存在差异。膝关节关节炎的等级分类是个有序分类问题,序列罚权损失函数将距离真实类别越远的等级赋予了更高的罚权,因此它更适合于膝关节关节炎的等级分类。然而,已有工作中的罚权一旦给定,就不再变化,导致其训练模型常常达不到期望的结果。本文针对序列罚权损失的不足,提出一种自适应序列罚权调整策略,通过对每一个阶段(epoch)得到的混淆矩阵,反向指导惩罚权重进行微调,使得罚权矩阵能够自适应调整。进一步地,本文利用来自骨关节炎倡议组织(OAI)的X射线图像数据,在ResNet, VGG, DenseNet以及Inception等几种经典的CNN模型上验证该方法的性能。实验结果表明在膝关节关节炎KL分级任务上,本文提出的自适应序列罚权调整策略在初始罚权分差较小时,能够有效地提升模型分类精度(AC)与平均绝对误差(MAE)。

关键词: 膝关节关节炎; KL评级; 自适应罚权; 卷积神经网络

中图分类号: TP391.41 TH7 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Grading scoring of knee osteoarthritis based on adaptive ordinal penalty weighted deep neural networks

Liu Weiqiang^{1,2,3}, Luo Linkai^{1,2,3}, Peng Hong^{1,3}, Zhang Qimin⁴, Huang Wei^{5,6}

(1. Department of Automation, School of Aerospace Engineering, Xiamen University, Xiamen 361005, China; 2. National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen 361005, China; 3. Xiamen Key Laboratory of Big Data Intelligent Analysis and Decision, Xiamen 361005, China; 4. Tongren Hospital of Wuhan University, Wuhan 430000, China;
5. Department of Orthopaedics, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430000, China; 6. Hefeng County Central Hospital, Enshi 445800, China)

Abstract: Knee osteoarthritis (OA) is one of the main causes of activity limitation and physical disability in the elderly. Early diagnosis and intervention of knee osteoarthritis can help patients slow down the deterioration of OA. At present, the early diagnosis of knee osteoarthritis is detected by X-rays and scored according to the Kellgren-Lawrence (KL) grade. However, doctors' scores are relatively subjective and vary from doctor to doctor. Grade classification of knee osteoarthritis is a matter of orderly classification. The ordinal penalty loss function assigns higher penalty weights to the classes that are further away from the ground truth, which is more suitable for knee osteoarthritis classification. In existing works, the penalty weights no longer change during training procedure, so the training model often fails to reach the expected results. In this paper, an adaptive ordinal penalty adjustment strategy is proposed to address the shortcomings of the ordinal penalty loss, in which the penalty weights are automatically tuned in reverse according to the confusion matrix obtained at each stage (epoch). Furthermore, the performance of the proposed method is validated on several classical CNN models such as ResNet, VGG, DenseNet and Inception by X-ray image data from Osteoarthritis Initiative (OAI). Experimental results show that the

adaptive ordinal penalty adjustment strategy proposed in this paper can effectively improve the classification accuracy (AC) and mean absolute error (MAE) of the model when the initial weight score difference is small.

Keywords: knee osteoarthritis; Kellgren and Lawrence grading; adaptive penalty weight; convolution neural network

0 引言

骨关节炎 (osteoarthritis, OA) 以关节软骨退变为主要特征, 是一种最为常见的关节炎形式, 也是老年人活动受限和身体残疾的主要因素之一^[1]。影像检查是确定 OA 进展的重要工具, 其中以 X 光片和磁共振成像 (magnetic resonance imaging, MRI) 最为常用。在 OA 问题中, 膝关节 OA 引起的疼痛和其他症状严重影响了老年人的生活质量。但遗憾的是, 没有任何治疗方案可以抑制膝关节 OA 进展的退行性结构变化。然而, 早期的发现和治疗可以帮助老年人延缓 OA 的发展, 提高生活质量。MRI 可以反映膝关节的三维结构。但是, MRI 只有在大型医疗中心才具备, 并且昂贵的检查费用使得 MRI 不适合用于常规膝关节 OA 的诊断。相反, X 光检查具有安全、成本低、普及性广等特点, 一直被作为膝关节 OA 检查的金标准。世界卫生组织于 1961 年使用的 Kellgren-Lawrence (KL) 分级标准是最常用的膝关节 OA 严重程度分级系统^[2]。KL 系统将膝关节 OA 严重程度分为 5 个等级, 分别以 0 级到 4 级来表示。

医师通常会检查扫描的膝关节 X 线图像, 然后在很短的时间内给出膝关节的 KL 等级。诊断的准确性高度依赖于医生的经验和细心。此外, KL 分级的标准相对模糊, 可能的骨质增生和可疑的膝关节间隙狭窄被作为 KL 分级的标准, 不同医生对同一膝关节图片可能会给出不同的分级结果。Culvenor 等^[3]进行的一项研究中, KL 内评分者的可靠性在 0.67~0.73 之间。这种医生分级的低可靠性根源在于, 由于标准的模糊性, 很大的概率将膝关节的 KL 等级误判为其附近的等级。在临床诊断中, 将膝关节的等级误判为附近的等级 (如 1 级至 2 级) 远没有将等级误判为远处的等级 (如 1 级至 4 级) 严重。因此, 将分级准确率作为唯一的评价指标是不够的。受 Niu 等^[4]的工作启发, 平均绝对误差 (mean absolute error, MAE) 可作为衡量膝骨关节炎 KL 等级的另一个指标, 该指标能够更好的反映错判的严重程度。

对于膝骨关节炎 KL 等级的分类, 早在 2009 年, Shamir 等^[5]就提出了一种加权最近邻算法, 但该方法需要人工设计特征, 包括纹理特征、切比雪夫统计量、haralic 特征等。近年来随着深度学习的发展, 已有 ResNet^[6]、VGG^[7]、Inception-V3^[8] 和 DenseNet^[9] 等较为成熟的卷积神经网络 (convolutional neural network, CNN) 分类网络模型。这些模型也已广泛应用到肝癌检测和糖尿

病视网膜病变检测等各种医学分类任务中^[10-13]。相比而言, 深度学习在骨关节炎影像诊断的研究仍处于起步阶段^[14]。Antony 等^[15-17]在 KL 分级这个任务上尝试了基于深度学习的方法, 他们设计了一个新的 CNN 模型, 并优化交叉熵损失和均方误差损失的加权组合, 得到了 63.6% 的识别率^[15]。Corriz 等^[18]提出了一种新颖的端到端的卷积神经网络框架, 利用 X 光图像自动量化膝关节炎的严重程度, 取得了 64.3% 的识别结果。Tiulpin 等^[19]利用深度孪生卷积网络, 对膝骨关节炎严重等级进行评分, 取得了 66.7% 的结果。Zhang 等^[20]提出了一种注意力机制下的 CNN 膝骨关节炎 KL 等级分类模型, 应用 ResNet 首先从 X 光片中提取膝关节特征, 然后与卷积注意力模块提取的特征相结合, 自动进行 KL 等级的预测。

对难分样本施加更大的惩罚, 是机器学习中的一种常见做法, 例如经典的 Adaboost 算法^[21]通过串行训练, 一步步地增大难分样本的惩罚权重, 得到最终的学习器, 以提高学习器的性能。Chen 等^[22]根据增大难分样本惩罚权重的思路, 提出了一个新的序列罚权损失, 通过对距离实际类别越远的等级分配更高的罚分, 使得预测的 KL 评分能够更容易的趋近实际评分, 并且利用不同的 CNN 网络模型验证其有效性。但该方式的不足之处在于, 罚权矩阵无法随着训练过程进行自适应调整。

针对已有工作在膝骨关节炎 KL 等级自动分类上的不足, 本文提出一种序列罚权调整策略, 通过对每一个阶段 (epoch) 得到的混淆矩阵结果, 反向指导惩罚权重的微调, 使得罚权矩阵能够进一步自适应的调整。然后, 我们利用了几类经典的 CNN 模型在膝骨关节炎 KL 分级任务上进行了验证, 实验表明, 所提出的自适应罚权调整策略, 在这些 CNN 模型上都表现出比交叉熵以及固定罚权更好的膝骨关节炎 KL 分级性能, 尤其在 MAE 指标上。

1 材料与方法

1.1 数据集与预处理

本文用于评估膝关节 X 光的图像来自骨关节炎倡议^[22] (osteoarthritis initiative, OAI), 这是一项针对膝关节骨关节炎的多中心、纵向、前瞻性观察研究, 旨在确定 OA 发病和进展的生物标志物。有 4 796 名年龄从 45 岁到 79 岁的参与者参加了这个检测。在我们的研究中, 我们使用了来自 Chen 等^[22]公开的数据集, 相关的数据可以从 Github 的网站上获取, 相关链接为:

<https://github.com/PingjunChen/GradingKneeOA>.

数据集包含 4 130 张 X 光图像,将所有膝关节 X 光图随机的拆分为训练集、验证集和测试集,比例为 7:1:2。拆分的过程保证训练集、验证集、测试集之间等

级分布相对稳定。拆分后,膝关节测试集包含 828 张 X 光图像,其中 0 级图像 639 张,1 级 296 张,2 级 447 张,3 级 223 张,4 级 51 张。各等级的样本和对应标准如图 1 所示。

Kellgren and Lawrence(KL) 等级系统

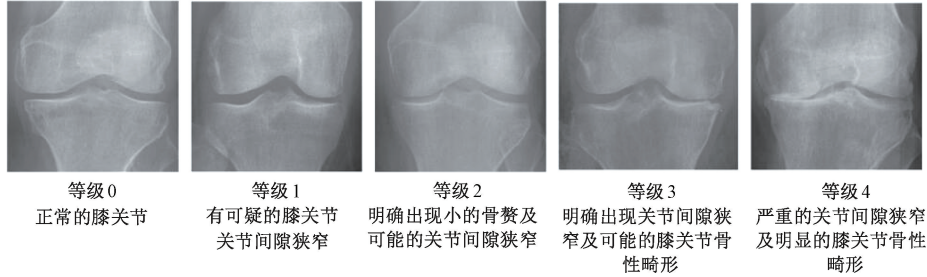


图1 所有 KL 等级的膝关节样本以及对应标准

Fig. 1 Knee joint samples of all KL grades and their corresponding criterion

1.2 序列罚权损失

KL 等级预测任务实质上是一个类别间具有渐进性的问题,不同等级间存在顺序关系。交叉熵损失作为深度学习网络分类模型中的默认损失函数,对所有类别同等对待,没有考虑到不同类别间的序列性。假设 CNN 分类模型中的 softmax 层输出 n 个类别概率为 $[p_1, p_2, \dots, p_n]$ 。在膝关节炎 KL 分级这样的序列分类任务中,给定一张真实等级为 m 的图像,我们期望输出的概率分布应当满足如下两个性质。1) p_m 尽可能的接近 1.0; 2) 对于任意 $k \in \{1, \dots, n\} \setminus \{m\}$, 如果 $|k - m|$ 较大时,那么相应的 p_k 应该更小。传统的交叉熵损失只满足第一条性质。因此,Chen 等^[8]提出了一种新的序列损失来满足这两个属性。

Chen 等^[22]设计了一个序列罚权矩阵 $\mathbf{W}$ 来表示预测等级与真实等级之间的惩罚程度。 $\mathbf{W}$ 是一个 $n \times n$ 的方阵,其中 $W_{ij} \in \mathbf{W}$ 代表预测等级 j 对真实等级 i 的惩罚权重,这里 $i, j \in \{1, \dots, n\}$, 本文膝关节炎 KL 分级任务当中, $n = 5$ 。图 2 左下角的矩阵是序列罚权矩阵的一个例子。在本设计中, $\mathbf{W}$ 的第 i 行代表真实等级 i 的惩罚向量,容易得到的是,每个等级对自身的惩罚权重为 1, 如果离真实类别等级较远的等级,其对等级的惩罚权重应当较高。根据这个序列矩阵和 softmax 层的输出概率,得出的序列损失可以定义为如下形式:

$$loss = \sum_{j=1}^n W_{ij} \cdot q_j \quad (1)$$

其中 i 是输入图像的真实等级,而 q_j 具有如下的定义:

$$q_j = \begin{cases} p_j, & j \neq i \\ 1 - p_j, & j = i \end{cases} \quad (2)$$

为了简化公式(1),使得模型能够按照 softmax 层输

出,重新对 $\mathbf{W}$ 进行修正,记为 $\mathbf{W}'$, 其定义如下:

$$W'_{ij} = \begin{cases} 0, & j = m \\ W_{ij} + 1, & \text{其他} \end{cases} \quad (3)$$

此时公式(1)等价于如下形式:

$$loss = \sum_{j=1}^n W'_{ij} \cdot p_j \quad (4)$$

在公式(4)中,可以直接使用 softmax 层输出的概率计算损失误差。图 2 展示了序列罚权损失的计算过程以及交叉熵损失的情况。由于惩罚矩阵 $\mathbf{W}$ 中的非对角线元素可以设定权重,故将惩罚矩阵 $\mathbf{W}$ 命名为序列罚权矩阵,损失命名为序列罚权损失。

1.3 自适应序列罚权矩阵调整策略

从上节可以看出,在 KL 等级预测任务中,序列罚权损失可以很好的弥补交叉熵损失的不足,通过加大惩罚权重的策略来得到更为理想的分类结果,这种思路既自然又合理。但存在的不足是,在网络的训练过程当中,序列罚权矩阵始终保持不变,有待进一步改进。序列罚权矩阵提出的目的,是为了让距离真实类别等级较远的等级误判较少,如图 2 左下角所示,其中 W_{ij} 表示将真实标签为 i 预测为 j 的惩罚权重。当真实标签是等级 1 时,我们希望将真实标签为 3 判错为 1 的个数远小于真实标签为 0 和 2 判错为 1 的个数,这不是固定序列罚权矩阵能做到的,因此有必要进一步对序列罚权矩阵做出自适应的调整,得到对应的调整策略。

混淆矩阵恰好可以反映权重调整后的预测情况,按照上面的描述,我们希望通过序列罚权矩阵得到更为有效的预测,这一点可以体现在混淆矩阵沿着对角线元素的左右方向呈现出单调递减的趋势。基于这一点考虑,我们提出了一种新的训练策略。

由于深度网络在一个 epoch 正好训练一遍所有的

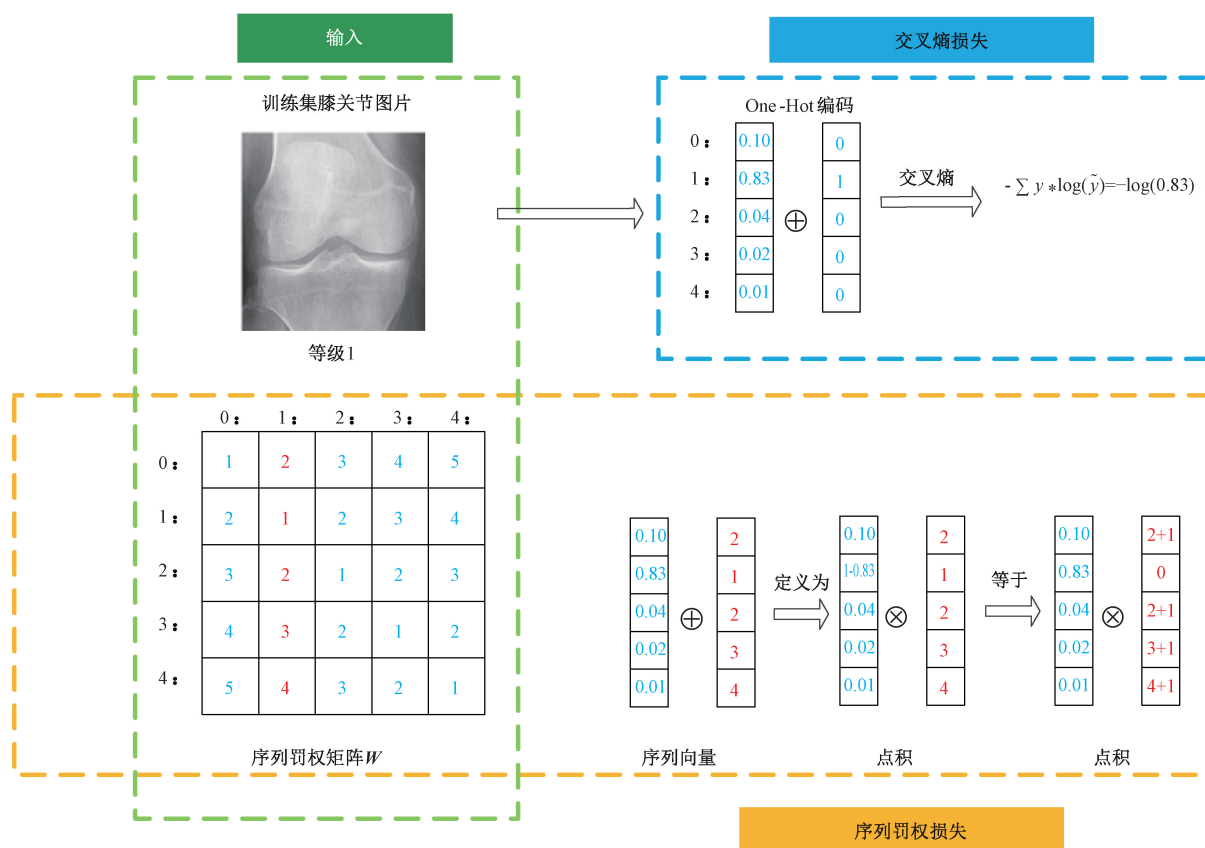


图2 序列罚权损失与交叉熵损失对比

Fig. 2 Comparison of ordinal penalty loss and cross-entropy loss

样本,所以在训练网络时,对一个 epoch 训练后实时统计对应的混淆矩阵结果。假设序列罚权矩阵以及对应的训练结果如图3所示,可以看到,画出圆圈的两个位置,没有达到预期的效果,距离真实类别等级越远的等

级,判错的个数呈现出递增的趋势。一种可能的原因是对应位置的惩罚权重不够大,此时继续固定序列罚权矩阵存在着不合理性。因此,应当给出对应的调整方案。

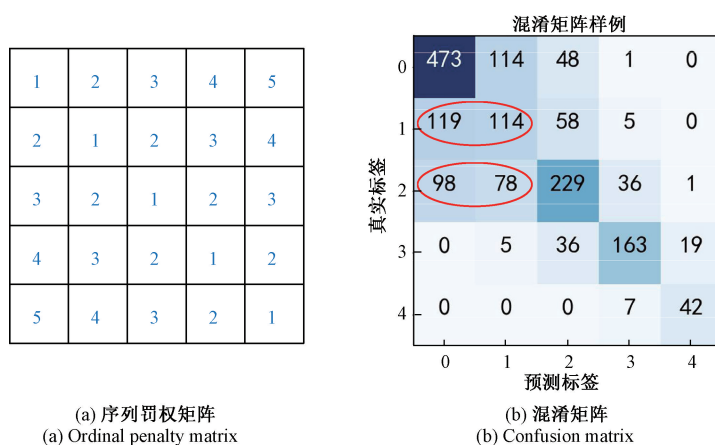


图3 序列罚权矩阵与混淆矩阵示例

Fig. 3 Example of ordinal penalty matrix and confusion matrix

假设初始序列罚权矩阵如图3(a)所示,当深度网络完成一个 epoch 的训练后,得到了图3(b)的结果。根据

上述的讨论,可以看到画圆圈的两个结果不符合沿着主对角元左右递减的规律,距离真实标签越远的判错个数

呈现出递增的趋势,此时有必要重新调整序列罚权矩阵。

在本文中,我们给出一种修改序列罚权矩阵的方案。首先通过混淆矩阵判断出异常元素的位置,如图 3(b) 所示的两个圆圈位置,其中的 119 和 98 两个元素违反了序列性。紧接着,对序列罚权矩阵中同样位置的权重进一步加大惩罚。我们认为,经过一个 epoch 训练后没有达

到预计效果的可能原因是由于对应位置的罚权不够大,需进一步提高。

权重的修改将直接影响到下一个 epoch 训练的结果,需要较为合理的方案。若修改幅度过大,可能会使得模型趋于震荡,因此,我们构思了调整序列罚权矩阵的一种方案,如图 4 所示。

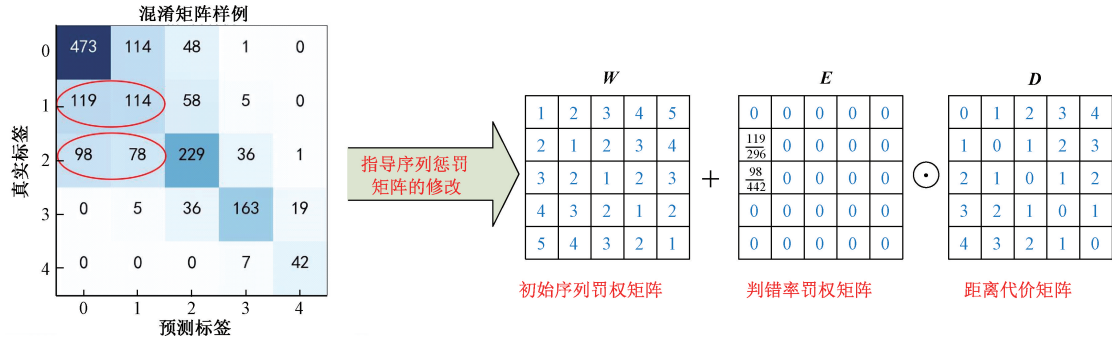


图 4 自适应序列罚权矩阵调整示例

Fig. 4 Example of adaptive ordinal penalty matrix adjustment

对于图 4 中圈出的元素,对其进行行和归一化,得到对应的判错率矩阵 E 。判错率能很好的将样本不平衡的类归一,把错误个数转换为相应权重比例。但考虑到将真实标签为 0 的图片错判为 2 的代价与错判为 1 的代价应该是不同的。因此,在判错率的基础上,我们定义了距离代价矩阵 D ,用来进一步区分距离真实标签不同位置的判错率。我们直接利用判错等级与真实等级差的绝对值作为矩阵 D 元素的定义,代表着判错的代价系数。图中的 + 和 $\odot$ 分别代表矩阵的相加以及矩阵对应元素的点乘,例如,将真实等级为 0 判错为

2 时,除了计算判错率 $98/442$ 外,还需乘以 2,代表着与判错为 1 代价的区别,这样加到初始的惩罚矩阵,完成一次罚权的更新。

进入到下一个 epoch 后,仍重复上述的步骤,先计算混淆矩阵,判断异常判错个数位置,利用判错率矩阵以及距离代价矩阵进行更新,重新训练网络。直至训练到最大的 epoch,停止对网络的训练。然后利用验证集,选择所有 epoch 中等级分类准确率最高的参数,作为最终的模型。最后将测试集带入模型进行准确率计算。图 5 展示了本文方法的流程图。

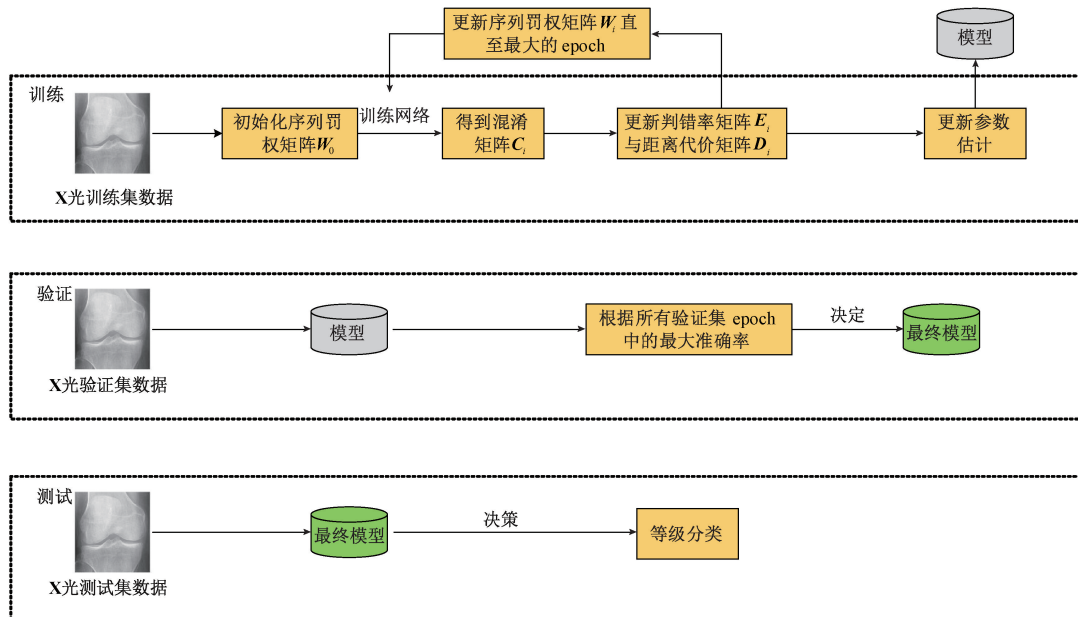


图 5 本文方法

Fig. 5 The proposed method

1.4 KL 等级分类的 CNN 分类器

在本文中,我们在膝骨关节炎 KL 分级任务上对上述这些经典的 CNN 分类网络进行微调,找到膝骨关节炎 KL 分级的最佳 CNN 模型。此外,我们在多个不同架构的 CNN 模型上比较所提出的自适应序列罚权损失与固定罚权损失以及交叉熵损失的性能,检验所提出损失的泛化能力。

2 实验分析与讨论

为了验证本文提出方法的有效性,我们对交叉熵损失,固定序列罚权损失以及本文的自适应系列罚权损失进行数值实验对比。

2.1 数据集预处理及相关参数设定

用于 KL 分级的膝关节图像统一进行尺寸的裁剪,然后进行均值为 0,方差为 1 的标准化处理。裁剪后的图像在 Inception-V3 中被调整为 299×299 ,而在 ResNet、VGG 和 DenseNet 中被调整为 224×224 。它们经过归一化处理后,被转换为三通图像,作为 CNN 分类器训练和测试的输入。在同一组网络的训练中,均设置了相同的随机种子,以确保 3 种方法的随机初始参数相同。

2.2 模型评估指标

为了对比 3 种方法的结果,本文主要采用以下两个评价指标,准确率 (accuracy, AC) 以及 MAE,下面详细介绍这两个指标。

$$AC = \frac{1}{n} \sum_{i=1}^n \delta(P_i, T_i) \quad (5)$$

式中: P_i 代表第 i 张图片的预测标签; T_i 代表第 i 张图片的真实标签; n 代表图片样本的总数。 δ 代表克罗内克函数,其定义如式(6)所示:

$$\delta(P_i, T_i) = \begin{cases} 1, & P_i = T_i \\ 0, & \text{其他} \end{cases} \quad (6)$$

准确率只能作为模型效果的标准衡量指标,在前文提到,对于 KL 等级分类任务,除了准确率外,医生更关心判错类别与真实类别之间的距离,比如真实标签为 0,将其判错为 4 的代价显然比判错为 1 的代价要高得多,因此有必要用另外一个指标来衡量判错的代价程度。平均绝对误差恰好能够很好的刻画这一点,其定义如式(7)所示:

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - T_i| \quad (7)$$

显然,平均绝对误差指标越小,代表着模型的判错代价越低,某种程度上比准确率指标更为重要。注意到,本文提出的自适应方法,首要的目标是为了增大错判样本

类别的罚权,进一步降低判错的代价程度。因此,本文将 MAE 指标作为最重要的评价指标。

2.3 膝骨关节炎 KL 等级分类评估

按照文献[22]的建议,用于训练与测试的 KL 分级膝关节图像分为人工标注图片以及自动检测图片。其中自动检测图片是根据目标检测 YOLO 算法进行定位得到。本文采用 AC 和 MAE 两个指标来评估分类结果。根据验证集上的分类精度,在训练中选择最佳的模型。

为了验证本文提出方法的有效性,我们分别取 4 组不同的初始序列罚权矩阵,其取值与文献[22]相同,其中 PD 代表分差如图 6 所示。

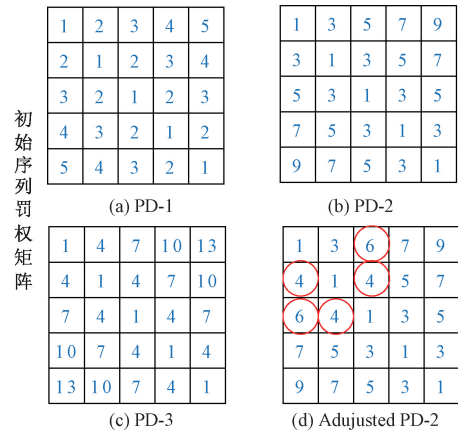


图 6 4 组初始罚权矩阵

Fig. 6 Four groups of initial ordinal penalty matrices

表 1~4 分别展示了图 6 的 4 组不同初始序列罚权对应的实验结果。其中 CE 代表交叉熵损失,Ordinal 代表 Chen 等^[22]提出的序列罚权损失,Proposed 代表本文提出的方法。从 4 个表的结果可以看出,MAE 指标结果的提升较为明显,而 AC 指标则在不同的初始罚权间存在着差异。

表 1 显示了所有模型在第一组初始序列惩罚矩阵的分类性能。在所有比较的 CNN 网络架构里,本文提出的方法获得了令人满意的结果。在人工标注数据方面,提出的方法在七组基线对比模型中的 AC 和 MAE 指标分别获得了六组与五组最优的结果;并且本文提出的 VGG-16-Proposed 方法的准确率为 68.6%,平均绝对误差为 0.478,为所有基线方法中的最佳。在自动标注数据方面,VGG-16-CE 获得了最好的准确率,但本文提出的 VGG-19-Proposed 获得了最小的 MAE。

表 2 显示了与表 1 类似的结果。尽管 VGG-19-CE 在人工标注数据获得了最好的准确率,但本文提出的 VGG-19-Proposed 获得了更小的 MAE。自动标注数据的结果比较一致,除了 DenseNet-161,剩余所有模型,本文

表 1 不同最佳网络在人工和自动膝关节数据上的
膝关节 KL 分级比较(第 1 组罚权矩阵)
Table 1 Knee KL grading comparison on artificial and
automatic knee data by different optimal networks
(Group 1 penalty matrix)

方法		人工图片		自动图片	
		AC/%	MAE	AC/%	MAE
ResNet	ResNet-34-CE	65.0	0.653	64.5	0.656
	ResNet-34-Ordinal	65.1	0.606	66.7	0.587
	ResNet-34-Proposed	65.8	0.559	65.7	0.556
	ResNet-50-CE	63.9	0.681	64.1	0.656
	ResNet-50-Ordinal	65.3	0.576	66.3	0.586
	ResNet-50-Proposed	66.1	0.520	65.6	0.534
DenseNet	DenseNet-161-CE	66.8	0.550	65.9	0.564
	DenseNet-161-Ordinal	66.3	0.481	65.8	0.500
	DenseNet-161-Proposed	68.1	0.490	67.7	0.489
	DenseNet-169-CE	65.8	0.608	65.4	0.623
	DenseNet-169-Ordinal	65.4	0.552	65.3	0.582
	DenseNet-169-Proposed	67.7	0.501	67.2	0.513
VGG	VGG-16-CE	68.4	0.523	69.9	0.506
	VGG-16-Ordinal	66.5	0.495	66.9	0.494
	VGG-16-Proposed	68.6	0.478	67.2	0.466
	VGG-19-CE	68.2	0.560	68.6	0.548
	VGG-19-Ordinal	67.7	0.491	68.7	0.477
Inception	VGG-19-Proposed	67.9	0.494	67.8	0.474
	Inception-V3-CE	66.3	0.589	67.6	0.547
	Inception-V3-Ordinal	65.5	0.514	65.9	0.479
	Inception-V3-Proposed	66.4	0.507	66.8	0.477

提出的方法均得到了更好的结果。从表 1 和 2 的结果可以看出,当初始罚权的分差较为接近时,本文提出的方法在 AC 与 MAE 指标上的提升较为稳定。

表 3 与 4 中的初始罚权分差间隔均较大,从它们的结果来看,本文提出方法在 AC 指标上的表现较不稳定,表 3 的人工标注数据仅三组 AC 结果最优,表 4 中仅两组最优。但可以看到,几乎所有 MAE 的最优结果均为本文提出的方法,说明本文提出的自适应策略能够有效的降低 MAE 指标。

综上所述,本文提出的自适应策略,更适合于初始权重分差较低的矩阵。究其原因,可能是由于初始权重的分差较小时,本文的方法可以加大权重惩罚的力度,进一步改进算法。而对于初始分差间隔较大的权重,序列损失已经对目标函数增加了足够的惩罚,再利用本文的方法进行调整,难以起到有效的改进。

表 2 不同最佳网络在人工和自动膝关节数据上的
膝关节 KL 分级比较(第 2 组罚权矩阵)
Table 2 Knee KL grading comparison on artificial and
automatic knee data by different optimal
networks (Group 2 penalty matrix)

方法		人工图片		自动图片	
		AC/%	MAE	AC/%	MAE
ResNet	ResNet-34-CE	63. 8	0. 679	62. 8	0. 664
	ResNet-34-Ordinal	65. 6	0. 613	64. 9	0. 649
	ResNet-34-Proposed	65. 4	0. 533	65. 3	0. 554
	ResNet-50-CE	63. 3	0. 618	66. 3	0. 599
	ResNet-50-Ordinal	65. 8	0. 548	64. 7	0. 575
	ResNet-50-Proposed	66. 1	0. 542	65. 5	0. 519
DenseNet	DenseNet-161-CE	65. 0	0. 556	66. 6	0. 560
	DenseNet-161-Ordinal	64. 8	0. 556	65. 7	0. 537
	DenseNet-161-Proposed	68. 4	0. 459	66. 3	0. 482
	DenseNet-169-CE	65. 8	0. 638	65. 7	0. 630
	DenseNet-169-Ordinal	64. 9	0. 566	66. 0	0. 577
	DenseNet-169-Proposed	66. 5	0. 496	68. 4	0. 499
VGG	VGG-16-CE	68. 1	0. 489	68. 7	0. 493
	VGG-16-Ordinal	68. 1	0. 484	69. 2	0. 474
	VGG-16-Proposed	68. 9	0. 454	70. 2	0. 447
	VGG-19-CE	69. 7	0. 531	69. 2	0. 537
	VGG-19-Ordinal	68. 2	0. 443	68. 4	0. 436
Inception	VGG-19-Proposed	69. 1	0. 424	69. 5	0. 431
	Inception-V3-CE	66. 1	0. 570	66. 4	0. 556
	Inception-V3-Ordinal	67. 0	0. 560	67. 1	0. 540
	Inception-V3-Proposed	67. 4	0. 487	68. 2	0. 474

表 3 不同最佳网络在人工和自动膝关节数据上的
膝关节 KL 分级比较(第 3 组罚权矩阵)
Table 3 Knee KL grading comparison on artificial and
automatic knee data by different optimal networks
(Group 3 penalty matrix)

方法		人工图片		自动图片	
		AC/%	MAE	AC/%	MAE
ResNet	ResNet-34-CE	62.9	0.733	62.6	0.726
	ResNet-34-Ordinal	63.2	0.556	64.7	0.551
	ResNet-34-Proposed	64.7	0.484	65.3	0.504
	ResNet-50-CE	63.7	0.639	64.0	0.626
	ResNet-50-Ordinal	66.7	0.582	66.2	0.565
	ResNet-50-Proposed	66.2	0.548	64.6	0.529
	DenseNet-161-CE	65.8	0.569	66.2	0.534
	DenseNet-161-Ordinal	65.1	0.536	64.4	0.487

续表 3

方法		人工图片		自动图片	
		AC/%	MAE	AC/%	MAE
DenseNet	DenseNet-161-Proposed	64.0	0.528	64.1	0.463
	DenseNet-169-CE	66.5	0.610	66.3	0.615
	DenseNet-169-Ordinal	64.9	0.608	66.2	0.614
	DenseNet-169-Proposed	67.5	0.603	68.0	0.611
VGG	VGG-16-CE	67.9	0.500	68.3	0.491
	VGG-16-Ordinal	66.4	0.500	69.2	0.463
	VGG-16-Proposed	68.8	0.435	68.9	0.438
	VGG-19-CE	68.4	0.498	68.0	0.523
	VGG-19-Ordinal	70.0	0.410	69.9	0.436
	VGG-19-Proposed	68.2	0.450	69.4	0.429
Inception	Inception-V3-CE	67.7	0.583	69.3	0.542
	Inception-V3-Ordinal	65.7	0.517	67.3	0.512
	Inception-V3-Proposed	67.5	0.481	66.8	0.502

表 4 不同最佳网络在人工和自动膝关节数据上的
膝关节 KL 分级比较 (第 4 组罚权矩阵)Table 4 Knee KL grading comparison on artificial and
automatic knee data by different optimal networks
(Group 4 penalty matrix)

方法		人工图片		自动图片	
		AC/%	MAE	AC/%	MAE
ResNet	ResNet-34-CE	66.4	0.585	65.4	0.652
	ResNet-34-Ordinal	65.8	0.550	64.9	0.580
	ResNet-34-Proposed	64.3	0.513	68.5	0.548
	ResNet-50-CE	65.2	0.618	65.3	0.598
	ResNet-50-Ordinal	62.7	0.591	62.8	0.590
	ResNet-50-Proposed	63.2	0.563	63.1	0.581
	DenseNet-161-CE	66.6	0.519	66.6	0.524
DenseNet	DenseNet-161-Ordinal	65.2	0.524	64.1	0.529
	DenseNet-161-Proposed	66.9	0.515	67.5	0.509
	DenseNet-169-CE	65.5	0.572	65.8	0.589
	DenseNet-169-Ordinal	61.2	0.570	61.9	0.567
	DenseNet-169-Proposed	64.0	0.481	64.4	0.490
VGG	VGG-16-CE	66.8	0.538	68.5	0.525
	VGG-16-Ordinal	68.0	0.431	68.6	0.430
	VGG-16-Proposed	69.1	0.428	69.6	0.422
	VGG-19-CE	66.6	0.511	67.3	0.496
	VGG-19-Ordinal	68.1	0.450	67.8	0.436
	VGG-19-Proposed	67.4	0.464	68.0	0.432
Inception	Inception-V3-CE	66.5	0.563	67.0	0.550
	Inception-V3-Ordinal	63.4	0.550	63.8	0.520
	Inception-V3-Proposed	65.8	0.506	66.0	0.511

3 结 论

本文提出了一种自适应调整策略,采用新颖的序列罚权损失进行膝关节关节炎的 KL 自动分级。在膝关节关节炎 KL 分级任务中,使用经典的 CNN 分类器模型进行验证。与传统的交叉熵损失以及固定的序列罚权损失相比,所提出的自适应策略进一步提高了分类性能,尤其能够有效的降低预测标签与真实标签之间的 MAE 值,说明该策略在 KL 分级任务中的潜力。实验结果表明,本文提出的自适应序列罚权调整策略在初始罚权分差较小时,能够有效地提升四类经典 CNN 网络的分类精度与平均绝对误差性能。与 ResNet、DenseNet 以及 Inception 的网络相比,微调后的 VGG 模型取得了最好的分类性能。

本文的自适应策略在初始权重分差较大时,准确率的提升不理想,后续我们将研究改进的策略,以获得更好的提升效果。此外,拟通过增加学习样本数量和模型组合等方法,进一步提高模型识别准确率,为膝关节关节炎的 KL 自动分级提供更好的辅助诊断。

致谢

感谢徐希德同学在论文撰写中所给予的支持与帮助。

参考文献

- [1] CONAGHAN P G, PORCHERET M, KINGSBURY S R, et al. Impact and therapy of osteoarthritis: The arthritis care OA nation 2012 survey[J]. Clinical Rheumatology, 2015, 34(9):1581-1588.
- [2] KELLGREN J H, LAWRENCE J S. Radiological assessment of osteo-arthritis [M]. Annals of the Rheumatic Diseases, 1957:494-502.
- [3] CNLVENOR A G, ENGEN C N, ØIESTAD B E, et al. Defining the presence of radiographic knee osteoarthritis: A comparison between the Kellgren and Lawrence system and OARSI atlas criteria [J]. Knee Surgery, SportsTraumatology, Arthroscopy 2015, 23 (12): 3532-3539.
- [4] NIU Z X, ZHOU M, WANG L, et al. Ordinal regression with multiple output CNN for age estimation[C]. IEEE Conference on Computer Vision & Pattern Recognition IEEE Computer Society, 2016:4920-4928.
- [5] SHAMIR L, LING S M, SCOTT W W, et al. Knee X-Ray image analysis method for automated detection of osteoarthritis [J]. IEEE Transactions on Biomedical

- Engineering 2008, 56(2): 407-415.
- [6] HE KM. Deep residual learning for image recognition[C]. IEEE Conference on Computer Vision & Pattern Recognition IEEE Computer Society, 2016:770-778.
- [7] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. ArXiv Preprint, 2014, ArXiv:1409.1556.
- [8] SZEGEDY C, WEI L, JIA Y, et al. Going deeper with convolutions[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1-9.
- [9] HUANG G, LIU Z, WEINBERGER K Q, et al. Densely connected convolutional networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.
- [10] LI QJ, YU B, TIAN X, et al. Deep residual nets model for staging liver fibrosis on plain CT images[J]. International Journal of Computer Assisted Radiology and Surgery 2020, 15(8): 1399-1406.
- [11] GULSHAN V, PENG L, CORA M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs[J]. Jama 2016, 316(22): 2402-2410.
- [12] 张永德, 彭景春, 刘罡, 等. 基于水平集的前列腺磁共振图像分割方法研究[J]. 仪器仪表学报, 2017, 38(2): 416-424.
- ZHANG Y D, PENG J CH, LIU G, et al. Research on prostate MRI image segmentation method based on level set[J]. Chinese Journal of Scientific Instrument, 2017, 38(2): 416-424.
- [13] 苑玮琦, 常乐, 张波. 基于多尺度颜色替换的角膜老年环分割[J]. 仪器仪表学报, 2017, 38(1): 236-244.
- YUAN W Q, CHANG L, ZHANG B. Corneal ageing ring segmentation based on multi-scale colour replacement[J]. Chinese Journal of Scientific Instrument, 2017, 38(1): 236-244.
- [14] 钟京谕, 姚伟武. 人工智能在骨关节炎影像诊断中的研究现状与进展[J]. 中华放射学杂志, 2019, 53(9): 790-793.
- ZHONG J X, YAO W W. Current status and progress of research on artificial intelligence in osteoarthritis imaging diagnosis[J]. Chinese Journal of Radiology, 2019, 53(9): 790-793.
- [15] ANTONY J, MCGUINNESS K, MORAN K, et al. Automatic detection of knee joints and quantification of knee osteoarthritis severity using convolutional neural networks[C]. International Conference on Machine Learning and Data Mining in Pattern Recognition, 2017: 376-390.
- [16] ANTONY J, MCGUINNESS K, CONNOR N, et al. Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks[C]. 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, 2016: 1195-1200.
- [17] ANTONY J, MCGUINNESS K, MORAN K, et al. Feature learning to automatically assess radiographic knee osteoarthritis severity[M]. Deep Learners and Deep Learner Descriptors for Medical Applications. Springer, Cham, 2020: 9-93.
- [18] GÓRRIZ M, ANTONY J, MCGUINNESS K, et al. Assessing knee OA severity with CNN attention-based end-to-end architectures[J]. International Conference on Medical Imaging with Deep Learning. PMLR, 2019: 197-214.
- [19] TIULPIN A, THEVENOT J, RAHTU E, et al. Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach[J]. Scientific Reports, 2018, 8(1): 1727.
- [20] ZHANG B F, TAN J M, CHO K H, et al. Attention-based CNN for KL grade classification: Data from the osteoarthritis initiative[C]. 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI) IEEE, 2020: 731-735.
- [21] FREUND, YOAV, ROBERT E. A decision-theoretic generalization of on-line learning and an application to boosting[J]. Journal of Computer and System Sciences 1997, 55(1): 119-139.
- [22] CHEN P J, et al. Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss[J]. Computerized Medical Imaging and Graphics 2019, 75: 84-92.

作者简介



Liu Weiqiang received his B. Sc. degree from Minnan Normal University in 2013, and received his M. Sc. degree from Xiamen University in 2016. He is currently a Ph. D. Candidate at

刘伟强, 2013 年于闽南师范大学获得学士学位, 2016 年于厦门大学获得硕士学位, 现为厦门大学博士研究生, 主要研究方向为人工智能与模式识别。

E-mail: 23220180155106@stu.xmu.edu.cn

Xiamen University. His research interests include artificial intelligence and pattern recognition.



罗林开(通信作者),1986 年于厦门大学

大学本科毕业,1991 年于厦门大学获得硕士学位,2007 年于厦门大学获得博士学位,现为厦门大学教授,主要研究方向为机器学习、模式识别、数据挖掘、计算机视觉、生物信息处理和金融数据分析。

E-mail: luolk@xmu.edu.cn

Luo Linkai (Corresponding author) received his B. Sc. degree from Xiamen University in 1986, received his M. Sc. degree and Ph. D. degree both from Xiamen University in 1991 in 2007, respectively. He is currently a professor at Xiamen University. His research interests include machine learning, pattern recognition, data mining, computer vision, bioinformation processing and financial data analysis.



章其敏,2017 年于湖北科技学院获得学士学位,2020 年于华中科技大学获得硕士学位,现为武汉大学附属同仁医院住院医师,主要研究方向为骨骼肌肉系统影像诊断。

E-mail: 352069570@qq.com

Zhang Qimin received her B. Sc. degree from Hubei University of Science and Technology in 2017 and received her M. Sc. degree from Huazhong University of Science and Technology in 2020. She is currently a resident at Tongren Hospital of Wuhan University. Her research interest is the imaging of the skeletal muscle system.



黄玮,2009 年于华中科技大学获得学士学位,2012 年于华中科技大学获得博士学位,现为华中科技大学同济医学院附属协和医院骨科医院主治医师,主要研究方向为骨关节疾病及运动损伤的诊治。

E-mail: drhuangwei@126.com

Huang Wei received his B. Sc. degree from Huazhong University of Science and Technology in 2009 and Ph. D. degree from Huazhong University of Science and Technology in 2012. He is currently an attending physician at the Department of Orthopaedics, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology. His research interests is the diagnosis and treatment of bone and joint diseases and sports injuries.