Research of Condition Health Sample Model based on Bootstrap Method for Hydropower Unit

Zhou Ye, Tang Shu, Cao Dengfeng and Pan Luoping

China Institute of Water Resource and Hydropower Research
zhouye@foxmail.com

Abstract

In this paper, long term trend data from condition monitoring systems of hydropower units was used to separate to grid cells based on the water head and active power, and to build the health sample model for abnormal evaluation. For the problem of insufficient monitoring samples on non-rated operating conditions, Bootstrap method is introduced for sample expansion and confidence interval calculation, thereby to achieve limit optimize settings of health sample of hydropower unit condition data, which is more approximate to practical engineering than normal approximation method, and has achieved a good result.

Keywords: Bootstrap Method, Condition Monitoring, Hydropower Unit, Fault Diagnosis

1. Introduction

As the low rotating speed machine, hydropower units rarely occurs sudden serious accidents, which are usually processed by monitoring and protection systems in real-time, but the development of degradation and faults of hydropower units operation condition is a gradual changing process[1]. Most hydropower stations have installed kinds of condition monitoring equipment, which store large units normal and healthy operation condition data. So, we can based on the units' status data under stable operation conditions, establish unit condition health sample, and achieve health standard value and threshold value by using the statistics method, then we can find out the hidden problems of unit in advance, avoid terrible fault, and improve the stability of hydropower units, which has a huge economic value and practical significance.

Combining with mature and widely used status monitoring systems in hydropower industry, and unit fault partitions, monitoring measurement points can be divided into four parts, turbine, generator, bearing and shafting, and main transformer, so monitoring data of every parts can be evaluated comprehensively, finally whole unit status can be got. Units' usual faults type can be divided into two parts, mechanical faults and electrical faults, the former can be evaluated by abnormal of vibration, run out and pressure fluctuation and the later can be evaluated comprehensively by oil temperature, pad temperature, air gap, partial discharge and gas content in main transformer oil.

For hydropower unit, there are not enough abnormal condition data, and the startup and shutdown time is not too long, so it is difficult to obtain the full sample data in all condition. If we want to use large amounts of historical trend data to establish the condition health sample model and evaluate the condition, we will inevitably meet the problem of small samples. The Bootstrap method [2, 3] is proposed by professor Efron, studied at Stanford University, based on others research achievements. It is a new augmented sample statistical method, and works well for solving the main inductive problem of small sample data. In recently years, many experts carried out the studies on application of Bootstrap methods [4-6].

So in this paper, long term trend data from condition monitoring systems of hydropower units was used to separate to grid cells based on the water head and active power, and build the health sample model for abnormal evaluation. For the problem of insufficient monitoring samples on non-rated operating conditions, Bootstrap method is introduced for sample expansion and confidence interval calculation, thereby to achieve limit optimize settings of health sample of hydropower unit condition data, which is more approximate to practical engineering than normal approximation method, and has achieved a good result.

2. Research of Health Sample Model

According to law of large numbers and central limit theorem, in condition of large sample, theory means value can be represented with sample mean value. So, the long term monitoring data can be accumulated according to unit normal operation condition, and we can consider the mean value of characteristic results in the sample space, the characteristic results can be mean value, amplitude, or frequency and so on, as the health value of its operation condition, also as the reference value and then, establish the relative evaluation criteria.

After analyzing monitoring data in hydropower units on site, taking unit active power and head as main two factors of operating status, and taking unit active power, head and current health criteria value as three-dimensional model, threshold value is determined by monitoring data statistic of current active power and head limit scope, and finally establish unit abnormal status evaluation model [7]:

$$\phi_i(P_m, H_n) = f(\mu_i, \delta_i, Z1_i, Z2_i) \tag{1}$$

In which, P_m is active power interval, H_n is head interval, μ_i is mean value, δ_i is mean square deviation/standard deviation, $Z1_i$ is health low limit, $Z2_i$ is health up limit, for a single characteristic quantity health sample, i is the product of active power interval partition number and head interval partition number.

Random error of hydropower unit status monitoring system data measurement is sum of small errors caused by many mutual independent factors, including operating conditions fluctuation, monitoring instrument sensitivity, data communication status and random interferences. Random error has obvious statistic distribution character, obey normal distribution. So, evaluation alarm value can be determined according to 3σ criterion(Rule of Lai), which is alarm limit value [8],

$$X_c = \overline{X} \pm 3\sigma \tag{2}$$

In the formula above, σ is standard deviation, which is

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2} \tag{3}$$

Feature extraction method, which is on basis of statistic method, is mainly applied in frequency domain characters after time domain process. In hydropower unit status monitoring and diagnosis, no matter unit is in normal or abnormal operating status, when it operates stably in this status, characteristic quantities coming from this status should be relatively stable, with no influence of operating parameters. At this time, characteristic indexes are independent identically distributed, and can be statistically simulated with method of Bootstrap, to determine the confidence interval and finally reflect unit status [9].

According to traditional mathematical statistics theory, instrument status characters need large numbers of samples to statistical analysis. For hydropower units, even monitoring system is installed in main engine and corollary instruments, actually data in normal status are collected mainly, so abnormal data is few. So, how to use small sample to establish health sample interval is also a difficulty in hydropower unit abnormal status discrimination.

3. Bootstrap Method

3.1. Mathematical Description of Bootstrap Method

Set random sample $X = [x_1, x_2, \cdots, x_n]$ is independent identically distributed, in which $x_i \sim F(x), i = 1, 2, \cdots, n$. $R(X, F)$ is some selected random variation in advance, and is the function of X and F. If according to monitoring sample $[x_1, x_2, \cdots, x_n]$ evaluate the distribution character of $R(X, F)$, set $\theta = \theta(F)$ parameter of total distribution F, F_n is experience distribution function of monitoring sample X, $\hat{\theta} = \hat{\theta}(F_n)$ is the evaluation of θ, among which evaluation is[10]:

$$R(X, F) = \hat{\theta}(F_n) - \theta(F) \square T_n \tag{4}$$

Now evaluate $R(X, F)$ distribution character by monitoring sample $X = [x_1, x_2, \cdots, x_n]$, Bootstrap method is actually a resample process, method of calculating $R(X, F)$ distribution character is as follows,

1) Establish experience distribution function F_n by monitoring sample $X = [x_1, x_2, \cdots, x_n]$.

2) Extract sample $X^* = [x_1^*, x_2^*, \cdots, x_n^*]$ from F_n, which is Bootstrap sample.

3) Calculate corresponding Bootstrap statistic $R^*(X^*, F_n)$, which is[11]:

$$R^*(X^*, F_n) = \hat{\theta}(F_n^*) - \hat{\theta}(F_n) \square R_n \tag{5}$$

In the formula, F_n^* is experience distribution function of Bootstrap sample, F_n^* is Bootstrap statistic of T_n.

4) Repeat the process 2) and 3) N times to get N possible values of Bootstrap statistic $R^*(X^*, F_n)$.

5) Use $R^*(X^*, F_n)$ distribution to approach $R(X, F)$ distribution, which is using R_n distribution to approach T_n distribution, and get N possible values of $\theta(F)$, which means distribution and characteristic value of parameter θ can be statistically got.

3.2. Calculation of Bootstrap Confidence Interval

In condition of unknown total distribution, use Bootstrap method provides interval evaluation which can hardly be achieved by traditional methods. For experience distribution function F_n^* of Bootstrap sample in 3.1, simplify $\hat{\theta}(F_n^*)$ as $\hat{\theta}^*(j)$, $j = 1, 2, \cdots, N$, usually use normal approximate method t-percentile method evaluate the confidence interval of location parameter θ [12].

Normal approximate method applies when $\hat{\theta}^*(j)$ distribution obey or approximately obey normal distribution, then N Bootstrap sample mean value $\overline{\theta}^*$ and mean value deviation $\mu(\theta^*)$ can be calculated, set significance level is α, then Bootstrap confidence interval $(1 - \alpha)$ is [13],

$$\left[\overline{\theta}^* - Z_{1-\alpha/2} \cdot \mu(\theta^*), \overline{\theta}^* + Z_{1-\alpha/2} \cdot \mu(\theta^*) \right] \qquad (6)$$

In the formula above, $Z_{1-\alpha/2}$ is $(1 - \alpha/2)$ percentage of standard normal distribution

$$\left[\overline{\theta}^* - t_{N \cdot (1-\alpha/2)}^* \cdot \sqrt{\mu(\hat{\theta})}, \overline{\theta}^* + t_{N \cdot (1-\alpha/2)}^* \cdot \sqrt{\mu(\hat{\theta})} \right] \qquad (7)$$

T-percentile method [14] is on basis of percentile Bootstrap method, they individually calculate original sample evaluation $\sqrt{\mu(\hat{\theta}^*)}$ and Bootstrap sample standard deviation evaluation $\sqrt{\mu(\hat{\theta}^*(j))}$, aiming at every Bootstrap sample, calculate t statistics, which is $t_j^* = (\hat{\theta}^*(j) - \hat{\theta}) / \sqrt{\mu(\hat{\theta}^*(j))}$, $j = 1, 2, \cdots, N$, then arrange t_j^* in order, and get $(1 - \alpha)$ Bootstrap confidence interval of parameter θ, which is

$$\left[\overline{\theta}^* - t_{N \cdot (1-\alpha/2)}^* \cdot \sqrt{\mu(\hat{\theta})}, \overline{\theta}^* + t_{N \cdot (1-\alpha/2)}^* \cdot \sqrt{\mu(\hat{\theta})} \right] \qquad (8)$$

## 4. Health Sample Threshold Calculation based on Bootstrap Method

With the long term condition monitoring data in the hydropower remote diagnosis center, which built by a certain large hydropower corporation, we can use the points of guide bearing displacement, bracket vibration, guide bearing bush temperature and *etc.*, to build the health sample models, including the health reference value and threshold limit value. Taking the thrust bearing bush No.1 temperature as an example, we can extract the long term feature trend values in 11 months, and divide the data by the active power and water head intervals, so we can get a large amount of data in one certain operation condition, and find the health reference value and threshold limit. Because the time periods of transient process of unit, just like start-up or shutdown process, are very short and have few data points.

To the Francis unit which rated active power is 700MW, we can divide the period with the 50MW interval in the range of 0 to 575MW, and with 10MW in the range of 575MW to 750MW. Also, divide the water head period with 2 meters interval in the range of 70

to 110 meters. Then, for the formula (1), $m = 30$, $n = 21$, and we can build $m \times n = 630$ groups of health sample model for one signal.

For the random variables X obey the normal distribution, the probability of which located within the threshold range of formula (2) is 99.74%, so we can consider the threshold values as the limit range of its health sample model.

Taking an example of water head in 80 to 82 meters interval, we can get the threshold range of health sample model, which shows the thrust bearing bush temperature changes with active power of unit. Here, we can use the median or mean value as the health reference line. It can be seen in Figure 1, when in a long time period, the data series of long term will weaken the effect of abnormal values to the mean value. In the stable and normal operation condition which has the large data points, like 650MW to 700MW, there are little differences between the mean and median value. So we can choose the mean or median, within the corresponding water head and active power range, as the health reference value.
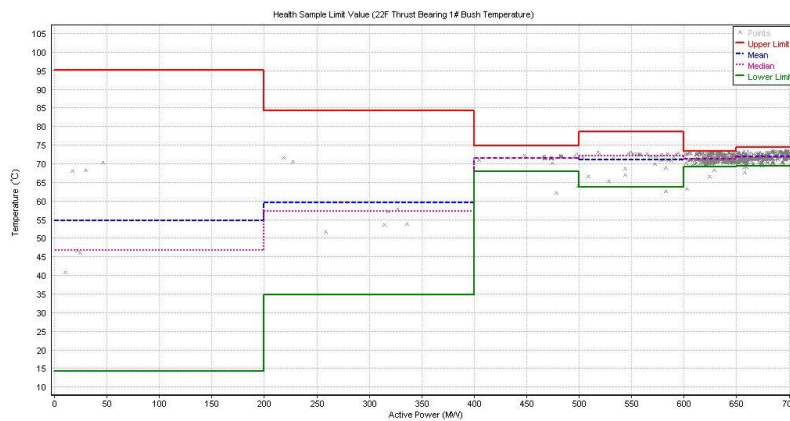
Figure 1. The Health Sample Model of Thrust Bearing Bush Temperature

If acquiring 1 feature value every 5 minutes, there are a total of 87843 points within 11 months' time interval. But after dividing the data by grid of active power and water head, the numbers of every grid cell has been greatly reduced. For instance, if we select the water head in 80-82 meters interval, the number of data points will be 8226, then divide by the active power interval, the amount of data will further reduce. At this time, if we still calculate and determine the health threshold by the normal distribution method, for the limit of data sample sizes, will lead to unreliable and not applicable results. Then, the Bootstrap method which can resample the data is effective to calculate the health sample threshold. The main steps are as follows:

1) Select the water head of interval 80-82 meters and active power in 650 to 700 MW, there are totally 3942 points for the No.1 thrust bearing bush temperature of unit, which marked as population parameter θ, the mean value is calculated as $\mu = 71.873$, the variance unbiased estimator $\delta = 0.835$, and median value is 71.7.

2) For the raw data sample $x = (x_1, x_2, \cdots, x_n)$, by the sampling method with replacement, extract the $B = 1000$ independent Bootstrap samples which single capacity $n = 1000$. Here $x^{*i} = \left(x_1^{*i}, x_2^{*i}, \cdots, x_n^{*i} \right)$, $i = 1, 2, \cdots, B$. For each Bootstrap sample, we can get the ith Bootstrap estimation of θ, which marked as $\hat{\theta}_i^*$.

3) Calculate mean value of $\hat{\theta}_1^*, \hat{\theta}_2^*, \cdots, \hat{\theta}_B^*$, $\overline{\mu}^*$ =71.815, and standard deviation $\overline{\delta}^*$ =0.02639.

4) If population parameter $X \sim N(\mu, \delta^2)$, then the mean value of Bootstrap sample $X = (x_1, x_2, \cdots, x_n)$, $\overline{\mu} \sim N(\mu, \delta^2 / n)$ [15], so we can calculate the mean value and standard deviation of population parameter θ , as mean value $\overline{\mu} = \overline{\mu}^*$ =71.815, standard deviation $\overline{\delta} = \sqrt{n} \cdot \overline{\delta}^*$ =0.8346.

5. Result

Finally, repeating the step 1) to 4), we can get all Bootstrap confidence ranges which water head values in 80 to 82 meters and all active power values included. The estimating values are shown in Table 1. It can be seen from Figure 1, for the traditional parameter estimation method, due to the limit of samples count when unit is not working in rated power condition, it has seriously deviated from the actual situation, and are no longer applicable to determination of health condition sample threshold.

Table 1. Bootstrap Estimation and Original Intervals with Water Head in 80-82m

Active Power（MW）	Mean	Bootstrap Mean	MSD*	Bootstrap MSD	Upper Limit	Bootstrap Upper Limit	Lower Limit	Bootstrap Lower Limit
0-200	54.751	54.694	13.474	12.743	95.174	95.117	14.329	14.271
200-400	59.531	59.471	8.216	8.117	84.182	84.121	34.880	34.820
400-500	71.388	71.387	1.169	1.156	74.896	74.895	67.880	67.879
500-600	71.134	71.120	2.461	2.366	78.519	78.505	63.748	63.734
600-650	71.227	71.196	0.695	0.691	73.312	73.281	69.142	69.111
650-700	71.873	71.815	0.835	0.834	74.378	74.320	69.367	69.309

* MSD: Mean Squared Deviation.

From the positive or negative of the deviation between the estimate and actual value in Table 1, we can see the upper limit values which based on Bootstrap estimation of normal approximation, are lower than the original values $\overline{X} + 3\sigma$. By several sample simulation calculations, we can find the estimate results are not only close to original values, but mostly located within the range $X_c = \overline{X} \pm 3\sigma$ of original ones. So, it is more conductive to determine the abnormal condition of hydropower unit.

Meanwhile, in different active power conditions, especially in non-rated active power condition with small data samples, the use of Bootstrap method reduces the threshold range of health model significantly, and makes the range value more consistent to the real condition. When unit working in rated active power condition, for the number of samples increases, the advantage is not obvious. So we can use Bootstrap and general estimation methods to make joint estimation for different operation and sample conditions, and finally get more accurate health sample model for the engineering practice.

6. Conclusion

The paper discussed the healthy sample modeling and calculation methods for condition monitoring data of hydropower unit in the small sample case, with the probability distribution density and non-parameter Bootstrap method. And it increased the sample size and confirmed the prior distribution characteristics of Bootstrap method without assuming parameters, by analyzing the examples with secondary sampling methods. Compared with the traditional statistical method, the application of Bootstrap method can increase the estimation accuracy of health threshold when unit is not working in rated operation condition, and it is significant for spreading and applying on fault diagnosis and abnormal evaluation.

Acknowledgement

References

[1] J. Liu, L. Pan, Z. Gui, Y. Zhou, "The current status of online state monitoring and fault diagnosis technologies for hydrogenerating unit", Large Electric Machine and Hydraulic Turbine, vol.45, no.2 (2010).

[2] B. Efron, R. Tibshirani, "An introduction to the Bootstrap", London:Chapman & Hall, (1993).

[3] B. Efron, "Jackknife-after-Bootstrap standard errors and influence functions", J.R Statist Soc.B, vol. 54, (1992), pp. 83-127.

[4] D. Xiaojun, W. Zhengming, "Bootstrap method on small sample", Journal of Ballistics, vol. 15, no. 3, (2003), pp. 1-5.

[5] T. Kijewski, A. Kareem, "On the reliability of a class of system identification techniques: Insights from Bootstrap Theory", Structural Safety, vol. 24, (2002).

[6] Y. Xu, C. Deng, J. Wu, "An reliability assessment method based on Bootstrap", Machinery Design & Manufacture, vol. 3, (2010).

[7] X. An, L. Pan, Z. Gui, Y. Zhou, "Abnormal condition detection model for pumped storage unit", Water Resources and Power, vol.31, no,1, (2013).

[8] L. Huipeng, H. Guo and M. Tingfeng, "Application of Bootstrap method in electromechanical equipment of vibrate threshold setting", Ship Science and Technology, vol.34, no.5 (2012).

[9] W. Xisen, "Pattern Recognition and Condition Monitoring", China Science Press, (2007).

[10] H. Zhao, "The Application in Confidence Intervals by Bootstrap", Jiangxi Science, vol 28, no.4, (2010).

[11] W. Dong, "Homology algebra", Beijing Gaodeng Press, (1998).

[12] Z. Jianjun, S. Jianyong, C. Haijuan, "Application research of Bootstrap Method in Small Sample Mesured Data Reduction of Random Vibration Environment", Equipment Environmental Engineering, vol. 10, no. 10, (2013), pp. 70-78.

[13] H M J. Al-Thani, "K-projective semmimodules", Kob J. Math, vol.13, no.1, (1996), pp. 49-59.

[14] M. Mudelsee, M. Alkio, "Quantifying effects in two-sample environmental experiments using bootstrap confidence intervals", Environmental Modelling & Software, vol. 22, vo.1 (2007), pp. 84-96.

[15] S. Ju, X. Shiqian and P. Chengyi, "Probability and mathematical statistics (the Second Edition)", Higher Education Press, Beijing, (1989).

Authors

Zhou Ye, he graduated as the post graduate student in 2006 from Huazhong University of Science and Technology in China. Now he is studying for PhD in China Institute of Water Resources and Hydropower Research. His current research interests include aspects of condition monitoring, fault diagnosis and pattern recognition for hydropower unit.