

PAPER • OPEN ACCESS

Application of XGboost Algorithm in Bearing Fault Diagnosis

To cite this article: Rongtao Zhang *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **490** 072062

View the [article online](#) for updates and enhancements.

Application of XGboost Algorithm in Bearing Fault Diagnosis

Rongtao Zhang¹, Binbin Li^{1,*} and Bin Jiao¹

¹School of Electrical Engineering, Shanghai Dianji University, Shanghai, China

*Corresponding author e-mail: libinbin2000@126.com

Abstract. This paper applies the XGboost(eXtreme Gradient Boosting) algorithm to the fault diagnosis of rolling bearing. XGboost is the realization of GBDT(gradient boosting decision tree). Generally speaking, the realization of GBDT(gradient boosting decision tree) is slow. XGBoost is characterized by fast computation and good performance of the model. At the end of this paper, we compare with other tree algorithms, and the results show that the XGboost algorithm is superior to other algorithms in accuracy and time.

1. Introduction

According to incomplete statistics, rolling bearing as one of the easily damaged parts in rotating machinery, caused accounts for about 30% of the faults of rotating machinery. Generally speaking, the bearing is mainly composed of cage, ball, inner ring and outer ring. The most widely method of detection and diagnosis is to collect the vibration data of the bearing through vibration sensor. Then the feature extraction is carried out for vibration data[14], and the extracted time-domain parameters[13] can easily distinguish the fault type, especially kurtosis[1-2]. When the bearing is in normal operation, the probability density distribution of vibration signal is close to the normal distribution under the influence of the complexity of the equipment and the variability of the environment in where the equipment is located. When fault occurs, The probability density value of the vibration point increases, the normal curve begins to deform, and the kurtosis value changes accordingly. The more serious the deformation, the greater the kurtosis value.

In this paper, we introduce the XGboost algorithm. It is an implementation of Gradient Boosting Machine by Chen Tianqi, who is studying machine learning at the University of Washington. In his research, he was deeply constrained by the computational speed and accuracy of the existing library, so he constructed the XGboost project. XGboost have many advantages: customizable loss function; canonical items; tree building and pruning; support for splitting point approximate search; sparse features. Processing; missing value processing; feature importance and feature selection; parallel computing; memory caching. In this paper, we apply the XGboost(eXtreme Gradient Boosting) algorithm to the fault diagnosis of rolling bearing. we compare XGboost with other tree algorithms, The results show that the XGboost algorithm is superior to other algorithms in accuracy and time.

2. XGboost Algorithm

As a non-parametric model for supervised learning, the selection of XGboost[4-6] parameters depends on the training data used in the model. The biggest difference with GBDT[3] is that the loss function only uses the first derivative of the loss function when calculating the objective function. XGBoost approximates the loss function with the second-order Taylor expansion:

$$f(x + \Delta x) \cong f(x) + f'(x)\Delta x + \frac{1}{2} f''(x)\Delta x^2 \quad (1)$$



The definition of a weak learner tree in XGboost is also different: splitting the tree into structural parts q and leaf fraction w .

$$\phi(x) = w_{q(x)}, w \in R^T, q: R^D \rightarrow \{1, \dots, T\} \quad (2)$$

Among them, the structure function q : The index number that maps the input to the leaf; w : it is given that the fraction of leaves corresponding to each index number; T is the number of leaf nodes in the tree; D is the characteristic dimension.

The complexity of the tree is defined as:

$$\Omega(\phi(x)) = \gamma T + \frac{1}{2} \lambda \sum_{t=1}^T w_t^2 \quad (3)$$

Among them: γ is the L1 regular coefficient; λ is the L2 regular coefficient. From the upper expression, γ And λ determine the complexity of the weak learning tree.

Therefore, the objective function is:

$$\begin{aligned} J(\theta) &= \sum_{i=1}^N L(f(x_i; \theta), y_i) + \Omega(\theta) \\ &\cong \sum_{i=1}^N [g_{m,i} \phi(x_i) + \frac{1}{2} h_{m,i} \phi(x_i)^2] + \gamma T + \frac{1}{2} \lambda \sum_{t=1}^T w_t^2 \\ &= \sum_{t=1}^T [\sum_{i \in I_t} g_{m,i} w_t + \frac{1}{2} \sum_{i \in I_t} h_{m,i} w_t^2 + \frac{1}{2} \lambda \sum_{t=1}^T w_t^2] + \gamma T \\ &= \sum_{t=1}^T [\sum_{i \in I_t} g_{m,i} w_t + \frac{1}{2} (\sum_{i \in I_t} h_{m,i} + \lambda) w_t^2] + \gamma T \end{aligned} \quad (4)$$

Among them: $g_{m,i} = [\frac{\partial^2 L(f(x_i), y_i)}{\partial^2 f(x_i)}]_{f=f_{m-1}}$ represents the The first derivative of the loss function;

$h_{m,i} = [\frac{\partial^2 L(f(x_i), y_i)}{\partial^2 f(x_i)}]_{f=f_{m-1}}$ represents the The second derivative of the loss function.

Assuming the structure of the known tree q , derivation of objective function and the derivative result is equal to 0:

$$w_t^* = - \frac{\sum_{i \in I_{T-t}} g_{m,i}}{\sum_{i \in I_t} h_{m,i} + \lambda} \quad (5)$$

Set $w_t = w_t^*$:

$$J(\theta) = - \frac{1}{2} \sum_{t=1}^T \frac{(\sum_{i \in I_{T-t}} g_{m,i})^2}{\sum_{i \in I_t} h_{m,i} + \lambda} + \lambda T \quad (6)$$

From the upper expression, the w_t smaller, the smaller the objective function is and the higher the classification accuracy is, the better the robustness of the model is due to the existence of regular items.

3. Experimental Process and Results

3.1. Evaluation criteria

Logloss function

$$\log loss = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{i,j} \log(p_{i,j}) \quad (7)$$

Among them: N is the number of samples; M is the number of categories.

3.2. Experimental process and results

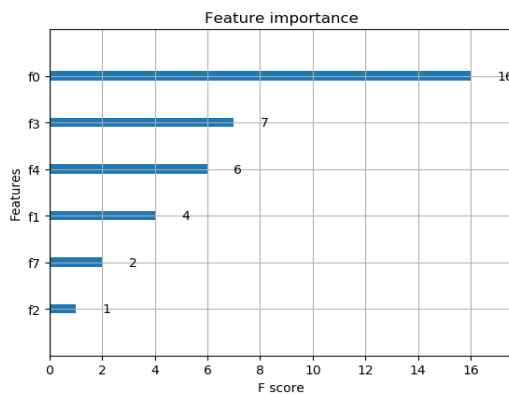
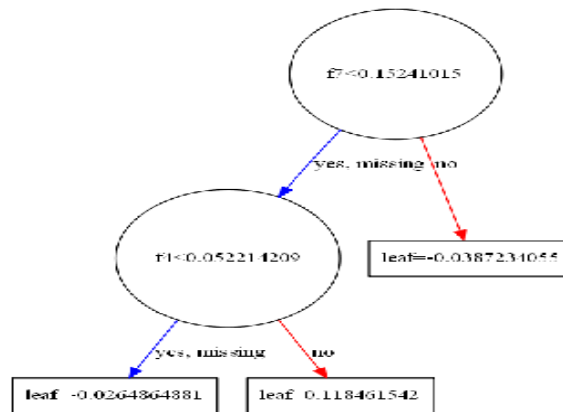
This paper's vibration data from the experimental data of rolling bearings from Western Reserve University. Selection of bearing data with fault diameter of 0.07inch as experimental data. After the bearing vibration data with 12kHz frequency is de-noised[7] by wavelet transform, the normal and original data of each fault are divided into 60 groups, each group have 2048 vibration points, extracting the time domain parameters of each group: maximum, minimum, average, etc. absolute mean, standard deviation, peak value, variance, kurtosis, root mean square, waveform factor, peak factor, kurtosis factor, pulse factor, margin. mark them as f_0, f_1, \dots, f_{13} . Corresponding bearing normal, ball fault, inner ring fault, outer ring fault label $L=[0,1,2,3]$. Respectively a piece of parameters data as a sample, a total of 240 samples. Table 1 is the partial time domain parameters and corresponding fault types.

Table 1. Partial time domain parameters and corresponding fault types

maximum	peak value	variance	kurtosis	Root square	mean Type
0.19818	0.44748	0.0055019	2.7639	0.074897	0
0.1792	0.3947	0.0048522	2.8519	0.070348	0
0.51216	1.0701	0.018042	3.1982	0.13503	1
0.36727	0.72462	0.015632	2.7614	0.12582	1
1.3212	2.4762	0.088443	5.6928	0.29761	2
1.3495	2.5597	0.086008	5.6411	0.29353	2
3.1322	6.2286	0.50369	7.5291	0.71015	3
3.2093	6.242	0.47494	7.4515	0.68965	3

All 240 samples are taken as input data of XGboost model. Because XGboost is a non-parametric model and the optimal parameters of the model vary with the input data, the Bayesian optimization method [10] is adopted to optimize the parameters of the XGboost model. After getting the optimal parameters, input the sample data to get the feature importance diagram as shown in Figure1. Figure 2 is a weak classifier (CART)[8-9] with kurtosis as its root node. Reference Table 1, When $f_7 > 0.1524$, the right leaf represents the outer ring of bearing have trouble. The right leaf weight is very small, indicating that the loss function is small and the classification accuracy is high, which is explained in Section 3.

Table 2 is a comparison of 4 species tree model. For the same sample data, when the leaf weight is the same as the tree depth, the number of weak learning devices, the training time and the accuracy rate which make the model reach the optimal cross-validation score is shown at Table 3. The most intuitive progress of XGboost is that the training time is much smaller, although the training time of adaboost [11] is also very small, but the number of weak learning devices is many. As we all know, the tree model is easy to over fit, the more the number of weak learning devices, This means that it is easier to over fit.

**Figure 1.** feature importance.**Figure 2.** CART tree.**Table 2.** Tree model comparison

	Random Forest adaboost	GBDT	XGboost
Number of weak learning devices	850	1308	730
Training time(s)	0.693329	0.002037	0.794217
Training set accuracy	1	0.9875	0.98932
Test set accuracy	0.9725	0.9523	0.99342

Table 3. Cross validation logloss comparison

Fault diameter (in)	load	Random Forest		Adaboost		GBDT		XGboost	
		time(s)	logloss	time(s)	logloss	time(s)	logloss	time(s)	logloss
0.007 (SKF)	0	8.063	0.00179	16.629	0.44476	8.103	0.00053	1.506	0.02857
	1	8.101	0.00178	16.715	0.34659	8.062	0.01173	1.488	0.01878
	2	8.188	0.00160	16.660	0.34659	8.202	0.00614	1.601	0.01553
	3	8.180	0.00190	16.643	0.34659	8.220	0.01237	1.593	0.02138
0.028 (NTN)	0	7.768	0.00111	15.780	0	6.100	0.01018	0.978	0.01222
	1	7.880	0.00109	15.928	0	6.124	0.01018	1.011	0.01151
	2	7.849	0.00074	15.850	0	5.986	0.00042	1.010	0.00909
	3	7.996	0.00506	15.751	0.03265	6.076	0.01995	0.981	0.02354

The sample data are extracted by the same method as above. After adjusting the model parameters, the logloss value of the model is calculated by 10 discount cross validation, in other words, all the sample data sets are randomly divided into 10 parts. First, one part is taken as test data, the other 9 parts are used as training data, the training model is trained with training data, and then the corresponding logloss value is obtained by test data, it is repeated 10 times in the same way. The final test logloss value is the average of the logloss value obtained from 10 tests. Random Forest have a low logloss value, but Table 2 shows that this algorithm is easily overfitted in a noisy environment. Adaboost have 0 logloss value, show that complete fitting on training set, but not necessarily on the test set have the same effect. XGboost compared with GBDT algorithm, not only has a higher logloss value, but also consumes less time. Because the control of model complexity and the pruning of the later stage are added XGboost to make the trained model more difficult to fit. In the complex industrial environment, the bearing data is complex and changeable, so it is necessary to extract a variety of

special features to make effective diagnosis. So XGboost have a higher logloss value compared with GBDT on the training set. Table 2 shows XGboost have a good effect on the test set.

4. Summary and Prospect

This paper compared XGboost with other tree models, it shows that XGboost algorithm has great research value in the field of bearing state monitoring and fault diagnosis.

There are still many unavoidable obstacles to the application of time-domain parameters in fault identification. Adjustment of Model parameters is important. Noise is unavoidable in practical work. Different noise in different environments leads to different time-domain parameters. Finding the parameters with strong robustness to noise, mastering or discovering new techniques for signal denoising and effective feature extraction are the most important research tasks at present.

References

- [1] Dong-DongXiao,Peng-FeiYan,Yu-Xuan Wang,Mohamed Saied Osman,Hong-Yang Zhao. Glioblastoma and primary central nervous system lymphoma: Preoperative differentiation by using MRI-based 3D texture analysis[J]. Clinical Neurology and Neurosurgery,
- [2] 2018,173.
- [3] Antoine Vallatos,Haitham F.I. Al-Mubarak,James M. Mullin,William M. Holmes. Accuracy of phase-contrast velocimetry in systems with skewed intravoxel velocity distributions[J]. Journal of Magnetic Resonance,2018.
- [4] Yanxu Liu. Gradient Boost Decision Tree Fingerprint Algorithm for Wi-Fi Localization[A].
- [5] WEIZENG WANG. Electricity Consumption Prediction Using XGBoost Based on Discrete Wavelet Transform[A]. Wuhan Zhicheng Times Cultural Development Co., Ltd.Proceedings of the 2nd International Conference on Artificial Intelligence and Engineering Applications(AIEA 2017)[C].
- [6] Zhengguang Xu. Complex production process prediction model based on EMD-XGBOOST-RLSE[A]. Kunming University of Science and Technology、 IEEE Control System Society Beijing Chapter、 IEEE Beijing Section.Proceedings of 2017 9th International Conference on Modelling, Identification and Control (ICMIC 2017)[C].
- [7] R.Urraca,E.Martinez-de-Pison,A.Sanz-Garcia,J.Antonanzas,F. Antonanzas-Torres. Estimation methods for global solar radiation: Case study evaluation of five different approaches in central Spain[J]. Renewable and Sustainable Energy Reviews,2016.
- [8] Peng Xuange,Huang ChunyingLi,Li Peipei.River Sediment Image Denoising Algorithm for Hydrological Monitoring Device[C].//The Second International Symposium on Test Automation & Instrumentation(ISTAI'2008).
- [9] Integrating CART Algorithm and Multi-source Remote Sensing Data to Estimate Sub-pixel Impervious Surface Coverage: A Case Study from Beijing Municipality, China[J].
- [10] Xiaowen Jin,Zhihua Chen,Fengqin Liu, et al.The Application of Classification and Regression Tree (CART) with Multi-feature Image on Land Cover Classification in Mining Area[C].
- [11] Xu Wei, Jiang Luo. Study on brain Age diagnosis based on XGBoost method in brain Wave data [J]. Journal of Wenzhou University (Natural Science Edition),2018,39(03):47-55.
- [12] Feature subset selection method for AdaBoost training[J]. 2011,(3):399-402.
- [13] Forest type identification by random forest classification combined with SPOT and multitemporal SAR data[J]. 2018,(5):1407-1414.
- [14] Fei Lin,Jian Jiang,Jin Fan,Shihua Wang. A stacking model for variation prediction of public bicycle traffic flow[J]. Intelligent Data Analysis,2018,22(4).
- [15] Lun Ma,Jian She Kang,Chun Yu Zhao.Research on Condition Monitoring of Bearing Health Using Vibration Data[C].//the 2012 International Conference on Vibration, Structural Engineering and Measurement .