

智能优化算法及其应用

实践和文献调研

遗传算法在神经网络和深度学习方面的应用

张蔚桐 2015011493 自 55

摘要

遗传算法（GA），进化算法（EA）和深度学习是三种目前常见的几乎不依靠专家经验进行学习和优化的算法和结构。同时，又由于神经网络的结构复杂，常规手段的结构调整往往基于一些不是很可靠的专家经验。因此，将进化算法（遗传算法）应用于深度神经网络（神经网络）的学习过程和结构优化过程便成为了一个比较自然的想法。本文调研了几篇有关此方向的文献，并粗略分析了使用进化算法对神经网络进行优化的发展和问题所在。

1 简介

（人工）神经网络的概念于 20 世纪中叶提出并完善，1958 年，Rosenblatt 等人提出了感知器模型，第一个有效的多层神经网络也在 1965 年代构造成功。然而，二十世纪中叶的神经网络研究的问题一般均是高层次的规则学习等问题，直到 20 世纪 80 年代，随着 BP 算法的提出，神经网络开始底层化。然而，由于当时计算资源受限，网络结构过于单一等原因，20 世纪 90 年代起以神经网络为代表的一系列人工智能方法开始进入寒冬期。本世纪初，随着 CNN 等方法的提出和并行计算的成熟化，神经网络模型获得更大的深度，并衍生出了很多复杂的模型。当前，深度学习和神经网络是人工智能领域的热点之一。

遗传算法和进化计算的概念也是于 20 世纪中叶被初次提及。在 20 世纪 70 年代，这个算法随着一些求取复杂工程问题的应用而广为人知。当神经网络方法正在低层次化，由之前的规则变为数值计算和优化问题时，正是遗传算法发展的比较成熟之时。因此，通过遗传算法来优化神经网络的思路被很快提出 [1]。然而不论是遗传算法还是神经网络，相比于之前的优化算法和机器学习算法，都需要较高的计算资源。因此这种尝试也随着人工智能的寒冬期而一起消失了。

21 世纪以来，深度学习方法首先成为了人工智能领域内的热点，然而，由于模型的增大，80 年代使用遗传算法来优化神经网络的思路变得不可行，因此学界开始采用新的子网络方式来研究深度学习网络，这方面的文章可见 [2],[3],[4],[5]。

虽然遗传算法在调整深度学习结构方面有了一定的应用，然而从目前的情况来看，遗传算法的模型调整收到计算资源的限制及其他原因仍然难堪重用。另外，近年来基于深

度网络的强化学习成为了人工智能领域的一个新热点，而遗传算法，进化计算等算法由于本身的“进化”特性，势必在强化学习领域有所作为。

2 遗传算法用于神经网络结构的调整及其问题

1989 年，David J. Montana 等人进行了一次有益的通过遗传算法来调整神经网络结构的测试，他们通过对神经网络神经元之间的连接关系进行编码，调整神经元之间的连接关系，得到性能更好的网络结构。他们的思想可以从两方面去理解。如果这个调整之后直接进行测试而不采用 BP 算法等来进一步训练优化结构，则他们的调整实际上是采用遗传算法 GA 来代替反向传播算法。同时，权重的选择由原来 BP 算法的基于梯度下降的连续值变为基于遗传算法的离散值。这种尝试在尤其是当时数据量不大的情况下是有意義的，他在一定程度上通过对权重的约数可以保证模型不会过学习，而采用 GA 算法可以保证模型不陷入局部极小。然而当数据量增大之后，首先样本的局部极小实际上已经和真正的最小值相差很小了（由于大量样本的平均效应，有随机噪声导致的方差明显减小）。其次在大数据情况下采用的基于 GA 的算法未免时间过长。

从另一个角度理解 GA 算法在神经网络结构优化中的作用可能更可行。通过遗传算法的优化，可能找出一个更适合采用 BP 算法进行优化的网络结构，在这个理解模式下，一个网络在经过 GA 算法的优化过后，还需要进一步的使用 BP 算法进行优化。GA 算法侧重对于结构的搜索而 BP 算法侧重对于已知结构的微调

然而，随着网络深度和宽度的逐步增大，这种全局范围内的编码方式出现了问题。首先，巨大的网络深度必然导致每一次的评估时间的增长。其次，巨大的网络深度和宽度必然导致编码长度明显增长，在这种情况下，通过变异等方式产生的优化效果将变得很不明显，而且计算代数相比较短的染色体长度明显提升。因此采用全局范围内的编码行为将导致大量的时间开支。

为了解决这个问题，目前主要的研究手段是将整个深度网络划分成几个相同的或相似的部分进行优化，相同的部分之间共享相似的结构。这种思路和 CNN 的权值共享的思路是致似的。只是 CNN 共享的是权值而当前手段共享的是模型的结构。这些问题在以 [2] 的文章中有着很多表述。然而困难之处在于，相比于 CNN 这种模型规则整齐的网络结构，GA 构造出来的新网络结构往往比较复杂且自由，因此利用反向传播的难度将明显上升。

为了保证网络结构模型的规整性，一个简单的思路是在当前的 CNN，全连接层的抽象高度基础上来调整网络结构。通过某种编码方式可以使得遗传算法可以对网络每一层的性质和对应的超参数进行优化。然而在这种情况下，如果我们只要求 GA 算法调整网络的超参数，则通过引文 [6] 所示，传统的贪心搜索方式已经有了很好的效果，没有采用遗传算法的必要。如果将层的性质（如是卷积层还是全连接层），则在编码上会遇到比较大的问题，包括交叉，变异的操作将变得不好理解。

上述就是遗传算法在神经网络结构优化上的应用的文献调研，通过调研我们知道，将遗传算法用于神经网络是一个比想象起来还要复杂的工程，而提出一种适用于网络形态的有效的简洁的编码是至关重要的一环。

3 遗传算法用于网络学习过程的实践

本节中，作为一个补充的实验，我们采用了 GA 算法代替传统的 BP 算法去学习一个两个隐层的全连接网络，每层网络的宽度为 100 个神经元，神经元激活函数选择为 relu 函数。在 GA 算法的优化过程中，我们遇到了如下的几个显著的困难点：

- 权重权值过小，GA 算法稍稍给出较大的权值易导致后面的网络的数值相当大 这个问题我们采用的处理方式是将 GA 算法给出的权值首先乘系数，其次做边界检测。保证 GA 环节权值的大小适合 GA 进行优化，而网络的权值适合进行比较合理的网络评估过程。
- 优化时间过长 本次优化中我们优化时间长达 1 个月，我认为这也是 GA 算法之所以没有取代 BP 算法的主要问题。

通过在 60000 个样本上进行优化，我们可以达到在训练集上的准确率远超使用 CNTK 深度学习框架利用 BP 算法进行的优化，在 10000 个样本上进行的先期实验表明其正确率甚至可以达到 100%，这也验证了模式识别等课程上提到的 2 隐层神经网络可以拟合任何集合边界的定理。在大小为 10000 个样本的测试集上，测试的正确率达到了 98%，稍稍高出使用 BP 算法的测试正确率，相比于训练集的正确率提高，测试集的正确率提高的相对小，这可能是因为出现了一定程度的过拟合导致的。但是通过一些测试（例如使用 BP 算法强制过拟合的对比）中我们发现，使用 GA 出现的过拟合情况要相对比 BP 的过拟合情况好，说明 GA 算法在优化的过程中确实通过染色体的交换等方式发现了数据集的基本特征，而不是类似 BP 算法在强制过拟合之后可能出现网络病态的问题。

然而，正如前面所说，这种算法实际上并不能投入到实际应用中，因为其时间开销远远大于传统的 BP 方法。这种算法仅能做一些理论层面上的解释。

具体的代码实现可见<https://github.com/ZeroWeight/IOAA/Optim>

参考文献

- [1] David J. Montana and Lawrence Davis. Training feedforward neural networks using genetic algorithms. *International Joint Conferences on Artificial Intelligence*, 1989.
- [2] Joshua D. L. Deep learning using genetic algorithms. Master's thesis, Rochester Institute of Technology, 2012.
- [3] Lingxi Xie and Alan Yuille. Genetic cnn, 2017.
- [4] Risto Miikkulainen, Jason Liang, Elliot Meyerson, Aditya Rawal, Dan Fink, Olivier Francon, Bala Raju, Hormoz Shahrzad, Arshak Navruzyan, Nigel Duffy, and Babak Hodjat. Evolving deep neural networks, 2017.
- [5] S.H. Ling F.H.F. Leung, H.K. Lam and P.K.S. Tam. Tuning of the structure and parameters of a neural network using an improved genetic algorithm. *IEEE Transactions on Neural Networks*, 2003.
- [6] Junqi Jin. *Deep Learning Model Structure Optimization and Applications in Image Understanding*. PhD thesis, Tsinghua University, 2016.