

基于模拟退火算法改进的 BP 神经网络算法

周爱武, 翟增辉, 刘慧婷

(安徽大学 计算机科学与技术学院, 安徽 合肥 230601)

摘 要: 提出一种基于模拟退火算法改进的 BP 神经网络. 该方法利用模拟退火算法寻找更优化的样本子集, 用来训练 BP 神经网络. 通过理论分析以及实验仿真证明, 在缩短训练时间以及迭代次数的基础上, 显著提高 BP 神经网络的分类准确性.

关键词: BP 神经网络; 样本选择; 主动学习; 模拟退火

中图分类号: TP301.6

文献标识码: A

文章编号: 1000-7180(2016)04-0144-04

Improved BP Neural Network Based on Simulated Annealing

ZHOU Ai-wu, ZHAI Zeng-hui, LIU Hui-ting

(College of Computer Science and Technology, Anhui University, Hefei 230601, China)

Abstract: The BP neural network based on simulated annealing algorithm is given out in this paper, the method uses the simulated annealing algorithm to find a more optimal subset of samples, to train the BP neural network. Through theoretical analysis and experimental simulation proved that not only the time and the number of iterations is shorten, the accuracy of BP neural network is significantly improved.

Key words: BP neural network; sample selection; active learning; simulated annealing

1 引言

BP 神经网络虽然在学术和应用上已经十分成熟. BP 神经网络也有其缺陷的地方: 网络结构难以确定、初始化权值取值困难、激励函数选择、学习方式不确定、样本筛选代价高等问题. 针对这些问题, 主要有采用构造法和剪枝法^[1] 确定网络结构; Cauchy 不等式和线性代数^[2] 的方法得到最优初始权值. 引入陡度因子^[3] 的方法改进误差曲面. 针对梯度下降法的问题, 有对应的如共轭梯度法^[4] 等方式来解决. 对于样本的筛选是许多监督学习算法都具有的问题, 一般的样本选择方式只能在牺牲算法性能的基础上去减少样本子集. 近来主动学习^[5] 的思想被提出并成为热点.

本文利用模拟退火算法^[6] 寻优的思想, 将样本子集看成是一个状态. 当模拟退火算法找到最优状态时, 该样本子集是当前最优的. 利用 UCI 数据集对

此进行建模, 大大提高了收敛性以及准确性, 这也证明了该方法的有效性.

2 改进的 BP 神经网络

2.1 BP 神经网络

BP 神经网络是神经网络中的一种, 由输入层、隐含层和输出层组成. 这种网络在训练的过程中有反馈信号来修改神经网络中连接节点的权值, 因此称为 BP 神经网络. 对于除输入层节点外的每个节点来说, 其输入和输出的关系为:

$$\text{net}_i = \sum_{j=1}^n W_{ij} X_j - \theta \quad (1)$$

$$Y_i = f(\text{net}_i) \quad (2)$$

式中, X 为输入节点, W 为节点连接权值, θ 为偏置, 而 f 则为激励函数, Y 则为该神经元的输出. 以上为神经网络正向传播的过程, 对于反馈的过程来说, 权值修改使用梯度下降法:

$$E = \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^m (I_j^i - O_j^i)^2 \quad (3)$$

E 为误差公式,也就是所有样本其输入输出差的平方和。

$$W_{ij} = W_{ij} - \eta \nabla E(W_{ij}) \quad (4)$$

式中, $\nabla E(W_{ij})$ 为误差的偏导,权值根据此误差偏导不断的调整,理论上能够得到权值的最优解。

设定函数 $\text{Accuracy}(i)$,接受训练样本集合 i 作为 BP 神经网络的训练集,在每一次迭代后都会得到输出,该输出与期望输出代入式(3)得到误差,并利用式(4)修正得到新的权值,最后计算本次迭代所得的准确率并返回。

2.2 模拟退火算法

模拟退火算法是从物理材料统计力学中的现象变化过来的算法。在材料统计学中温度环境对粒子的结构以及其他特性影响较大,在温度较高的条件之下,粒子的能量较高,粒子的活性加大。而温度较低时,能量较低。其中缓慢降低材料所承受的温度即为退火的过程,最终材料会形成较低能量的晶体。

粒子的能量是由粒子当前的状态所决定的, Metropolis 算法使用一个数学模型来模拟这个过程。首先假设材料的状态为 i , 那么其能量表示为 $E(i)$, 外界温度 T 时, 状态从 i 改变为 j , 对应的能量也从 $E(i)$ 变化为 $E(j)$, 此时根据一定条件判断该能量的变化到底接受与否:

如果 $E(i) \leq E(j)$, 说明能量降低, 那么接受状态的改变, 否则能量增加, 接受与否是按照以下概率来判断的:

$$\exp \frac{E(i) - E(j)}{KT} \quad (5)$$

式中, K 为物理学中玻尔兹曼常数, 而 T 为材料所处的温度。

在特定的温度 T 下, 材料所处的状态 i 的概率满足玻尔兹曼分布:

$$P_T(X = i) = \frac{\exp - \frac{E(i)}{KT}}{\sum_{j \in S} \exp - \frac{E(j)}{KT}} \quad (6)$$

式中, S 为整个状态空间, 易得到:

$$\lim_{T \rightarrow \infty} \frac{\exp - \frac{E(i)}{KT}}{\sum_{j \in S} \exp - \frac{E(j)}{KT}} = \frac{1}{|S|} \quad (7)$$

说明温度很高时所有的状态概率都相同, 在温度下降的情况下:

$$\lim_{T \rightarrow \infty} \frac{\exp - \frac{E(i) - E_{\min}}{KT}}{\sum_{j \in S} \exp - \frac{E(j) - E_{\min}}{KT}} = \begin{cases} \frac{1}{|S_{\min}|}, & i \in S_{\min} \\ 0, & \text{其他} \end{cases} \quad (8)$$

式中, $E_{\min} = \min E(j) j \in S$ 并且为能量处于最小状态的个数。该公式表示当温度趋向与 0 的时候, 很大概率能够得到最小能量的状态。

根据上述的数学模型来解决寻找最优值问题, 在给定温度 T_K 以及其对应状态 i 的情况下, 状态改变为 i' , 接受新状态作为下一个状态的概率为:

$$P(i \rightarrow i') = \begin{cases} 1, E(i) > E(i') \\ \exp - \frac{E(i') - E(i)}{T_K}, & \text{其他} \end{cases} \quad (9)$$

在此温度时进行一次或多次转换之后再降低温度, 逐步进行这个过程, 就是模拟退火算法寻优的过程。

2.3 基于模拟退火算法改进的 BP 神经网络算法

本文的主要思想是利用模拟退火算法来寻找最有样本子集合, 为了让模拟退火算法能够用在样本选择上, 需要一些条件。

由于模拟退火算法针对于有限的状态的寻优, 因此需要将训练样本集合分成 n 份 ($x_1, x_2, \dots, x_n$), 然后在此 n 份样本子集中随机选取其中的 m 份作为状态集合中的一个状态 i 。

状态空间并不是简单的 C_n^m , 由于算法需要状态的变化, 为了能够让状态 i 改变, 将 i 中的 m 份子集中 k 个部分与总样本集合中剩余的 $(n-m)$ 份子集中的 k 个进行交换, 将该方法称之为 k 交换法。这样一来状态空间的数目则可以总结为: $C_n^m C_m^k C_{n-m}^k$ 。状态空间于是已经被决定下来了, 可以使用随机交换的方式来改变状态, 也可以按照一定的规则来进行遍历该状态空间。

至此还有能量函数 $E()$ 未定, 将能量函数 $E(i)$ 设定为 2.1 节中的函数 $\text{accuracy}(i)$ 。根据式(9), 重新设定接受新状态为下一状态的概率:

$$P(i \rightarrow i') = \begin{cases} 1, \text{accuracy}(i) < \text{accuracy}(i') \\ \exp - \frac{\text{accuracy}(i) - \text{accuracy}(i')}{T_K}, & \text{其他} \end{cases} \quad (10)$$

根据以上规则, 算法具体步骤如下所示。

步骤 1: 初始化温度 T_0 , 交换计数 $\text{count} = 0$ 以及其预设值 c , 训练样本集 A 按照随机抽样的方式分成 n 份: ($x_1, x_2, \dots, x_n$)。决定 m 的大小, 产生第一个状态 $i_0(y_1, y_2, \dots, y_m)$ 。决定 k 的大小, 生成 k

交换法函数 $\text{swap}(k, \text{state}, \text{sample})$.

步骤 2: 使用当前状态 i 调用 $\text{accuracy}(i)$ 并保存返回值, 比较其与准确率阈值, 如果达到阈值则退出算法.

步骤 3: 为当前状态 i 调用 $\text{swap}(k, i, A)$ 得到 i' .

步骤 4: 根据状态 i' 调用 $\text{accuracy}(i)$, 比较保存起来的 $\text{accuracy}(i)$, 根据式(10) 决定是否接受状态 i' 作为下一状态, 如果接受 $\text{count} = \text{count} + 1$, 若 $\text{count} + 1 > c$ 则跳转步骤 5, 否则跳转步骤 2. 如果不接受为下一状态则跳转步骤 3.

步骤 5: 跳转到该步骤说明同一温度下状态变换足够, 则让当前温度 $T - \eta T$, 并将结果保存为当前温度. 跳转到步骤 2.

3 实验及结果分析

为了检测算法的效果, 实验采用 MATLAB R2013b 版本的集成开发环境, win7 32 位操作系统, AMD Phenom(tm) II N830 Triple-Core 处理器 2.1 GHz, 4 GB 内存.

本文使用基于模拟退火算法改进的 BP 神经网络算法和标准 BP 神经网络两种算法分别在 UCI 数据集中的 Abalone 以及 Adult 数据集上进行实验仿真, 从准确率、训练时间和空间需求三个方面对两种算法和两种数据集进行比较.

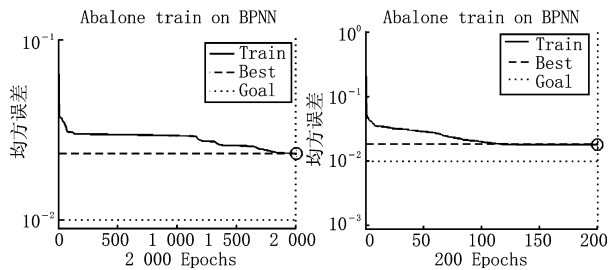
设定目标误差为 0.01, 三分类和二分类阈值为 82% 和 90%, 设定标准 BP 神经网络的迭代次数为 2 000, 而基于模拟退火算法的则设定 BP 神经网络迭代次数为 200, 模拟退火算法寻优的状态次数设为 10, 以让总的迭代次数相同. 只要够达到误差或准确率阈值跳出程序, 收敛完成.

3.1 Abalone 数据集

该数据集包含 4 177 条数据, 每条数据包含 8 个属性, 鲍鱼年龄属性标签, 由于该属性具有连续行, 实验将其分成 3 类. 设定 BP 神经网络的输入层节点为 7, 隐含层两层分别为 10 和 5, 输出层节点为 3. 使用 sigmoid 函数作为输出层的激励函数.

在以上参数都既定的情况下, 使用标准 BP 神经网络和改进的 BP 神经网络对该数据集进行训练和测试, 共进行 5 次实验. 统计的均方误差结果对比图 1 所示, 测试数据的准确率如表 1 所示.

图 1 中虚线为目标, 在 2 000 次迭代中能够看出其经过缓慢的下降之后最后误差停止下降, 达到的误差为 0.024 左右. 而相对比的右图实际上在模拟退火算法未达到 10 次的寻优的过程中其误差达



(a) 标准 BPNN (b) 改进的 BPNN

图 1 Abalone 数据均方误差图

到 0.018 左右, 比较于标准 BPNN 算法, 后者更接近于预定的目标误差值. 并且迭代的次数小于标准 BPNN 的迭代次数, 所需要的训练时间更短. 而两种算法训练 Abalone 数据所达到的准确率以及需要的时间如表 1 所示.

表 1 Abalone 数据标准 BPNN 算法与改进的 BPNN 算法的准确率和时长比较

	SP/%	IP/%	ST/s	IT/s
一	75.465	79.445	249	235
二	74.020	81.200	320	230
三	74.970	80.545	251	244
四	76.785	82.520	331	188
五	76.660	81.080	298	260

表 1 中 SP 和 IP 分别为 BPNN 准确率和改进的 BPNN 准确率, ST 和 IT 为两者时间, 表 2 也相同. 五次实验虽然针对的数据相同, 但是每次改进的 BPNN 算法总是能够在较短的时间内完成数据的训练并且在数据的准确率上超过标准 BPNN. 甚至在改进的 BPNN 算法中有一次超过了实验设定的准确率阈值 82%, 之后程序直接退出完成收敛, 时间大量的少于平均所需要的训练时间.

3.2 Adult 数据集

该数据集共有 48 842 条数据, 每条数据包含 15 个属性, 其中最后一个属性收入是否大于 50 k 为实验标签, 该数据集明确用于二分类问题上. 因此设定 BP 神经网络的输入层为 14 个节点, 隐含层 2 层分别是 15 和 5, 输出节点为 2. 同样使用 sigmoid 函数作为输出层激励函数.

在以上参数都既定的情况下, 使用标准 BP 神经网络和改进的 BP 神经网络对该数据集进行训练和测试, 共进行 5 次实验. 统计的均方误差结果对比图 2 所示, 测试数据的准确率如表 2 所示.

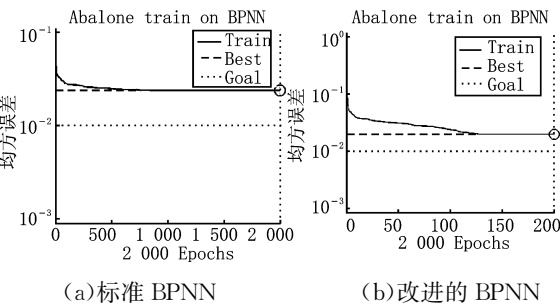


图 2 Adult 数据均方误差图

图 2 中同样将目标误差设定为 0.01,而 BPNN 对 Adult 数据训练基本上在迭代 800 次的时候停止,最终其误差达到 0.023 左右.对于改进的 BPNN 达到的目标则明显优于 BPNN 的训练,其达到的误差在 0.017 左右.而对于五次实验中统计的准确率以及消耗的时间则如表 2 所示.

表 2 Adult 数据标准 BPNN 算法与改进的 BPNN 算法的准确率和时长比较

	SP/%	IP/%	ST/s	IT
一	83.120	89.225	744	655
二	79.865	88.400	830	637
三	82.275	90.240	738	664
四	84.940	87.785	751	803
五	82.505	89.080	844	704

对于二分类非线性问题来说,比较于三分类的问题理论上也是稍微简单的.表 2 也表明准确率比较于二类问题来说也更高一些.而训练由于数据集的数量较大,因此时间也较长.

在相同条件下改进 BPNN 的训练时间长度更小一些,这些时间主要是在模拟退火算法筛选出的样本数量平均比 BPNN 小 $(n-m)/n$.针对 Adult 数据集,有附属其他算法进行训练所得到的错误率统计,在此处也将改进的 BPNN 的平均错误率加入图 3 中进行对比.

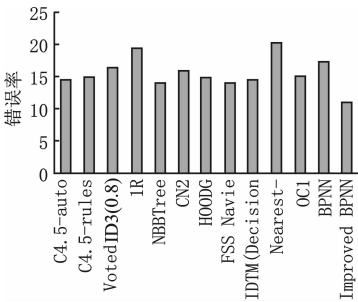


图 3 各算法针对 Adult 数据集的测试错误率

从图 3 能够看出 BPNN 算法针对二类问题的 Adult 数据集分类结果只处于各分类算法性能的中

间层次,而改进的 BPNN 的效果加上了训练数据集寻优的部分后,错误率大大的降低,收敛的能力以及收敛所需要的时间代价也相对降低.

4 结束语

本文提出了基于模拟退火算法改进的 BP 神经网络算法,从理论分析以及实验验证两方面对于改进的 BPNN 算法和 BPNN 算法进行比较,尤其是分别经过二分类问题以及三分类问题的实验说明,改进的 BPNN 算法不仅在准确率上具有很大的优势,同时在训练时间上也小于 BPNN 算法.因此本文提出的基于模拟退火算法改进的 BP 神经网络算法达到预期效果.

在将样本集合分成 n 份,并将其中的 m 份作为模拟退火算法的一个状态,以及算法的 k 交换法并没有深入讨论.一方面将针对 n,m 以及 k 的设定会影响整体状态集合的大小的改变.另一方面,如果 n,m 以及 k 的设定已经固定,则该状态集合就已经定下,针对既定状态集合中寻找较优的状态 k 交换法可能并不能满足需求.因此本文后续研究将从 n,m 以及 k 这三个参数的初始值来展开,并研究更为高效的 k 交换法.

参考文献:

[1] 魏海坤. 神经网络结构设计的理论与方法[M]. 北京: 国防工业出版社,2005.

[2] Yam J Y F, Chow T W S. A weight initialization method for improving training speed infeedforward neural network[J]. Neurocomputing, 2000, 30(1): 219-232.

[3] SimonHayKin. 神经网络原理[M]. 北京: 机械工业出版社,2004:109-173.

[4] 杨斌,夏耀先. (2002). 基于改进共轭梯度法的前馈网络快速监督学习算法[J]. 电子学报, 2002,30(12): 1845-1847.

[5] 胡静,高隽,杨静. 模糊神经网络分类器的主动学习方法[J]. 中国科学技术大学学报, 2008, 38(3): 241-246.

[6] 康丽山. 非数值并行算法(第 1 册)——模拟退火算法[M]. 北京: 科学出版社,1997:22-55.

作者简介:

周爱武 女,(1965-),硕士,副教授. 研究方向为数据库与 WEB 计数. E-Mail:943619770@qq. com.

翟增辉 男,(1991-),硕士研究生. 研究方向为数据挖掘.

刘慧婷 女,(1978-),副教授,硕士生导师. 研究方向为数据挖掘.