ATLANTIS PRESS

Advances in Intelligent Systems Research, volume 134

2nd International Conference on Control, Automation, and Artificial Intelligence (CAAI 2017)

Fault Diagnosis of Gearbox of Wind Turbine Based on Improved Decision Tree Algorithm

Siwen Zhu and Bin Jiao

Shanghai Dianji University; Lane 300, Shuihua Road, Pudong New District, Shanghai, China.

***Abstract*—The most significant feature of the decision tree algorithm is to transform the complex decision-making process into a number of simple decision-making processes and then accumulate it. It's a tree structure similar to the flow chart. The decision tree can be applied to the fault diagnosis of wind turbine gearbox to data mining for gearbox and then find rules and reflect in the form of rules. Experiments show that the use of decision tree method to extract rules can be faster and more accurate.**

Keywords-component; formatting; style; styling; insert

I. INTRODUCTION

As an important component of a wind turbine, the gearbox takes on a key role of transmitting wind turbine power to the generator and gaining a certain speed. At once the gearbox broke down, it will take serious economic losses to the wind turbine industry. Therefore, it is of great practical significance to monitor the status of wind turbine effectively and study the fault diagnosis.

There are massive amounts of data brought out during the daily operation of the gearbox monitoring process, these data contain some potential rules, mastering these rules means grasping the gearbox fault diagnosis. Data mining can get regular knowledge from this massive amount of data. As the most commonly used classification mining method in data mining, the decision tree can find rules effectively and quickly by traversing the tree.

The earliest decision tree originated from CLS (Concept Learning System), which was developed by Hunt in 1966[1], and J.R.Quinlan formally proposed ID3 algorithm in 1979. [2] Decision tree classify data into different groups or brunches by accumulating s series of rules, and then realize the quick classification and easy understanding. It is a tree structure similar to the flow chart[3]. There are some other decision tree algorithm such as C4.5, CART and assistant. Introducing the decision tree into the fault diagnosis of gearbox of wind turbine can form a fault tree effectively and classify the faults by accurate and simple discrimination rules.

II. THE BASIC PRINCIPLE AND OPTIMIZATION OF THE DECISION TREE

A. *The Basic Principle of The Decision Tree*

Decision tree algorithm is the main technology used to classify and predict, is a typical case-based induction learning algorithm. [4] The decision tree often uses the top-down recursion method to compare the attributes in the internal nodes of the decision tree, and judge the branch of this node according to the different attributes. Finally, the conclusion is made at the leaf node. It's a rule from root to the leaf node, the whole tree is a set of rules. The basic algorithm of decision tree induction is greedy algorithm. The input of the decision tree construction algorithm is a set of examples with a class tag, and the result of the construction is a binary tree or a multi-frame. The inner node (non-leaf node) of the binary tree is often expressed as a logical judgment, such as (a = b), where a is the attribute and b is a value of the attribute. The edge of the tree is the result of the branching of the logical judgment, and the leaf nodes of the tree are the category markers. The core of constructing a good decision tree is choosing good logical judgments and attributes. The results show that the smaller the tree, the ability to predict is stronger. Choosing the appropriate logical judgment or attribute is important for constructing a decision tree as small as possible. The FIGURE Ⅰ shows the generation process of decision tree.

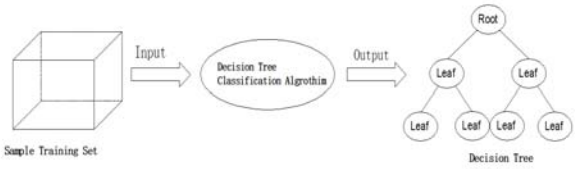

FIGURE I. GENERATION PROCESS OF DECISION TREE

Set a sample collection is made up of s sample data. Suppose the class label property has m different values, define m different Ci (i=1, 2, ... , m), si is the sample amounts of Ci. For a given sample, the required information entropy is classified as

$$E(s)=-\sum_{i=1}^{m} p_i \log_2(p_i) \quad (1)$$

Among them, Pi is the probability of random sample belongs to Ci, which is estimated by "s" _"i" /"s" . The logarithm is based on 2 because the information is encoded in binary.

Setting A has t different values $\{a_1,a_2,\cdots a_t\}$, A divides the sample set S into t subsets $\{S_1,S_2,\cdots,S_t\}$. Let the number of samples containing the j class of Si be P_{ij}(j=1, 2, … , m), The information entropy of the subset Si is

$$E(S_i)=\sum_{j=1}^{m}\frac{p_{ij}}{|S_i|}\log_2\frac{p_{ij}}{|S_i|} \quad (2)$$

The information entropy of the subset which divided by A as the root is

$$E(S|A)=\sum_{i=1}^{t}\frac{|S_i|}{|S|}E(S_i) \quad (3)$$

As can be seen from the above, each time a split node is selected, the algorithm has to perform several logarithmic operations. In the environment of large data, this operation obviously affects the tree generation efficiency.

B. Improvement of Decision Tree

One kind of Weighted simplified entropy has been proposed. The entropy method firstly combines the Taylor formula and the McLaughlin formula to change the test standard of the attribute, to simplify the formula and reduce the computational cost and finally accelerate the decision tree generation. Secondly, the method gives the simplified information entropy a weight to improve the algorithm.

Take (2) into (3) :

$$E(S|A)=\sum_{i=1}^{m}\sum_{i=1}^{t}\frac{P_{ij}\ln\frac{p_{ij}}{|S_i|}}{s\ln 2}$$

Since $|S|\ln 2$ is a constant, it is assumed that $e(S|A)$ satisfies the following formula

$$e(S|A)=\sum_{i=1}^{m}\sum_{i=1}^{t}P_{ij}\ln\frac{p_{ij}}{|S_i|}$$

From the Taylor formula and the McLaughlin formula, when $x\to 0$, $\ln(1+x)\approx x$, so $e(S|A)$ can be simplified as

$$e(S|A)=\sum_{i=1}^{m}\sum_{i=1}^{t}\frac{P_{ij}(|S_i|-p_{ij})}{S_i}$$

Assume that N is the number of attribute values, and N is brought in, and the $E(S|A)$ is equivalent

$$\sum_{j=1}^{m}\sum_{i=1}^{t}\frac{p_{ij}(|S_i|-p_{ij})}{S_i}N \quad (4)$$

Formula (3) contains only the addition, multiplication, division of the operation, the operation time compared to $E(S|A)$ in the logarithmic operation is shorter. At the same time, after using the number of attributes to weight, the new attribute selection standard makes up the error caused by the Taylor formula, thus improved the classification accuracy of the decision tree classifier.

III. RULE EXTRACTION

All the fault samples are stored in tabular form, and then the improved decision tree algorithm is used to generate a fault tree. Each path from the root to the leaf corresponds to a rule of discrimination. A fault tree generated by training a fault sample with rich and complete data can serve as a standard discriminant rule fault tree. For a given decision tree, a union of the regular elements of the rule is formed along each "attribute-value" on a given path. The leaf nodes of the decision tree contain class predictions that form the rule post.

It is known that (X, Y) is a joint random distribution variable, where q dimension vector X is a pattern or eigenvector; Y is a related class of X; The X component is some feature information. Y vector values 1, 2, 3, which involve J classes. The purpose of the decision tree classification is usually to estimate Y by observation.

IV. APPLICATION OF IMPROVED DECISION TREE ALGORITHM IN GEARBOX FAULT DIAGNOSIS

There are five typical faults in the operation of the wind turbine, including the rolling bearing inner ring, the bearing outer ring peeling, the bearing cage is damaged, the gear tooth surface wear, the gear teeth. Here, the time domain waveform is obtained by continuous sampling, and then the original vibration signal containing a large amount of noise is subjected to noise reduction pretreatment. Which includes four kinds of time domain characteristic parameters including peak index, kurtosis index, margin index and skewness index, and three frequency domain indexes including power spectrum center of gravity index, power spectrum variance and harmonic factor.

Due to the different dimension of the different characteristic parameters, the characteristic parameters are normalized before modeling. The selected 30 sample training set data are shown in Table Ⅰ, where $\{f_1, f_2, f_3, f_4, f_5, f_6\}$ indicate the normal state of the gearbox and five fault states. The improvement of decision tree algorithm is based on the classic ID3 algorithm.

TABLE I. TRAINING SET OF SAMPLES

No.	1	2	3	4	5	…	30
peak index	0	0.9079	0.9769	0.5199	0.6029	…	0.2269
kurtosis index	0.0027	0.3501	1	0.26	0.2359	…	0.171
margin index	0	0.6515	1	0.4122	0.4465	…	0.1759
skewness index	0.0494	0.3487	1	0.3433	0.3133	…	0.2476
power spectrum center of gravity index	0.7707	0.8396	0.614	0.5489	0.592	…	0.4585
power spectrum variance	0.7203	0.8554	0.8732	0.9952	0.9724	…	0.994
harmonic factor	0.0506	0.4868	0.0819	0.1874	0.118	…	0.0586
Fault types	f1	f3	f6	f5	f5	…	f4

In this paper, the fuzzy c-means (FCM) algorithm is used to discretize the training sample set data. The discretization interval of the feature index is [0.345, 0.764], [0.335, 0.765], [0.299, 0.695] 0.341, 0.735], [0.584, 0.705], [0.838,0.945],

[0.202, 0.529], {0,1,2} denote low, medium and high, TABLE II shows the discrete training set of samples, where c1, c2, ..., c7 represent seven characteristic indicators.

TABLE II. DISCRETE TRAINING SET OF SAMPLES

No.	c1	c2	c3	c4	c5	c6	c7	Fault
1	0	0	0	0	2	0	2	f1
2	0	0	0	0	1	1	1	f3
3	0	0	0	0	1	2	1	f6
4	0	0	0	0	1	2	0	f5
5	2	1	2	1	0	1	0	f5
…	…	…	…	…	…	…	…	…
30	1	1	1	1	2	0	1	f4

And finally generate a decision tree with the characteristic index c4 as the root, the c4 parameters are divided into three parts, 0 branches, 1 branch, 2 branches, and then on different branches, continue to use the same attribute selection rule algorithm recursively to construct the decision tree, the constructed decision tree is shown in Figure II. The data sample used in this paper is small and the decision tree is simple. When the data set is large and the fault category is very rich, the decision tree will be richer and the discriminant rules will be more accurate.

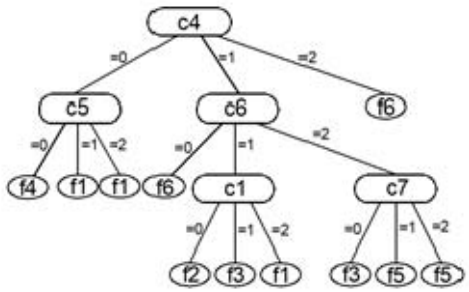

FIGURE II. THE CONSTRUCTED DECISION TREE

From the constructed decision tree, the depth of the decision tree is 3, c4 is the root, c1, c5, c6, c7 for the child nodes, there are a total of 9 rules. These 9 rules, as shown in TABLE Ⅲ, reflect the rules of the fan gear box failure characteristics, and can be stored in the wind turbine diagnostic knowledge library to diagnose the fault of wind turbine gearbox.

TABLE III. DECISION TREE RULES

RULES
IF c4≤0.342 AND c5≤0.584 THEN f4
IF c4≤0.342 AND 0.584＜c5≤0.705 THEN f1
IF c4≤0.342 AND c5＞0.795 THEN f1
IF 0.342＜c4≤0.735 AND c6≤0.837 THEN f6
IF 0.342＜c4≤0.735 AND 0.837＜c6≤0.945 AND c1≤0.345 THEN f2
IF 0.342＜c4≤0.735 AND 0.837＜c6≤0.945 AND 0.345＜c1≤0.764 THEN f3
IF 0.342＜c4≤0.735 AND 0.837＜c6≤0.945 AND c1＞0.764 THEN f2
IF 0.342＜c4≤0.735 AND c6＞0.945 AND c7≤0.201 THEN f3
IF 0.342＜c4≤0.735 AND c6＞0.945 AND 0.201＜c7≤0.529 THEN f5
IF 0.342＜c4≤0.735 AND c6＞0.945 AND c7＞0.529 THEN f5
IF c4＞0.735 THEN f6

In this paper, 20 sets of gearbox failure data are used to test the decision tree generated by the new algorithm and compared with the original ID3 algorithm. The results are compared in time and accuracy. The test results are shown in TABLE Ⅳ.

TABLE IV. COMPARISON OF TEST RESULTS

Algorithm	Classification time	Accuracy
ID3 algorithm	0.9s	88.1%
Weighted simplified entropy ID3 algorithm	0.69s	89.7%

As can be seen from TABLE Ⅳ, in the classification time, the new algorithm faster than the original algorithm, the accuracy rate compared to the original algorithm has also been improved. The experimental results show that the weighted simplified entropy ID3 algorithm improves the accuracy of the algorithm while reducing the computational complexity. This advantage will be more pronounced when the amount of data is larger.

V. CONCLUSION

In this paper, the failure characteristics of the common fan gearbox are given and the fuzzy c-means algorithm is used to discretize it. The improved algorithm has accomplished the classification problem of the wind turbine gearbox, the calculation is simplified to shorten the classification time, and the accuracy of judgment is improved by the way of weighting. The rules obtained from the sample training set can be used either for diagnosing faults or for determining fault decisions and can be used in fault diagnosis of wind turbine gearboxes.

REFERENCES

[1] Porter B W, Bareiss R, Holte R C. Concept Learning and Heuristic Classification in WeakTtheory Domains. [M]. University of Texas at Austin, 1989.Jiawei Han, Micheline Hamber.

[2] Quinlan J R. Discovering rules by induction from large collections of examples[J]. In Expert System in the Micro Electronic Age, 1979, 26~37.

[3] Concept and Technology of Data Mining[M]. Fan Ming, Meng Xiaofeng. Beijing: Machinery Industry Press, 2001.

[4] Huang Yiwen, Lu Shijun. Research on Decision - making Tree Based on Information Entropy[J]. Computer and digital engineering, 2016(5):878-883.

[5] Wang Guangfa. Analysis of Achievement Factors of Five-year Higher Vocational Students Based on ID3 Algorithm [J]. Software Engineer, 2015(7):16-18.

[6] Research on Decision Tree Classification Algorithm in Data Mining[D]. Zhao Xiang. Zhengjiang: Jiangsu University of Science and Technology, 2005.

[7] Wu Dehui. A fault diagnosis method based on support vector machine for gearbox [M]. Vibration, testing and diagnosis, 2008, 28(4):338-342.