

用于自动超参数设置和卷积神经网络架构设计的统计驱动珊瑚礁元启发式算法

亚历杭德罗·马尔蒂内斯^{1,*}, 劳尔·拉拉·卡布雷拉^{2,*}, Víctor Manuel Vargas¹,
佩德罗·安东尼奥·古铁雷斯¹, Martí David Camacho¹, María Jesús Rodríguez¹
{alejandro.martin,raul.lara,david.camacho}@upm.es ¹ Universidad de
Córdoba, 科尔多瓦, 西班牙电子邮件: {vvargas,pagutierrez,cherवास}@uco.es

摘要 调整卷积神经网络 (CNN) 的超参数和网络结构是朝着构建有效但仍然高效的机器学习模型迈出的重要一步。最佳配置的选择是一项与问题相关的任务,涉及探索巨大而复杂的搜索空间。由于这个原因,基于启发式搜索的使用非常适合这项任务,寻求在复杂和大的探索空间中获得接近最优的解决方案。本文介绍了 SCRODeep,这是一种基于统计驱动的珊瑚礁优化算法 (SCRO) 的自适应算法,用于在特定领域中选择最合适的 CNN 架构。这种元启发式设计用于在搜索空间中导航,其中架构 (定义特定的层集,包括卷积层或池化层) 和网络的超参数 (即激活函数、单元数或内核初始化程序等) 表示,但其中连接权重和偏差是使用典型的 CNN 优化算法推断的。与其他方法相比,元启发式的使用反过来意味着固定一系列超参数 (即遗传算法中的突变概率),我们的方法遵循自参数化的观点,从而消除了固定这些值的必要性。该方法已经在用于图像分类的 CNN 设计中进行了测试,表明 SCRODeep 能够找到有竞争力的解决方案,同时发现的架构的复杂性受到限制。

以及当前降低的计算成本。由于其特性,DNN 能够解决包含数千个特性的复杂问题。事实上,DNN 适应非线性空间的能力以及构建强大的分类和预测系统的能力都在其流行的源泉之内。然而,与支持向量机或随机森林等其他机器学习算法相比,训练过程更为复杂。

另一个不受欢迎的特征是 DNN 建立的模型不是不言自明的,因此无法对收集到的知识进行推理。相同的优点和缺点可以与卷积神经网络 (CNN) 相关联,专门用于视觉相关任务的架构。

一般来说,DNN 或 CNN 的优化可以在三个不同的层次上进行:架构、参数 (连接权重和偏差) 和超参数,例如激活函数或批量大小。这配置了很大的可能性空间,在卷积神经网络的情况下尤其复杂。这些模型涉及多层架构,每个架构都有自己的超参数和参数,最后一层有大量神经元。虽然训练过程中参数 (连接权重) 的定义通常使用反向传播算法执行,但架构和超参数是手动固定的,基于反复试验过程或基于以前的经验。已经采用了几种方法来改进此过程,其中许多方法依赖于进化计算。然而,新的功能和规范以及增加的复杂性需要新的专门方法。

索引词 卷积神经网络,珊瑚礁
基于优化,架构定义,优化

一、引言

最近,机器学习已率先成为解决复杂问题的方法,尤其是在处理大量数据时。在信息社会时代,我们在日常活动中不断产生数据。这种情况暴露出大量问题,由于难度大,无法用特定算法有效解决。另一方面,机器学习算法已被证明可以有效地解决这些问题。

深度神经网络 (DNN) 可以在机器学习生态系统中找到。它们可以被视为经典人工神经网络的改进,它利用并行和分布式计算的新范式作为

在燕子人工神经网络中,通常需要多次训练和评估网络以找到合适的架构和超参数集,并考虑随机初始权重。目前,超参数搜索通常是根据经验法则和一般准则进行的,并结合手动实验 [1],但在使用 CNN 和 DNN 时,它仍然是一个主要挑战 [2]、[3]。此外,最佳超参数值集在训练过程中发生变化 (即学习率

减少了),这需要动态优化 [4]。
因此,CNN 架构的设计可以表示为超参数优化问题 [5]、[6],其中层数、神经元或步幅等可以使用进化算法等技术进行优化 [2] 或粒子群优化 [7] 等。

面临的**最大挑战之一在于评估给定 CNN 配置所需的很长时间,因为这意味着要训练整个模型。在每个卷积层上具有潜在大量过滤器的深度模型尤其如此。**因此,CNN 架构的自动设计最近成为一个相关的研究课题,探索将进化算法用于 CNN 的设计 [8]。另一方面,许多当代机器学习算法的性能关键取决于超参数 [6] 的特定初始化。CNN 中的学习算法需要从业者手动设置许多超参数的值,例如一般架构、学习率、正则化参数、批量大小、dropout 层中要丢弃的单位分数、过滤器的数量或步幅大小在卷积层中,仅举几例。为了评估特定配置的性能,必须训练和评估具有该配置的模型。然而,在 CNN 中,模型的训练可能需要相当长的时间,**而且搜索空间通常是非常高维的。**

作为上述示例,Simonyan, K. 和 Zisserman [9] 中使用的 CNN 模型架构与具有非常小感受野的过滤器相关联:3×3,但在其中一种配置中,他们也使用 1×1 个卷积滤波器。对于 3×3 卷积,卷积步幅固定为 1 个像素,空间填充为 1 个像素。层。空间池化由五个最大池化层执行(在 2×2 像素窗口上执行,步幅为 2),但并非所有卷积层后都跟有最大池化。这些超参数的确定是乏味的,但找到接近最佳的组合对于达到高性能水平至关重要。

在本文中,我们提出了 SCRODeep,这是一种用于**卷积神经网络超参数优化的自适应算法,但可以轻松扩展以处理循环神经网络等其他架构。**

SCRODeep 基于统计驱动的珊瑚礁优化算法 [10],一种对深度学习模型的超参数和网络结构进行编码的个体编码,一组专门设计的复制和变异算子以及有限状态机,负责在层之间建立有效的转换,并允许建立可行的学习模型。SCRO 元启发式的自参数化能力允许避免固定额外超参数的必要性。

最后,本文分为以下几个部分:第二部分介绍了问题的背景,第三部分介绍了 SCRODeep。第四节涉及实验装置和获得的结果。最后,第五节列举了得出的结论以及未来可能的工作路线。

二.背景

根据提到人工神经网络(ANN)概念的第一作者的说法,它们是一种模糊地受生物神经元启发的计算模型 [11]。自诞生以来,ANN 已被用于解决范围广泛的不同问题,包括控制系统 [12]、分类 [13]、人脸识别 [14] 和金融 [15]。由于计算能力的进步,随着 DNN [16] 和 CNN 等特定架构的出现,在 90 年代自然过渡到更多神经元和更复杂的架构

与 ANN 一样,这种计算模型在广泛的科学领域中非常流行,主要是因为该模型能够取得良好的结果,以及可以找到用于开发 DNN 的大量框架和编程库。Tensorflow [17] 和 PyTorch [18] 等框架,仅举几例,通过让开发人员和科学家更轻松地完成任务,为这些模型的兴起做出了贡献。从**深度网络中获益最多的领域之一是计算机视觉和图像识别 [19]、[20]。**在此范围内,图像通过形成深层架构的多个层,以便从图像中提取特征并对其进行分类。

深度网络占主导地位的其他例子**包括时间序列预测 [21]、视频识别 [22]、恶意软件检测和分类 [23] 以及音频识别 [24]。**

无论在哪个领域使用它们,配置和设计深度神经网络时最关键的时刻之一就是正确选择网络的超参数和架构。虽然这项任务通常是根据反复试验方案手动执行的,但一个看似合理的替代方法是使用元启发式算法和仿生算法来进行最佳参数选择。例如,Xin Yao [25] 为此目的研究了进化计算的应用,得出的结论是这些技术可以应用于构成人工神经网络训练的不同阶段:突触权重、架构和学习规则。

在处理经典神经网络的架构和参数调整时,可以在文献中找到许多方法 [26]。例如,[27] 的方法包括架构的定义以及权重的调整。此外,应该制定问题,以便可以通过进化算法来解决,其中神经元的数量包含在群体的个体中 [28]。与进化算法保持一致,[29] 提出了一种基于进化算法的超启发式方法来调整层的多项式类型、每个节点的数量以及体系结构中的层数。这种类型的方法也被用来提高性能 [30] 以及在进化搜索 [31] 中包括连接权重。

EvoDeep [32] 是一种致力于进化超参数和卷积神经网络架构的方法

网络为了最大化其分类精度,如

以及维护有效的层序列。它的核心是一种进化算法,其种群由个体组成,这些个体不仅编码 CNN 的超参数,还编码层的序列及其超参数。此外,该算法包括一个有限状态机,以确保生成的所有架构都是有效的层序列,因为它们对其输入和输出类型有限制。实验结果表明,EvoDeep 能够构建有效的 CNN 架构,进而在使用手写图像数据集时实现良好的准确性。然而,EvoDeep 还需要调整进化算法的不同超参数(例如变异和交叉概率),这为其应用带来了困难。

在本文中,我们提出了一种利用统计驱动的 CRO (SCRO) [10] 算法优化卷积神经网络的超参数和网络结构的新方法。这种方法包括对进化过程的超参数进行自适应微调,以自动化整个优化过程。

三、 SCRODEEP:卷积神经网络

使用统计驱动的参数化
珊瑚礁优化算法

深度学习 (DL) 模型的优势被其固有的复杂性和可变架构所抵消。虽然训练传统的机器学习算法(例如流行的随机森林)是一项简单而直接的任务,通常只需调整树的数量,但 DL 需要一种通常依赖于问题的不同特殊性的架构。例如,应用于计算机视觉的 DL 包括按规定顺序对多个层进行规范,负责评估多个数据更改。

该体系结构由许多分布的层组成,以表征解决特定问题的非线性关系。每层都有未定义数量的神经元、输出和用于初始化和激活的不同方法。这些超参数必须在神经网络的训练过程之前设置,从而创建一个大的搜索空间,其中包含未知且依赖于问题的最佳设置组合。

多层架构的设计存在各种限制。例如,特定层的输入形式必须与前一层的输出相匹配。当一个层期望输入向量(即完全连接的层)限制前面的层时,可能会发生这种情况。在这种情况下,重塑层不能留在全连接层,因为它始终提供至少 2 维的输出。模型的每个附加层的参数化也必须满足某些属性,这是另一个需要考虑的限制。

神经网络中的超参数和架构形成了一个广泛的搜索区域,其中定义了许多配置。

由于期望在特定任务的性能中最大化网络准确性,因此可以将正确配置的选择视为优化过程。一个

本文使用进化算法来执行元启发式搜索,以达到一种配置,从而最大限度地提高精确分类。

A. SCRO 算法概述

SCRO 元启发式 [10] 是对原始珊瑚礁优化 (CRO) 方法 [33]、[34] 的修改。CRO 模拟珊瑚繁殖(考虑有性和无性繁殖操作者)和珊瑚礁形成(珊瑚争夺空间)的过程。这些运算符在循环中执行,直到达到停止条件。

该算法的 SCRO 版本包括统计驱动的参数化,以避免手动调整进化过程中涉及的不同超参数。

SCRO 的主要步骤包含在算法 1 中,将在以下小节中进行描述。此外,SCRO 用于处理 CNN 训练配置的具体改编包含在第 III-B 和 III-E 节中。

算法 1统计珊瑚礁优化算法

输入:数据集。
输出: CNN。
1:算法的初始化。 2:同时不停止条件做3:无性繁殖。

4:有性生殖(外部和内部)。
5:幼虫定居。
6:评估新种群(珊瑚礁)。
7:捕食过程。 8: end
while 9:返回Best solution。

在 SCRO 中,主要成分是算法分析适应度值以确定如何应用不同的算子,动态地使行为适应种群状态,避免在每个优化问题中调整超参数。这样,解决方案的质量由 $f \in [0, 1]$ 给出。所有珊瑚的适应性决定了珊瑚礁中的种群状态(问题的潜在解决方案), $\{f_1j, f_2j, \dots, f_{Nj}j\}$,其中在第 j 代算法中种群数量为 N_j 个珊瑚。假设这是一个随机样本。

主要假设是适应度分布近似为高斯分布。这样,总体的方差可以估计为:

$$s_{fj}^2 = \frac{\sum_{i=1}^{Nj} (f_{ij} - \bar{f}_j)^2}{Nj - 1}, j = 1, \dots, \text{米}, \tag{1}$$

其中 f_{ij} 为第 j 代第 i 个个体的适应度, N_j 为第 j 代 N_j 个个体的个数, $\bar{f} = \sum_{i=1}^{Nj} f_{ij}/N_j$ 为所有个体的平均适应度值的一代。

B. Coral 编码和适应度函数

关于前者,有一个全局运算符O指定使用六个优化器中的哪一个。全局算子B表示神经网络一次接收多少样本,相应地更新其权重。请注意,已应用动态停止标准,如果在固定次数的迭代中没有提高获得的最佳准确度,则训练过程将停止。

C. 进化算子

$$f_{i1} \in (-f_1 - Sf_1, 1] \quad (2)$$
$$f_{ij} \in (-f_j + Sf_j, 1]. \quad (3)$$

1)外部有性繁殖类似于进化算法中的交叉算子。在 SCRO 中,该算法使用那些适应性验证的珊瑚:

$$f_{ij} \in (-f_j - Sf_j, 1]. \quad (4)$$

$$f_{ij} \in [0, \bar{f}_j - S f_j]_0. \quad (5)$$

5)掠夺:定居后,那些人
其适应度函数验证:

$$f_{ij} \in [0, \bar{f}_j - 2Sf_j], \quad (6)$$

D. Operators to evolve CNN architectures 由于 CNN

在这种情况下, EVODeep 中使用的运算符已被改编为所使用的 SCRO 算法的一部分。值得注意的是,运算符在两个级别上起作用:全局超参数和图层级别。原因在于个体的编码,它可以由不同数量的层组成,每个层具有不同的超参数,具体取决于它们的类型。以下是突变和重组算子的描述。

1)突变 :正如已经提到的,突变在两个层面上起作用:全局超参数和层。首先,每个人的全局超参数根据一致的方法发生变化,其中每个值重置为概率 $p = 0.5$ 的随机值 (前提是满足全局突变概率)。在数值和分类超参数的情况下,新的随机值分别基于值的范围和可能值的列表创建。

对于层,突变执行如下:1)两个连续的随机插入点

图层被选中。
2) 运算符插入一个有效的 n 层序列,其中 $n = \{1, 2, 3\}$ 并随机均匀选择,前提是最终层序列的长度不超过最大层数。

3)每一层的每个超参数都以与全局超参数类似的方式发生变异,即以一定的概率将值重置为随机值。

2) 重组 :再一次,重组在两个层面上起作用:全局和层。在全球层面,个体在统一交叉后重新组合。也就是说,超参数以 $p = 0.5$ 的概率成对交换。

另一方面,层按照切割和拼接方法重新组合,其工作方式如下:它随机选择满足以下条件的两个点 p_1 和 p_2 : $1 < p_1 < n$ 和 $1 < p_2 < m$,其中 n 和 m 是每个个体的层数,因此第一层和最后一层始终位于正确的位置。交换部件后层的顺序仍然有效。之后不超过最大层数

交换
最后,有必要对层的内部超参数进行重组,因为它们没有受到上述机制的影响。在这种情况下,它应用层级交叉,其工作方式如下:从第一层开始,交换位于相同位置 l 的两层 L_l 和 L 的两个超参数 p_i 和 p_j ,直到最短个体的倒数第二层是到达。

每个人的最后一层也被交叉,因为它们总是共享相同的层类型。

E. 限制算法 :构建有效的层序列与 EVODeep 类似,层之间所有可能的转换都通过有限状态机 (FSM) 建模。

当个体交叉和变异时,FSM 用于避免无效结构或合并新的一致层。此外,该 FSM 还用于生成具有最小和最大长度的所有可能路径,从中随机选择一个以初始化第一个种群。

四、实验

一系列实验允许测试 SCRODeep,证明其能够自动调整两个超参数

范围	取值范围
内核初始化程序Kli	Uniform, LeCun uniform, Normal, Zero, Glorot 正常,Glorot 制服,He 正常,他制服
激活函数Ai	ReLU, Softmax, Softplus, Softsign, Tanh, Sigmoid, Hard Sigmoid, Linear
单位Ui	{10, 20, 30, ..., 500} {0.1, 0.2, 0.3, ..., 0.8}
Rate (dropout) Pi	{5, 10, 15, ..., 50} {3, 5, 7, ..., 15} {2, 3, 4,
No. filters Fi Kernel	5, 6} {2, 3, 4, 5, 6}
size KSi Pool size P	
Si Strides Si	
优化器Oi	亚当,SGD,RMSProp,Adagrad Adamax,Nadam {100, 200,
批量大小Bi/层数N	300, ..., 1000} {3, 4, 5, ..., 9}

表 1:启发式搜索中涉及的每个超参数的值范围。表中显示了层级和全局超参数。

和卷积神经网络模型的架构。

此外,统计驱动的方法允许将整个工作流程程留在算法手中。以下小节描述了实验中使用的数据集、实验设置和获得的结果。

A. 数据集

众所周知的 MNIST 数据集 [35] 用于运行实验。这一决定的原因在于文献中提出的大量方法,可以对结果进行客观比较。此外,该数据集还用于训练和测试 EvoDeep [32],从而可以与 SCRODeep 进行进一步比较。MNIST 数据集包含一个包含 60,000 个训练示例和 10,000 个用 28x28 像素图片中的灰度级表示的手写数字测试示例的数据库。

B. 实验设置

SCRODeep 运行了 30 次以获得具有代表性的解决方案,以公平地评估所提出的方法。启发式搜索负责自动提供超参数和网络架构的关系,从而最大化模型的结果

为了精度,需要为涉及的每个超参数提供一个取值范围,以缩小搜索空间,从而避免达到可以提前排除的值,并限制执行整个搜索算法所涉及的时间。

表 1 总结了所考虑的不同层的每个超参数以及全局超参数的测试值范围。在第一种情况下,不同的内核初始化器、激活函数、单元数或过滤器数是在搜索过程中探索不同值的一些超参数。对于全局超参数、不同的优化器、批量大小 (由于内存限制,限制为最大 1,000)和网络的层数大小。在里面

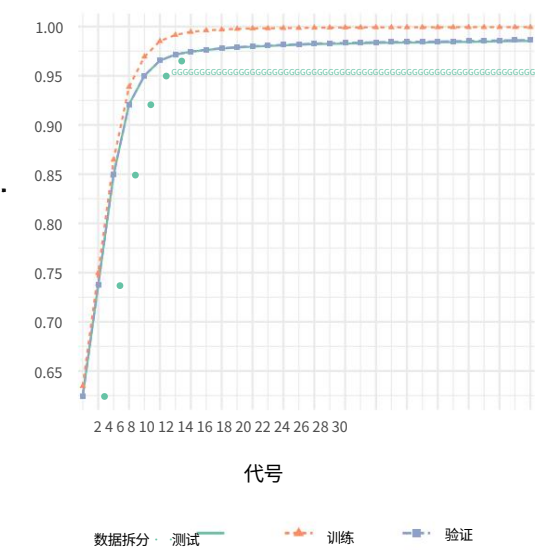


图 2:针对所有执行的三个数据拆分的代际适应度演变。

对于 epochs 的数量,设置了一个停止标准,以便在最后 10 个 epochs 中验证精度没有提高时停止训练步骤,建立最多 100 个 epochs。通常,为每个超参数建立了广泛的值,试图不限制可行解决方案的范围来处理手头的问题。

为了进行评估,MNIST 数据集被分成两部分,提供 50% 的示例用于训练,50% 用于测试目的。对训练数据集进行了进一步划分,将这批实例的 10% 分配给验证集。最后一组实例用作特定珊瑚或模型质量的外部度量,并用作适应度函数。

关于 SCRO 参数化,虽然所有的超参数都是统计确定的,但珊瑚礁的大小需要手动固定。在这些实验中,它是 6×6 ,这个大小足以让算法在适当的时间内给出答案。换句话说,通过寻求一种在计算速度和个体质量之间折衷的解决方案,珊瑚礁的大小已经确定。然而,如果需要,可以考虑更大的值。

C. 结果

SCRODeep 寻求获得卷积神经网络模型的超参数。在本节中,描述了使用 MNIST 数据集执行该算法后获得的结果,展示了所提出的方法如何能够定义超参数和层序列,从而最大限度地提高模型的性能。

首先,对验证集的使用进行了评估,以检查它是否是指导启发式搜索的有效程序,以及这部分获得的结果是否可以外推到测试集。图 2 包括三个珊瑚礁在每一代产生的平均适应度值

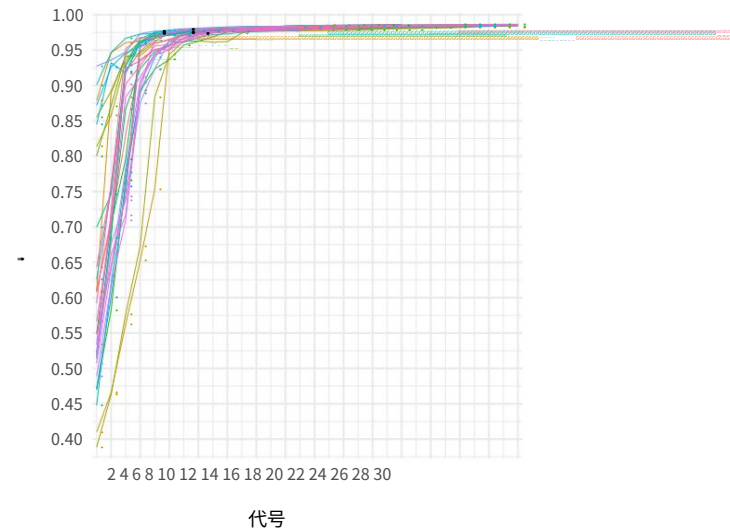


图 3:显示所有执行运行数据的测试拆分中的适应度进化。

数据拆分。它们显示出非常相似的趋势,在测试和验证集中实现的准确性几乎相同。这证明所使用的验证集构成了一个有用且无偏见的工具,用于指导启发式搜索朝着能够实现高精度值的强大模型发展。

执行的所有执行都在图 3 中表示。在该图中,可以观察到仅 6 代就足以收敛平均准确度 (在测试集中)超过 92% 的总体。从这一点来看,准确率略有提高,接近 99%。类似地,图 4 以箱形图表示测试数据集的准确性,可以更好地了解解决方案在搜索过程中的分布情况。前三代将各种各样的个体引入珊瑚礁,表现出高偏差水平,并且相当多的解决方案无法达到良好的准确性水平。从这一点来看,就准确性水平而言,多样性降低了,整个珊瑚礁由在以下迭代中成功改进的准确解决方案组成。

寻找新的和改进的解决方案也有望增加它们的复杂性,从而增加评估每一代新生成的个体所需的时间。SCRODeep 执行增量搜索过程,从评估时间较短的低复杂性解决方案开始。在接下来的几代中,运营商允许以可控的方式增加解决方案的复杂性 (即增加层数) 。

达到适得其反的复杂性且没有改善迹象的解决方案将被丢弃,限制启发式搜索过程以避免不必要的复杂模型。

从最佳执行 (在验证集中达到最高值的那个)获得的结果如表二所示。SCRODeep 的平均准确率达到 98.36%,最佳运行达到 98.67%。如果这些值显示出改进

运行测量		训练99.82%	验证	测试
		最小值	平均值	98.11%
SCRO深30				98.11%
		99.95% ± 0.06	98.38% ± 0.16	98.36% ± 0.15
		99.78% ± 0.21	98.26% ± 0.36	98.24% ± 0.36
深进化	30			97.77%
				98.50%
				98.71%

表 II:SCRODeep 获得的结果总结,考虑到每次执行的验证集中获得的最佳个体的准确性。下半部分允许与 EvoDeep 进行比较。

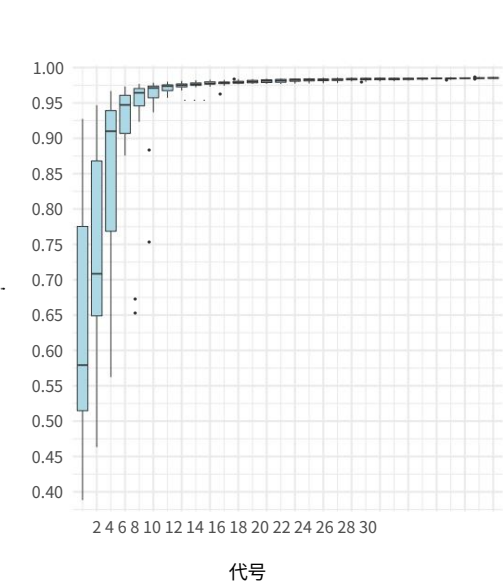


图 4:针对所有执行分组的测试拆分中几代人的适应度演变。

与 EvoDeep 相比。

SCRODeep 获得的方差明显低于 EvoDeep 使用 Levene 检验获得的方差 [36] (p 值 = 0.001,F = 14.873)。由于 EvoDeep 和 SCRODeep 的正确分类率 (CCR) 都不正常,我们对具有不同方差和平均值之间微小差异的小非正常样本进行了 Wald-Wolfowitz 检验 [37]。根据该测试,SCRODeep 与 EvoDeep 的累积分布存在统计学显著差异 (p 值 = 0.038,z = -2.073)。这个结果强调了 SCRODeep 提供的增强功能,自动调整超参数和层,而不需要定义一组元启发式超参数。

最后,图 5 表示在验证集的准确性方面找到的最佳执行的层结构和超参数。它由两个 Convolution2D 和 MaxPooling2D 循环组成,一个扁平层和一个最终的 10 个单元的密集层,直接连接到输出。在全局超参数级别,使用 300 个实例的批量大小和 Adamax 优化器。与使用相同层类型的最先进模型相比,该架构类似于 LeNet-4 [35],尽管存在一些差异。

实现的精度也相似,因为 LeNet-4 以 1.1 的错误率执行分类。

五. 结论和未来工作

深度学习被认为是一种强大的工具,能够处理各种复杂的大型问题。属于这一范围的广泛技术可以解决精确有效模型的问题。然而,它们高度依赖于特定问题的定义。因此,大量的架构已被证明在特定领域是成功的。这些模型的结构或超参数的设计并非易事。SCRODeep 旨在解决通过执行元启发式搜索来决定这些因素的任务,其中个体代表潜在的超参数集和网络结构,以寻求最大化模型的准确性。所提出的方法通过自动化整个搜索过程而无需调整搜索算法本身的超参数,从而改进了先前提出的称为 EvoDeep 的方法。通过使用珊瑚礁优化算法的统计驱动版本,SCRODeep 允许自动定义要使用的最合适的模型。

1 Reshape
Output shape: 28x28
2 Convolution2D
Activation: ReLU Kernel initializer: He normal
40 filters Kernel size: 9
3 MaxPooling2D
Pool size: 3
4 Convolution2D
Activation: ReLU Kernel initializer: He normal
40 filters Kernel size: 12
5 MaxPooling2D
Pool size: 4
6 Flatten
7 Dense
Activation: Softmax Units: 10

图 5:找到的最佳个体的架构和超参数。

实验表明,所提出的方法能够

提供有竞争力的模型,能够达到具有统计差异的高准确率。在未来的工作中,我们的目标是通过包含能够处理更复杂问题的新层以及实施能够减少评估次数的技术来增强 SCRODeep 的可能性,因为这是算法中计算成本最高的元素。

致谢

这项工作由以下赠款共同资助:Co munidad Autonoma de Madrid under grant S2018/TCS-4566 (CYNAMON: Cybersecurity, Network Analysis and Monitoring for the Next Generation Internet);西班牙科学教育和竞争力部 (MINECO) 和欧洲区域发展基金 (FEDER) 资助 TIN2017-85727-C4-3-P (DeepBio);TIN2017-85887-C2-1-P 和 TIN2017-90567-REDT 和Consejer´ a de Econom´ a, Conocimiento, Empresas y Universidad of the Junta de An daluc´ a (西班牙)资助 UCO-1261651。

参考

[1] JS Bergstra,R. Bardenet,Y. Bengio and B. Kegl,“超参数优化算法”,神经信息处理系统进展 24,J. Shawe-Taylor,RS Zemel,PL Bartlett,F. Pereira and KQ Weinberger 编着。Curran Associates, Inc.,2011 年,第 2546–2554 页。

[2] I. Loshchilov and F. Hutter,“用于深度神经网络超参数优化的 Cma-es”,arXiv 预印本 arXiv:1604.07269,2016 年。

[3] H. Mendoza,A. Klein,M. Feurer,JT Springenberg 和 F. Hutter,“To wards automatically-tuned neural networks”,自动机器学习研讨会,2016 年,第 58-65 页。

[4] M. Jaderberg,V. Dalibard,S. Osindero,WM Czarnecki,J. Don ahue,A. Razavi,O. Vinyals,T. Green,I. Dunning,K. Simonyan 等人,“基于人口的培训神经网络”,arXiv 预印本 arXiv:1711.09846, 2017 年。

[5] J. Bergstra and Y. Bengio,“超参数优化的随机搜索”,机器学习研究杂志,卷。13,没有。2 月,第 281–305 页,2012 年。

[6] J. Bergstra,D. Yamins and DD Cox,“使模型搜索成为科学:视觉架构数百个维度的超参数优化”,第 30 届国际机器学习国际会议论文集 - 卷28,爵士。

ICML 13。JMLR.org, 2013, p. l–115–l–123。

[7] F. Ye,“深度神经网络基于粒子群优化的自动参数选择及其在大规模高维数据中的应用”,PLOS ONE, 卷。12,没有。12,页。e0188746,2017 年 12 月。

[8] T. Hinz,N. Navarro-Guerrero,S. Magg and S. Wermter,“加速深度卷积神经网络的超参数优化”,国际计算智能与应用杂志,第一卷。17,没有。02,页。1850008, 2018。

[9] K. Simonyan 和 A. Zisserman,“用于大规模图像识别的超深卷积网络”,arXiv 预印本 arXiv:1409.1556,2014 年。

[10] AM Duran-Rosal,PA Gutierrez,S. Salcedo-Sanz and C. Herv as-Mart´ nez,“一种统计驱动的珊瑚礁优化算法,用于优化时间序列的尺寸缩减”,Applied Soft Computing,卷。63,第 139–153 页,2018 年。

[11] WS McCulloch and W. Pitts,“神经活动中内在思想的逻辑演算”,数学生物物理学公报,卷。5,没有。4,第 115–133 页,1943 年 12 月。

[12] K. Hunt,D. Sbarbaro,R. Zbikowski and P. Gawthrop,“控制系统的神经网络 一项调查”,Automatica,卷。28,没有。6,第 1083 – 1112 页,1992 年。

[13] GP Zhang,“用于分类的神经网络:调查”,IEEE Transactions on Systems, Man, and Cybernetics, C 部分 (应用和评论),卷。30,没有。4,第 451–462 页,2000 年 11 月。

[14] E. Hjelmas and BK Low,“人脸检测:一项调查”,计算机视觉和图像处理,卷。83,没有。3,第 236–274 页,2001 年。

[15] A. Bahrammirzaee,“金融人工智能应用的比较调查:人工神经网络,专家系统和混合智能系统”,神经计算与应用,卷。19,没有。8,第 1165–1195 页,2010 年 11 月。

[16] J. Schmidhuber,“神经网络中的深度学习:概述”,神经网络,卷。61,第 85–117 页,2015 年。

[17] M. Abadi,P. Barham,J. Chen,Z. Chen,A. Davis,J. Dean,M. Devin,S. Ghemawat,G. Irving,M. Isard 等人,“Tensorflow:a大规模机器学习系统。”在 OSDI 中,卷。2016 年 16 日,第 265–283 页。

[18] A. Paszke,S. Gross,S. Chintala and G. Chanan,“Pytorch”,2017 年。

[19] K. Simonyan 和 A. Zisserman,“用于大规模图像识别的非常深的卷积网络”,CoRR,卷。abs/1409.1556,2014 年。

[20] K. He,X. Zhang,S. Ren and J. Sun,“用于图像识别的深度残差学习”,IEEE 计算机视觉和模式识别会议,2016 年,第 770–778 页。

[21] Z. Wang,W. Yan and T. Oates,“使用深度神经网络从头开始进行时间序列分类:强大的基线”,2017 年国际神经网络联合会议 (IJCNN),2017 年 5 月,第 1578 页– 1585。

[22] T. Afouras,JS Chung,A. Senior,O. Vinyals and A. Zisserman,“深度视听语音识别”,IEEE 模式分析和机器智能汇刊,第 1–11 页,2018 年。

[23] Z. Yuan, Y. Lu, and Y. Xue, “Droiddetector: android malware characterization and detection using deep learning,” 清华科技, vol. 21,没有。1,第 114–123 页,2016 年 2 月。

[24] W. Xiong,L. Wu,F. Alleva,J. Droppo,X. Huang 和 A. Stolcke,“Microsoft 2017 会话语音识别系统”,2018 年 IEEE 声学、语音和信号处理国际会议(ICASSP),2018 年 4 月,第 5934–5938 页。

[25] X. Yao,“进化的人工神经网络”,IEEE 会刊,卷。87,没有。9,第 1423–1447 页,1999 年 9 月。

[26] M. Srinivas and L. Patnaik,“使用遗传算法学习神经网络权重 - 通过搜索空间缩减提高性能”,神经网络,1991 年,1991 年 IEEE 国际联合会议。IEEE,1991 年,第 2331–2336 页。

[27] JR Koza and JP Rice,“神经网络的权重和架构的遗传生成”,神经网络,1991 年,IJCNN 91-西雅图国际联合会议,卷。2。IEEE,1991 年,第 397–404 页。

[28] FH-F.梁,H.-K.林,S.-H.玲和 PK-S. Tam,“使用改进的遗传算法调整神经网络的结构和参数”,IEEE 神经网络交易,卷。14,没有。1,第 79–88 页,2003 年。

[29] J. Gascon-Moreno,S. Salcedo-Sanz,B. Saavedra-Moreno,L. Carro- Calvo and A. Portilla-Figueras,“一种基于进化的超启发式方法,用于最优构造组方法数据处理网络”,信息科学,卷。247,第 94–108 页,2013 年。

[30] X. Yao and Y. Liu,“用于进化人工神经网络的新进化系统”,IEEE 神经网络交易,卷。8,没有。3,第 694–713 页,1997 年。

[31] A. Abraham,“元学习进化人工神经网络”,神经计算,卷。56,第 1–38 页,2004 年。

[32] A. Mart´ n,R. Lara-Cabrera,F. Fuentes-Hurtado,V. Naranjo and D. Ca macho,“EvoDeep:自动深度神经网络参数化的新进化方法”,Journal of Parallel and分布式计算,卷。117,第 180–191 页, 2018 年。

[33] S. Salcedo-Sanz,J. Del Ser,I. Landa-Torres,S. Gil-Lopez and A. Portilla-Figueras,“珊瑚礁优化算法:解决困难优化问题的有效元启发式算法”,载于第 15 届应用随机模型和数据分析国际会议论文集 (ASMDA2013),Matar´ o,2013 年,第 751–758 页。

[34] S. Salcedo-Sanz,J. Del Ser,I. Landa-Torres,S. Gil-Lopez and J. Portilla-Figueras,“珊瑚礁优化算法:一种有效解决优化问题的新型元启发式”,科学世界杂志,卷。2014 年,2014 年。

[35] Y. LeCun,L. Bottou,Y. Bengio,P. Haffner 等人,“基于梯度的学习应用于文档识别”,IEEE 会刊,卷。86,没有。第 11 页,第 2278–2324 页,1998 年。

[36] H. Levene,“方差相等性的稳健检验”,对概率和统计的贡献。纪念哈罗德·霍特林 (Harold Hotelling) 的论文,第 279–292 页,1961 年。

[37] RC Magel and SH Wibowo,“比较 wald wolfowitz 和 kolmogorov-smirnov 测试的能力”,Biometrical Journal,vol。39,没有。6,第 665–675 页,1997 年。