

基于动态加权集成学习的遥测数据预测方法

梅玉航, 贾海艳

(中国人民解放军 92941 部队, 辽宁 葫芦岛 125000)

摘要: 飞行任务中的遥测数据是飞行器中各功能子系统监测模块顺序产生的多维时间序列, 其反应各子系统功能是否正常, 对遥测数据的精准预测是研判飞行器运行状态的重要依据; 针对已有时间序列预测算法会随时间劣化的缺点, 提出基于集成学习原理的动态加权神经网络集成算法; 该方法通过神经网络强数据拟合能力, 集成学习算法具有的泛化特性和动态加权算法适应数据的漂移变化特性, 提升算法的整体预测精度; 选择多层感知机神经网络作为基学习器, 给出神经网络基学习器结构设计方法和优化方法, 以及动态加权算法的具体过程, 将其与静态加权算法进行比较实验, 该算法对预测精度提高效果显著, 一定程度抑制数据的漂移, 结果表明采用动态加权集成学习适合对遥测数据的预测。

关键词: 遥测数据; 集成; 神经网络; 动态加权; 预测

Telemetry Data Prediction Method Through Dynamic Weighted Ensemble

MEI Yuhang, JIA Haiyan

(Unit 92941 of PLA, Huludao 125000, China)

Abstract: Telemetry data of flight task are multi-dimensional time-series data streams sequentially produced by functional subsystem, It reflects whether the function of each subsystem is normal, accurate prediction of telemetry data is an important basis for the study and judgment of aircraft running state. Aiming at the disadvantage that the existing time series prediction algorithm will deteriorate over time, a dynamic weighted neural network integration algorithm based on ensemble learning principle is proposed. In this method, the strong data fitting ability of neural network, the generalization characteristic of learning algorithm and the dynamic weighting algorithm are integrated to adapt to the drift and change characteristics of data, so as to improve the overall prediction accuracy of the algorithm. The multilayer perceptron neural network is chosen as the base learner, and the structure design and optimization method of the neural network base learner and the specific process of the dynamic weighting algorithm are given, the experimental results are compared with the static weighted algorithm, it show that this algorithm can improve the prediction accuracy remarkably, and restrain data drift to some extent, the results show that dynamic weighted ensemble learning is suitable for telemetry data prediction.

Keywords: telemetry data; ensemble; neural networks; random weights; prediction

0 引言

遥测数据主要用于获取飞行器上各子系统的运行状态监测数据, 通过分析遥测数据以监测各子系统功能正常与否并对其进行安全控制^[1]。遥测数据本质是多维时间序列, 同时遥测数据易受测量设备、空间环境和运行状态等影响产生不明显的数据漂移变化, 其表现可能是数据的突变形式, 也有可能是缓慢的线性变化等, 变化趋势不易判断, 对其进行数据分析的主要问题之一是能进行高精度的预测, 从而判定其监测的子系统状态是否正常。

目前用于时间序列预测的方法有两类: 一类为数学和统计方法, 如线性回归, ARIMA 方法和指数平滑等^[2-3]; 另一类基于机器学习方法, 如支持向量机回归^[4], 随机森林和神经网络等^[5-6]。对某种模式时间序列的预测都存在特定对应的方法, 但在时间序列存在数据漂移的情况下, 随着时间的推移, 这些方法预测的精度会不断降低。集成学习的优势在于即使一些基学习模型不能准确预测时间序列

模式的变化, 但另一些基学习模型能实现精确预测, 通过对各基学习模型预测输出的加权求和从而实现预测误差相互抵消, 使总体的预测精度得到补偿。因此通过集成多样性的基学习器能解决传统预测方法随时间性能下降的问题^[7]。

在遥测数据分析处理过程中不可避免伴随着噪声, 而噪声源通常不易确定, 因此人工智能神经网络方法更适合对遥测数据的预测。神经网络借助有监督学习方法能建立输入变量与输出变量间复杂的函数关系^[8], 目前广泛用于解决预测问题, 其预测精度和效率取决于神经网络的结构和训练方法^[9]。另外, 神经网络对于遥测数据预测还具有一些特殊的优势: 神经网络具有自学习能力, 有多个可供选择的转换函数和神经网络结构; 通过多层神经网络的隐层神经元的非线性特性可实现非线性映射功能; 高度的并行机制可提高预测的实时性。凭借以上优点, 神经网络算法可实现具有复杂算法的数据预测。但是, 传统的神经网络方法在确定复杂神经网络的参数时需要占用一定时间和

收稿日期: 2021-08-03; 修回日期: 2021-09-09。

作者简介: 梅玉航(1975-), 男, 辽宁兴城人, 硕士研究生, 主要从事数据处理与安全控制方向的研究。

引用格式: 梅玉航, 贾海艳. 基于动态加权集成学习的遥测数据预测方法[J]. 计算机测量与控制, 2021, 29(10): 144-147.

资源, 因此针对多变量时间序列预测问题需要开发能自动确定网络参数和结构的工具, 以找到参数最优解。

考虑到如前述使用单一预测模型存在不稳定性, 而集成算法模型能有效解决模型的泛化性和可信度^[10-11], 因此提出一种基于动态加权神经网络 (NNRW) 集成学习方法^[12], 通过集成多个较简单神经网络来替换单一复杂的神经网络, 从而减少预测模型建立的时间和复杂性, 保证遥测数据预测的实时性。选择神经网络作为基学习器, 根据各基学习器目前和过去的性能, 动态调整输出权值, 实现对遥测数据的预测, 并提高预测的总体精度。

1 集成预测原理

集成预测通过集成学习方法获取数据预测模型, 该方法的主要原则是利用基模型的多样性, 通过对不同基模型输出的适当联合进而获取到相比于其中最优的单个基模型更精确的结果。集成学习过程具体分为 3 个阶段: 集成产生, 集成修剪和集成整合。在集成产生阶段, 训练完成一组具有多样性的预测基模型。在集成修剪阶段, 实现消除基模型集合的冗余, 剔除高误差模型, 从而提高最终预测的总体精度。最后在整合阶段, 利用各基模型计算结果的线性联合实现对各基模型预测结果的整合, 以获取最优的预测结果。

为解决飞行器上多个子系统遥测数据预测问题, 针对各子系统配置相应的神经网络集成学习器, 遥测数据经预处理后, 各子系统相关的遥测数据集分别输入对应训练完成的集成学习器, 并输出下一时刻的状态参数预测值。其工作原理如图 1 所示。

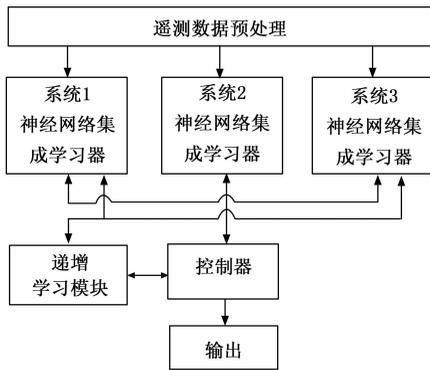


图 1 神经网络集成预测工作原理图

子系统识别 ID 通过通信模块传递到控制器, 控制器可控制不同子系统对应的遥测数据预处理数据集依次输入相应的神经网络集成学习器, 并将集成学习器预测结果传回控制器进行输出, 当控制器发现某个集成学习器预测性能下降时, 向递增学习模块发出学习信号, 该递增学习模块进行训练初始化, 完成训练数据集准备和新的神经网络的训练, 训练完成后对集成学习器中性能变差的神经网络基学习器进行替换。

2 神经网络 (NNRW) 集成学习器

飞行器遥测数据本质是一个多维时间序列, 该时间序

列为按照严格的采样间隔获取的数据集, 而时间序列预测是利用已知数据 $Y_{t-m}, Y_{t-3}, Y_{t-2}, Y_{t-1}, Y_t$, 预测得到 Y_{t+1} ^[10]。其中: Y_{t+1} 为集成神经网络预测得到的目标值; Y_t 为已知的观测值; t 为时间轴上的序列号。

2.1 集成学习器原理

基于神经网络的预测, 要求具有较高的预测精度以及计算效率, 可以满足遥测数据处理的要求。考虑到单一神经网络存在不稳定性, 而集成学习方通过联合多个同质或异质神经网络进行预测, 不仅可实现预测模型的泛化性能, 而且预测精度可高于单一最优神经网络, 神经网络集成学习器由多个同质神经网络构成, 这样可获得稳定的预测输出, 极大减少神经网络的设计参数数量, 简化了参数选择和优化过程, 提高了集成学习器的效率。

集成学习器原理如图 2 所示。

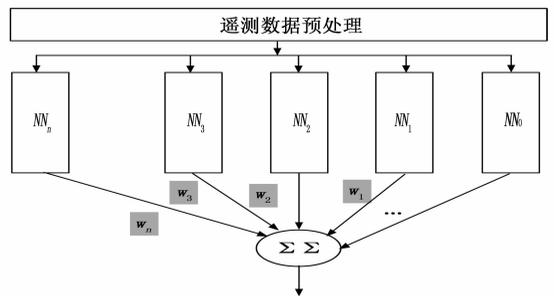


图 2 神经网络集成学习器

一个神经网络集成学习器包含一组神经网络基学习器, 其中的每个神经网络基学习器都独立对输入的遥测数据进行预测, 神经网络基学习器的结构为单隐藏层的前向反馈多层感知机, 为简化集成学习器结构和提高预测性能, 将各神经网络基学习器输出的加权和作为集成学习器的输出预测值。该方法可保证预测的精度优于其中最优化神经网络的预测精度并能在数据存在漂移及扰动时对预测结果进行平滑处理。

2.2 多层感知机神经网络

神经网络基学习器为多层感知机神经网络 (multilayer perceptron) 结构, 该结构设计具有一个输入层, 一个隐藏层和一个输出层结构。

其功能函数如式 (1) 所示:

$$\hat{y}_{t+1} = \beta_0 + \sum_{h=1}^H \beta_h g(\gamma_{0h} + \sum_{i=1}^I \gamma_{hi} p_i) \quad (1)$$

其中: I 表示神经网络输入 p_i 的数量; H 为神经网络隐藏层节点数量; $g(\cdot)$ 为非线性激活函数, 这里为双曲正切函数。

$\beta = [\beta_1, \dots, \beta_H]$ 为输出层权值; $\gamma = [\gamma_{11}, \dots, \gamma_{HI}]$ 为隐藏层权值; β_0, γ_{0h} 为单个神经元的偏置。

定义神经网络输入层为 N_I , 为子系统监测参数数量和用于预测的时间窗口之积。输出层 N_O 由子系统检测器数量确定。隐藏层神经元输出连接到输出节点, 如式 (1) 所示, 通过线性运算产生预测输出。预测值 \hat{y}_{t+1} 根据已知观测值序列运算得到。

由于基学习器中隐层神经元的数量对预测性能起关键作用,为获得隐层节点数量的最优值,以最小均方误差(MSE)为判定标准,采用搜索方法获取。定义隐藏层神经元数量的搜索间隔 [3, 31]。训练位于间隔内的 10 个具有该隐藏层数量的神经网络,评估集成神经网络的精度,之后转换到该间隔内其他值,直至数量间隔搜索完成。选择具有最小 MSE 的集成神经网络,其隐藏层数量为最优。

2.3 神经网络训练算法

由于反向传播训练算法通常收敛速度较慢,选择 Levenberg-Marquard 算法,该算法适合于性能指标为均方误差(MSE)的神经网络训练。由于该方法不需要计算正定矩阵,其正定矩阵可近似为:

$$H = J^T J \quad (2)$$

梯度为:

$$g = J^T e \quad (3)$$

其中: J 为 Jacobian 矩阵,包含网络误差对权值的一阶导, e 为网络误差向量。Jacobian 矩阵可通过标准化的反向传播算法求取的,降低了计算正定矩阵的复杂性。

如果权值和偏差更新等于 δ :

$$\delta = \Delta W^m(k) + \Delta b^m(k) \quad (4)$$

因此 Levenberg-Marquard 算法近似如下函数:

$$(J^T J + \mu I) \delta = J^T e \quad (5)$$

其中: J 为神经网络的 Jacobian 矩阵; μ 为阻尼因子; δ 为权值更新向量; e 为误差向量。

每次成功的迭代都会减小 μ 值,当迭代中性能函数增大时, μ 值增加,这样性能函数在每次迭代循环中总是递减,当性能指标达到时算法收敛。

2.4 动态加权算法

以往集成学习的加权算法是一种各基学习器输出权值为静态和固定值的方法,其权值确定方法如式(6)所示:

$$w_i = \frac{mse_i}{\sum_{i=1}^n mse_i} \quad (6)$$

其中: mse_i 为第 i 个神经网络在验证数据集上的测量均方差。

动态加权方法中每个神经网络基模型的权值根据其过去和现在的预测性能改变,该方法具有优于静态加权集成方法的性能。神经网络集成学习器输出为:

$$\hat{y}_k = w_k, 1 \hat{y}_k^{(1)} + w_k, 2 \hat{y}_k^{(2)} + \dots + w_k, n \hat{y}_k^{(n)} \quad (7)$$

$$\forall k = 1, 2, \dots, N$$

w_k, i 为第 i 个模型的在数据 k 上的权值,取决于第 i 个模型和时间索引 k 。设权值为非负且无偏。

每个基模型的过去性能由样本内验证数据集评估,目前性能由样本外预测评估,通过利用预测误差作为评估标准,对基模型输出权值进行增减。具体算法如下:

$(\bar{x}, \bar{y})_n^1$: 训练数据集

β : 缩减因子,

θ : 神经网络基学习器移除时权值对应的门限值

p : 神经网络基学习器替换周期

e, w_m : 神经网络基学习器和权值集合

$\Lambda, \lambda\{1, \dots, c\}$: 总体和局部预测;

m : 集成学习器中神经网络基学习器数量

ϵ_1, \dots, m : 预测误差

步骤:

(1) 训练获得神经网络基学习器

(2) $w_1, \dots, m \leftarrow 1/m$

(3) for $i=1, \dots, m$ do

(4) $\Lambda = 0$

(5) for $j \leftarrow 1, \dots, m$ do

(6) $\lambda \leftarrow \text{predict}(e_j, x_i)$

(7) $\epsilon_j \leftarrow \text{MAPE}(\lambda, y_i)$

(8) $\Lambda \leftarrow \Lambda + \lambda * w_j$

(9) $w \leftarrow \text{lowerweights}(w, \epsilon)$

(10) if $\text{mod } p = 0$ then

(11) for $j \leftarrow 1, \dots, m$ do

(12) if $w_j < \theta$ then

(13) $e, w \leftarrow \text{replacewithnewNN}()$

(14) 归一化 w

(15) 输出 Λ

该算法初始训练并创建一组 m 个多层感知机神经网络作为基学习器,各基学习器具有相等的权值。迭代开始获取训练样本的预测值,这些局部预测值与权值之积与全局预测值 Λ 进行累加,计算每个基学习器的预测误差,如式(8)所示,以向量形式保存:

$$\text{MAPE} = \frac{100}{n} - \sum_i^n \left| \frac{A_i - F_i}{A_i} \right| \quad (8)$$

其中: n 为样本数量, A_i 为实际值, F_i 为预测值。

算法中通过转换函数实现权值动态权值更新,如式(9)所示,对预测性能最差的基学习器权值分配缩减因子 β ,而对其它基学习器逐渐增加乘子,直到理论最大值 1。

$$\text{mult}_i = (1 - \frac{\epsilon_i}{100})^{1/\log_2(1 - \max(\epsilon)/100)} \quad (9)$$

3 实验验证及分析

遥测数据集是由飞行器上各子系统功能传感器产生输出的一组连续监测数据,表征了各子系统的特定功能和状态。各子系统的监测参数依赖于飞行器的运行状态,同时各监测参数间具有一定的相关性。针对遥测数据预测问题,利用多层感知机神经网络具有的数据预测特性,采用动态加权算法集成多个感知机神经网络做为基学习器,通过对各基学习器预测结果的加权计算,实现对各基学习器的预测结果进行一种优化平均运算,以减小计算结果的分散性,满足遥测数据预测精度要求。飞行器遥测数据有距离、速度、俯仰角以及内部各传感器采集的数据,为验证集成学习预测的性能,选用飞行器某气体单元子系统内压力参数的遥测数据作为实验样本集进行集成学习算法的验证,并对集成学习模型的性能进行评估。

3.1 数据处理

由于多层感知机神经网络为有监督学习,为了保证训

训练的神经网络预测性能最优, 需要保证一定规模的数据进行训练, 首先需要产生具有输入向量, 输出向量的训练数据集。为此采用滑动窗口技术获取训练数据集, 具体过程为从遥测数据的时间序列中抽取对应窗口的片段作为输入向量, 输出值的期望值为对应该窗口的下一顺序观测值。通过在遥测数据上逐个数据滑动窗口, 随着时间从一个位置变换到下一位置, 串行连接每个片段构成训练数据集。

将获取的数据样本集划分 70% 为训练集, 15% 为验证集, 15% 为测试集三部分, 训练集和验证集用于训练, 估计和搜索最优神经网络结构, 最后, 使用测试集评估所获得神经网络的预测性能。

多层感知机神经网络结构为包含两个输入的一个输入层, 具有 15 个神经元节点的隐藏层和一个输出层。

3.2 动态加权算法评估

遥测数据集首先需要进行一定的预处理, 包括去除噪声, 对数据进行归一化处理。在集成学习预测方法中, 神经网络基学习器的性能对最终的预测性能起着决定性作用, 因此要求对神经网络基学习器进行最优化, 最优化参数包括隐层节点数量、隐层输出权值。动态加权集成学习预测算法实际的预测效果如图 3 所示。

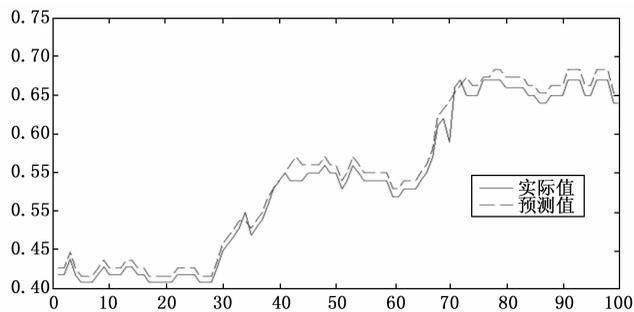


图 3 实际数据和预测

为实现动态加权集成算法的性能评估, 将其与静态加权算法进行比较, 两种算法具有完全相同的网络结构参数和基学习器数量, 静态加权算法只对各基学习器预测结果进行简单的平均运算, 将两种算法输出结果的 MAPE 指标作为算法性能的判断标准。其性能比较结果如图 4 所示。

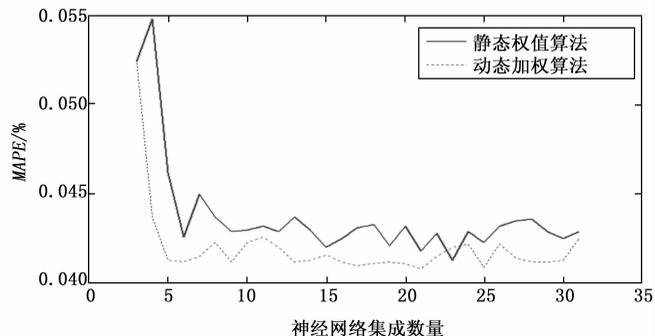


图 4 动态和静态加权算法性能比较

为 5~9 个, 包含更多数量的神经网络基学习器对集成学习预测精度提升效果有限, 而且增加了计算复杂性。同时 MAPE 指标也表明推荐的动态加权方法同静态加权方法相比具有更小的预测误差。

4 结束语

针对飞行器遥测数据预测中要求具有实时性和高精度的需求, 提出基于多层感知机神经网络的动态加权集成学习方法, 该方法通过集成多个多层感知机神经网络作为预测基学习模型降低了预测模型的复杂度, 提高了预测模型的实时性, 实现了更加精确的预测, 给出了多层感知机神经网络的设计、优化方法和动态加权算法的过程。对于遥测数据存在异常漂移情况, 通过给出的动态加权算法, 能识别优秀的神经网络基学习器, 增加其输出权值, 对性能差的神经网络基学习器减弱其对预测输出的影响, 甚至对其进行更换, 从而提升预测结果的精度。

采用飞行器某子系统内的遥测数据对集成学习方法性能进行了验证, 确定了感知神经网络隐藏层的最优节点数和神经网络基学习器数量的最优值, 通过与静态加权集成学习方法的比较, 结果表明提出的动态加权集成方法预测精度更高。该集成学习预测方法也取决于神经网络基学习器的参数选取, 通过对隐藏层节点数量和学习速度进行最优化设计, 可以保证最终的集成结果也是预测最优值。

参考文献:

- [1] 房红征, 马好东, 罗 凯, 等. 基于遥测数据的航天器长期性能预示方法研究 [J]. 计算机测量与控制, 2013, 21 (7): 71-75.
- [2] 何志勇, 张志峰, 王 珺. AR 模型在火箭发动机遥测数据中的应用 [J]. 固体火箭技术, 2019, 42 (5): 673-678.
- [3] 朱丽莎, 姜 斌, 程月华, 等. 一种卫星遥测参数动态 ARMA 建模方法 [J]. 航天控制, 2017, 35 (2): 37-43, 50.
- [4] 陈 冰, 鲁 刚, 房红征, 等. 基于最小二乘支持向量机的卫星异常检测方法 [J]. 计算机测量与控制, 2014, 22 (3): 690-692.
- [5] 闫谦时, 崔广立. 基于时间序列的航天器遥测数据预测算法 [J]. 计算机测量与控制, 2017, 25 (5): 188-191.
- [6] 任国恒, 李 纲, 贾超超. 一种基于 RBF 神经网络的遥测数据中长期预测方法 [J]. 鄂州大学学报, 2014 (7): 94-96.
- [7] DING J, WANG H, LI C, et al. An online learning neural network ensemble with random weights for regression of sequential data stream [J]. Soft Comput., 2017, 21 (20): 5919-5937.
- [8] 乔俊飞, 李凡军, 杨翠丽. 随机神经网络研究现状与展望 [J]. 智能系统学报, 2016, 11 (6): 758-767.
- [9] 张天伦. 多隐层前馈神经网络的随机赋权训练算法研究 [D]. 保定: 河北大学, 2016.
- [10] 李善营, 高占宝, 刘紫阳, 等. 基于遥测数据的无人机气动参数辨识方法研究 [J]. 飞行力学, 2017 (3): 32-36.
- [11] 李 敏, 李 华, 程茂华. 一种自适应子融合集成多分类器方法 [J]. 计算机测量与控制, 2019, 27 (8): 120-124.
- [12] 成红艳, 舒传华, 崔俊峰. 通过数据处理提高遥测精度的方法 [J]. 无线电工程, 2015 (8): 10-14.

结果表明集成学习器中合理的神经网络基学习器数量