基于支持向量回归集成的陀螺仪参数漂移预测

刘勇志
(海军潜艇学院导弹兵器系，青岛　266071)

摘　要:泛化能力是智能方法用于参数预测的最重要的问题之一,提出了支持向量回归集成方法。为了增加个体之间的差异性,提出了基于聚类方法的个体生成方法。首先利用聚类方法将样本分为若干子类,然后用不同结构的支持向量回归学习不同的样本子类,权值由个体在验证集上的泛化误差决定。将ESVR陀螺仪参数飘移数据的预测,并与单支持向量回归,单神经网络,神经网络集成以及组合预测方法进行比较。结果证实,ESVR的预测精度总体高于其他方法。

关键词:参数预测;支持向量回归;集成;神经网络;泛化能力

中图分类号: TP206+.3　　**文献标识码**:A　　**文章编号**:1009-3516(2007)04-0049-04

陀螺仪是导弹控制系统中的十分重要的惯性器件,陀螺仪的精度对导弹的命中精度有直接的重要影响。陀螺仪的参数漂移是导弹落点偏差的重要误差源,如何减小陀螺仪的参数漂移进而提高导弹命中精度是一个非常重要的课题。陀螺仪参数漂移预测是解决这一问题的有效途径之一,主要思想是通过预测陀螺仪的参数漂移进而对其进行补偿修正,达到提高陀螺仪精度的目的。

陀螺仪参数漂移预测本质上是一个非线性时间序列预测问题。基于知识的非线性时间序列预测方法是比较有效的方法,这类方法包括:神经网络,支持向量机等。神经网络和支持向量机用于时间序列预测的关键问题之一是方法的泛化能力,即方法对未训练样本正确输出的能力。神经网络集成被认为是泛化能力较强的神经网络模型[1],其主要思想是用若干个神经网络对同一个问题进行学习,集成在某输入下的输出由所有神经网络共同决定[2]。理论证明,集成的泛化能力强于单神经网络[3]。集成模型能够提高泛化能力的主要原因在于:集成中个体之间存在的差异性减小了个体的泛化误差。但是,神经网络集成中的个体全部为神经网络,存在神经网络固有的缺陷,如预测结果的难以解释,神经网络以及集成的结构设计问题,个体的泛化能力不强等。从下面的论述我们可以知道,集成的泛化能力公式不仅适用于神经网络而且适用于其他方法。支持向量机(SVM)是一种泛化能力较强的通用学习方法,它较好地解决了小样本、非线性、高维数、局部极小点等实际问题[4]。但是,单个支持向量机的泛化能力是有限的。本文提出支持向量机集成方法,即用若干个SVM模型对同一个问题进行学习,集成在某输入下的输出由所有SVM共同决定。

1　基于K-means的个体生成方法

算法首先随机选取k个点作为初始聚类中心,然后计算各个数据对象到各聚类中心的距离,把数据对象归到离它最近的那个聚类中心所在的类;对调整后的新类计算新的聚类中心,如果相邻两次迭代的聚类中心没有任何变化,说明数据对象调整结束,聚类准则函数已经收敛,至此算法结束。该算法具体如下[5]:

1)给定大小为n的样本集$X=\{x_1,x_2,\cdots,x_n\}$,选取k个初始聚类中心center(i,t),t为迭代次数,$1\leq i\leq k$;

2)计算每个样本对象与聚类中心的距离:$d(x_i,\text{center}(j,t))=||x_i-\text{center}(j,t)||$。$||x||$为向量范

收稿日期:2007-03-10
作者简介:刘勇志(1957-),男,黑龙江哈尔滨人,教授,主要从事导弹测试,故障诊断,智能测试与控制等研究.

数,一般用欧式距离表示。

3) 将 x_i 归为与其距离最小的类别,直到所有样本均被聚类;

4) 重新计算聚类中心:

$$\text{center}(j,t) = \frac{1}{n}\sum_{i=1}^{n} x_i^{(j)}, 1 \leqslant j \leqslant l \tag{1}$$

式中:n 为第 j 类的样本数, $x_i^{(j)}$ 表示属于第 j 类的第 i 个样本。

5) 如果 $\text{center}(i, t) \neq \text{center}(i, t+1)$,则返回 2),直到 $\text{center}(i, t) = \text{center}(i, t+1)$。

从上述算法可以看出,K - means 算法将样本分为互斥的 k 个子类。同一子类内的样本距离最小,而样本子类之间的距离最大。如果不考虑结构和激活函数的差异,利用不同的个体网络学习不同的样本子类,将会得到差异性最大的个体,这样有利于提高 NNE 的泛化能力。但是,由于个体的学习样本过于集中于任务的某一方面,所以个体的泛化误差会增大,不利于 NNE 的泛化误差的减小。

为了在个体泛化能力与差异性之间达到一个平衡,本文提出样本插入方法,即在每一子类样本中插入其他子类的样本,具体如下:

假设样本被分为 k 类:$X = \{X_1, X_2, \cdots, X_k\}$, $X_i = \{x_{i,1}, x_{i,2}, \cdots, x_{i,mi}\}$, $\sum_{i=1}^{n} m_i = n$,$1 \leqslant i \leqslant k$。样本选择因子为 λ,$0 < \lambda < 1$,各子类插入样本数为

$$\text{num\_insert}(i) = \text{round}(\lambda \times \text{size}(X_i)) \quad 1 \leqslant i \leqslant k \tag{2}$$

式中:round 为取整函数,size 为样本数函数。从各个样本子类中随机选择 num_insert(i) 个样本,形成插入样本子类:$X_i^{\text{Insert}} = (x_i^1, x_i^2, \cdots, x_i^{\text{num\_insert}(i)})$, $1 \leqslant i \leqslant k$

新的样本子类为

$$X'_i = \{X_i, X_1^{\text{Insert}}, X_2^{\text{Insert}}, \cdots, X_{i-1}^{\text{Insert}}, X_{i+1}^{\text{Insert}}, \cdots, X_k^{\text{Insert}}\} \tag{3}$$

新的样本子类不仅完全包含了任务的某一剖面,而且包含了其他剖面的信息。从样本的角度考虑,KBAGI 既满足了个体之间的差异性,也满足了个体的泛化能力。样本选择因子 λ 决定了各子类包含的其它子类的信息量,从而影响个体网络的泛化能力。如果 λ 太小,各子类样本过于偏向某一剖面,导致个体的泛化能力差;如果 λ 太大,各子类样本信息差异性减小,导致个体的差异性较差,从而影响 NNE 的泛化能力。一般情况下,$\lambda = 1/k$, k 为聚类数。

由于 K - means 方法产生了 k 个样本子类,所以个体网络数为 k,这样就确定了 NNE 的集成规模。但是,k 的选择对聚类结果有很大影响,从而对 NNE 的泛化能力影响很大。为了说明此问题,在试验中将样本分别分为不同数目的子类,通过实验找到最优的分类数。

2 支持向量回归机(SVR)

给定训练样本(x_i, y_i), $i = 1,2,\cdots,l$,在线性情况下,我们可以利用下式对函数进行拟合[6]:

$$f(x,\alpha) = (\omega \cdot x) + b \tag{4}$$

上式的损失函数定义为 ε 不敏感损失函数:

$$L(y,f(x,\alpha)) = L(| y - f(x,\alpha) |_{\varepsilon}) \tag{5}$$

ε 不敏感损失函数的优点是可以使估计具有鲁棒性,而且得到的解是稀疏的。本文采用线性 ε 不敏感损失函数为

$$L(y,f(x,\alpha)) = | y - f(x,\alpha) |_{\varepsilon}) = \begin{cases} 0 & | y - f(x,\alpha) | \leqslant \varepsilon \\ | y - f(x,\alpha) |_{-\varepsilon} & | y - f(x,\alpha) | > \varepsilon \end{cases} \tag{6}$$

对于上述优化问题我们引入 Lagrange 算子,化为无约束问题,极小化下面泛函:

$$\begin{aligned} \Phi(\omega,\xi,\xi^*;\alpha,\alpha^*;\gamma,\gamma^*) = {} & \frac{1}{2}(\omega \cdot \omega) + c\Big(\sum_{i=1}^{l}\xi_i + \sum_{i=1}^{l}\xi_i^*\Big) - \sum_{i=1}^{l}\alpha_i[(\omega \cdot x_i) + b - y_i + \varepsilon + \xi_i] \\ & - \sum_{i=1}^{l}\alpha_i^*[y_i - (\omega \cdot x_i) - b + \varepsilon + \xi_i] - \sum_{i=1}^{l}(\gamma_i\xi_i + \gamma_i^*\xi_i^*) \end{aligned} \tag{7}$$

由上面的优化问题可以得最优参数 α_i, α_i^*, $i = 1,2,\cdots,l$,由 ω, b 与 α_i, α_i^* 的关系,可以得到最优估计函

数:

$$\omega = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)x_i, b = y_i - (\omega \cdot x_i) - \varepsilon, \alpha_i \in (0, C) \tag{8}$$

将式(9) 带入式(4) 就可以得到估计的回归函数。由上式可知, ω 的最优估计值仅与 $\alpha_i \neq \alpha_i^*, i = 1,2,\cdots,l$ 所对应的样本矢量 x_i 有关,这样的样本矢量 x_i就称为支撑矢量,支撑矢量只占总样本的一小部分,所以利用支撑矢量可以使训练过程大大简化,而且使得到的估计函数具有很好的推广能力。

上述支持向量回归机是针对样本线性可分的情况,而对样本为非线性的情况,只要选择合适的核函数,用核函数取代线性情况下的内积运算,便可以得到将输入空间投影到高维特征空间,在高维空间中可利用线性回归估计方法解决。

在非线性情况下,我们可以构造如下形式的待估计函数:

$$f(x,\beta) = \sum_{i=1}^{l}\beta_i K(x,x_i) + b \tag{9}$$

其中 $K(\cdot,\cdot)$ 是核函数,系数$\beta_i = \alpha_i - \alpha_i^*, i = 1,2,\cdots,l$ 。同上面的线性情况一样,其中的 α_i, α_i^* 是下面问题的解:

$$\begin{aligned}\max W(\boldsymbol{\alpha},\boldsymbol{\alpha}^*) &= -\varepsilon\sum_{i=1}^{l}(\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_i^*) + \sum_{i=1}^{l}y_i(\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_i^*) \\ &- \frac{1}{2}\sum_{i,j=1}^{l}(\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_i^*)(\boldsymbol{\alpha}_j - \boldsymbol{\alpha}_j^*)K(x_i,x_j)\end{aligned}, \text{subject}\begin{cases}\sum_{i=1}^{l}\boldsymbol{\alpha}_i = \sum_{i=1}^{l}\boldsymbol{\alpha}_i^* \\ 0 < \boldsymbol{\alpha}_i < C, i = 1,2,\cdots,l \\ 0 < \boldsymbol{\alpha}_i^* < C, i = 1,2,\cdots,l\end{cases} \tag{11}$$

支撑矢量为 $\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_i^* \neq 0, i = 1,2,\cdots,l$ 所对应的样本矢量。

3　SVR 集成(Ensemble of SVR, ESVR)

ESVR 的主要思想是利用若干个SVR对同一个问题进行学习,最后集成输出结果。假设有N个SVR共同学习一个任务,学习后的回归函数为:

$$f_i(x,\alpha) = (\omega_i \cdot x) + b_i, 1 < i < N \tag{12}$$

式中:f_i, ω_i, b_i分别表示第 i 个回归函数的输出和参数。假设第 i 个回归函数的权值为 w_i,且 w_i 满足式(1)。ESVR 的输出为

$$f = \sum_{i=1}^{N} w_i f_i \tag{13}$$

本文中权值的选择方法是按照每个 SVR 在验证集上的泛化误差决定的。假设 SVR(i) 在验证集上的泛化误差为 error(i),则 SVR(i) 的权值表示为

$$w_i = \frac{1/\mathrm{error}(i)}{\sum_{i=1}^{N} 1/\mathrm{error}(i)} \tag{14}$$

上式表示泛化误差越大的 SVR,其权值越小,反之亦然。这种权值分配方法比较客观的反映了 SVR 的泛化能力。ESVR 算法步骤如下:

1) 首先按照本文第三部分将样本分为若干子类;

2) 分别利用不同的 SVR 学习不同的样本子类;

3) 根据 SVR 在验证集上的误差确定个体 SVR 的权值;

4) 集成输出。

4　实验

本文将用于陀螺仪参数飘移预测,并与单 SVR、单神经网络、神经网络集成以及组合预测方法做比较验证 ESVR 的优越性。本文利用实测的一组陀螺仪飘移数据对 ESVR 进行测试,数据共 50 组,其中的 42 组作为训练数据,其余作为测试数据。利用 5 倍交叉验证方法训练 SVR 和 NN。本文利用输入为 5 维的 SVR 和

NN 模型对陀螺仪参数飘移数据进行一步预测。为了充分比较 ESVR 的性能,本文还比较了 ARIMA 模型和 RBF 网络的组合模型的预测结果[7],如表 2:

表 2 陀螺仪参数飘移预测结果比较

ESVR	SVR	NN	NNE	组合
0.090 2	0.094 5	0.368 2	0.119 4	0.094

在陀螺仪飘移数据预测中,ESVR 的泛化能力也要好于其他模型,而单神经网络的泛化能力最差。上述实验验证了 ESVR 在泛化能力方面的优越性。但是,在试验中我们仅仅选择了 BP 网络作为比较模型,而没有利用其他性能更好的神经网络。

5 结束语

本文提出了一种泛化能力较强的机器学习模型 ESVR。通过基准数据和实测数据的实验,我们可以看出,ESVR 的泛化能力要优于其他比较模型。由于篇幅所限,以下问题需要进一步研究:

1) 个体生成中,需要确定聚类数,而这也是集成中个体的数目。如果聚类数太多,个体的泛化能力下降,而且计算复杂性增大;如果聚类数太小,个体之间的差异性减小,不利于泛化能力的增强。在 ESVR 中,如何确定最优的聚类数是个重要的问题。

2) 在实测数据试验中,我们选择了输入为 5 维,输出为 1 维的模型,即用 5 个前时刻数据预测后一时刻数据。如果用更多或更少的数据预测能否得到更好的预测效果,是个值得研究的问题。

参考文献:

[1] Granitto P M, Verdes P F, Ceccatto H A. Neural Network Ensemble: Evaluation of Aggregation Algorithms[J]. Artificial Intelligence, 2005 ,163: 139 - 162.

[2] 周志华,陈世福. 神经网络集成[J]. 计算机学报. 2002,25(1): 1 - 8.

[3] Krogh A, Vedelsby J. Neural Network Ensembles, Cross Validation, and Active Learning[A]. Advances in Neural Information Processing Systems 7[C]. CamBridge: MIT press, 1995. 231 - 238.

[4] 罗公亮. 从神经网络到支撑矢量机(上)[J]. 冶金自动化, 2001,(5): 1 - 5.

[5] 李翠霞,于 剑. 一种模糊聚类算法归类的研究[J]. 北京交通大学学报, 2005, 29(2): 17 - 21.

[6] 李国正,王 猛,曾华军. 支持向量机导论[M]. 北京:电子工业出版社,2004.

[7] 吕瑛洁. 基于神经网络的惯性器件故障预报[D]. 西安:第二炮兵工程学院硕士学位论文,2005.

(编辑:姚树峰)

Parameter Drift Forecasting of Gyro Based on Ensemble of Support Vector Regression

LIU Yong - zhi

(Missile Department, Navy Submarine Academy, Qingdao 266071, China)

Abstract: Generalization performance is one of the most important problems of intelligent approaches for parameters forecasting. This paper presents an ensemble of support vector regression (ESVR) which has better generalization performance than other intelligent approaches. To increase the diversity among individuals of ensemble, the paper proposes an individual generating approach based on clustering technique. Firstly, ESVR is used to classify training samples as several subclasses that are used to train different individuals with different kernel functions. The ensemble weights of individuals are determined by the generalization errors on the validation sets. ESVR is tested on the parameter drift data of gyroscope. By comparing single SVR, single neural network, neural network ensemble and combination approach with ESVR in generalization performance, the results reveal that ESVR has better generalization performance than other intelligent approaches in most cases.

Key words: parameter forecasting; support vector regression; ensemble; neural network; generalization performance