

无线传感网络监测数据分类算法仿真分析

洪月华

(广西经济管理干部学院 计算机系, 广西 南宁 530007)

摘 要: 研究无线传感网络数据的准确分类问题, 针对无线传感网络监测得到的数据属性呈现高冗余特征, 用传统的 BP 神经网络进行分类易陷入局部最优解、泛化能力差、收敛速度慢与精度低等问题, 造成数据很难被准确分类, 提出用免疫算法优化 BP 神经网络的分类算法. 该算法利用免疫算法所具有的全局收敛特性和个体多样性保持机制, 全局搜索优化 BP 神经网络的权值, 再用 BP 算法对其开展局部搜索工作. 仿真实验结果证明, 该算法能够有效克服训练经典 BP 神经网络时容易陷入局部极值的不足, 加快了网络的收敛速度, 传感器网络数据分类识别的准确率也得到了大幅度提高.

关键词: 免疫算法; BP 神经网络; 无线传感器网络; 数据挖掘

中图分类号: TP183

文献标识码: A

文章编号: 1000-7180(2013)10-0149-04

Simulation Analysis of the Monitoring Data Classification Algorithm for Wireless Sensor Networks

HONG Yue-hua

(School of Computer, Guangxi Economic Management Cadre College, Nanning 530007, China)

Abstract: The problem of data classification for wireless sensor network is researched. The data monitored by wireless sensor network is high redundant, so if traditional BP neural network is used for classification will result in local optimal solution, poor generalization ability, slow convergence speed and low accuracy problem. Therefore data is difficult to be accurately classified. In order to solve the problem immune algorithm is used to optimize BP neural network classification algorithm. Based on the global convergence and individual diversity maintaining mechanism of immune algorithm, the algorithm is used to search globally and optimize BP neural network weights, and then carry out local search with BP algorithm. The simulation results show that, the algorithm can effectively overcome the shortcoming that training classic BP neural network can easily fall into local extreme, so it can accelerate the network convergence speed and the accurate rate of sensor network data classification has been greatly improved.

Key words: immune algorithm; BP neural network; wireless sensor network; data mining

1 引言

数据分类是数据挖掘研究领域中的一个重要方向. 分类的目的是从数据库样本数据中找到同类数据对象的共性, 然后构建一个分类模型(别称分类器)对其进行描述和刻画, 该模型能把未知类别的数

据对象映射到已知类别中的某一个. 在众多数据分类算法中, 已在分类领域广泛应用神经网络分类方法具有记忆联想、高度的分布式并行处理和非线性映射等其他分类方法所没有的优点. 然而, 神经网络存在容易陷入局部最优解、泛化能力差、收敛速度慢与精度低等问题, 限制了其在分类领域的应用. 针

收稿日期: 2012-12-20; **修回日期:** 2013-03-01

基金项目: 广西自然科学基金青年项目(2012GXNSFBA053178); 广西混杂计算与集成电路设计分析重点实验室开放基金课题(2012HCIC05)

对这些不足,许多学者提出了解决方案.文献[1-3]提出了遗传算法对 BP 网络学习算法进行改进.文献[4-5]提出了粗糙集与神经网络结合的分类算法.本文根据这些研究基础,同时鉴于传感器网络自身的特点,提出一种免疫算法优化 BP 神经网络的分类器模型来解决传感器网络数据分类的问题.

2 免疫 BP 神经网络算法的实现

在构造好分类器模型的基础上形成免疫 BP 神经网络的分类算法.该算法的主要工作是用免疫算法优化 BP 神经网络的权值.

通常情况下,免疫算法将待求解的问题作为抗原,问题的对应解作为抗体.本文在免疫算法优化 BP 神经网络权值的过程中,将 BP 神经网络误差作为抗原,权值进行编码作为抗体,对 BP 网络进行训练时,首先将抗体解码后得到一组权值矩阵,接着输入训练样本,对抗原、抗体与抗原之间的亲和力以及抗体的抑制与促进进行计算后,将优秀抗体选择出来进行交叉和变异操作,再从中选出亲和力高的抗体,然后和新抗体一起构成下一代抗体群体.

算法中用到的几个重要概念公式如下:

(1) 抗原

网络训练的的目的是使均方误差趋于最小,其作为免疫算法抗原的定义如下:

$$E_i = \frac{1}{n} \sum_{i=1}^n (z_i - z'_i)^2 \quad (1)$$

式中,用 n 代表训练样本总数,用 z_i 和 z'_i 分别代表网络的实际和期望输出值.

(2) 抗体编码

作为免疫算法抗体的权值和网络结构是对应的,同时权值是用实数表示,因此将隐含层节点数和一组网络权值结合起来用实数编码^[7],这样就达到串的长度减少和算法的运算速度加快的目的.每个抗体编码 $kt_i (i = 1, 2, \dots, N)$ 如图 1 所示.

h	w_{11}	v_{11}	$\dots$	w_{ji}	v_{kj}	$\dots$	w_{jn}	v_{nh}
-----	----------	----------	---------	----------	----------	---------	----------	----------

图 1 抗体编码方式

图中, h 代表网络隐含层节点数目; w_{ji} 代表输入层节点 j 到隐含层节点 i 的权值; v_{kj} 代表隐含层节点 k 到输出层节点 j 的权值.

(3) 抗体与抗原的亲和力

抗原与抗体间的亲和力 $F(i)$ 对应于遗传算法的适应度概念,用来描绘可行解逼近最优解的程度.定义亲和力 $F(i)$ 为 BP 网络均方误差 E_i 的函数,即:

$$F(i) = \frac{1}{E_i + \tau} \quad (2)$$

式中, $\tau \in [0, 1]$ 是常量,当亲和度 $F(i)$ 的值为 1 时抗体处于最佳情况,即抗体是最优解.

(4) 抗体浓度

用抗体的浓度描述抗体与群体 KT 中另外抗体的相象程度.在本文,抗体 kt_i 的浓度是按照其与另外个体间的欧几里德距离之和来度量:

$$d_i = \sum_{n=1}^N \| kt_i - kt_n \| \quad i = 1, 2, \dots, N \quad (3)$$

这样抗体 kt_i 的浓度 den_i 定义为

$$den_i = \frac{1}{N} \sum_{n=1}^N \frac{1}{d_i} \quad i = 1, 2, \dots, N \quad (4)$$

从式(4)、(5)可知距离越小则抗体的浓度越大,说明抗体与群体 KT 中其他抗体就越相象,当抗体的浓度达到最大值 1 时,表明群体 KT 中的全部抗体都相同.

(5) 抗体的促进与抑制

免疫系统的特性是促进高亲和力抗体的产生,抑制高浓度抗体的分泌.免疫算法根据抗体的选择概率对抗体进行促进和抑制^[8].综合前面式(3)~(5)得到选择抗体的概率公式如下所示:

$$p(i) = \delta(1 - den_i) F(i) \quad (5)$$

式中, $\delta \in [0, 1]$ 是可调参数.由式(6)知,亲和力低,浓度高的抗体被选中的概率小;反之亲和力高,浓度低的抗体被选中的概率大.因此优秀个体得到了保留,同时选择相似抗体又可减少,维持了抗体的多样性(即解的多样性).

(6) 交叉

交叉是为了在下次迭代过程中获得新的优良的抗体 kt_{new} , 替换重组两个父抗体 (kt_{old1}, kt_{old2}) 的部分结构的操作.

$$kt_{new} = kt_{old1} + \alpha \times (kt_{old2} - kt_{old1}) \quad (6)$$

式中, α 是随机数,在 $[0, 1]$ 上均匀分布.

(7) 变异

抗体的变异通过正态分布进行.当抗体的性能差时,就在大范围内搜索,反之,则在小范围搜索微调,以保证找到抗体的全局最优值.

$$kt'_i = kt_i + \sqrt{1 - F(i)} \cdot M_i(0, 1) \quad (7)$$

$$i = 1, 2, \dots, N$$

式中, $M_i(0, 1)$ 是标准偏差为 1, 期望值为 0 的正态分布, N 是 BP 网络权值的总数,即群体规模 N .

(8) 加入新抗体^[9]

模拟生物机体的新陈代谢功能.淘汰 KT 中

5%亲和力低的抗体并用随机产生的新抗体代替。

免疫算法优化 BP 神经网络权值的流程如图 2 所示,主要步骤如下:

Step 1:对 BP 网络进行初始化:经过编码形成抗体的初始群体 $KT = \{kt_i | i=1, 2, \dots, N\}$, 群体规模 N , 最大迭代次数 $g_{\max}$, 训练样本总数 n , 以及 τ 和 δ , 统计抗体个数 i 、统计训练样本个数 j 和统计迭代次数 g 这三个计数器赋初值 0。

Step 2:迭代计数器计数 $g = g + 1$ 。

Step 3:抗体计数器计数 $i = i + 1$ 。

Step 4:对抗体 kt_i 解码后得到相应网络的权值矩阵 W 与 V 。

Step 5:训练样本计数器计数 $j = j + 1$ 。

Step 6:用第 j 个训练样本对 BP 神经网络进行训练,根据实际输出与期望输出计算得到网络的误差 E_j ,累加 E_j 得到总误差 $E = E + E_j$ 。

Step 7:误差函数沿梯度下降修正第 j 个训练样本对应的权值调整量。判断是否训练完所有的样本 ($j > n$),成立则转至 Step 8,否则转至 Step 5。

Step 8:所有训练样本输入网络训练后得到的权值调整量的均值作为权值的修正量。

Step 9:按照修正量对权值做调整后,得到满足精度要求的网络误差(即抗原) $E_i = E / (h * m)$ 。

Step 10:由式(3)计算得到抗体 kt_i 与抗原 E_i 的亲和力。

Step 11:当处理过的抗体总数小于初始群体规模 N ,则转至 Step 3,否则代表初始群体 KT 中的所有抗体处理完毕,接着对群体中的抗体可以根据式(4)、(5)、(6)进行促进和抑制。

Step 12:用式(7)、(8) 对抗体群体 KT 进行交叉和变异后,权值得到更新,产生新的抗体群体 KT' 。

Step 13:两个抗体群体 KT 和 KT' 合并,把其中与 E_i 亲和力值最大的前 N 个抗体赋给 KT ,同时找出最好的抗体并标记为 $kt_{\max}$ 。

Step 14:将 KT 中 5%的亲和力值小的个体删除并以随机产生的新个体代替。

Step 15:判断迭代次数是否大于最大迭代次数 $g_{\max}$,成立则转至 Step 16,否则转至 Step 2。

Step 16:对 $kt_{\max}$ 解码得到最优权值矩阵 W 和 V 。

3 仿真实验与结果分析

本文以文献[10]中无线传感器网络监测得到的

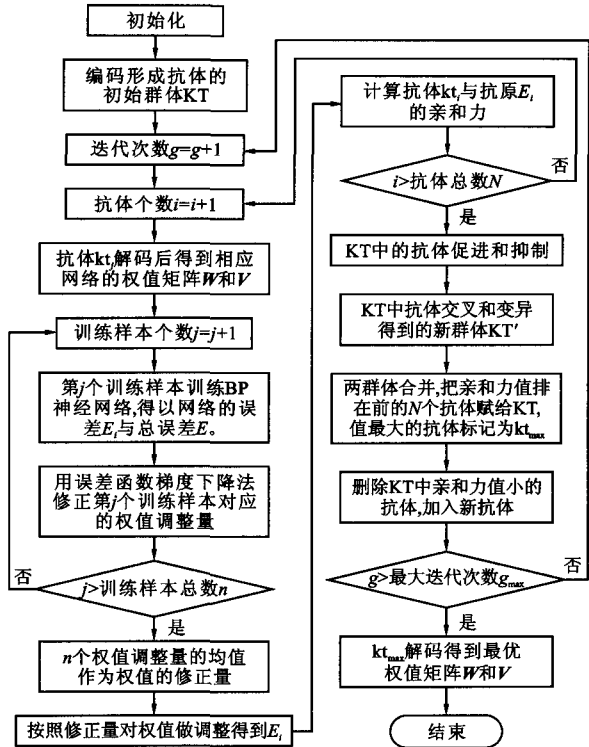


图 2 免疫算法优化 BP 神经网络流程图

矿井环境数据样本为例用 Matlab 7.0 进行了仿真实验。根据文献[10],实验使用的 BP 神经网络的输入层节点个数 n 取值 6,对应煤矿 6 种气体;输出层节点个数 m 定义为 3,将不安全、较不安全、一般不安全、较安全和安全这 5 种煤矿状态的期望输出值分别编码为 001,010,011,100,101;隐含层节点个数 h 是 9;因此网络结构是 6-9-3。其他各项参数设置是:抗体初始化个数 $N=18$,最大迭代次数 $g_{\max}=2000$,最优解误差 $E=0.0001$, $\tau=1$, $\delta=0.38$ 。

本文免疫 BP 神经网络算法和对比算法经典 BP 算法在相同的网络结构和参数条件下,对文献[10]的数据进行分类识别。以下是两者从算法训练收敛过程、数据识别准确率和算法性能方面进行比较分析。

由图 3 中两个算法的神经网络训练误差曲线可知,经典 BP 算法在未达到目标误差(即最优解误差)之前就已经 3 次陷入了局部极值,直到迭代 1531 次才跳出极值达到目标误差。而本文免疫 BP 神经网络算法能避免陷入局部极值,同时只迭代了 369 次就收敛于目标误差,训练速度显然比经典 BP 算法有了明显提高。

用文献[10]中表 4.7 的样本数据作为测试数据,测试样本序号对应表 4.7 中的序号,经过本文免疫 BP 神经网络和经典 BP 网络两种网络进行煤矿

状态识别后的实际输出和期望输出如表 1 所示。

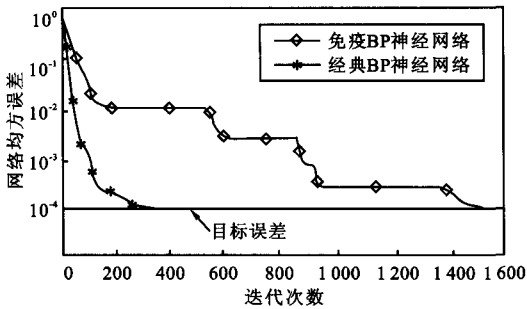


图 3 两种算法的训练收敛曲线

从表 1 的分类识别结果可知,经典 BP 神经网络对煤矿状态进行识别的准确率不高,用 20 组测试数据做测试时的准确率仅大约 75%,而免疫 BP 神经网络能有效地识别煤矿状态,其准确率有 90% 左右。

表 1 两种算法的识别准确率比较

测试样 本序号	实际输出						期望 输出
	免疫 BP 神经网络输出	经典 BP 神经网络输出	免疫 BP 神经网络输出	经典 BP 神经网络输出	免疫 BP 神经网络输出	经典 BP 神经网络输出	
1	0.031	0.981	0.127	0.212	0.913	0.364	010
2	0.997	0.038	0.965	0.893	0.327	0.296	101
3	0.981	0.025	0.973	0.798	0.699	0.401	100
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
20	0.008	0.018	0.996	0.322	0.375	0.509	001

4 结束语

仿真结果表明,本文提出的免疫 BP 神经网络的数据分类器,充分发挥了 BP 算法的局部和免疫算法的全局搜索寻优能力,有效避免了经典 BP 神经网络极易陷入局部极值的现象,且 BP 神经网络

经过免疫算法优化后进行传感器数据识别时,在准确率、网络性能和收敛速度等方面都比经典 BP 神经网络优势明显。

参考文献:

[1] 王智文. 基于改进 BP 神经网络的车牌字符识别[J]. 微电子学与计算机, 2011, 28(9): 66-69.

[2] 李松, 罗勇, 张铭锐. 遗传算法优化 BP 神经网络的混沌时间序列预测[J]. 计算机工程与应用, 2011, 47(29): 52-55.

[3] 杨梅, 卿晓霞, 王波. 基于改进遗传算法的神经网络优化方法[J]. 计算机仿真, 2009, 26(5): 198-201.

[4] 洪月华. 基于粗糙集与 BP 神经网络的分布式数据挖掘算法[J]. 玉林师范学院学报, 2011, 32(5): 133-136.

[5] 王宏图, 黄振华, 范晓刚, 等. 粗糙集—神经网络理论在矿井通风系统评价中的应用[J]. 重庆大学学报, 2011, 34(9): 90-94.

[6] Chun J S, Kim M K, Jung H K, et al. Shape optimization of electronic devices using immune algorithm[J]. IEEE Trans on Magnetics, 1997, 33(2): 1876-1879.

[7] 倪红梅, 王维刚. 免疫神经网络在固井质量预测中的应用研究[J]. 计算机仿真, 2009, 26(7): 267-269.

[8] 罗菲, 何明一. 基于免疫遗传算法的多层前向神经网络设计[J]. 计算机应用, 2005, 25(7): 1161-1165.

[9] 陈伟根, 齐辉. 王有元, 等. 一种气体传感器阵列检测模式识别新方法[J]. 重庆大学学报, 2010, 33(1): 61-66.

[10] 曹亮. 基于多源信息融合的煤矿安全无线监测系统研究[D]. 西安: 西安建筑科技大学, 2008.

作者简介:

洪月华 女, (1973—), 硕士, 副教授. 研究方向为数据挖掘、人工智能、无线传感器网络。

(上接第 148 页)

[5] Wang S, Gupta S K. DS-LFSR: a bist tpg for low switching activity[J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2002, 21(7): 842-851.

[6] Girard P, Guiller L, Landrault c, et al. A test vector ordering technique for switching activity reduction during test operation[C]// Ninth Great Lakes Symposium on VLSI. Ypsilanti, MI: IEEE, 1999: 24-27.

[7] 何蓉晖, 李晓维, 宫云战. 一种低功耗 BIST 测试产生器的方案[J]. 微电子学与计算机, 2003, 20(2): 36-39.

[8] 谈恩民, 宋胜东, 施文康. 单/双跳变向量插入式低功耗 BIST 设计方法[J]. 微电子学与计算机, 2007, 24(8): 183-185.

[9] 王义, 傅兴华. 低功耗单输入跳变测试理论的研究[J]. 微电子学与计算机, 2009, 26(2): 5-7.

[10] Zhang X, C I H Chen, Chakravarthy A. Structure design and optimization of 2-D LFSR-based multi-sequence test Generator in built-in self-test [J]. IEEE Transactions Instrumentation and Measurement, 2008, 57(3): 651-663.

作者简介:

李 鹏 男, (1987—), 硕士研究生. 研究方向为集成电路测试、可测性设计。

颜学龙 男, (1962—), 教授. 研究方向为集成电路测试、可测性设计、故障诊断和边界扫描技术等。