

文章编号: 1005-5630(2023)01-0032-06

DOI: 10.3969/j.issn.1005-5630.2023.001.005

XGBoost 机器学习在光电编码器误差 补偿中的应用

李映铮, 李志斌, 金磊, 胡珍珍, 康晔斐, 李庚白
(上海电力大学自动化系, 上海 200082)

摘要: 光电编码器检测系统的误差主要受基准光电编码器测角误差、数据采集误差、检测系统同轴误差影响。其中, 基准光电编码器的测角误差可进行补偿。因此设计了一种基于极度梯度提升树(extreme gradient boosting, XGBoost)机器学习的算法用来补偿基准光电编码器的误差。经该算法补偿后, 静态精度提高了35倍, 标准差由3.62"减小至0.13", 最大误差值由5.53"降低至0.39"。与传统的误差反传(back propagation, BP)神经网络算法以及径向基函数(radial basis function, RBF)神经网络算法补偿效果相比, XGBoost的补偿效果更优。XGBoost机器学习算法有效降低了基准光电编码器的测量误差, 提高了光电编码器检测系统的检测精度。

关键词: 光电编码器; 误差补偿; XGBoost; 检测精度
中图分类号: TP 274 **文献标志码:** A

Application of XGBoost machine learning in error compensation of photoelectric encoder

LI Yingzheng, LI Zhibin, JIN Lei, HU Zhenzhen, KANG Yefei, LI Gengbai
(Department of Automation, Shanghai University of Electric Power, Shanghai 200082, China)

Abstract: The error of photoelectric encoder detection system is mainly affected by the angle measurement error of the reference photoelectric encoder, data acquisition error and coaxial error. The angle measurement error can be compensated. In this paper, an algorithm based on extreme gradient boosting (XGBoost) machine learning is designed to compensate the error of the reference photoelectric encoder. After compensation, the static accuracy is improved by 35 times. The standard deviation is decreased from 3.62" to 0.13", and the maximum error value is reduced from 5.53" to 0.39". Compared with the traditional back propagation (BP) neural network algorithm and radial basis function (RBF) neural network algorithm, XGBoost's compensation is better than the others. XGBoost machine learning algorithm compensation effectively reduces the measurement

收稿日期: 2022-07-05

基金项目: 上海市电站自动化技术重点实验室项目(13DZ2273800)

第一作者: 李映铮(1998—), 男, 硕士研究生, 研究方向为编码器检测。E-mail: 278430591@qq.com

通信作者: 李志斌(1974—), 男, 博士, 教授, 研究方向为测控技术及自动化装置。

E-mail: thermal_li@163.com

error of the reference photoelectric encoder and improves the detection accuracy of the photoelectric encoder detection system.

Keywords: photoelectric encoder; error compensation; XGBoost; accuracy of detection

引 言

光电编码器是一种集光、机、电于一体的高分辨率角位移传感器,它以光电转换技术为核心,输出与角位置信息相对应的角度代码来实现角位移的测量^[1-2],广泛应用于航空航天、工业控制等领域^[3-4]。出厂前的光电编码器都需要经过性能指标检测,其性能检测准确性则由编码器检测系统的检测精度决定。该检测精度主要又由检测系统中基准光电编码器的精度决定。随着科学技术的进步,许多提高光电编码器精度的方法应运而生^[5]。Hou等^[6]研究了光电编码器的非线性误差,提出一种基于谐波分析的方法对误差进行补偿。陈洪月等^[7]用粒子群优化算法对误差反传(back propagation, BP)神经网络进行优化,将优化后的网络用于误差训练。洪喜等^[8]采用径向基函数(radial basis function, RBF)神经网络提高基准光电编码器的精度,将神经网络的训练值用于补偿。董凯炎等^[9]利用傅里叶技术拟合测角误差,其拟合结果可对基准光电编码器的误差进行补偿。利用BP神经网络算法进行非线性误差拟合,容易出现局部极小值且存在过拟合的现象,会影响网络训练效果。利用RBF神经网络算法的非线性自适应滤波器对输出信号进行滤波,如果拟合中心点为样本中的数据,就不能反映正式样本的状况。利用极度梯度提升树(extreme gradient boosting, XGBoost)算法进行误差修正,其正则化项可防止过拟合,且当训练数据较少或者存在缺失时,损失函数能够进行优化,处理缺失或者稀疏的数据,大大提升算法的效率。实验结果显示, XGBoost能够高精度逼近非线性误差曲线,具有较强的非线性拟合性,且修正光电编码器误差的效果优于神经网络。

1 光电编码器检测系统的误差构成

传统的编码器对比检测法系统框图如图 1 所

示。该系统主要由基准光电编码器、待测光电编码器、电机、联轴节、支架、电机驱动器以及上位机构成,其误差来源主要有:基准光电编码器的既定误差、装调误差以及数据采集卡带来的误差^[10-11]。其中基准光电编码器带来的误差要远大于装调误差以及数据采集卡带来的误差^[12]。故考虑从基准光电编码器的误差出发,对其进行补偿,进而提高整个编码器检测系统的精度^[13]。

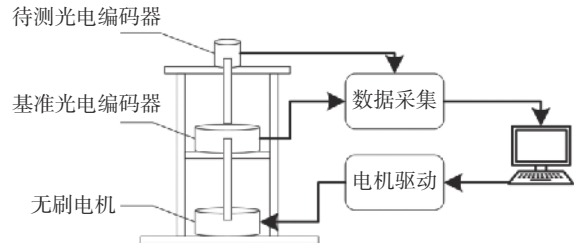


图 1 编码器精度检测装置

Fig. 1 Precision measurement system for photoelectric encoder

用静态精度 ε 、标准差 σ 以及误差最大值 e_{\max} 来衡量基准光电编码器的精度。编码器的静态精度用于衡量静态或低速运行情况下,实际值和设定值之间重复的平均偏差量值。标准差反映了数据值偏离算术平均值的程度,用以衡量误差的离散程度及仪器的稳定性。最大误差可以反映仪器的计量性能。静态精度的计算公式为

$$\varepsilon = \frac{\sum_{i=1}^n |e_i|}{n}, \quad i = 1, 2, 3, \dots, n \quad (1)$$

式中: e_i 表示编码器检测第 i 次的误差; n 表示传感器的测量点点数。

标准差的计算公式为

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (e_i - \bar{e})^2}{n-1}}, \quad i = 1, 2, 3, \dots, n \quad (2)$$

式中, \bar{e} 表示编码器的误差平均值。

误差最大值的计算公式为

$$e_{\max} = \max |\theta_{i,T} - \theta_{i,L}|, \quad i = 1, 2, 3, \dots, n \quad (3)$$

式中： $\theta_{i,T}$ 表示第 i 个点的角度基准值， T 为基准编码器； $\theta_{i,L}$ 表示第 i 个点的被测角度值， L 为待测编码器。

2 XGBoost 基本原理以及补偿方法

XGBoost 算法是在梯度提升决策树 (gradient boosting decision tree, GBDT) 的基础上提出的机器学习系统，具有防过拟合性好、计算效率高、样本适应度高等特点。其学习器包括分类树以及线性分类器。本文利用分类树作为学习器。其核心为根据样本特征生长出新的叶子节点，分裂生长出新的树。每次添加一棵树，就学习一个新函数，去拟合上次预测的残差^[14]。如训练完成的系统由 K 棵树组成，当预测某个样本的值时，就根据这个样本的特征，在每棵树中落到对应的叶子节点处。每个叶子节点有唯一的权重值，即分数对应。一棵树的得分用 F 表示，将样本值在每棵树的得分加起来就是该样本的预测值 \hat{y} ，其表达式为：

$$\hat{y} = \Phi(x_i) = \sum_{k=1}^K f_k(x_i) \quad (4)$$

$$F = \{f(x) = \omega_{q(x)}\} (q: \mathbf{R}^m \rightarrow T, \omega \in \mathbf{R}^T) \quad (5)$$

式中： $f_k(x_i)$ 表示第 K 棵树第 i 个叶子节点的权重值，即所得分数； $\omega_{q(x)}$ 为叶子节点 q 的分数； $f(x)$ 为其中一棵回归树；一棵完整的决策树其叶子节点数为 T ； $\omega \in \mathbf{R}^T$ 表示该棵决策树是由所有叶子节点组成的向量； $q: \mathbf{R}^m \rightarrow T$ 中 q 表示结构函数， \mathbf{R}^m 表示输入叶子节点分数向量，整体表示将叶子节点分数映射到叶子节点索引。

模型的偏差和方差决定模型的精度。模型偏差由损失函数表示，正则项则用于防止过拟合现象。目标函数 Obj 可以由损失函数以及正则项的和组成，其表达式为

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (6)$$

式中：第一部分表示回归树预测分数和真实分数的差值即训练误差；第二部分为正则化项，即每棵树的复杂度之和，目的是控制模型复杂度，防止过拟合。

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \omega^2 \quad (7)$$

正则化项由两项组成，式中第一项 T 表示叶子节点数目； ω 表示叶子节点的分数； γ 和 λ 是两个可以调节的系数，其目的为防止叶子节点的分数过大产生过拟合。

在构建每一棵树模型时，采用贪心算法来决定叶子节点的分裂以及树的最大深度，确保每棵树的增益值最大。在叶子节点分裂过程中，叶子节点的权值用 $w(g_i, h_i)$ ， g_i 和 h_i 的表达式为

$$g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)}) \quad (8)$$

$$h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)}) \quad (9)$$

式中： y_i 表示目标值； \hat{y}_i 表示前面 $t-1$ 轮的预测值。为使分裂后增益最大，利用增益 $Gain$ 来衡量分裂后与分裂前的增益差。分裂后的增益为

$$Gain = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad (10)$$

式中： $\frac{G_L^2}{H_L + \lambda}$ 代表分裂后左子树的分数； $\frac{G_R^2}{H_R + \lambda}$ 代表分裂后右子树的分数； $\frac{(G_L + G_R)^2}{H_L + H_R + \lambda}$ 代表不分裂可得分数； γ 代表加入新叶子节点引入的复杂度代价。若 $Gain$ 大于 0，则引入新的节点。

XGBoost 在原有的决策树上进行优化，使用梯度提升算法来确保较小的损失值，若构成一棵新的树模型损失值要大于原有的树模型，则停止生成新的树。XGBoost 每进行一次迭代就是增加一棵树， K 棵树的线性组合可以表示为

$$\hat{y}_i^{(t)} = \sum_{k=1}^K f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i), \quad f_k \in F \quad (11)$$

式中： F 表示所有树的模型； $f_k(x_i)$ 表示第 i 个样本在第 K 棵树中被划分至当前叶子节点的权重值； $\hat{y}_i^{(t-1)}$ 表示前面 $t-1$ 轮的模型预测值； $f_t(x_i)$ 表示新加入的预测函数的值。

利用 XGBoost 修正基准光电编码器误差，先随机选取一定量的样本数据作为训练集进行训练，其训练结果决定了误差补偿的效果。完成训练后，再将剩余的样本数据输入训练分类树，将其预测值与真实值进行比较，验证 XGBoost 机器学习补偿系统的泛化推广能力^[15]。

3 基准编码器误差补偿实验

3.1 编码器检测系统装置

为获得 XGBoost 误差补偿的训练样本, 采用自准直仪和多面棱体对基准光电编码器进行误差检测^[16]。基准光电编码器选型为 SIMOTICS S-1FL6 伺服电机所自带的 21 位增量式编码器, 其出厂静态精度 $\sigma \leq 5$ 。光电自准直仪装置的测角精度可达到 $0.25''$, 其精度远高于编码器的精度。系统结构包括: 光电自准直仪装置、金属多面棱体、伺服电机、联轴节、电机驱动结构、支架以及上位机。其结构简图如图 2 所示。

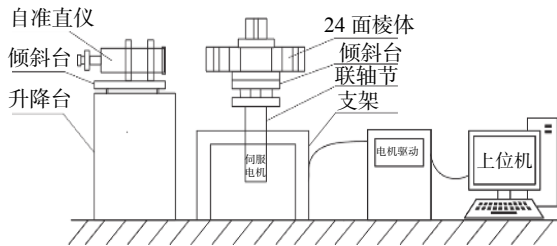


图 2 基准光电编码器误差检测系统结构

Fig. 2 Structure of error detection system for reference photoelectric encoder

检测方法为: 首先将多面棱体与基准光电编码器同轴连接, 调节初始位置, 然后进行测量^[17]。多面体每转过一面, 自准直仪对角度进行一次标定。为了提高样本数据的值使其更接近真实值, 用金属 24 面棱体进行 3 次测量, 共得 72 个数据集, 取 3 次测量误差的平均值, 检测平均误差结果见表 1。

3.2 基于 XGBoost 误差补偿建模及数据处理

在 MATLAB 上建立模型并验证。由于数据集的个数较少, 将训练数据集与验证数据集的比例划分为 7:3, 并将数据集打乱。经过多次训练调试发现, 将输入数值映射到 0~100 之间, 学习率设置为 0.15, 树的最大深度设置为 6, 训练效果最好。其训练能力如图 3 所示。

可以看出, 经过 XGBoost 学习的曲线高度逼近编码器的实际检测误差。

使用学习样本数据外的数据作为补偿系统的测试数据, 用以验证测试样本的泛化能力, 其泛

表 1 编码器误差检测结果

Tab. 1 Encoder error detection results

角度/(°)	转角误差均值/(")	角度/(°)	转角误差均值/(")
0	0	180	-4.44
15	-4.81	195	-3.70
30	-3.02	210	-4.94
45	5.28	225	-4.12
60	4.11	240	2.19
75	-3.99	255	-0.12
90	1.31	270	3.09
105	4.59	285	4.25
120	4.93	300	3.11
135	3.30	315	1.27
150	-0.07	330	5.53
165	-2.16	345	-1.20

化能力如图 4 所示。

由图 4 可以看出, 经过学习的分类树网络能够较好地拟合误差曲线, 其泛化能力很好。补偿前后误差对比如图 5 所示。

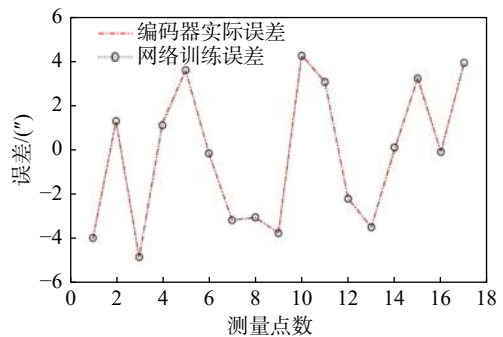


图 3 XGBoost 机器学习训练能力

Fig. 3 XGBoost machine learning training ability

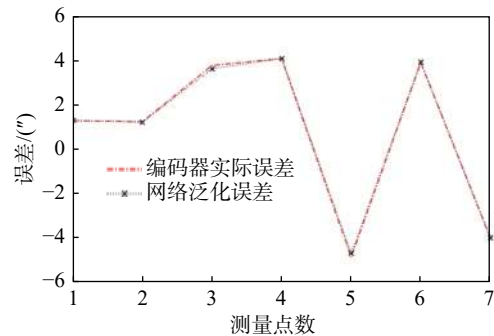


图 4 XGBoost 机器学习泛化能力

Fig. 4 XGBoost machine learning generalization ability

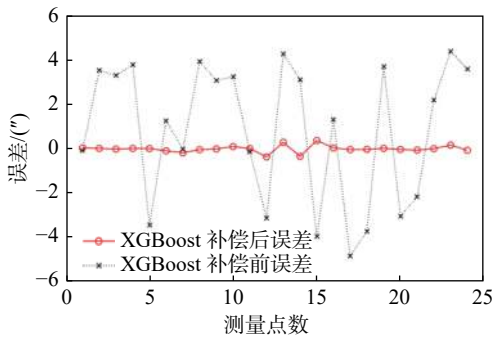


图 5 补偿前后编码器误差对比

Fig. 5 Comparison of encoder error before and after compensation

可以看出, XGBoost 具有较好的非线性拟合能力, 使用补偿系统后, 编码器的误差在 0 附近波动, 其补偿效果很好。

为了验证 XGBoost 训练补偿方法的有效性, 利用 BP 神经网络与 RBF 神经网络进行误差补偿。BP 神经网络与 RBF 神经网络的拓扑结构均设置为 $1 \times 6 \times 1$, 两种网络补偿前后的误差图像如图 6、图 7 所示。

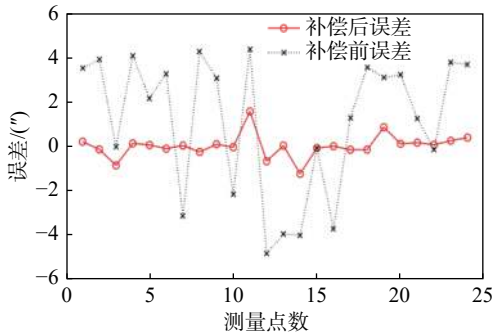


图 6 BP 神经网络补偿对比图

Fig. 6 Comparison diagram of BP neural network compensation

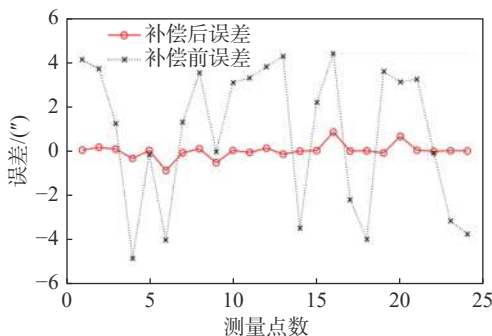


图 7 RBF 神经网络补偿对比图

Fig. 7 Comparison diagram of RBF neural network compensation

可以看到, 两种神经网络的误差补偿效果不如 XGBoost 机器学习的补偿效果。其误差补偿效果的对比结果如表 2 所示。

表 2 各补偿系统补偿效果对比
Tab. 2 Comparison of compensation effect of each compensation system

系统	静态精度/ (")	标准差/ (")	最大误差值/ (")
补偿前	3.15	3.62	5.53
BP神经网络补偿系统	0.51	0.77	1.59
RBF神经网络补偿系统	0.19	0.31	0.87
XGBoost补偿系统	0.09	0.13	0.39

分析表中数据可知, 3 种补偿系统均使得静态精度、标准差、最大误差值有较大的改善, 其中 XGBoost 的补偿效果最好。经过 XGBoost 补偿系统补偿优化, 静态精度由 $3.15''$ 降低至 $0.09''$, 标准差由 $3.62''$ 降低至 $0.13''$, 最大误差值由 $5.53''$ 降低至 $0.39''$ 。

4 结 论

利用 XGBoost 算法设计了一个误差补偿系统。首先, 对编码器检测系统的误差来源进行分析, 确立了误差补偿效果评价的标准。使用 24 面棱体对基准光电编码器进行精度标定, 将得到的数据集在 MATLAB 上使用 XGBoost 机器学习系统进行补偿。补偿后的结果表明, 该方法具有较好的误差补偿效果, 理论上能够进一步提高光电编码器检测系统的精度。在实际建立 XGBoost 机器学习系统模型时, 通常需要根据样本数据进行不同的参数调节以及算法优化, 以获得最好的补偿效果。将 XGBoost 用于对光电编码器误差的补偿, 补偿过后的光电编码器具有更高的精度, 可将此误差补偿方法用于其他测量系统当中。

参考文献:

- [1] 汤天瑾, 曹向群, 林斌. 光电轴角编码器发展现状分析及展望 [J]. 光学仪器, 2005, 27(1): 90 - 96.
- [2] INCZE I I, SZABÓ C, IMECS M. Incremental encoder

- in electrical drives: modeling and simulation[M]// RUDAS I J, FODOR J, KACPRZYK J. Computational Intelligence and Informatics: Principles and Practice. Berlin, Heidelberg: Springer, 2010: 287 – 300.
- [3] 吴文峰, 李志斌, 杨勇, 等. 基于 LabVIEW 的光栅编码器动态特性检测系统 [J]. 微特电机, 2019, 47(3): 25 – 28.
- [4] LU L, WANG L, WANG W, et al. A self-calibration method for error of photoelectric encoder based on gyro in rotational inertial navigation system[J]. *Microsystem Technologies*, 2019, 25(6): 2145 – 2152.
- [5] BHANGU B S, RAJASHEKARA K. Control strategy for electric starter generators embedded in gas turbine engine for aerospace applications[C]//Energy Conversion Congress & Exposition. Phoenix: IEEE, 2011: 1461 – 1467.
- [6] HOU B, ZHOU B, LI X, et al. Periodic nonlinear error analysis and compensation of a single-excited petal-shaped capacitive encoder to achieve high-accuracy measurement[J]. *Sensors*, 2019, 19(10): 2412.
- [7] 陈洪月, 张坤, 刘治翔, 等. 基于 PSO-BP 神经网络光电编码器误差补偿研究 [J]. 传感技术学报, 2017, 30(8): 1182 – 1186.
- [8] 洪喜, 续志军, 杨宁, 等. RBF 网络在编码器误差补偿中的应用 [J]. 光电工程, 2009, 36(8): 139 – 142.
- [9] 董凯炎, 史姝倩, 刘杰, 等. 一种光电编码器精度自动检测系统 [J]. 传感技术学报, 2019, 32(3): 463 – 468.
- [10] 刘丰文, 邓文和. 25 位绝对式编码器 [J]. 光电工程, 2000, 27(6): 66 – 68.
- [11] JIA H K, YU L D, ZHAO H N, et al. A new method of angle measurement error analysis of rotary encoders[J]. *Applied Sciences*, 2019, 9(16): 3415.
- [12] YU H. Angle measurement based on in-line digital holographic reconstruction[J]. *Optics and Lasers in Engineering*, 2021, 137: 106385.
- [13] 王义文, 焦环宇, 刘献礼, 等. 绝对式编码器输出信号的误差自动补偿 [J]. 光电工程, 2015, 42(1): 89 – 94.
- [14] 李占山, 刘兆赓. 基于 XGBoost 的特征选择算法 [J]. *通信学报*, 2019, 40(10): 101 – 108.
- [15] 李叶紫, 王振友, 周怡璐, 等. 基于贝叶斯最优化的 Xgboost 算法的改进及应用 [J]. 广东工业大学学报, 2018, 35(1): 23 – 28.
- [16] 杨金利, 魏晓蓉. 绝对式光电编码器测角误差自动检测系统 [J]. 光学与光电技术, 2021, 19(5): 30 – 35.
- [17] 董静, 万秋华, 于海, 等. 小型绝对式光电编码器误码自动检测系统 [J]. 中国光学, 2016, 9(6): 695 – 703.

(编辑: 李晓莉)