

一种实现结构风险最小化思想的结构自适应神经网络模型*

陈 果

(南京航空航天大学民航学院 南京 210016)

摘要: 本文提出了一种实现结构风险最小化思想的结构自适应神经网络学习模型,该方法运用遗传算法进行神经网络结构参数的学习,运用BP算法进行神经网络内部权值学习,有效地实现了结构风险最小化思想。与传统的基于经验风险最小的神经网络模型相比,它具有更强的自适应能力,能够弥补学习方法本身的缺陷,充分保证了模型的泛化能力。最后,将本文方法应用于非线性时间序列预测和模式识别,并与基于结构风险最小原则的支持向量机学习模型进行了比较,算例充分表明了本文方法的正确有效性。

关键词: 机器学习; 结构风险最小化; 神经网络; 遗传算法; 支持向量机

中图分类号: O329 F201 **文献标识码:** A **国家标准学科分类代码:** 460.1540

Structure self-adaptive neural network model realizing structural risk minimization principle

Chen Guo

(College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: In this paper, the structure self-adaptive neural network learning model, which can realize SRM principle, is put forward. In the model, GA is used to implement neural network structure parameter learning, and back propagation algorithm (BP) is used to carry out inner weight learning, and the SRM principle is realized effectively. Comparing with traditional neural network model based on ERM, structure self-adaptive neural network model possesses stronger self-adaptive ability, and it can remedy the shortcomings of single learning method and fully assure model generalization. In the end, the proposed new method is applied to non-linear time series forecast and pattern recognition, and is compared with support vector machine (SVM) learning model that is based on SRM. Examples show the correctness and validity of the proposed method.

Key words: machine learning; structural risk minimization (SRM); artificial neural network (ANN); genetic algorithm (GA); support vector machine (SVM)

1 引言

基于数据的机器学习是现代智能技术中十分重要的一个方面,主要研究如何从一些观测数据(样本)出发得出目前尚不能通过原理分析得到的规律,利用这些规律去分

析客观对象,对未来数据或无法观测的数据进行预测。目前,神经网络已经成为机器学习方法的研究热点,并且得到了广泛的应用。但是,基于经验风险小的学习原则使神经网络的泛化能力不能得到保证,因此性能优越的神经网络结构往往获取比较困难;基于结构风险最小学习原则的支持向量机^[1],在理论上充分保证了模型的泛化能力,但

收稿日期:2006-11 Received Date:2006-11

* 基金项目:国家自然科学基金(50705042),航空科学基金(2007ZB52022)资助项目

其模型参数(如惩罚因子 C 、损失函数 ε 、核函数及其相关参数等)对模型的识别精度均具有很重要的影响^[2]。

文献[3]研究了神经网络预测模型的影响因素,并用遗传算法进行了结构参数优化,本文在此基础上,提出一种实现结构风险最小化思想的机器学习新方法——结构自适应神经网络学习模型。该模型运用遗传算法实现神经网络结构参数的学习,运用 BP 算法实现网络内部连接权值的学习。与支持向量机相比,新方法在泛化能力上与其基本相当,但是本文方法能够实现结构参数的自适应选取,不存在模型参数选择问题,因此更加具有适用性。

2 结构风险最小原则

统计学习理论中关于经验风险和实际风险之间关系的结论,即推广性的界^[1],是分析学习机器性能和发展新的学习算法的重要基础。从推广性的界可以得出以下结论:

(1) 当样本数较小时(通常小于 20),用经验风险近似真实风险就有较大的误差,用经验风险最小化取得的最优解可能具有较差的推广能力;样本数较多时,置信范围就会很小,经验风险最小化的最优解就接近实际的最优解。

(2) 对于一个特定的样本数,学习机器(分类器)的 VC 维越高(即复杂性越高),置信范围就越大,导致真实风险与经验风险之间可能的差就越大。

传统机器学习方法中普遍采用的经验风险最小化原则是首先确定选定分类器模型,然后最小化经验误差。由于缺乏对模型推广能力的认识,这种选择往往是依赖先验知识用另一种和经验进行的,造成了神经网络方法对使用者“技巧”的过分依赖。

结构风险最小化思想首先把函数集分解为一个函数子集序列,使各个子集能够按照复杂性大小排列,也就是按照 VC 维大小排列,这样在同一个子集中置信范围就相同;在每个子集中寻找最小经验风险,通常它随着子集复杂度的增加而减少。选择最小经验风险与置信范围之和最小的子集,就可以达到期望风险的最小,这个子集中使经验风险最小的函数就是要求的最优函数。

结构风险最小化原则为我们提供了一种不同于经验风险最小化的更科学的学习机器设计原则,但是,实施这一原则却非常困难,关键是如何构造函数子集结构。遗憾的是,目前尚无关于如何构造预测函数子集结构的一般性理论。支持向量机是一种比较好的实现了结构风险最小化思想的方法。

3 实现结构风险最小化思想的结构自适应神经网络模型

本文构造了一种能够有效实现结构风险最小化思想

的结构自适应神经网络学习模型。该算法的基本思想是:将神经网络模型分为外部结构参数和内部权值参数。利用样本集中的一部分数据作为训练样本,在给定神经网络模型初始外部结构参数的情况下,利用神经网络的 BP 算法进行学习,获取模型的内部连接权值参数,再利用学习模型实现对全体样本的预测或识别,计算预测或识别误差,形成遗传算法的适应度函数,然后运用遗传算法的学习机制,自动调节神经网络模型的外部结构参数,在新的外部结构参数下,应用神经网络 BP 调整算法再得到模型内部连接权值参数,同时计算预测误差和适应度值,再进行调整直到达到遗传算法停止条件。最后输出具有最优泛化能力的神经网络模型。整个学习过程均自动完成。因此,该机器学习算法具有很强的自适应能力,同时有效地实现了结构风险最小化思想。

结构自适应神经网络学习模型的学习流程如图 1 所示,该学习过程包括内部连接权值参数的学习和外部结构参数的学习,在图 1 中,虚线框为内部参数的学习过程,双点划线框中为外部参数的学习。下面将详细阐述该方法的学习过程。

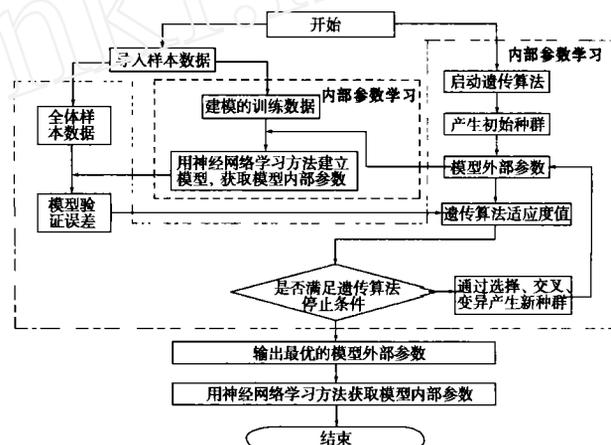


图 1 结构自适应神经网络模型学习流程

Fig. 1 Learning flowchart of structure self-adaptive NN model

(1) 算法开始后,读取样本数据,同时启动遗传算法,对模型外部结构参数进行编码,产生初始种群,对种群解码后得到模型的外部结构参数;

(2) 将样本数据的一部分作为神经网络的训练样本,对于模式识别,通常随机选取一半作为训练样本,对于时间序列预测问题,通常为前一半数据。

(3) 在获取了模型的初始外部结构参数和训练样本数据后,将启动神经网络的 BP 算法,通过对训练样本的学习,算法收敛后将得到模型的内部连接权值参数。对所得到的神经网络模型,用全体样本数据进行验证,以衡量模型的泛化能力。通常用模型的预测值与样本的实际值进行比较,得到模型的误差,并转化为遗传算法的适应度函数值。

(4) 依据种群中各染色体的适应度值,对种群中的

个体进行选择、交叉和变异以获得新一代的种群,再对新一代种群中的染色体解码,获取模型的新的外部参数,然后启动神经网络/支持向量机在给定训练样本学习得到新的内部参数。再运用全体样本对内部参数进行验证,并对外部参数进行调整,进行下轮的外部参数学习,直到达到遗传停止条件。

(5)最后输出模型最优的外部参数,同时在给定的训练样本下,通过神经网络的学习也将得到最优的内部参数。

由于时间序列预测和模式识别是2类典型的机器学习问题,因此本文将结构自适应神经网络模型应用于这2方面的机器学习问题进行研究。

4 基于结构自适应神经网络模型的非线性时间序列预测

设所研究的时间序列为 $\{x(t)\}, t = 1, 2, \dots, N$, 则当前状态的信息可以表示成 m 维的延迟矢量: $x(t + \tau) = f\{x(t), x(t - \tau), \dots, x(t - (m - 1)\tau)\}$, 其中, m 为嵌入维数, τ 为时间延迟, 通常取为采样间隔。已经证明: 假设动力系统的维数为 d , 如果 $m \geq 2d + 1$, 则这种映射产生的伪相空间和系统的状态空间微分同胚, 即拓扑等价, 它们的动力学特性定性意义上完全相同。

由此可见, 对时间序列的预测, 关键在于根据已知时间序列数据, 对非线性系统相空间的重构, 找出从 m 维空间映射到一维空间的映射函数。由于多步预测可以由单步预测迭代而成, 因此, 不失一般性, 可以以单变量单步预测为例进行研究。设一个单变量时间序列 $\{x_1, x_2, \dots\}$, 对它进行预测的前提是认为其未来值与其前面的 m 个值之间有着某种函数关系, 可描述为:

$$x_{n+k} = F(x_n, x_{n-1}, \dots, x_{n-m+1}) \quad (1)$$

由此可见, 时间序列预测本质上仍然是一个函数逼近问题。关键在于是否能通过对序列的学习获得式(1)所表达的函数^[4]。通常衡量预测值与实测值差别的变量采用平均相对变动值(average relative variance, ARV)^[5], 其定义为:

$$ARV = \frac{\sum_{i=1}^N [x(i) - \hat{x}(i)]^2}{\sum_{i=1}^N [x(i) - \bar{x}]^2} \quad (2)$$

式中: N 为比较数据个数; $x(i)$ 为实测数据值; \bar{x} 为实测数据平均值; $\hat{x}(i)$ 为预测值。显然, 平均相对变动值 ARV 越小, 也表明预测效果越好, $ARV = 0$ 表示达到了理想预测效果, 当 $ARV = 1$ 时, 表明模型仅达到平均值的预测效果。

设样本总数为 N , 将其分为训练样本集和测试样本集, 用训练样本集建立预测模型, 用测试样本集测试模型

的泛化能力。式(2)中的比较数据个数如果为样本总数 N , 则得到整个样本的平均相对变动值 ARV_{ALL} , 如果比较的数据个数为测试样本数, 则得到测试样本的相对评价指标 ARV_{TEST} , 显然, 当 ARV_{ALL} 与 ARV_{TEST} 接近时, 表明模型具有很好的泛化能力, 当二者相差很大时, 表明模型的泛化能力很差。

利用神经网络进行预测研究的基础是它能拟合任意的非线性函数, 并且具有很强的泛化能力^[7]。通常利用多层前向神经网络。由于只有一个隐层的神经网络, 只要隐层节点足够多, 就可以以任意精度逼近一个非线性函数。因此, 不失一般性, 可以用3层BP网络来进行时间序列的预测。

基于神经网络的时间序列预测模型中, 模型的内部参数和外部参数分别为:

(1)内部参数: 网络连接权值, 通过对训练样本用神经网络的BP算法学习成功后自动产生。

(2)外部参数: ①输入层节点数(嵌入维数): 影响到能否重构非线性系统的相空间; ②隐层节点数: 直接影响到所逼近的非线性函数复杂程度和模型的泛化能力; ③训练步数: 对网络泛化能力具有很大的影响, 网络训练步数越多, 网络结构越复杂, 泛化能力往往将变得很差。

5 基于结构自适应神经网络的模式识别

由于在模式识别中, 多类问题往往可以转化为两类问题来处理, 而且这样做往往会使网络结构简单、训练样本要求少、训练时间大大缩短, 因此本文不失一般性, 仅仅考虑两类模式识别问题。并以3层BP网络为研究对象。

设 $\{x_1, x_2, \dots, x_n\}$ 为输入 n 维输入向量, 输入向量需要归一化处理, 将数值变换到0和1之间; y 为输出, 训练样本的取值为“+1”或“-1”, 分别代表不同的2类。神经元激活函数均为双极性 Sigmoid 函数, 即 $f(x) = \frac{1 - e^{-x}}{1 + e^{-x}}$ 。

设神经网络模型的隐层节点数为 q , 隐层节点的激活函数为 f_H 、阈值为 $\theta_j (j = 1, 2, \dots, q)$; 输出层节点数为 1, 输出层节点的激活函数为 f_O 、阈值为 β 。第 i 个隐节点与第 j 个输入节点的连接权为 w_{ij} , 则, 第 i 个隐节点与输出节点的连接权为 w_i 。则:

$$y = f_O \left\{ \sum_{i=1}^q w_i f_H \left[\sum_{j=1}^n w_{ij} x_j + \theta_i \right] + \beta \right\} \quad (3)$$

由2类分类器原理, 根据输出 y 的值来对样本进行判别分类, 即: 如果 $y > 0$, 则判别为“+1”类; 如果 $y < 0$, 则判别为“-1”类;

基于神经网络的分类器模型中, 模型的内部参数和

外部参数分别为:

(1)内部参数:网络连接权值,通过对训练样本用神经网络的 BP 算法学习成功后自动产生。

(2)外部参数:①初始权值:影响网络的收敛位置;②隐层节点数:直接影响到所逼近的非线性函数复杂程度和模型的泛化能力;③训练步数:对网络泛化能力具有很大的影响,网络训练步数越多,网络结构越复杂,泛化能力往往将变得很差。

6 算 例

6.1 时间序列预测算例

本文针对 1700~1900 年的太阳黑子数据,分别建立结构自适应神经网络和支持向量机的学习模型,并对其预测结果进行比较,在模型中,训练样本比例均为 0.2。结构自适应神经网络模型中遗传算法的参数为:交叉率和变异率分别为 0.50 和 0.05,种群数为 30,进化代数数为 10。图 2 为实测值与预测值的比较结果。

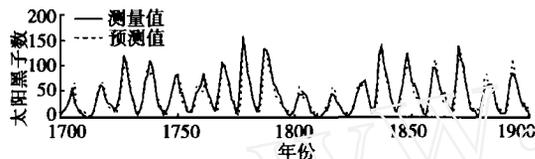


图 2 结构自适应神经网络学习模型获取的最优模型预测值与实测值比较

Fig. 2 Comparison of forecast value and measured value (Optimum NN)

通过结构自适应神经网络学习后,优化参数均用二进制编码。得到的模型外部参数为:输入层节点数为 3,隐层节点数为 21,训练步数为 9,在该外部参数下,神经网络通过对 0.2 的样本数据进行训练后,得到的模型对整个样本的预测误差和测试样本的预测误差分别为: $ARV_ALL = 0.1298$, $ARV_TEST = 0.1318$ 。图 3 为实测值与预测值的比较结果。

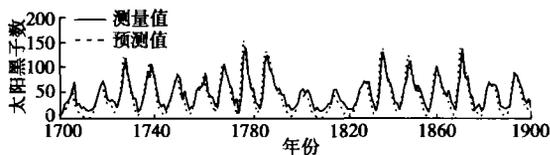


图 3 支持向量机在最优参数下学习获取的预测值与实测值比较

Fig. 3 Comparison of forecast value and measured value (Optimum SVM)

由于支持向量机也是基于结构风险最小的学习机器,为了验证本文结构自适应神经网络模型有效地贯彻了结构风险最小化思想,本文将结构自适应神经网络模

型与其进行对比。支持向量机通过人工选择最优模型参数为:惩罚因子 C 为 3、多项式核函数的次数为 3、嵌入维数为 2,在该参数下,支持向量机通过对 0.2 的样本数据进行训练后,得到的模型对整个样本的预测误差和测试样本的预测误差分别为: $ARV_ALL = 0.2026$, $ARV_TEST = 0.2002$ 。图 4 为实测值与预测值的比较结果。通过比较可以得出以下结论:

(1)在结构自适应神经网络模型中,遗传算法起到了调节模型泛化能力的作用,是实现结构风险最小化思想的关键,因此在很大程度上弥补了神经网络模型的不足,从而大大提高了神经网络的泛化能力。从预测结果来看,神经网络的效果要比支持向量机好,这有力地证明了结构自适应神经网络有效地实现了结构风险最小思想,具有良好的泛化能力和对参数的自适应能力。

(2)结构自适应神经网络模型在学习过程中自动选取,支持向量机尽管理论上充分保证了模型的泛化能力,但其泛化能力也需要适当的模型参数才能体现出来,而支持向量机的核函数及其参数的选择目前尚无理论指导,需要人工进行选择。因此,支持向量机模型的自适应能力较差。

6.2 模式识别算例

(1)数据 1:圆分类问题

在直角平面 XOY 上,在圆 $x^2 + y^2 = 16$ 内定义为一类,标记为“-1”,在圆外定义为一类,标记为“+1”,随机产生 60 个训练样本,其中 30 个“+1”类,30 个“-1”类;随机产生 800 个测试样本进行分类精度测试。

(2)数据 2:双螺旋问题

训练样本数为 200 个,在直角平面 XOY 上,设 $i = 0, 1, \dots, 99$, $\theta(i) = i \times \pi/16$, $r(i) = 6.5 \times (104 - i)/104$, $x(i) = r(i) \times \sin \theta(i)$, $y(i) = r(i) \times \cos \theta(i)$ 。由点 $(x(i), y(i)) (i = 0, 1, \dots, 99)$ 组成第一类,标记为“+1”;由点 $(-x(i), -y(i)) (i = 0, 1, \dots, 99)$ 组成第二类,标记为“-1”。

测试样本数为 160 个,在直角平面 XOY 上,设 $i = 0, 1, \dots, 80$, $\theta(i) = (i + 0.5) \times \pi/16$, $r(i) = 6.5 \times [104 - i - 0.5]/104$, $x(i) = r(i) \times \sin \theta(i)$, $y(i) = r(i) \times \cos \theta(i)$ 。由点 $(x(i), y(i)) (i = 0, 1, \dots, 80)$ 组成第一类,标记为“+1”;由点 $(-x(i), -y(i)) (i = 0, 1, \dots, 80)$ 组成第二类,标记为“-1”。

为了验证本文方法的有效性,分别用结构自适应神经网络和支持向量机对分类数据 1 和 2 进行算例分析。

表 1 为结构自适应神经网络的模型参数及计算结果。在遗传算法中,对中间隐层节点数 M 和网络训练步数 S 以及产生初始随机种群的随机种子 R 进行二进制编码,种群数为 30,进化代数数为 10,交叉率为 0.5,变异率为 0.05。表 2 为支持向量机学习模型参数及计算结果。

表1 结构自适应神经网络模型参数及计算结果

Table 1 The recognition results and structure parameters of NN

分类数据	隐层节点 M 编码位数	训练步数 S 编码位数	随机种子 R 编码位数	隐层节点 M 优化结果	训练步数 S 优化结果	随机种子 R 优化结果	误差率(%)	识别率(%)
数据集1	5	5	5	28	13	200	1.375	98.625
数据集2	5	10	5	27	258	10	8.75	91.25

表2 支持向量机模型参数及计算结果

Table 2 The recognition results and parameters of SVM

分类数据	惩罚系数 C	高斯核函数 σ	误差率(%)	识别率(%)
数据集1	0.1	0.65	9.2	90.8
数据集2	100	0.235	0	100

图4为对数据集1,利用结构自适应神经网络学习模型得到的最优分类器对样本进行训练所得到的分类曲线;图5为对数据集1,利用支持向量机学习模型得到的最优分类器对样本进行训练所得到的分类曲线;图6为对数据集2,利用结构自适应神经网络学习模型得到的最优分类器对样本进行训练所得到的分类曲线;图7为对数据集2,利用支持向量机学习模型得到的最优分类器对样本进行训练所得到的分类曲线。从表1、表2、图4、图5、图6及图7可以得出以下结论:

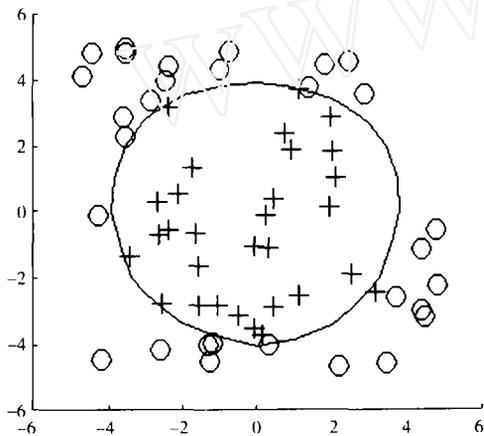


图4 数据集1的分类曲线(结构自适应神经网络)

Fig. 4 Classification curve of data set 1 by self-structure NN

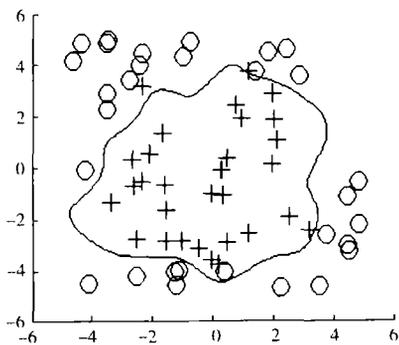


图5 数据集1的分类曲线(支持向量机)

Fig. 5 Classification curve of data set 1 by SVM

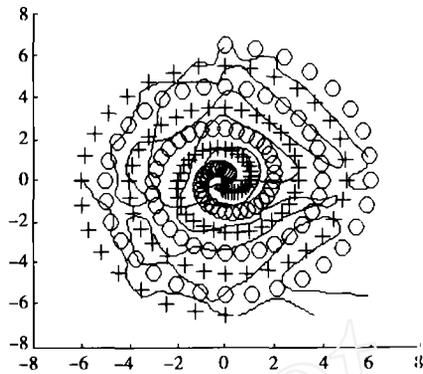


图6 数据集2的分类曲线(结构自适应神经网络)

Fig. 6 Classification curve of data set 2 by self-structure NN

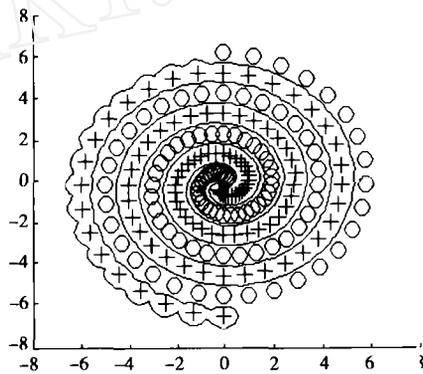


图7 数据集2的分类曲线(支持向量机)

Fig. 7 Classification curve of data set 2 by SVM

(1)对数据集1和2,结构自适应神经网络模型和支持向量机模型均获得了非常好的计算结果,对数据集1和2的识别率均达到了90%以上,整个学习过程不需要人工干预,大大提高了模式识别的自动化程度和精度。

(2)对于双螺旋这样的经典分类难题,许多文献均进行了详细的研究,通常认为用常规的BP网络很难获得好的泛化特性,Baum等人试图用一个2-50-1的BP网络来解决,未能获得良好的结果^[6];Chen等人提出生成一个收缩算法来训练一个2-20-20-1的BP网络,经过3000次迭代后对测试样本的识别率只有89.6%^[7];文献[8]用优化神经元激活函数的方法使识别率达到了100%。而本文提出的结构自适应神经网络学习模型对双螺旋问题所得到的常规BP网络模型,模型结构为2-27-1,训练步数为258,网络初始权值的随机种子为10,最后对双螺旋问题的识别率达到了91.25%。整个

过程不需要人工设置神经元函数、网络层数、每层节点数以及设计复杂的学习算法等繁琐程序,因此,本文的算法在自动化、适应性和识别精度等方面较其他方法具有明显优势。

(3)在结构自适应神经网络学习模型中,遗传算法起到了调节模型泛化能力的作用,因此在很大程度上弥补了神经网络模型的不足,有效地实现了结构风险最小化思想,从而大大提高了神经网络的泛化能力。从识别结果可以看出,结构自适应神经网络模型与支持向量机基本相当,均表现出很强的泛化能力。对数据集 1,结构自适应神经网络模型识别效果更好,对数据集 2,SVM 更好。

从模型参数选取来看,结构自适应神经网络学习模型在学习过程中,能够自动选取模型参数,支持向量机尽管是基于结构风险最小原则的学习方法,但是模型的泛化能力需要适当的核函数及其参数才能表现出来,因此本文提出的结构自适应神经网络模型比支持向量机更具适用性。

7 结 论

(1)提出了一种实现结构风险最小思想的结构自适应神经网络学习模型。利用遗传算法实现神经网络外部结构参数的学习,运用 BP 算法实现神经网络连接权值的学习,有效地实现了结构风险最小化思想。

(2)将结构自适应神经网络模型应用于时间序列预测和模式识别,并与支持向量机进行了比较分析,有力地验证了结构自适应神经网络学习模型的正确性、有效性和适用性。

参考文献

- [1] VAPNIK V. The nature of statistical learning[M]. New York: Springer, 1995.
- [2] 尉询凯. 基于支持向量机的航空发动机滑油监控分析[J]. 航空动力学报, 2003, 18(6):393-397.
WEI X K. Aeroengine oil monitoring analysis based on support vector machine[J]. Journal of Aerospace Power, 2003, 18(6):393-397.
- [3] 陈果. 神经网络模型的预测精度影响因素分析及其优化[J]. 模式识别与人工智能, 2005, 18(5): 528-534.
CHEN G. Analysis of influence factors for forecasting precision of artificial neural network model and its optimizing[J]. PR & AI, 2005, 18(5): 528-534.
- [4] CHOLEWO T, ZURADA J M. Sequential network construction for time series prediction[C]. Proceedings of the IEEE International Joint Conference on Neural Networks, 1997: 2034-2039.
- [5] FARBER L A. Nonlinear signal processing using neural network: Prediction and system modeling[R]. Technical Report LA-UR-87-2662, Los Alamos National Laboratory. Los Alamos. NM, 1987.
- [6] BAUM E B, LANG K L. Constructing hidden units using examples and queries[A], In: Advances in Neural Information Processing system 3[C]. San Maeto. CA: Morgan Kaufmann, 1991: 904-910.
- [7] CHEN Q C. Generating-Shrinking algorithm for learning arbitrary classification[J]. Neural Networks, 1994(7): 1477-1489.
- [8] 吴佑寿, 赵明生, 丁晓青. 一种激励函数可调的新人工神经网络及应用[J]. 中国科学: E 辑, 1997, 27(1): 55-60.
WU Y SH, ZHAO M SH, DING X Q. A new artificial neural network with adjustable excitation function and its application[J]. Chinese Science: E, 1997, 27(1): 55-60.

作者简介



陈果,男,1972 年出生,分别于 1994、1997 和 2000 年在西南交通大学获得学士、硕士和博士学位,现为南京航空航天大学副教授,主要研究方向为航空发动机状态监测与故障诊断、智能诊断与专家系统、机器学习与知识获取、图像处理及模式识别、非线性转子动力学等。

地址:南京航空航天大学民航学院,210016

电话:025-84891850;E-mail:cgzyx@263.net

Chen Guo, male, born in 1972. He received bachelor degree in 1997, master degree in 1997, and doctor degree in 2000, all from Southwest Jiaotong University. Now he is an associate professor in Nanjing University of Aeronautics and Astronautics, his main research fields include aeroengine condition monitoring and fault diagnosis, intelligent diagnosis and expert system, machine learning and knowledge acquisition, image processing and pattern recognition, nonlinear rotor dynamics, and etc.

Address: College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, Jiangsu, China

Tel: +86-25-84891850; E-mail:cgzyx@263.net