

基于 LM 算法的 MLP 模型及其应用*

李南星^{1,2} 盛益强¹ 倪宏¹

(¹ 中国科学院声学研究所 国家网络新媒体工程技术研究中心 北京 100190

²中国科学院大学 北京 100049)

摘要: 准确率是衡量 MLP 模型性能的典型且广泛采用的指标。为了提高 MLP 模型的准确率,本文分别设计 9 种激活函数组合、11 种模型隐层宽度的 MLP 模型,通过对比分析 MLP 模型的验证集 MSE 确定最优的 MLP 模型结构;分别采用 LM、BFGS、Adam 和 SGD 算法训练 MLP 模型,通过对比分析 MLP 模型的测试集 MSE,确定最优训练算法。实验结果显示,基于 LM 算法训练的 MLP 模型测试集 MSE 达到 $2.756e-4$,比 BFGS 算法降低了 23.4%,比 Adam 算法降低了 85.6%,比 SGD 算法降低了 97.2%。超声治疗作为一种 MLP 模型典型应用,获取颅骨的准确声学参数是非侵入性经颅聚焦超声治疗成功的关键,为此提出了一种使用 MLP 模型基于 CT 图像准确预测颅骨声参数的方法。

关键词: 多层感知器, Levenberg - Marquardt 算法, 反向传播算法, 经颅聚焦超声仿真

A Multilayer Perceptron Model Based on Levenberg - Marquardt Algorithm with its Applications

LI Nanxing^{1,2}, SHENG Yiqiang¹, NI Hong¹

(¹National Network New Media Engineering Research Center, Institute of Acoustics,
Chinese Academy of Sciences, Beijing, 100190, China,

²University of Chinese Academy of Sciences, Beijing, 100049, China)

Abstract: Accuracy is a typical and widely used measure of the performance of MLP models. In order to improve the accuracy of MLP model, MLP models of nine different activation function combinations and MLP models of eleven different hidden layer widths were designed in this paper. The optimal MLP model structure was determined by comparing the MLP model validation set MSE. The LM, BFGS, Adam and SGD algorithms are respectively used to train the MLP model, and the optimal training algorithm is determined by comparing the test set MSE of the MLP model. The experimental results show that the MSE of MLP model based on LM algorithm reaches $2.756e-4$, which is 23.4% lower than BFGS algorithm, 85.6% lower than Adam algorithm and 97.2% lower than SGD algorithm. As a typical application of MLP model, obtaining accurate acoustic parameters of skull is the key to success of transcranial focused ultrasound. To solve this problem, we propose a method to accurately predict skull's acoustic parameters based on CT images using MLP model.

Keywords: Multi-layer Perceptron, Levenberg - Marquardt Algorithm, Error Back Propagation Algorithm, Transcranial focused ultrasound

0 引言

多层感知器(multilayer perceptron, MLP)也叫深度前馈网络,是一种典型且被广泛应用的深度学习模型。

主要将其用于拟合一个映射函数,在采用非线性激活函数情况下能拟合非线性映射函数^[1]。若要 MLP 拟合比较复杂的映射函数或者以较高的准确率拟合映射函数,需要解决两个关键问题:模型结构的设计和模型的训练。

模型结构主要是指模型的深度和每一层的宽度以及激活函数。MLP 模型能自动从数据中提取特征表示及抽象级别^[2],通常来说深度越深使得每层宽度能够使用更少节点数,但是由于梯度消失或梯度爆炸会导致难以训练,即使模型深度较浅仍能通过增加模型宽度提高准确率;一般而言宽度越宽模型拟合能力越强,但越容易过拟合;各种激活函数的优劣之分并没有定论,需要根据应用场景、所用数据集选取。因此针对特定应用场景为了得到较理想的模型结构,需要通过多次实验进行调整。卷积神经网络典型的模型结构如 AlexNet^[3]、VGGNet^[4]、Google Inception Net^[5]和 ResNet^[6],这些模型结构均在 Image Net 大规模视觉识别挑战赛 ILSVRC 分类项目中取得了令人惊艳的成绩。其中 AlexNet 获得了 2012 年冠军(8 层神经网络),VGGNet 获得了 2014 年亚军(19 层神经网络),Google Inception Net 获得了 2014 年冠军(22 层神经网络),ResNet 获得了 2015 年冠军(152 层神经网络)。

模型的训练主要是指选取优化算法训练模型。目前的优化算法主要有二阶优化算法和一阶优化算法。在一阶优化算法中,随机梯度下降(stochastic gradient descent,SGD)算法与其变种是深度学习里应用最广泛的算法,随机梯度下降算法十分简单且计算复杂度低,但是在一些情况下收敛十分缓慢,比如在误差曲面的平坦区域由于梯度较小导致收敛缓慢,在较深的神经网络中由于梯度消失导致收敛缓慢等。由于学习率对优化算法的性能有显著影响,因此有研究学者基于随机梯度下降算法,提出了几种自适应学习率算法,其中自适应矩估计(adaptive moment estimation,Adam)^[7]算法由于在实践中性能较好从而十分流行。

相比较于一阶优化算法,二阶优化算法使用二阶导数进行优化。其中 LM(Levenberg-Marquardt)^[8,9]算法通过引入阻尼项对 Hessian 矩阵正则化,使得 Hessian 矩阵正定不仅保证 Hessian 矩阵可逆,而且在靠近鞍点处避免梯度朝错误方向移动。BFGS 算法通过梯度迭代近似 Hessian 矩阵的逆矩阵,从而减少计算量。

本文在经颅聚焦超声的应用场景下,设计 MLP 模型预测颅骨声参数。本文贡献有如下三点:①在 2.1 中通过实验对比分析模型隐层宽度对模型预测准确率的影响,确定最优的模型隐层宽度;②在 2.2 章节中通过实验对比分析模型隐层激活函数对模型预测准确率的影响,确定最优的模型激活函数组合;③在 2.3 章节中通过实验对比分析优化算法对模型预测准确率的影响,确定最优的训练算法。最终将模型预测结果应用于声场模拟中,通过对比分析发现,无论在收敛速度、模型预测准确率上,LM 算法均优于 BFGS、Adam 以及 SGD 算法,从而验证了 LM 算法训练 MLP 模型的有效性。

1 模型结构

本文所构造的多层感知器整体结构如图 1 所示,由一个输入层、一个隐层和一个输出层构成。输入层由 7×7 个神经元(即由 7×7 大小窗口采样的 CT 值)构成,隐层是 9 个神经元组成的全连接层,输出层是 3 个神经元组成的全连接层。输入和输出都经过归一化预处理。

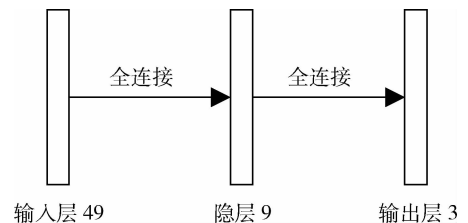


图 1 模型整体结构

1.1 数据集

采用中科院深圳先进研究院的 176 份颅骨 CT 扫描文件作为输入,基于 Abury 等人提出的模型^[10]计算三种颅骨声参数 ρ 、 c_p 、 α_p 作为输出,如式(1)所示。

$$\begin{cases} \Phi = 1 - \frac{H}{1000} \\ \rho = \Phi \times \rho_{skull, \min} + (1 - \Phi) \times \rho_{skull, \max} \\ c_p = \Phi \times c_{p, skull, \min} + (1 - \Phi) \times c_{p, skull, \max} \\ \alpha_p = \alpha_{p, skull, \min} + \Phi^{0.5} \times (\alpha_{p, skull, \max} - \alpha_{p, skull, \min}) \end{cases} \quad (1)$$

其中, H 是颅骨 CT 值, Φ 是孔隙率, $\rho_{skull, \min}$ 、 $c_{p, skull, \min}$ 和 $\alpha_{p, skull, \min}$ 分别为颅骨中的最小密度、最小纵波声速、最小纵波吸收系数, $\rho_{skull, \max}$ 、 $c_{p, skull, \max}$ 和 $\alpha_{p, skull, \max}$ 分别为颅骨中的最大密度、最大纵波声速、最大纵波吸收系数。

随机选取 8 个样本用作测试集, 余下 168 个样本按 1: 9 比例随机划分为验证集和训练集。其中训练集用于训练 MLP 模型, 验证集用于在训练过程中验证 MLP 模型的泛化能力防止过拟合。

2 实验

2.1 MLP 模型不同激活函数对比分析

隐层及输出层的激活函数选取能在很大程度上影响 MLP 模型预测声参数的误差及其收敛速度。本文构造了 9 个初始化权值矩阵相同但激活函数不同的 MLP 模型, 采用均方误差 (mean-square error, MSE) 评估 MLP 模型预测声参数准确率。表 1 列出了每个 MLP 模型最小 MSE。

由表 1 可知, 选取合适的输出层激活函数十分重要, 各激活函数按 MSE 升序排列依次是 tanh - > identity - > sigmoid。tanh 激活函数的 MSE 比 sigmoid 减小了约 3 个数量级, 但当隐层激活函数为 identity 时例外, 因为 identity 是线性函数, 使得多层模型等效于单层模型, 所以 MSE 没有得到显著的改善。

当输出层激活函数为 tanh 或 identity 时, 将隐层激活函数设为 tanh 使得训练集 MSE 最小即模型获得最优的拟合能力; 将隐层激活函数设为 sigmoid 使得验证集 MSE 最小即模型获得最优的泛化能力。综合考虑模型的拟合能力和泛化能力, 应该将隐层及输出层激活函数设为 tanh。

2.2 隐层规模性能对比分析

理论上只要隐层规模足够大, 多层感知器能够拟合任意连续函数^[11,12]。换言之, 若隐层节点数越多则准确率越高, 但多层感知器的训练时间也会越长, 并且过多的隐层节点数可能导致过拟合。为了验证隐层节点数对准确率的影响以及选取合理的隐层节点数, 本文通过 LM 算法分别训练多个多层感知器, 输出层和隐层激活函数均设为 tanh。表 2 给出了不同隐层节点数的多层感知器预测声参数的均方误差。

由表 2 可以看出, 随着隐层节点数增加, 训练集 MSE 越小, 说明模型的拟合能力越强。但是隐层节点数超过 15 后, 验证 MSE 明显增加, 说明模型的泛化能力变差。综合考虑模型的拟合能力与泛化能力, 隐层节点数应该设为 15 较好。

2.3 训练算法性能对比分析

SGD 算法、Adam 算法是目前主流的一阶优化算法, LM、BFGS 是主流的二阶优化算法。本文分别通过 LM、BFGS、Adam 和 SGD 算法训练同一个初始化 MLP 模型, 采用 MSE 评估 MLP 模型准确率。各训练算法均方误差如图 2 所示。

LM 算法的预测 MSE 比 BFGS 算法降低了 23.4%, 比 Adam 算法降低了 85.6%, 比 SGD 算法降低了 97.2%, 印证了二阶优化算法相比一阶优化算法在收敛速度和泛化能力上有明显优势。在一阶优化算法中, Adam 算法通过修正梯度的一阶矩和二阶矩估计实现自适应学习率, 所以比随机梯度收敛快, 并有更高的准

表 1 不同激活函数的均方误差

隐层 激活函数	输出层 激活函数	训练集 MSE	验证集 MSE
tanh	sigmoid	1.1039e-1	1.106e-1
sigmoid	sigmoid	1.1143e-1	1.108e-1
identity	sigmoid	1.3234e-1	1.322e-1
tanh	tanh	2.5416e-5	3.194e-4
sigmoid	tanh	1.0186e-4	2.182e-4
identity	tanh	5.9242e-2	5.967e-2
tanh	identity	9.4141e-5	2.893e-4
sigmoid	identity	1.3601e-4	1.908e-4
identity	identity	5.9182e-2	5.846e-2

表 2 不同隐层节点数的均方误差

隐层节点数	训练集 MSE	验证集 MSE
1	1.6894e-2	1.719e-2
2	2.1527e-3	2.255e-3
3	2.2378e-4	3.404e-4
4	1.2674e-4	1.651e-4
5	1.0149e-4	2.027e-4
7	1.0106e-4	3.48e-4
9	2.5416e-5	3.194e-4
11	2.0184e-5	1.45e-4
13	5.7628e-6	1.808e-4
15	1.6553e-9	1.396e-4
17	1.4495e-9	2.383e-4

准确率;在二阶算法中 LM 算法给 Hessian 矩阵增加阻尼项,确保 Hessian 矩阵正定,避免在鞍点处朝错误的方向更新,而 BFGS 虽然通过迭代近似 Hessian 矩阵的逆减少了计算量,但是近似误差导致 BFGS 算法在收敛速度和准确率上不如 LM 算法。需要说明的是,SGD 和 Adam 算法在训练过程中 MSE 收敛速度已经十分缓慢,但由于横坐标采用对数坐标,无法在图中凸显这一特征。

2.4 预测结果分析

8 个测试集样本的误差分析如图 3 所示。对于整个测试集,尽管不同测试样本的最大相对误差有一些波动,但这其中最大的一个相对误差也非常小,其值仅为 0.00355%。此外,平均相对误差均非常小,而且几乎没有波动,这表明基于 LM 算法的多层感知器能准确地基于 CT 图像预测颅骨声参数。

3 结束语

本文构造了不同网络宽度、不同激活函数组合的多层感知器神经网络,并分别采用两种经典二阶优化算法 LM 算法和 BFGS 算法,以及两种经典一阶优化算法 SGD 算法和 Adam 算法进行训练,通过对比分析在训练集和测试集下的均方误差确定最优方案,并将最优方案预测的颅骨声参数用于声场模拟。通过对比分析发现,无论在收敛速度、声参数预测准确率以及声场模拟的准确率指标上,二阶优化算法均比一阶优化算法有明显优势。本文的实验证明了基于 LM 训练算法的 MLP 模型能够有效应用于经颅超声聚焦场景,需要强调的是,该模型能被运用在更加广泛的应用场景下。

虽然 LM 算法训练速度远快于其他训练算法,但是在应用于大型神经网络时仍无法替代现有的一阶优化算法。LM 算法需要计算 Hessian 矩阵的逆,计算复杂度为 $O(n^3)$,所以当网络参数量达到百万级别时 LM 算法计算量远超其他算法。因此我的下一步工作是改进 LM 算法,使其适用于大型神经网络。

致谢

感谢中国科学院声学研究所的博士生王祥达对该研究工作的贡献与协助。

参 考 文 献

- [1] Cybenko G. Approximation by superpositions of a sigmoidal function[J]. Mathematics of Control, Signals, and Systems (MCSS), 1989,2(4):303-314.
- [2] 王亚,朱明. 基于卷积神经网络的人脸检测方法[J]. 网络新媒体技术, 2016,5(6):24-29.
- [3] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the Acm, 2012,60(2):2012.
- [4] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large - Scale Image Recognition [J]. Computer Science, 2014.

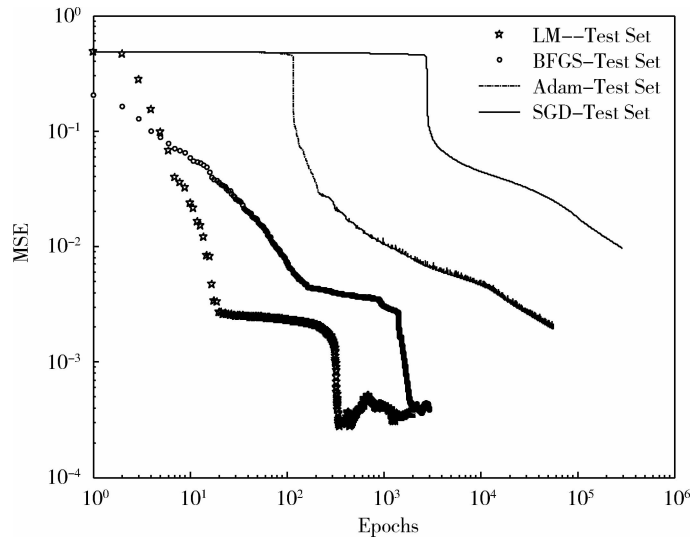


图 2 各训练算法均方误差

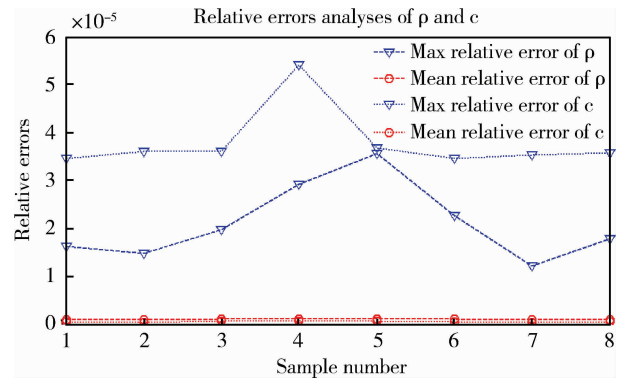


图 3 LM 算法测试集相对误差

- [5] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Computer Vision and Pattern Recognition. IEEE, 2015:1-9.
- [6] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. 2015:770-778.
- [7] Kingma D P, Ba J. Adam: A Method for Stochastic Optimization[C]. Computer Science, 2014.
- [8] Levenberg K. A method for the solution of certain non-linear problems in least squares[J]. Quarterly of applied mathematics, 1944,2(2):164-168.
- [9] Marquardt DW. An algorithm for least-squares estimation of nonlinear parameters[J]. Journal of the society for Industrial and Applied Mathematics, 1963,11(2):431-441.
- [10] Aubry J F, Tanter M, Pernot M, et al. Experimental demonstration of noninvasive transskull adaptive focusing based on prior computed tomography scans[J]. The Journal of the Acoustical Society of America, 2003,113(1):84-93.
- [11] Kreinovich V Y. Arbitrary nonlinearity is sufficient to represent all functions by neural networks; a theorem[J]. Neural networks, 1991,4(3):381-383.
- [12] Hartman E J, Keeler J D, Kowalski J M. Layered neural networks with Gaussian hidden units as universal approximations[J]. Neural computation, 1990,2(2):210-215.

作者简介

李南星,男,博士研究生,主要研究方向:新媒体技术。

盛益强*,男,博士,主要研究方向:智能集成系统、分布式机器学习、未来网络,推荐与传播。

倪宏,男,研究员,主要研究方向:网络通信、网络新媒体技术。

* 为通信作者。

(上接第 15 页)

择算法,并分析了该算法的多种并行化方案,实现了算法并行化框架。最终将 MNS、P-MNS 以及传统选点算法应用在网络测量选点问题和非线性函数求解极值问题进行效果、性能对比。结果显示基于蚁群算法的选点算法在节点平均权重、节点个数、网络覆盖率等方面较于传统选点算法都有很大提升,更适用于测量节点选取的应用场景。与非并行算法相比,并行化算法处理大规模、复杂网络环境选点问题时,算法执行效率有很大提升,降低了对计算资源的占用时间,有效的提高了系统响应速度。

参 考 文 献

- [1] Jin Yue-hui. Large-scale network and business performance monitoring system V1.0 [Z]. Beijing University of Posts and Telecommunications, 2013.
- [2] Dorigo M, Birattari M, Stutzle T. Ant colony optimization[J]. IEEE Computational Intelligence Magazines, 2006, 1(4):36-39.
- [3] Zhu Chang-hua, Pei Chang-xing, Li Jian-dong, et al. Network measurement and its key technologies[J]. Journal of Xidian University, 2002, 29(6): 813-818.
- [4] M Dorigo, V Maniezzo, A Colorni. Ant system: optimization by a colony of cooperating agents[J]. IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics: a publication of the IEEE Systems, Man, and Cybernetics Society, 1996, 26(1):29-35.
- [5] Dorigo M, Gambardella L M. Ant colony system: a cooperative learning approach to the traveling salesman problem[J]. IEEE Transactions on Evolutionary Computation, 1997, 1(1):53-66.
- [6] Dorigo M, Bonabeau E, Theraulaz G, Ant algorithms and stigmergy[J]. Future Generation Computer Systems, 2000, 16(9): 851-871.

作者简介

郝航,男,(1992-),硕士研究生在读,主要研究方向为互联网性能测量、网络技术与应用。

金跃辉,女,(1965-),教授,主要研究方向为互联网性能测量与用户体验、社会化网络与推荐系统。

杨谈,女,(1982-),讲师,主要研究方向为移动互联网、社会化网络。