Water Science and Engineering 2015, 8(2): 158–163

HOSTED BY

ELSEVIER

CrossMark

Available online at www.sciencedirect.com

Water Science and Engineering

journal homepage: http://www.waterjournal.cn

Hydraulic metal structure health diagnosis based on data mining technology

Guang-ming Yang [a,b,*], Xiao Feng [c], Kun Yang [c]

[a] *College of Energy and Electrical Engineering, Hohai University, Nanjing 210098, PR China*

[b] *Research Center for Renewable Energy Generation Engineering, Ministry of Education, Hohai University, Nanjing 210098, PR China*

[c] *Dayu College, Hohai University, Nanjing 210098, PR China*

Received 14 November 2013; accepted 31 October 2014
Available online 8 May 2015

Abstract

In conjunction with association rules for data mining, the connections between testing indices and strong and weak association rules were determined, and new derivative rules were obtained by further reasoning. Association rules were used to analyze correlation and check consistency between indices. This study shows that the judgment obtained by weak association rules or non-association rules is more accurate and more credible than that obtained by strong association rules. When the testing grades of two indices in the weak association rules are inconsistent, the testing grades of indices are more likely to be erroneous, and the mistakes are often caused by human factors. Clustering data mining technology was used to analyze the reliability of a diagnosis, or to perform health diagnosis directly. Analysis showed that the clustering results are related to the indices selected, and that if the indices selected are more significant, the characteristics of clustering results are also more significant, and the analysis or diagnosis is more credible. The indices and diagnosis analysis function produced by this study provide a necessary theoretical foundation and new ideas for the development of hydraulic metal structure health diagnosis technology.

Keywords: Hydraulic metal structure; Health diagnosis; Data mining technology; Clustering model; Association rule

1. Introduction

Hydraulic metal structure health diagnosis is based on investigation and analysis of the status of operating equipment, site safety testing, review of the structural safety calculation results, and comprehensive analysis of every diagnosis index, eventually obtaining the final equipment health diagnosis through fuzzy comprehensive diagnosis analysis. At present, there are many health diagnosis methods in engineering fields, such as the reliability evaluation method, analytic hierarchy process, expert judgment method, and neural network technology. Each method has its advantages and disadvantages, applicable in different occasions. The hydraulic metal structure health diagnosis system needs to consider various factors, and perform comprehensive diagnosis with the multi-layer, multi-standard, and multi-factor analysis model. Yang (2011) established an expert system for evaluating the safety of hydraulic metal structures. Yang (2012) developed a multi-layer fuzzy comprehensive evaluation index system for hydraulic metal structure health diagnosis.

Data mining is the process of extracting or mining useful information from large amounts of data. Because data mining technology has significant advantages in the processing of large amounts of data, many industries, especially scientific research, finance, and education, have been utilizing data mining technology. In the fields of water resources and hydropower engineering, many scholars have used data mining technology to research dam safety monitoring system (Xiang et al., 2003; Zhu et al., 2007). However, few experts have conducted research on hydraulic metal structure diagnosis

This work was supported by the Key Program of the National Natural Science Foundation of China (Grant No. 50539010) and the Special Fund for Public Welfare Industry of the Ministry of Water Resources of China (Grant No. 200801019).
* Corresponding author.
E-mail address: gmyang@hhu.edu.cn (Guang-ming Yang).
Peer review under responsibility of Hohai University.

http://dx.doi.org/10.1016/j.wse.2015.04.010

with data mining technology. Combining data mining with information theory technology (Han and Kamber, 2006; Mao et al., 2007; Hu et al., 2012), this study used data mining technology to establish a clustering data mining model for hydraulic metal structure health diagnosis, and examined key technology for hydraulic metal structure health diagnosis, which includes importance analysis of the indices, correlation analysis of the indices, and analysis of the diagnosis conclusion. The significant problem in hydraulic metal structure health diagnosis is ensuring the accuracy and consistency of the original data. First, in conjunction with association rules for data mining, the connections between testing indices and strong and weak association rules were obtained by making full use of the historical data, and new derivative rules could be obtained by further reasoning. Based on correlation analysis of indices with association rules, checking the consistency between index data was proposed. Based on that, this study analyzed the reliability of diagnosis with clustering technology. Using this method, data mining technology can be applied in real-world hydraulic metal structure health diagnosis.

2. Association rules of testing indices

Association rules (Su et al., 2004; Lou et al., 2003) are an important subject in data mining; they are used to dig out valuable correlations between data. Using the gate testing as an example, we studied the use of association rules to mine the possible connection between different testing indices (Li and Li, 2003; Cover et al., 2003; Liang, 2006; Sha, 2008). It is difficult to determine whether there is a correlation between indices by tested data directly, so multi-level association rules are used in mining. For two testing indices, X and Y, there exists an association rule between them: X: $x \rightarrow$ Y: y. The association rule can be classified into the following three types: (1) a strong association rule: when the grade of index X is x, the grade of index Y is most likely to be y; (2) a weak association rule (even a non-association rule): when the grade of index X is x, the grade of index Y is most unlikely to be y; and (3) a derivative rule, the other weak association rule, deduced by the weak association rule: when the grade of index X is x, the grade of index Y is most unlikely to be y and less likely to be grades lower than y.

2.1. Strong association rules

We selected 18 samples from historical testing data records of gates, and numbered them from 1 to 18. To analyze correlation between six indices (also referred to as six items in association rules), the strength of main components, depth of the corrosion pit, area of corrosion, maintenance, years of operation, and performance, the calculation process of mining strong association rules was as follows:

(1) Based on each data sample, 1-itemsets were obtained, as shown in Table 1.

(2) Assuming that the minimum support value was 5/18, grades of testing indices with support values less than the minimum value were deleted, and frequent 1-itemsets were obtained, as shown in Table 2. The support value can be calculated with the following formula:

$$\mathrm{Support}(X:x \rightarrow Y:y) = \frac{m_1}{M} \tag{1}$$

where m_1 is the number of same samples with the testing index X at grade x and the testing index Y at grade y, and M is the total number of all samples.

(3) Connecting different indices in frequent 1-itemsets in pairs, candidate 2-itemsets were obtained. The 2-itemsets with support values less than the minimum value in the candidate 2-itemsets were deleted, and frequent 2-itemsets were obtained, as shown in Table 3, where the testing indices X_1 and X_2 are two items of a frequent 2-itemset.

(4) Connecting the items in Table 3, candidate 3-itemsets were obtained. The 3-itemsets with support values less than the minimum value in the candidate 3-itemsets were deleted, and frequent 3-itemsets were obtained, as shown in Table 4, where the testing indices X_1, X_2, and X_3 are three items of a frequent 3-itemset.

Table 1
1-itemsets.

Testing index	Grade	Gate sample	Testing index	Grade	Gate sample
Strength of main components	A	9, 11	Maintenance	A	10, 11, 14, 15, 16
	B	4, 14, 15, 16		B	5, 6, 7, 8, 9, 13, 17, 18
	C	10, 18		C	1, 2, 3, 4
	D	1, 2, 3, 5, 6, 7, 8, 12, 13, 17		D	12
Depth of corrosion pit	A	6, 16	Years of operation	A	10, 11, 12, 13, 17, 18
	B	10, 11, 14, 15, 17		B	14, 15, 16
	C	2, 4, 5, 7, 9, 13, 18		C	1, 2, 3, 4, 5, 6, 7, 8, 9
	D	1, 3, 8, 12		D	
Area of corrosion	A	6, 10, 11, 15, 16	Performance	A	5, 9, 15, 16
	B	13, 14, 17, 18		B	1, 2, 3, 4, 7, 8, 10, 11, 12, 13, 14
	C	4, 5, 7, 8, 9		C	6, 17, 18
	D	1, 2, 3, 12		D	

Note: A, B, C, and D are different grades of the testing index, signifying that the tested values of the index are good, qualified, basically qualified, and unqualified, respectively.