

基于聚粒子群算法的神经网络权值优化方法

邓文杰

(四川大学 计算机学院, 四川 成都 610065)

摘要:神经网络作为机器智能分支中一种优秀的分类算法,在图像分类、人脸识别等领域中都有非常广泛的应用。但由于其参数过多,所以容易陷入局部最优解。针对 BP 神经网络易陷入局部最优的问题,提出了一种粒子群算法和聚类算法结合的优化神经网络权值的方法。该方法通过把神经网络的权值作为粒子群算法的初始粒子并利用粒子群算法的随机性全局搜索神经网络的待选初始权值,然后利用 C 均值算法找出包含权值较多的那一类,并把其聚类中心作为 BP 神经网络的初始权值。仿真结果表明,利用这种新的融合算法在防止 BP 神经网络易陷入局部最优的问题上能比普通的粒子群算法更加优秀。

关键词:BP 神经网络;粒子群优化;聚类;权值;局部最优

中图分类号:TP183

文献标识码:A

文章编号:1673-629X(2017)10-0016-03

doi:10.3969/j.issn.1673-629X.2017.10.004

A Neural Network Weights Optimization Method Based on Clustering Particle Swarm Optimization

DENG Wen-jie

(School of Computer Science, Sichuan University, Chengdu 610065, China)

Abstract:Neural network is a kind of excellent classification algorithm in the branch of machine intelligence, which has a wide range of applications in the field of image classification, face recognition and so on. However, because of its excessive parameters, it is easy to fall into the local optimal solution. According to this problem, a method combining particle swarm algorithm and clustering algorithm to optimize the weights of neural networks is proposed, which takes the neural network weights as the initial particle of particle swarm algorithm and uses the random of particle swarm algorithm to search the initial weights of neural network. Then the class contains more weight is found using C-means algorithm and its clustering center is regarded as the initial weights of BP neural network. The simulation results show that it is more excellent than the conventional particle swarm optimization algorithm in preventing the BP neural network from falling into local optimum.

Key words:BP neural networks; particle swarm optimization; clustering; weight; local optimum

0 引言

人工神经网络(Artificial Neural Networks, ANNs)的研究最早可以追溯到 19 世纪 50 年代末。F · Rosenblatt 设计的“感知机”^[1]在当时便引发了神经网络的研究热潮,各大实验室纷纷制作出感知机模型并应用在文字识别、声音识别等领域。但是,感知机只有输出层神经元进行激活处理,即只拥有一层功能神经元(functional neuron),其学习能力非常有限。于是科学家们开始尝试训练多层网络,但感知机的学习规则就捉襟见肘了。为了解决这个问题,以 Rumelhart 和 McClelland 为首的科学家小组于 1986 年提出了误差逆

传播(Error Back Propagation)算法^[2],现已成为最热的神经网络学习算法。

BP 算法通过对网络的均方误差求导得到权值和阈值的梯度,并依梯度方向修正权值和阈值,简单有效却存在局部最优和初值敏感问题^[3-4]。为了解决这些问题,学者们提出通过遗传算法^[5-6]、粒子群算法^[7-9]和蚁群算法^[10]改进神经网络的参数和结构。

针对 BP 神经网络易陷入局部最优的问题,文中提出一种融合 K-means^[11]与传统粒子群算法特性的群智能方法—聚粒子群算法,直接得到接近全局最优的网络权值,并利用 BP 使网络快速收敛。

收稿日期:2016-10-24

修回日期:2017-02-15

网络出版时间:2017-07-19

基金项目:国家自然科学基金资助项目(61571312)

作者简介:邓文杰(1993-),男,硕士研究生,CCF 会员(200051867G),研究方向为神经网络优化。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20170719.1109.026.html>

1 基本 BP 算法原理

给定训练集 (x_k, y_k) , $k = 1, 2, \dots, N$, 组成了拥有 n 个输入神经元, m 个隐藏神经元, l 个输出神经元的典型 BP 神经网络。其中隐藏层第 i 个神经元的阈值为 b_{1i} , 输出层第 i 个神经元的阈值为 b_{2i} , 输入层第 i 个神经元和隐藏层第 j 个神经元的权值为 w_{ij} , 隐藏层第 i 个神经元和输出层第 j 个神经元的权值为 v_{ij} , 第 j 个隐藏层神经元接收的输入和第 j 个输出层神经元接收的输入分别为:

$$\alpha_j = \sum_i^n w_{ij} x_i \quad (1)$$

$$\beta_j = \sum_i^m v_{ij} b_i \quad (2)$$

假设神经网络采用 Sigmoid 函数作为激活函数, 且对于训练集 (x_k, y_k) 的对应输出为 $\hat{y} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_l)$, 即有 $\hat{y}_j = f(\beta_j - b_{2j})$, 则网络在 (x_k, y_k) 上的均方误差为:

$$E_k = \frac{1}{l} \sum_{j=1}^l (\hat{y}_j - y_j)^2 \quad (3)$$

因为 BP 算法基于梯度下降策略, 以目标梯度的方向对参数进行调整, 给定学习率 η , 有:

$$\Delta v_{ij} = -\eta \frac{\partial E_k}{\partial v_{ij}} \quad (4)$$

而根据求导的链式法则有:

$$\frac{\partial E_k}{\partial v_{ij}} = \frac{\partial E_k}{\partial \hat{y}_j} \cdot \frac{\partial \hat{y}_j}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial v_{ij}} \quad (5)$$

根据式(1)、(2)和式(5)有:

$$h_j = -\frac{\partial E_k}{\partial \hat{y}_j} \cdot \frac{\partial \hat{y}_j}{\partial \beta_j} = \hat{y}_j(1 - \hat{y}_j)(y_j - \hat{y}_j), \quad \frac{\partial \beta_j}{\partial v_{ij}} = b_i \quad (6)$$

将式(6)带入式(5)再带入式(4), 就得到了 BP 算法中的权值更新公式:

$$\Delta v_{ij} = \eta h_j b_i \quad (7)$$

同理, 可计算出:

$$\Delta b_{2j} = -\eta h_j, \Delta b_{1j} = -\eta e_j, \Delta w_{ij} = -\eta e_j x_i \quad (8)$$

其中

$$e_j = b_j(1 - b_j) \sum_{k=1}^l v_{jk} h_k \quad (9)$$

BP 神经网络任意参数 θ 的更新公式为:

$$\theta = \theta + \Delta \theta \quad (10)$$

从以上推导可知, BP 神经网络算法是根据均方误差的梯度方向修正权值和阈值, 这就意味着如果初始权值选择不当, BP 神经网络在学习过程中很容易陷入局部最优, 由于均方误差曲面比较复杂, 故满足条件的解不唯一, 这就使得选择初始权值非常困难。

2 聚粒子群算法优化

通过 PSO 确定网络权值的一般思路是: 先将神经网络的性能函数作为粒子群的性能函数, 粒子代表网络中的权值和阈值^[12], 然后通过式(11)和式(12)分别对粒子的位置和速度进行更新。

$$v_{ij}(t+1) = v_{ij}(t) + c_1 \times r_{1j}(t) \times (p_{ij}(t) - x_{ij}(t)) + c_2 \times r_{2j}(t) \times (p_{gj}(t) - x_{ij}(t)) \quad (11)$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \quad (12)$$

其中, i 表示第 i 个粒子; j 表示第 i 个粒子的维度; c_1 和 c_2 表示学习因子; r_1 和 r_2 表示 $[0, 1]$ 范围内的随机数; $v_{ij}(t)$ 和 $x_{ij}(t)$ 分别表示 t 时刻 i 粒子的第 j 维的速度和位置。

PSO 算法虽然能利用自己的随机性以大概率将神经网络的初始权值定位在全局最优值附近, 但是其计算出的权值依然会有部分概率收敛到局部最优值^[13]。不过经以上实验不难发现, PSO 得到的网络权值要大概率靠近全局最优, 小概率靠近局部最优。所以, 文中采用改进的 PSO 算法(CPSO)。将这种使用 CPSO 确定神经网络权值结合 BP 算法的方法称为 CPSO-BP 算法。

CPSO-BP 算法流程为:

Step1: 将网络的权值向量化作为粒子群每个粒子个体;

Step2: 初始化粒子群的各项参数与适应度函数: 学习因子 c_1 和 c_2 ; 随机数 r_1 和 r_2 ; t 时刻 i 粒子的第 j 维的速度和位置 $v_{ij}(t)$ 和 $x_{ij}(t)$, 并将网络的均方误差函数

设为粒子的适应度函数 $F = E_k = \frac{1}{l} \sum_{j=1}^l (\hat{y}_j - y_j)^2$;

Step3: 按 PSO 算法更新, 并记录下更新结束后最优的粒子个体;

Step4: 重复 N 次 Step2-Step3 并且得到 N 个粒子个体;

Step5: 将 N 个粒子个体用 C 均值方法进行聚类, 并设聚类中心数 $C = \log_2 N + 1$ 。将包含结果较多的类别的聚类中心作为神经网络的初始权值和阈值;

Step6: 利用 Step5 计算出的权值和阈值创建神经网络, 然后用 BP 算法训练神经网络, 对其权阈值进行微调;

Step7: 输入测试数据, 验证网络误差。

3 验证和仿真

为了说明 CPSO-BP 的优越性, 对其进行仿真实验, 并与 BP 算法进行比较。图 1 是一个 1-2-1 的网络, 具有如下权值和阈: $w_1 = 10, w_2 = 10, w_3 = 1, w_4 = 1, w_5 = 5, w_6 = -5, w_7 = -1$ 。输入 P 在 $[-2, 2]$ 中采样。

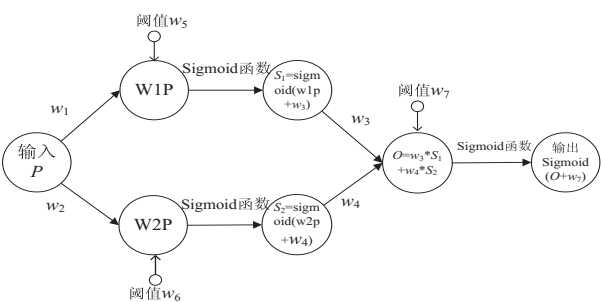


图 1 1-2-1 神经网络

为做出权值分布图,只调整参数 w_1 和 w_3 。设 BP 算法最大运算步数为 5 000,学习速率 $\eta=0.5$,目标误差为 $1e-9$ 。CPSO-BP 算法选用的粒子数为 30,迭代次数为 30。图 2 显示了不同算法优化出的权值所处的位置分布,(a)是标准 PSO 优化出的权值位置,可见依然会陷入局部最优,而(b)是 CPSO 算法优化出的权值分布,其很明显地分布在全局最优周围。

表 1 是对参数 w_1 和 w_2 进行 20 次训练后的误差平均值和平均迭代次数。

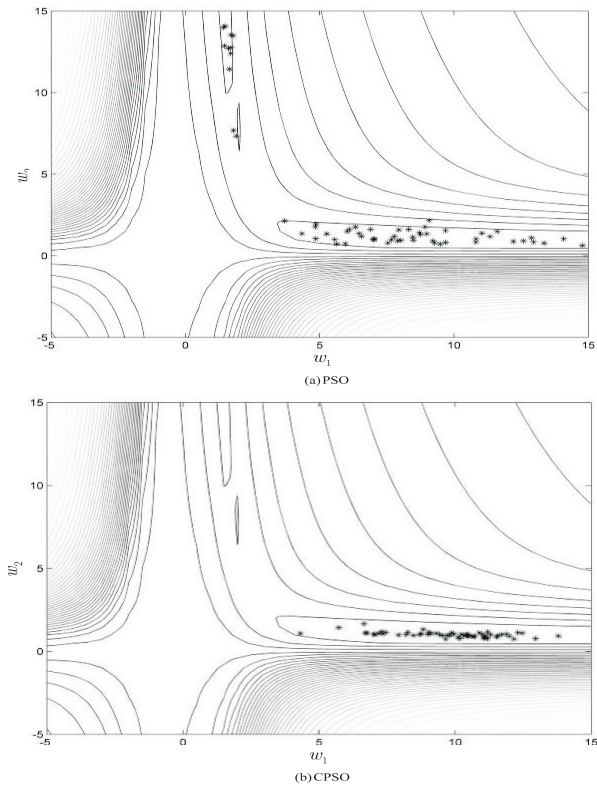


图 2 使用不同算法优化出的初始权值

表 1 对参数 w_1 和 w_2 进行 20 次训练后的平均误差和平均迭代次数

算法	最大误差	平均误差	迭代次数
无优化 BP	0.108 8	0.042 6	3 828
PSO-BP	0.106 1	0.020 8	2 664
CPSO-BP	9.996 5e-10	9.972 3e-10	1 954

步数为 10 000,学习速率 $\eta=0.5$,目标误差为 0.001。CPSO-BP 算法选用的粒子数为 30,迭代次数为 30。表 2 是对整个网络进行 20 次训练后得到的误差平均值和平均迭代次数,图 3 为不同优化结果的输出和期望输出。

表 2 对整个网络 20 次训练得到的误差平均值和迭代次数

算法	最大误差	平均误差	迭代次数
无优化 BP	0.590 6	0.252 2	10 000
PSO-BP	0.174 1	0.019 9	9 078
CPSO-BP	0.068 80	0.003 8	4 808

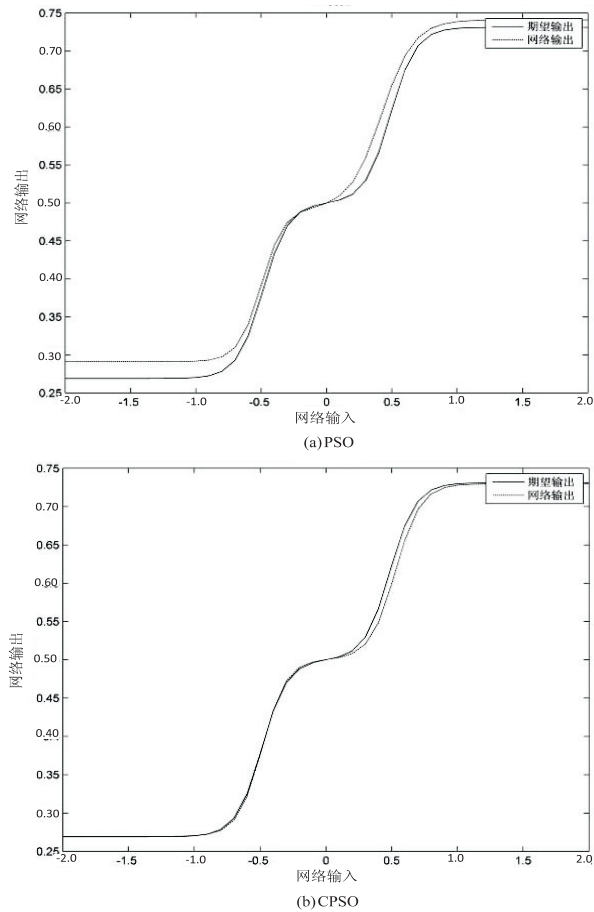


图 3 使用不同的算法对网络的拟合

由上述对比结果可以看出,CPSO 选出的初始权值明显优于 PSO,其训练后的误差和收敛速度也优于普通 PSO 算法。

4 结束语

针对 BP 神经网络容易陷入局部最优等缺点,利用 CPSO 算法将神经网络权值和阈值作为粒子进行全局优化,然后通过 C 均值算法将粒子族群数较多的族群的聚类中心作为神经网络的初始权值。仿真实验表明,CPSO-BP 算法能最大限度地克服 BP 神经网络对

(下转第 23 页)

当训练整个网络的参数时,设 BP 算法最大运算

[7] Nguyen T M, Wu Q M J. Gaussian mixture model based spatial neighborhood relationships for pixel labeling problem [J]. IEEE Transaction on System Man and Cybernetic, 2012, 42 (1):193–202.

[8] 赖裕平,丁洪伟,周亚建,等. 有限贝塔刘维尔混合模型的变分学习及其应用[J]. 电子学报, 2014, 42 (7): 1347–1352.

[9] Ma Zhanyu, Leijon A. Bayesian estimation of beta mixture models with variational inference [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33 (11): 2160–2173.

[10] Bdiri T, Bouguila N. Positive vectors clustering using inverted Dirichlet finite mixture models [J]. Expert Systems with Applications, 2012, 39(2): 1869–1882.

[11] Biernacki C, Celeux G, Govaert G. Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models [J]. Computational Statistics and Data Analysis, 2003, 41 (3–4): 561–575.

[12] Grimmer J. An introduction to Bayesian inference via variational approximations [J]. Political Analysis, 2011, 19 (1): 32–47.

[13] Bishop C M. Pattern recognition and machine learning [M]. New York: Springer-Verlag, 2006.

[14] Teschendorff A E, Wang Y, Barbosa-Morais N L, et al. A variational Bayesian mixture modelling framework for cluster analysis of gene – expression data [J]. Bioinformatics, 2005, 21 (13): 3025–3033.

[15] Bali B, Mohammad-Djafari A. A variational Bayesian algorithm for BSS problem with hidden Gauss–Markov models for the sources [C]//Proc of ICA. [s. l.]: [s. n.], 2007: 137–144.

[16] Snoussi H, Mohammad-Djafari A. Estimation of structured Gaussian mixtures; the inverse EM algorithm [J]. IEEE Transactions on Signal Processing, 2007, 55(7): 3185–3191.

[17] Skolidis G, Sanguinetti G. Bayesian multitask classification with Gaussian process priors [J]. IEEE Transactions on Neural Networks, 2011, 22 (12): 2011–2021.

[18] Wang B, Titterton D M. Convergence properties of a general algorithm for calculating variational Bayesian estimates for a normal mixture model [J]. Bayesian Analysis, 2006, 1 (3): 625–650.

[19] Corduneanu A, Bishop C M. Variational Bayesian model selection for mixture distributions [C]//Proceedings of eighth international conference on artificial intelligence and statistics. [s. l.]: [s. n.], 2001: 27–34.

[20] Fan Wentao, Bouguila N, Ziou D. Variational learning for finite Dirichlet mixture models and applications [J]. IEEE Transaction on Neural Networks and Learning Systems, 2012, 23 (5): 762–774.

[21] 梅素玉,王 飞,周水庚. 狄利克雷过程混合模型、扩展模型及应用 [J]. 科学通报, 2012, 57 (34): 3243–3257.

[22] MacKay D, Peto L C B. A hierarchical Dirichlet language model [J]. Natural Language Engineering, 1994, 1 (3): 1–19.

[23] Bouguila N, Ziou D. A Dirichlet process mixture of Dirichlet distributions for classification and prediction [C]//IEEE workshop on machine learning for signal processing. [s. l.]: IEEE, 2008: 297–302.

[24] Ishwaran H, James L F. Gibbs sampling methods for stick-breaking priors [J]. Journal of the American Statistical Association, 2001, 96 (453): 161–173.

[25] Blei D M, Jordan M I. Variational inference for Dirichlet process mixtures [J]. Bayesian Analysis, 2005, 1 (1): 121–144.

(上接第 18 页)

初值敏感,容易陷入局部极小值等缺点。

参考文献:

[1] Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain [J]. Psychological Review, 1958, 65 (6): 386–408.

[2] Rumelhart D E, Hinton G E, Williams R J. Learning internal representation by back-propagation errors [J]. Nature, 1986, 323 (323): 533–536.

[3] 余本国. BP 神经网络局限性及其改进的研究 [J]. 山西农业大学学报: 自然科学版, 2009, 29 (1): 89–93.

[4] 王美玲,王念平,李 晓. BP 神经网络算法的改进及应用 [J]. 计算机工程与应用, 2009, 45 (35): 47–48.

[5] 王 晶. 蚁群算法优化前向神经网络的一种方法 [J]. 计算机工程与应用, 2006, 42 (25): 53–55.

[6] Trelea I C. The particle swarm optimization algorithm: convergence analysis and parameter selection [J]. Information Processing Letters, 2003, 85 (6): 317–325.

[7] 高海兵,高 亮,周 驰,等. 基于粒子群优化的神经网络训练算法研究 [J]. 电子学报, 2004, 32 (9): 1572–1574.

[8] 刘洪波,王秀坤,孟 军,等. 神经网络基于粒子群优化的学习算法研究 [J]. 小型微型计算机系统, 2005, 26 (4): 638–640.

[9] 王崇骏,于汶滢,陈兆乾,等. 一种基于遗传算法的 BP 神经网络算法及其应用 [J]. 南京大学学报: 自然科学版, 2003, 39 (5): 459–466.

[10] 翟宜峰,李鸿雁,刘寒冰,等. 用遗传算法优化神经网络初始权重的方法 [J]. 吉林大学学报: 工学版, 2003, 33 (2): 45–50.

[11] 孙吉贵,刘 杰,赵连宇. 聚类算法研究 [J]. 软件学报, 2008, 19 (1): 48–61.

[12] 杨 维,李歧强. 粒子群优化算法综述 [J]. 中国工程科学, 2004, 6 (5): 87–94.

[13] 吴晓军,杨战中,赵 明. 均匀搜索粒子群算法 [J]. 电子学报, 2011, 39 (6): 1261–1266.