

随机占优在图像分割神经网络对比中的应用^①

张 施, 庄玮玮

(中国科学技术大学 管理学院, 合肥 230026)
通信作者: 庄玮玮, E-mail: weizh@ustc.edu.cn



摘 要: 在过去半个多世纪中, 随着计算机技术的发展, 神经网络已经在图像、语音、决策等众多领域取得了广泛的应用. 不同学者为了提高神经网络的准确率设计了大量的网络结构, 神经网络也变得越来越复杂和多参数化. 这使得神经网络的训练过程具有很强的非凸性, 相同的网络不同的初始参数往往会训练出不同的模型. 为了更精准地描述两个网络的表现, 前人提出通过统计学方法——随机占优 (stochastic dominance) 评估不同随机种子对同一网络训练出的不同模型的表现的分布. 本文在此基础上认为, 不同模型在测试集中不同样本上的表现的分布同样值得关注, 并将随机占优方法应用到不同模型在不同样本表现分布的对比中. 通过对图像分割应用中的网络进行实验, 本文关注到不同网络训练出的两个模型其中一个尽管在表现分数上具有一定的优势, 但是其在测试集中不同样本中表现出的离散度可能更强. 实际应用需要表现分数更好同时离散度尽可能小的神经网络模型, 随机占优方法可以有效地对不同模型进行比较从而筛选出更适合的模型.

关键词: 神经网络; 图像分割; 随机占优; 几乎随机占优; 网络评价

引用格式: 张施, 庄玮玮. 随机占优在图像分割神经网络对比中的应用. 计算机系统应用, 2022, 31(12): 383-389. <http://www.c-s-a.org.cn/1003-3254/8829.html>

Application of Stochastic Dominance in Comparison of Neural Network for Image Segmentation

ZHANG Shi, ZHUANG Wei-Wei

(School of Management, University of Science and Technology of China, Hefei 230026, China)

Abstract: In the past half-century, with the development of computer technology, neural networks have been widely used in many fields such as images, speeches, and decision-making. To improve the accuracy of neural networks, different scholars have designed a large number of network structures, and thus neural networks have become more and more complex and multi-parametric. As a result, the training process of neural networks has strong non-convexity, and different initial parameters of the same network often train different models. To more accurately describe the performance of two networks, predecessors proposed to evaluate the distribution of the performance of different random seeds on different models trained by the same network through the statistical method of stochastic dominance. On this basis, this study believes that the distribution of the performance of different models on different samples in a test set is also worthy of attention, and thus the stochastic dominance method is applied to compare the distribution of the performance of different models on different samples. Through the experiments on the networks applied in image segmentation, this study finds that for the two models trained by different networks, although one of them has certain advantages in the performance score, it may show stronger dispersion on different samples in the test set. The practical application, however, requires the neural network model with a better performance score and dispersion as small as possible. The stochastic dominance method can effectively compare different models for the selection of a more suitable one.

Key words: neural network; image segmentation; stochastic dominance; almost stochastic dominance; network evaluation

^① 基金项目: 国家自然科学基金面上项目 (71971204)

收稿时间: 2022-03-30; 修改时间: 2022-04-22; 采用时间: 2022-04-29

1 概述

神经网络是机器学习中的重要技术手段,在图像处理、语音识别、时间序列预测等领域有着广泛的应用^[1].在过去20年中,随着计算机技术的发展,神经网络的结构、规模以及训练方法都得到了飞速的发展.诸如卷积神经网络、循环神经网络等一系列技术的引入使得神经网络的准确率取得了快速的提升.但是由于引入了大量非线性的结构,神经网络的训练过程存在非常强的非凸性,这使得对神经网络准确性的评估也变得更加复杂,尤其是当神经网络的准确率越来越高时,其评分也都越来越接近极限^[2].

对神经网络的评估主要是通过对神经网络在训练集以外的测试集上的表现.通常会用损失(loss)、准确率(accuracy)等函数作为神经网络的表现分数.为此在很多问题上,大量学者和机构提供了很多公开的训练集、验证集以及测试集(如对于手写字体分类的mnist^[3],对于自然分割问题的Oxford-IIIT^[4]).由于神经网络的训练是一个非凸问题,不同的神经网络初始参数可能最终得到的网络模型并不相同,因此在评估神经网络时,通常会选择不同次训练中表现最好的模型的表现分数来与其他网络对比.但这个过程通常具有较大的偶然性和不可复制性,Reimers等^[5,6]指出通过分析同一网络不同随机种子训练得到模型的表现分数的分布能更稳定地比较分数接近的不同网络之间的优劣.在此基础上,Dror等^[7]引入了随机占优(stochastic dominance)和几乎随机占优(almost stochastic dominance)来分析两个分数相近的网络分数分布的优劣.随机占优是一种基于不同随机变量的分布函数间关系得到排序的检验.最早明确地提出随机占优概念的是Quirk等^[8],直到上世纪60年代末70年代初,Hadar等^[9]、Hanoch等^[10]一系列随机占优的经典论文发表后,这一理论的相关发展和应用逐步推广到金融、医学、工程等不同领域中^[11].随机占优理论弥补了均值-方差模型判断随机变量分布的局限性.通常评价两个神经网络时,我们希望其中一个神经网络分数的期望优于另一网络,同时其标准差也比另一网络小.但实际上可能存在期望较优,但是标准差更大的情况,随机占优可以很好地应对这种情况.

除了不同初始模型训练得到不同模型的分布值得关注,我们认为不同模型对于测试集中不同样本的表现的分布也值得关注.尤其是在对于神经网络预测精度的稳定性有要求的领域(如图像分割、股票预测)^[12,13],我们要求训练出的模型在对测试集进行预测

时具有较高的准确率且较小的标准差.因此本文中进一步将随机占优引入到图像分割问题中不同网络模型在测试集中不同样本表现分布的比较中.以图像分割中常用的卷积编码器-解码器网络(convolutional encoder decoder,以下简称CED)^[14]和U-Net网络^[15]为例,本文先对比了两个网络在500个不同初始模型训练之后表现的分布,再进一步地对比了这些训练出的模型在测试集中不同样本表现的分布.结果显示尽管不同初始模型训练得到的结果U-Net占优于CED网络,但是CED网络在测试集不同样本中表现的标准差要小于U-Net网络,这使得在本文的抽样测试对比中,半数以上的CED网络模型在不同样本表现中的分布占优于U-Net网络模型.因此本文认为,不同模型在测试集中表现的标准差应该引起网络对比中的重视,通过随机占优方法可以更合理地对比不同的模型,并给出合适的选择.

2 随机占优和几乎随机占优理论

2.1 随机占优的定义

令随机变量 X 、 Y ,累积分布函数分别为 F 、 G ,且 F 、 G 都是定义在 \mathbb{R} 上的连续函数.当且仅当对所有的 $a \in \mathbb{R}$,都有 $F(a) \leq G(a)$ (不等号在某些 a 处严格成立)时,称 X 随机占优于 Y (记作 $Y > X$).即对所有 a , X 中有更高比例大于 a 的样本.完全随机占优成立的条件非常严格,见图1(a),针对 a 的每一个值,不等式 $F(a) \leq G(a)$ 都要成立,这在应用中几乎难以实现.因此,随机占优的条件的放宽引起了更多关注^[16].

Leshno等^[17]提出了几乎随机占优的概念,允许 G 在一些 a 处位于 F 的下方,见图1(b).del Barrio等^[18]则采用 L_2 -Wasserstein距离来定义和检验几乎随机占优.接下来从分位数的概念开始介绍.

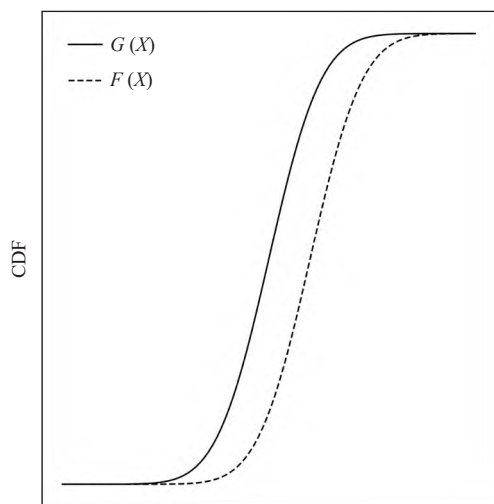
令 $F^{-1}(t) = \inf\{x : t \leq F(x)\}$, $t \in (0, 1)$ 为概率密度函数 F 对应的分位数函数.随机占优可表示为: $X > Y \Leftrightarrow F^{-1}(t) \geq G^{-1}(t), \forall t \in (0, 1)$.可以看出当 $F^{-1}(t) < G^{-1}(t)$ 时, X 和 Y 间随机占优不成立.用分位数函数定义 A_X 和 A_Y :

$$\begin{cases} A_X = \{t \in (0, 1) : F^{-1}(t) < G^{-1}(t)\} \\ A_Y = \{t \in (0, 1) : F^{-1}(t) > G^{-1}(t)\} \end{cases}$$

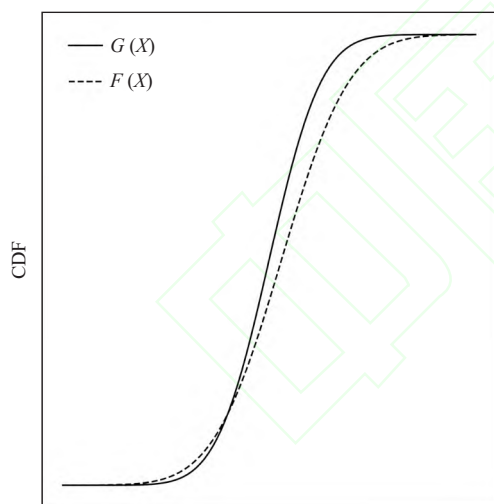
del Barrio等^[18]基于这些概念提出占优指数 ε_{w_2} 来定义一个随机变量占优于另一个随机变量的程度:

$$\varepsilon_{w_2}(F, G) := \frac{\int_{A_X} (F^{-1}(t) - G^{-1}(t))^2 dt}{(w_2(F, G))^2}$$

其中, $w_2(F, G) = \sqrt{\int_0^1 (F^{-1}(t) - G^{-1}(t))^2 dt}$ 是分布间的单变量 L_2 -Wasserstein 距离. 这个指数满足 $0 \leq \varepsilon_{w_2}(F, G) \leq 1$, 其中 0 表示 X 完全占优于 Y , 1 表示 Y 完全占优于 X . 同时有 $\varepsilon_{w_2}(F, G) = 1 - \varepsilon_{w_2}(G, F)$ 成立. 较小的指数值 (一般定义在 0 到 0.5 之间) 表示越接近完全随机占优.



(a) 随机占优



(b) 几乎随机占优

图1 随机占优和几乎随机占优的示意图

同时, 随机占优具有传递性, 这意味着若分布 F 占优于 G , G 占优于 H , 那么 F 占优于 H . 这样在比较多个分布时, 可以大大减少比较次数^[19].

2.2 几乎随机占优的统计检验

使用占优指数 ε_{w_2} , 可以制定以下假设检验问题, 来检验几乎随机占优:

$$\begin{cases} H_0 : \varepsilon_{w_2}(F, G) \geq \epsilon \\ H_1 : \varepsilon_{w_2}(F, G) < \epsilon \end{cases}$$

对于提前给定的 $\epsilon > 0$, 若占优指数 $\varepsilon_{w_2}(F, G)$ 比 ϵ 小, 则拒绝原假设, 这意味着分布 F 以一定的偏差几乎随机占优于 G .

累积分布函数 F 、 G 可由经验分布函数估计得到: $F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(x_i \leq t)$; $G_m(t) = \frac{1}{m} \sum_{i=1}^m \mathbb{I}(y_i \leq t)$, 其中 $\mathbb{I}(\cdot)$ 为示性函数.

del Barrio 等^[18] 证明了占优指数 $\varepsilon_{w_2}(F, G)$ 的渐近正态性, 并指出在没有更多假设的前提下, 若有:

$$\sqrt{\frac{nm}{n+m}} (\varepsilon_{w_2}(F_n, G_m) - \epsilon) < \hat{\sigma}_{n,m} \Phi^{-1}(\alpha)$$

则在显著性水平 α 下拒绝 H_0 (本文中 α 取 0.05). 其中, ϵ 为违规水平, Φ 是标准正态分布函数, $\hat{\sigma}_{n,m}$ 是式 $\sqrt{\frac{nm}{n+m}} (\varepsilon_{w_2}(F_n^*, G_m^*) - \varepsilon_{w_2}(F_n, G_m))$ 的方差估计值, 其中 F_n^* 、 G_m^* 是由原始样本 X_n 、 Y_m 混合后抽取得到的样本 X_n^* 、 Y_m^* 计算而来.

此外, 在置信水平为 $1 - \alpha$ 下 F 几乎随机占优于 G 的最小 ϵ 为:

$$\epsilon^{\min}(F_n, G_m, \alpha) = \varepsilon_{w_2}(F_n, G_m) - \hat{\sigma}_{n,m} \Phi^{-1}(\alpha)$$

若 $\epsilon^{\min}(F_n, G_m, \alpha) < 0.5$, 可以称 F 优于 G , $\epsilon^{\min}(F_n, G_m, \alpha)$ 的值则表明分布间的最大差距, 当 $\epsilon^{\min}(F_n, G_m, \alpha) = 0$ 时, F 完全随机占优于 G . 然而, 若 $\epsilon^{\min}(F_n, G_m, \alpha) \geq 0.5$, F 几乎随机占优于 G 不成立, 接受原假设.

在 $\hat{\sigma}_{n,m}$ 估计准确的前提下, $\epsilon^{\min}(F_n, G_m, \alpha) = 1 - \epsilon^{\min}(G_m, F_n, \alpha)$ 成立. 因此, 对于给定的 α 值, 除非 $\epsilon^{\min}(F_n, G_m, \alpha) = \epsilon^{\min}(G_m, F_n, \alpha) = 0.5$, 我们都能得到一个算法更优的结论.

2.3 几乎随机占优的算法流程

现有来自两个神经网络模型 A 、 B 的评分集合: $A = \{x_1, x_2, \dots, x_n\}$, $B = \{y_1, y_2, \dots, y_m\}$, 几乎随机占优模拟算法的伪代码如下:

- (1) 将两个集合中分值从小到大排列得到: $A = [x_{(1)}, \dots, x_{(n)}]$, $B = [y_{(1)}, \dots, y_{(m)}]$, 其中 $x_{(i)}$ 表示第 i 小的值.
- (2) 计算评分的经验分布函数 F_n 、 G_m : $F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(x_i \leq t)$.
- (3) 用式 $F^{-1}(t) = \inf\{x : t \leq F(x)\}$, $t \in (0, 1)$ 构造得分的经验分布逆函数 $F_n^{-1}(t)$ 、 $G_m^{-1}(t)$.
- (4) 计算随机占优指数 $\varepsilon_{w_2}(F, G)$, 实践中, 我们用黎

曼积分计算积分,也就是将积分区间 $(0, 1)$ 分为长度为 Δ 的小区间,每个区间的积分乘以 Δ 再求和。

(5) 估计 σ : 从经验分布 F_n 、 G_m 中抽取足够的样本 X_n^* 、 Y_m^* , 对每一对样本计算表达式 $\sqrt{\frac{nm}{n+m}}(\varepsilon_{w_2}(F_n^*, G_m^*) - \varepsilon_{w_2}(F_n, G_m))$ 的值, 将这些值的方差作为 σ^2 的估计, 样本越多, 精度越高。

(6) 在置信水平 $1-\alpha$ 下模型 A 几乎随机占优于 B 的最小 ϵ 为:

$$\epsilon^{\min}(F_n, G_m, \alpha) = \varepsilon_{w_2}(F_n, G_m) - \hat{\sigma}_{n,m} \Phi^{-1}(\alpha)$$

当 $\epsilon^{\min}(F_n, G_m, \alpha) < \epsilon$ 时, 模型 A 几乎随机占优于模型 B 。

3 图像分割

3.1 图像分割简介

图像分割是神经网络的重要应用, 其主要任务是

把图片分割为若干个具有独特特征的区域, 是图像分析以及计算机视觉等领域中的关键步骤^[20]。高精度的图像分割在工业生产、医学图像处理以及科学研究等诸多领域都具有重要的应用空间。卷积神经网络(CNN)是图像分割任务中的重要实现手段。本文以卷积神经网络发展而来的 CED 和 U-Net 网络在 Oxford-IIIT 3.0.0 数据集^[4]中的表现为例, 展示随机占优在神经网络比较中的应用。两者的网络结构如图 2 所示, 其中 Conv Block 表示由两个包含 3×3 卷积核的卷积层组成卷积块; Pooling Layer 为池化层; trConv Block 表示由两组转制卷积(transposed convolution)组成的转制卷积块; Upsampling Layer 表示上采样层。需要注意的是, 为通过随机占优直观地对比两组结果相近的神经网络模型, 我们这里仅选用这两种在图像分类任务中经典的神经网络结构, 且无需关注在网络参数和结果精度上的优化。

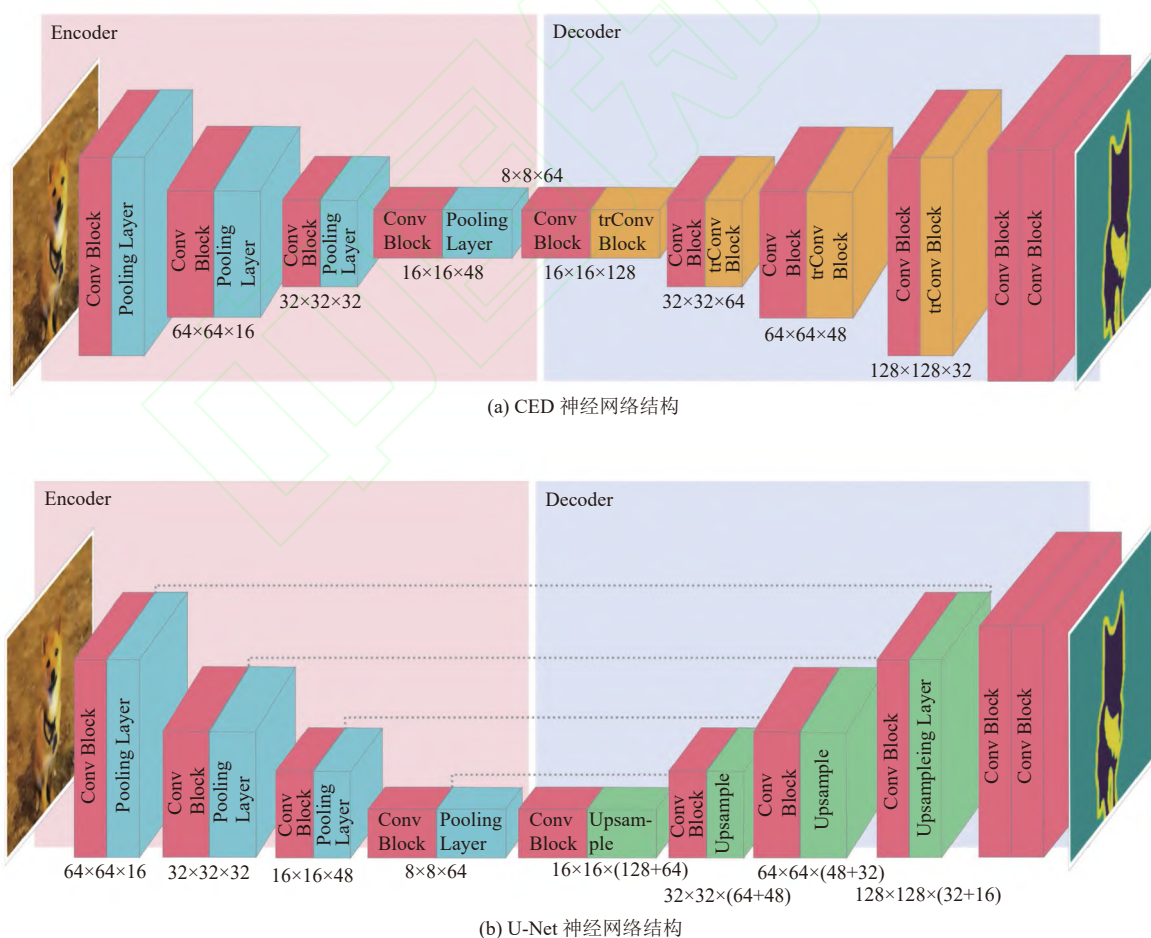


图2 本文使用的 CED 神经网络结构和 U-Net 神经网络结构

Oxford-IIIT 数据集是一个宠物图像数据集, 包含了 37 种宠物的图片, 每种宠物约 200 张图片, 并包含了每张图片的宠物轮廓的标注信息. 我们分别以 0, 1, 2 标注没有宠物的区域, 宠物的轮廓以及宠物所在的区域. 在训练时, 我们把每一张图片转化为 128×128 像素的三通道图片, 并定义每一张图片的准确率 (accuracy) 为和标注完全一致的像素点占总的像素点的比例作为本文对网络评分的主要参数. 如图 3 所示, 图 3(a) 为需要进行图像分割的宠物图片, 图 3(b) 为标注后的轮廓图片, 图 3(c)、图 3(d) 则分别是两个模型预测的结果.

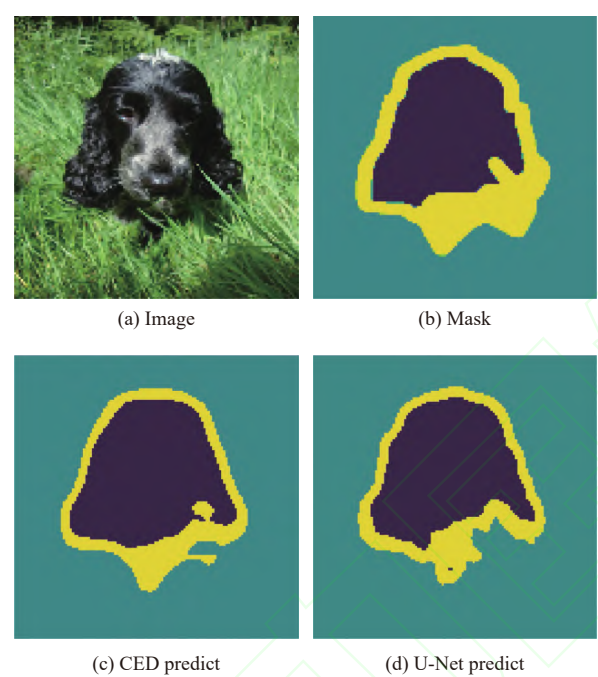


图3 宠物图片实例及其轮廓标注

3.2 应用随机占优对比网络模型预测结果的流程

传统的网络对比方法通常只关注网络预测结果与正确标签的误差. Dror 等^[7] 将随机占优引入到网络对比中, 其主要流程为: 对于两个不同的网络结构 A 和 B , 分别由 n 个和 m 个随机初始模型训练得到 n 个和 m 个模型 $\{A_1, A_2, \dots, A_n\}$ 和 $\{B_1, B_2, \dots, B_m\}$; 分别计算模型 $\{A_1, A_2, \dots, A_n\}$ 和 $\{B_1, B_2, \dots, B_m\}$ 在测试集中的评分 (对应本文中的评分函数为测试集中每一图片预测准确率的均值) 为 $\{x_1, x_2, \dots, x_n\}$ 和 $\{y_1, y_2, \dots, y_m\}$; 通过第 2.3 节中几乎随机占优算法评估两者间的占优关系即为 A 网络和 B 网络的占优关系.

在实际应用中, 比较训练好的每个模型的表现同样值得关注, 因此本文提出随机占优应用于比较两个训

练好的图像分割网络模型 A 和 B , 流程如下: 分别计算 A 和 B 在测试集中每张图片中的准确率 $\{x_1, x_2, \dots, x_n\}$ 和 $\{y_1, y_2, \dots, y_m\}$; 再通过几乎随机占优评估两个准确率集合间的占优关系即为 A 模型和 B 模型的占优关系.

4 实验与分析

4.1 多次训练 U-Net 和 CED 网络之间的比较

我们首先重复 Reimers 等^[5,6] 以及 Dror 等^[7] 的方法对这两个网络进行评估. 我们分别使用 500 组不同的随机种子生成 500 组不同的初始模型对 U-Net 和 CED 网络各进行了 500 次训练. 每次训练使用的优化器均为 Adam 优化器^[21], 且参数完全相同, 每次训练最大训练次数为 50 次, 并设置了 accuracy 稳定之后停止训练, 每次训练最终保留验证集表现最好模型.

图 4 展示了通过训练得到的 500 个 U-Net 和 CED 模型在测试集中准确率的分分布对比, 表 1 展示了两个网络得到 500 个模型准确率的均值、标准差、最大值、最小值. 可以看出 U-Net 网络表现的均值优于 CED 网络, 同时其标准差较小. 通过随机占优比较我们计算出 U-Net 相对于 CED 的占优指数为 0, 这表明 U-Net 网络得到的模型的表现完全占优于 CED 网络.

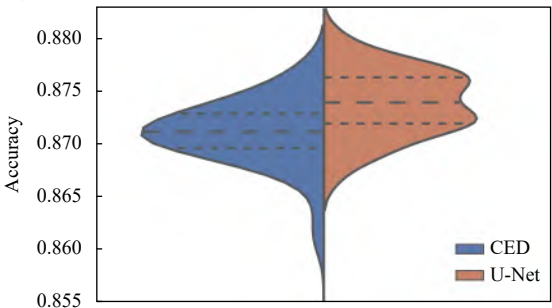


图4 在 Oxford-IIIT 数据集中 500 次初始模型训练对比

表1 U-Net 和 CED 在测试集 500 次得分的统计特征

统计特征	U-Net	CED
均值	0.873 830	0.870 666
标准差	0.003 122	0.004 296
最大值	0.880 595	0.877 801
最小值	0.866 203	0.848 110

4.2 U-Net 和 CED 不同模型在测试集不同样本表现分布的对比

从不同随机种子的训练结果来看 U-Net 网络的表现明显是优于 CED 网络的表现的. 但是我们发现, U-Net 网络训练出的模型对于训练集中不同样本的准确率的

标准差 (0.8 附近) 要比 CED 网络训练出的模型对于训练集中不同样本的准确率的标准差 (0.7 附近) 更大. 这意味着 U-Net 网络训练出的模型尽管表现分数更好, 但是也更加离散. 这里我们从 U-Net 网络和 CED 网络中分别选出一个模型进行对比. 表 2 展示了这两个模型在测试集不同样本中准确率的均值、标准差、最大值和最小值. 从均值上看, U-Net 模型的表现仍然优于 CED 模型的表现, 但是在标准差上 U-Net 模型的表现的离散程度要比 CED 模型的离散程度更大. 通过几乎随机占优, 我们可以得到 CED 相对于 U-Net 网络的占优指数为 0.081 928, 此时 CED 模型占优于 U-Net 模型.

表 2 U-Net 和 CED 同一测试集得分的统计特征

统计特征	U-Net	CED
均值	0.876 711	0.869 920
标准差	0.088 147	0.067 944
最大值	0.981 140	0.984 955
最小值	0.060 302	0.575 836

为了排除这种占优是一种偶然现象, 我们从训练出的模型中有放回地随机选取了 500 对模型进行对比. 图 5 展示了这 500 对模型中 CED 网络模型相对于 U-Net 网络模型的占优指数的分布. 根据几乎随机占优理论占优指数小于 0.5 时 CED 网络模型占优于 U-Net 网络模型, 而当占优指数小于 0.1 时这种占优关系尤为明显, 相反则 U-Net 占优于 CED 网络. 从抽样结果中, 共有 267 对 CED 网络占优于 U-Net 网络, 163 对占优指数小于 0.1, 而占优指数大于 0.9 的仅有 137 对. 这显然与网络之间的对比的结论相反. 因此我们认为不同模型对于测试集中不同样本的表现的离散程度值得在神经网络评估时的重视.

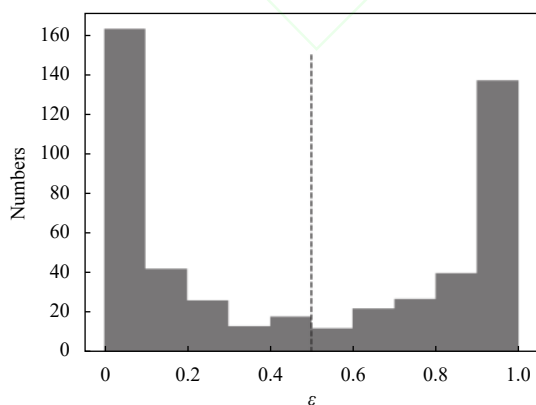


图 5 500 组抽样的 CED 和 U-Net 模型对的占优指数分布
在对神经网络结构和训练的研究中, 网络的占优

关系可以有效地减少随机因素对研究的影响. 但在实际解决问题中, 模型的占优与否往往要比网络的占优与否更重要. 尤其是随着神经网络的发展准确率越来越接近 100% 的今天, 不同模型之间的差距越来越小, 模型在测试集中表现的离散度应引起足够的重视. 但仍然需要值得注意的是, 通常在表现相近的网络或模型 (如准确率差距小于 3%) 中, 占优关系才可能和均值的表现相反, 而在表现差距较大时, 占优关系大概率与均值的表现是一致的.

4.3 通过几乎随机占优筛选出更合适的模型

在上文中我们已经验证了随机占优可以有效地对不同模型进行比较. 更进一步我们可以利用随机占优的传递性筛选出更合适的模型, 对于 n 个训练好的模型其具体流程如下:

- (1) 挑选模型集合中排序前两个模型分别命名为 A 和 B ;
- (2) 通过随机占优对比 A 和 B 模型在测试集中评分的表现, 保留其中占优的模型并命名为 A ;
- (3) 将集合序列中下一个模型命名为 B ;
- (4) 重复步骤 (2) 和步骤 (3) 直至所有模型都被比较, 最后保留下的 A 模型即为 n 个模型中最优模型.

通过上述流程, 可以得到本文训练得到的 500 个 U-Net 模型中最优模型的评分均值为 0.880 197, 标准差为 0.071 627; 本文训练得到的 500 个 CED 模型中最优模型的评分均值为 0.877 677, 标准差为 0.062 972; 这两者中 CED 模型占优于 U-Net 模型, 计算得到的占优指数为 0.187 514.

5 结论与展望

如何精确地评估神经网络的表现是当下神经网络的重要问题. 本文在 Reimers 等^[5,6] 以及 Dror 等^[7] 的工作基础之上进一步将几乎随机占优引入到神经网络模型在测试集中不同样本表现分布的对比中. 随机占优可以有效地解决在评估不同分布的对比时均值和方差结论不统一的问题. 通过图像分割的实验对比, 本文关注到不同神经网络模型在测试集中不同样本表现的离散性对于评估不同神经网络的重要性. 对于一些网络, 尽管不同次训练出来的模型总体的表现可能更好, 但是其每个模型对于测试集不同样本表现的方差仍然可能较大. 本文希望通过随机占优方法的研究进一步提高神经网络对比的有效性.

参考文献

- 1 周志华. 机器学习. 北京: 清华大学出版社, 2016.
- 2 Dauphin YN, Pascanu R, Gulcehre C, *et al.* Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2014. 2933–2941.
- 3 Deng L. The MNIST database of handwritten digit images for machine learning research [best of the Web]. IEEE Signal Processing Magazine, 2012, 29(6): 141–142. [doi: [10.1109/MSP.2012.2211477](https://doi.org/10.1109/MSP.2012.2211477)]
- 4 Parkhi OM, Vedaldi A, Zisserman A, *et al.* Cats and dogs. 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence: IEEE, 2012. 3498–3505.
- 5 Reimers N, Gurevych I. Optimal hyperparameters for deep LSTM-networks for sequence labeling tasks. arXiv: 1707.06799, 2017.
- 6 Reimers N, Gurevych I. Reporting score distributions makes a difference: Performance study of LSTM-networks for sequence tagging. arXiv: 1707.09861, 2017.
- 7 Dror R, Shlomov S, Reichart R. Deep dominance—How to properly compare deep neural models. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence: ACL, 2019. 2773–2785.
- 8 Quirk JP, Saposnik R. Admissibility and measurable utility functions. The Review of Economic Studies, 1962, 29(2): 140–146. [doi: [10.2307/2295819](https://doi.org/10.2307/2295819)]
- 9 Hadar J, Russell WR. Rules for ordering uncertain prospects. The American Economic Review, 1969, 59(1): 25–34.
- 10 Hanoch G, Levy H. The efficiency analysis of choices involving risk. Stochastic Optimization Models in Finance. New York: Academic Press, 1975. 89–100.
- 11 唐爱国. 广义随机占优理论: 群体决策、收入分配与风险管理. 北京: 北京大学出版社, 2005.
- 12 Minaee S, Boykov Y, Porikli F, *et al.* Image segmentation using deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(7): 3523–3542.
- 13 Chong E, Han C, Park FC. Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. Expert Systems with Applications, 2017, 83: 187–205. [doi: [10.1016/j.eswa.2017.04.030](https://doi.org/10.1016/j.eswa.2017.04.030)]
- 14 Yang JM, Price B, Cohen S, *et al.* Object contour detection with a fully convolutional encoder-decoder network. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 193–202.
- 15 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich: Springer, 2015. 234–241.
- 16 Zhuang WW, Hu BY, Chen J. Semiparametric inference for the dominance index under the density ratio model. Biometrika, 2019, 106(1): 229–241. [doi: [10.1093/biomet/asy068](https://doi.org/10.1093/biomet/asy068)]
- 17 Leshno M, Levy H. Preferred by “all” and preferred by “most” decision makers: Almost stochastic dominance. Management Science, 2002, 48(8): 1074–1085. [doi: [10.1287/mnsc.48.8.1074.169](https://doi.org/10.1287/mnsc.48.8.1074.169)]
- 18 del Barrio E, Cuesta-Albertos JA, Matrán C. An optimal transportation approach for assessing almost stochastic order. The Mathematics of the Uncertain. Cham: Springer, 2018. 33–44.
- 19 Levy H, Robinson M. Stochastic Dominance: Investment Decision Making Under Uncertainty. New York: Springer, 2006.
- 20 Shapiro LG, Stockman GC. Computer Vision. New Jersey: Prentice Hall, 2001.
- 21 Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv: 1412.6980, 2014.

(校对责编: 牛欣悦)