Fault Diagnosis System Based on Rough Set Theory and Support Vector Machine*

Yitian Xu and Laisheng Wang**

College of Science, China Agricultural University, 100083, Beijing, China
{xytshuxue, wanglaish}@126.com

Abstract. The fault diagnosis on diesel engine is a difficult problem due to the complex structure of the engine and the presence of multi-excite sources. A new kind of fault diagnosis system based on Rough Set Theory and Support Vector Machine is proposed in the paper. Integrating the advantages of Rough Set Theory in effectively dealing with the uncertainty information and Support Vector Machine's greater generalization performance. The diagnosis of a diesel demonstrated that the solution can reduce the cost and raise the efficiency of diagnosis, and verified the feasibility of engineering application.

1 Introduction

In order to raise the efficiency and reduce the cost of fault diagnosis, intelligent identification of faults is desired in engineering application. Some theories or methods in computational intelligence are applicable to this task, such as neural networks, fuzzy set theory, genetic algorithm and so on. Considering the vagueness and uncertainty information in the process of fault diagnosis, a kind of hybrid fault diagnosis system based on Support vector machine (SVM) and Rough Set Theory(RS) is proposed in the paper.

Support vector machine is a new and promising machine learning technique proposed by Vapnik and his group at AT Bell Laboratories, It is based on VC dimensional theory and statistical learning theory. Classification is one of the most important applications. It is widely applied to machine learning, data mining, knowledge discovery and so on because of its greater generalization performance. But there are some drawbacks that it doesn't distinguish the importance of sample attributes, computation rate is slow and takes up more data storage space because of a large number of sample attributes. Moreover, It doesn't effectively deal with vagueness and uncertainty information. In order to resolve those problems, A kind of SVM fault diagnosis system based on Rough set pre-processing is proposed in the paper, Making great use of the advantages of Rough Set theory in pre-processing large data, eliminating redundant information and overcoming the disadvantages of slow processing speed causedby SVM approach. A

* This work was Supported by the National Natural Science foundation of China (No.10371131).

** Corresponding author.

L. Wang and Y. Jin (Eds.): FSKD 2005, LNAI 3614, pp. 980–988, 2005.
© Springer-Verlag Berlin Heidelberg 2005

hybrid fault diagnosis system based on Rough set and Support Vector Machine is presented in the paper too. It is more suitable to multi-classification. It may decrease fault diagnosis system complexity and improve fault diagnosis efficiency and accuracy.

2 Support Vector Machine[1][2][3]

Consider the problem of separable training vectors belonging to two separate classes,

$$T = \{(x_1, y_1), \cdots, (x_l, y_l)\}, x_i \in R^n, y_i \in \{-1, 1\}, i = 1, \cdots, l \tag{2.1}$$

with a hyperplane

$$(w \cdot x) + b = 0 \tag{2.2}$$

The set of vectors is said to be optimally separated by the hyperplane if it is separated without error and the distance between the closest vectors to the hyperplane is maximal. where the parameters w, b are constrained by

$$min|(w \cdot x) + b| = 1 \tag{2.3}$$

we should find a linear function:

$$f(x) = (w \cdot x) + b \tag{2.4}$$

that is to say ,we should make the margin between the two classes points as possible as big , it is equal to minimize $\frac{1}{2}\|w\|^2$, we should be according to structure risk minimum principle not experiential risk minimization principle, that is to minimize equation (2.5) upper bound with probability $1 - \sigma$,

$$R[f] \leq Remp[f] + \sqrt{\frac{1}{2}(h\ln(\frac{2l}{h} + 1)) + ln(\frac{4}{\sigma})} \tag{2.5}$$

the optimal classification function is transformed into a convex quadratic programming problem:

$$\min \quad \frac{1}{2}\|w\|^2 \tag{2.6}$$

$$\text{s.t.} \quad y_i((w \cdot x_i) + b) \geq 1, i = 1, 2, \cdots, l \tag{2.7}$$

when the training points are non-linearly Separable, (2.6)-(2.7) should be transformed into(2.8)-(2.9).

$$\min \quad \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{l}\xi_i \tag{2.8}$$

$$\text{s.t.} \quad y_i((w \cdot x_i) + b) \geq 1 - \xi_i, i = 1, 2, \cdots, l \tag{2.9}$$

The solution to the above optimization problem of equation (2.8)-(2.9) is transformed into the dual problem(2.10)-(2.12) by the saddle point of the Lagrange functional

$$\min_{\alpha} \quad \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j K(x_i, x_j) - \sum_{j=1}^{l} \alpha_j \tag{2.10}$$

$$\text{s.t.} \quad \sum_{i=1}^{l} y_i \alpha_i = 0 \tag{2.11}$$

$$0 \le \alpha_i \le c\ ,\ i = 1, 2, \cdots, l \tag{2.12}$$

We can get the decision function:

$$f(x) = \sum_{i=1}^{l} y_i \alpha_i K(x_i, x) + b \tag{2.13}$$

kernel function $K(x_i, x) = (\Phi(x_i) \cdot \Phi(x))$ is a symmetric function satisfying Mercer's condition, when given the sample sets are not separate in the primal space, we can be used to map the data with mapping Φ into a high dimensional feature space where linear classification is performed.

There are three parameters in svm model that we should choose, they make great impact on model's generalization ability, It is well known that svm generalization performance (estimation accuracy) depends on a good setting of hyperparameters C, the kernel function and kernel parameter. moreover, kernel function and kernel parameter's selection connects with feature selection in svm, so feature selection is very important.

3 Rough Set Theory[4]

Rough sets theory has been introduced by Zdzislaw Pawlak (Pawlak, 1991) to deal with imprecise or vague concepts. It has been developed for knowledge discovery in databases and experimental data set, It is based on the concept of an upper and a lower approximation of a set.

Rough set theory deals with information represented by a table called an information system. This table consists of objects (or cases) and attributes. The entries in the table are the categorical values of the features and possible categories. An information system is composed of a 4-tuple as following:

$$S = \langle U, A, V, f \rangle \tag{3.1}$$

where U is the universe, a finite set of N objects$\{x_1, x_2, \cdots, x_N\}$(a nonempty set), $A = C \bigcup D$ is condition attribute and decision attribute. V is attribute value. $f : U \times A \longrightarrow V$ is the total decision function called the information function.

For a given information system S, a given subset of attributes $R \subseteq A$ determines the approximation space $RS = (U, ind(A))$ in S, For a given $R \subseteq A$

and $X \subseteq U$ (a concept$X$), the R-lower approximation $\underline{R}X$ of set X in RS and the R upper approximation $\overline{R}X$ of set X in RS are defined as follows:

$$\underline{R}X = \{x \in U : [x]_R \subseteq X\}, \overline{R}X = \{x \in U : [x]_R \bigcap X \neq \phi\} \tag{3.2}$$

where $[X]_R$ denotes the set of all equivalence classes of $ind(R)$ (called indiscernibility relation). The following ratio defines an accuracy of the approximation of $X(X \neq \phi)$, by means of the attributes from R:

$$\alpha_R = \frac{|\underline{R}X|}{|\overline{R}X|} \tag{3.3}$$

where $|\underline{R}X|$ indicates the cardinality of a (definite) set $\underline{R}X$. Obviously $0 \leq \alpha_R \leq 1$. If $\alpha_R = 1$, then X is an ordinary (exact) set with respect to R; if $\alpha_R < 1$, then X is a rough (vague) set with respect to R.

Attribute reduction is one of the most important concept in RS. the process of finding a smaller set of attributes than original one with same classification capability as original sets is called attribute reduction. A reduction is the essential part of an information system (related to a subset of attributes) which can discern all objects discernible by the original information system. Core is the intersection of all reductions. Given an information system S, condition attributes C and decision attributes D, $A = C \bigcup D$, for a given set of condition attributes $P \subseteq (C)$, we can define a positive region $pos_p(D) = \bigcup\limits_{X \in U/D} \underline{P}X$, The positive region $pos_p(D)$ contains all objects in U, which can be classified without error (ideally) into distinct classes defined by $ind(D)$ based only on information in the $ind(P)$. Another important issue in data analysis is discovering dependencies between attributes. Let D and C be subsets of A. D depends on C in a degree denoted as

$$\gamma_C(D) = |pos_C(D)|/|U| \tag{3.4}$$

It was shown previously that the number $\gamma_C(D)$ expresses the degree of dependency between attributes C and D , It may be now checked how the coefficient $\gamma_C(D)$ changes when some attribute is removed. In other words, what is the difference between $\gamma_C(D)$ and $\gamma_{C-\{\alpha\}}(D)$. Attribute importance $\{\alpha\}$ about decision attribute is defined by

$$\sigma_{CD}\{\alpha\} = \sigma_C(D) - \sigma_{C-\{\alpha\}}(D) \tag{3.5}$$

4 Fault Diagnosis System Based on Rough Set Theory and Support Vector Machine[5][6]

In support vector machine, the solution of the model is transformed into a quadratic programming problem and we will achieve global solution but not local solution. it will produce good generation performance, But it is difficult to resolve a large number of training sample sets and not to deal with the vagueness and uncertainty information. Rough Set Theory is a data analysis tool in preprocessing imprecise or vague concepts. It is only based on the original data and does not need any additional information about data like probability in statistics or grade of membership in the Fuzzy set theory, it can reduce the attributes without decreasing its discriminating capability.

Integrating the advantages of RS and SVM, a kind of support vector machine fault diagnosis systemon the Rough Sets pre-processor is presented in the paper. When given a training sample set, we firstly discretize them if the sample attributes values are continuous and we can get a minimal feature subset that fully describes all concepts by attribute reduction, constructing a support vector machine fault diagnosis system . When given a testing set, we reduce the corresponding attributes and then put into SVM fault diagnosis system, then acquire the testing result. The whole process as fig 1.

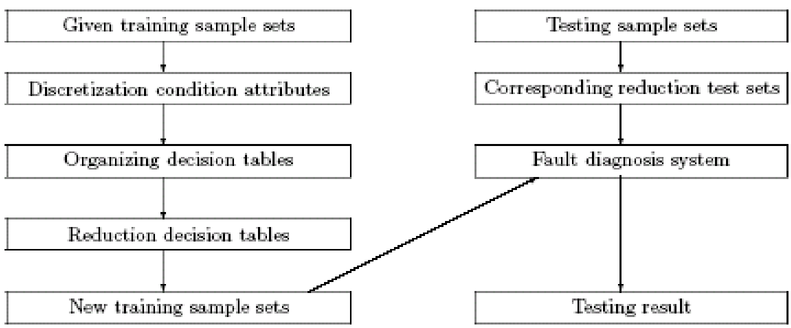

Fig. 1. Fault diagnosis system based on Rough set theory and support vector machine

When the training samples are the two classes separable, we can achieve them by above the fault diagnosis system. At first, we may preprocess training samples sets by Rough set theory, then classify them by support vector machine. Only need one classifier. When the training samples are multi-class (such as k classes), we often resolve them according to blow 3 method.

(1) One versus the rest: one class sample are signed "+1", the rest classes samples are signed "-1", Its need to construct k classification hyperplane to achieve them, that is to need to resolve k quadratic programming problem, but there is drawback that is produce multi classes to some samples, or some samples don't belong to any classes.

(2) One versus one: we can select 2 classes from all classes at random, thus it needs k(k-1)/2 classifier, calculating capacity is very larger. but there is drawback that may produce multi-classes to some samples.

(3) Layer classification method: This method is a improved One versus one method, we may combined K classes into 2 classes at first, and so on, different layers at last, we can classify by support vector machine in each layer.

In order to avoid multi-class to some samples, a kind of hybrid fault diagnosis system based on RS and SVM is presented in the paper. Utilizing the advantages of Rough set theory in extraction of rules, in order to improve classification accuracy, We can classify them exactly by support vector machine.

5 Fault Diagnosis About 4153 Diesel Engine[7][8]

Fault diagnosis on machinery has been researched presently. In this paper, we will take the 4153 diesel engine fault diagnosis for example. The fault diagnosis on diesel engine is a difficult problem due to the complex structure of the engine and the presence of multi-excite sources. The vibration signal of a 4135 diesel under normal and fault states is acquired(i=1,2,3) is the symptom, and represents the waveform complexity in frequency domain, center frequency of spectrum, waveform complexity in time domain, nonperiod complexity, variance of time series, and kurtosis of time series of the signals from measurement point 1, 2 and 3 respectively. They are the first cylinder head, the second cylinder head and another one that is at the center of the piston stroke, on the surface of the cylinder block. D is the fault reason, and the associated integers 1, 2, 3, and 4 represent normal state, intake valve clearance is too small, intake valve clearance is too large, and exhaust valve clearance is too large respectively.

5.1 Continuous Attributes Discretization Based on Fuzzy K-Means

Rough set theory only analyzes the discrete data, but the fault diagnostic data is continuous. so they must be quantized before extraction of rules from the original data. continuous attribute discretization directly affects the analysis result. Considering the vagueness and uncertainty diagnosis data in the process of fault diagnosis, Fuzzy K-means discretization method is proposed in the paper, it is an objective function based on fuzzy clustering algorithm, this algorithm typically converges to the local minimums, and possesses better robustness. In the course of clustering, the number selection of clusters is important. If the clusters are few, incompatible decision system will be resulted, and decision can not be made in the applications. If the clusters are too much, overdiscretization will be resulted, so that match of the condition for every rule will become too complicate. In this paper, corresponding to the states of the engine, 4 clusters are determined for each attribute. we select sample 8,9,10,11,12,13,25,26,27,28,29,30 as testing set, and the rest samples as training set, and getting the discretization decision Table 1.

5.2 Attributes Reduction and Rules Attraction Based Rough Set Theory

By attributes reduction, we can get a reduction of fault diagnosis system decision table as Table 2. certainly it isn't the only reduction table. In Table 2, more redundant values can be reduced, from decision and more concise rules can be

Table 1. Training sample continuous attributes values discretization besed on Fuzzy k-means, the number of clustering is 4

U	a1	b1	c1	d1	e1	f1	a2	b2	c2	d2	e2	f2	a3	b3	c3	d3	e3	f3	D
1	1	3	1	1	1	1	1	1	3	3	3	1	3	1	1	3	1	1	1
2	1	3	1	1	1	1	2	1	3	3	3	1	3	1	1	3	1	1	1
3	1	3	1	1	1	2	3	2	3	1	3	1	1	1	1	4	2	1	1
4	1	3	1	1	1	2	2	2	3	1	3	1	2	1	1	3	2	1	1
5	1	1	3	1	1	2	1	1	3	2	3	1	1	1	1	4	1	1	1
6	1	1	3	1	1	2	1	1	3	2	3	1	1	1	1	3	1	1	1
7	3	3	1	1	1	2	4	3	1	4	1	1	4	1	1	4	2	1	1
14	3	3	1	1	3	4	1	2	3	3	3	1	4	2	1	4	2	1	2
15	3	4	2	1	4	3	1	2	3	3	3	3	4	2	2	4	2	2	2
16	3	3	1	1	4	3	1	2	3	3	3	4	4	2	3	4	2	2	2
17	3	4	2	2	3	1	1	3	1	3	1	2	4	3	4	1	4	4	2
18	3	4	2	1	3	1	1	3	1	3	1	2	4	4	4	1	4	4	2
19	1	3	1	1	1	4	1	4	4	3	4	1	4	1	1	4	3	1	3
20	1	3	1	3	1	4	1	4	4	3	4	1	4	1	1	4	3	1	3
21	1	3	1	1	1	3	1	3	1	1	1	1	1	3	1	1	1	1	3
22	1	3	1	1	1	1	1	3	1	3	1	1	4	1	1	4	3	1	3
23	1	3	1	1	1	3	1	3	1	3	1	1	4	1	1	4	3	1	3
24	1	4	2	1	1	3	1	3	1	3	1	1	4	1	1	4	3	1	3
31	1	1	3	1	1	1	1	3	1	1	1	1	1	3	1	1	1	1	4
32	1	3	1	1	1	1	1	3	1	3	1	1	4	4	4	4	2	1	4
33	1	3	1	1	1	1	1	3	1	3	1	1	4	1	1	4	2	1	4
34	1	1	3	2	2	1	1	3	2	3	2	1	4	2	1	4	2	1	4
35	1	1	3	2	2	4	1	3	1	3	1	1	4	2	1	4	2	1	4
36	3	2	4	4	1	1	1	3	1	3	1	1	4	2	1	4	2	1	4
37	3	1	3	4	1	1	1	3	1	3	1	1	4	2	1	4	2	1	4

Table 2. The fault diagnosis decision system table after attributes reduction

U	e1	f1	e3	D
1	1	1	1	1
3	1	2	2	1
5	1	2	1	1
15	4	3	2	2
17	3	1	4	2
19	1	4	3	3
21	1	3	3	3
31	1	1	2	4
34	2	1	2	4
35	2	4	2	4

generated. we can know that attribute $\{e1, f1, e3\}$ are the most important attributes in the fault diagnosis system. we will get the same fault diagnosis result without losing any information by decision Table 2.

We can achieve some decision rules by the decision Table 2, such as

(1) If $e1 = 1$, $f1 = 1$ and $e3 = 1$ then $D = 1$;
(2) If $e1 = 1$, $f1 = 2$ and $e3 = 2$ then $D = 1$;
(3) If $e1 = 1$, $f1 = 2$ and $e3 = 1$ then $D = 1$;
(4) If $e1 = 1$, $f1 = 1$ and $e3 = 2$ then $D = 4$;

and so on.

As we can learn that the decision rules will be different because of different attribute value about f1 or e3. It will cause much difficult in the course of fault diagnosis. The rules generated by the Rough set theory are often unstable and have low fault diagnosis accuracy. In order to improve the fault diagnosis accuracy, we will further diagnose them by support vector machine because of it's greater generalization performance.

For example, if e1=3 or 4 then we can learn D= 2, if e1=2 then D=4. such can reduce the diagnosis time, but if e1=1, it will cause difficult to our diagnosis because of being to D=1 or D=3 or D=4. It need us to further diagnosis, we can classify them by the second or third attribute values, but we can't classify them exactly sometimes, we can classify them by support vector machine and by using the first multi-class method.

Certainly, we can construct the support vector fault diagnosis system based on the attribute reduction by Rough set theory. There are 18 attributes in the fault diagnosis system before the attribute reduction, but there are only 3 attributes after attribute reduction, it will bring us convenience to fault diagnosis and overcome effectively the drawback of support vector machine.

5.3 Multi-classification Based on Support Vector Machine

We can separately construct the fault diagnosis system on the conditions of before and after attribute reduction of the original data. we select one against the rest method in the multi-classes method. In 25 training samples, we separately train and test. Firstly, sample 1-7 as "+1" and the rest samples as "-1"; second, sample 14-18(original data) as "+1" and the rest samples as "-1"; third, sample 19-24(original data) as "+1" and the rest samples as "-1"; the last, sample 31-37(original data) as "+1" and the rest samples as "-1".

Choosing the parameter C=10, kernel function $K(x_i, x) = e^{-\gamma\|x-x_i\|^2}$ and kernel parameter $\gamma = 0.05$, getting decision function (2.15). then testing the 12 testing sets, fault diagnosis results(average accuracy) as Table 3.

Table 3. The fault diagnosis result comparative table about multi-classification, first-rest is the first class versus the rest classes

methed	first-rest	second-rest	third-rest	fourth-rest
Svm(%)	100	100	91.6	100
Rsvm(%)	100	100	100	100

As we can learn that great diagnosis accuracy has been produced only by support vector machine, and only one sample testing error. but fault diagnosis accuracy based on support vector machine and Rough set pre-processing is 100%. At the same time, reducing fault diagnosis system complexity, reducing training time and data storage space. generally speaking, it contributes us to diagnose the fault on time and reduce the cost for machine fault.

6 Conclusions

On the one hand, on the condition of keeping with same diagnosis ability, making great use of the advantage of Rough set theory in pre-processing, eliminating redundant information and reducing the training sample's dimension, a kind of support vector machine fault diagnosis System based on Rough set pre-processing is proposed in the paper. On the other hand, utilizing the advantage of Rough sets theory in acquiring diagnosis rules and combining with support vector machine greater generalization performance, a kind of hybrid fault diagnosis system based on Rs and SVM is proposed too in the paper. The diagnosis of a diesel demonstrated that the solution can reduce the cost and raise the efficiency of diagnosis.

References

1. Kecman,v. Learning and Soft Computing, support vector machine, Neural Networks and Fuzzy Logic Models.The MIT Press, Cambridge, MA (2001)
2. Deng Naiyang, Tian Yingjie. A new method of Data Mining-Support Vector Machine, Science Press (2004)
3. Wang, L.P.(Ed.): Support Vector Machines: Theory and Application. Springer, Berlin Heidelbrg New YorK (2005)
4. Zhang Wenxiu, Wu Weizhi. Rough Set Theory and Application, Science Press (2001)
5. Li Bo, Li Xinjun. A kind of hybrid classification algorithm based on Rough set and Support Vector Machine, Computer Application. 3 (2004) 65-70
6. Renpu Li, Zheng-ou Wang. Mining classification rules using rough sets and neural networks. European Journal of Operational Research. 157 (2004) 439-448
7. Lixiang Shen, Francis E.H. Tay, Liangsheng Qu, Yudi Shen. Fault diagnosis using Rough Sets Theory. Computers in Industry. 43 (2000) 61-72
8. FENG Zhi-peng, DU Jin-lian, SONG Xi-geng, CHI Zhong-xian, GE Yu-lin, SUN Yuming. Fault diagnosis based on integration of roughsets and neural networks.Journal of Dalian University of Technology. 1 (2003) 70-76