FAULT DIAGNOSIS IN ROTATING MACHINERY USING ROUGH SETS AND ROSETTA: EXTENDED ABSTRACT

Torgeir R. Hvidsten, Marte S. Bjanger and Jan Komorowski
Knowledge Systems Group, Department of Information and Computer Science
Norwegian University of Science and Technology
7491 Trondheim, Norway
Phone: +47 73593440, Fax: +47 73594466
e-mail: {torgeihv, marte, janko}@idi.ntnu.no

Maurice F. White and Bai Guanglai
Division of Marine Engineering
Norwegian University of Science and Technology
7034 Trondheim, Norway
e-mail: {Maurice.White, Bai.Guanglai}@imm.ntnu.no

ABSTRACT: The rosetta system, a toolkit for rough set analysis, is being used to classify different fault states in a large diesel engine. The data measured from this engine was collected from six different states, five fault states and the normal state, and from 15 different measurement points. The approach addresses two important questions; How much data is needed to induce an effective model? How can the amount of data in consideration be reduced using knowledge about how the engine works?

KEYWORDS: Fault diagnosis, rough set theory, rosetta, rotating machinery.

INTRODUCTION

Experts play an important role in a machinery environment to ensure both a correct and effective operation of the process. Since these experts can not be present 24 hours a day, it would be highly desirable to support human operators with computer-based systems (White, 1991). Due to non-linearity and high complexity, the problems of using mathematical models for fault diagnosis in machinery are severe. Thus expert systems have been built for this purpose, using mathematical modelling (or inverse modelling) and neural networks (Cholewa, 1993), or both mathematical modelling, fault matrix analysis, machine specific experience and computer simulation (Steinebach, 1990).

Figure 1: The diesel engine considered in this paper

Pawlaks rough set theory and the rosetta system (Øhrn, 1999; Komorowski, 1999; Øhrn, 1998) have also been used for fault diagnosis in rotating machinery. The rosetta system is a toolkit for rough set data analysis and takes a machine learning approach to the problem of fault diagnosis. By inducing rules on the basis of measured data from a machine in different fault states, a classifier is obtained which can diagnose the same machine given a situation where the state of the machine is unknown.

PROBLEM DESCRIPTION

Data for the experiments considered in this paper was collected from the large diesel engine shown in Figure 1. Two

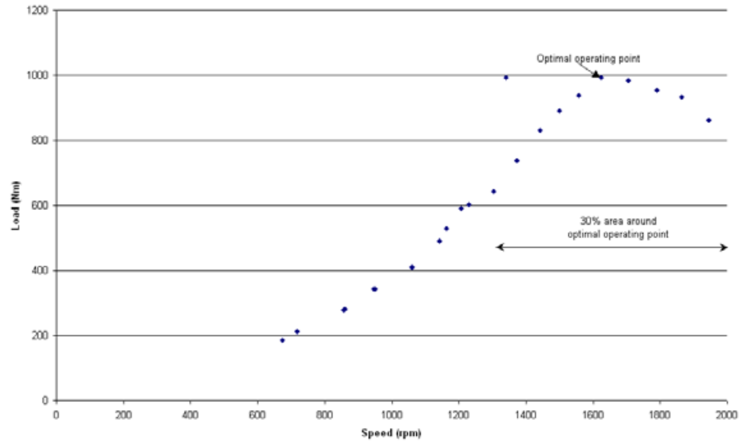

Figure 2: The engine operating in normal condition

important parameters for this machine are speed (rpm) and load (Nm). As speed increases, so does load until a point where it starts decreasing. This point is called the optimal operating point. Figure 2 shows the relation between speed and load for the diesel engine in normal condition. From each point along the speed/load graph, measurements were done from 15 different measurement points (including pressure, temperature etc). The same measurements were then done six times, every time with a different known fault present in the machine. Each set of measurements given a fixed speed, load and fault, are considered to be an object with 16 attributes; 15 conditional attributes, including speed and load, and one decision attribute, the fault. In rough set theory the table consisting of all objects measured is called the decision table or the decision system. The set of objects having the same decision is consequently called a decision class. The decision system used in this report includes the following decision classes (where the number of objects available from each decision class are pointed out in brackets):

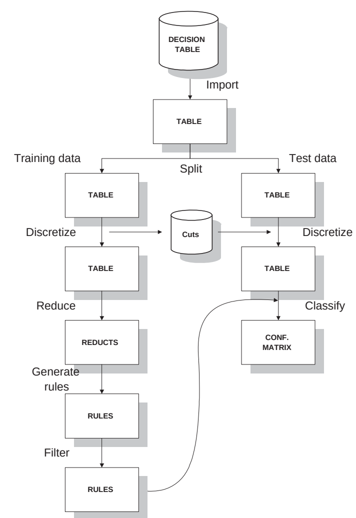

Figure 3: How to use the rough set method and the rosetta system to induce and validate a rule set or a classifier

1. Normal (25 objects).
2. Insufficient air (50 objects).
3. Insufficient air & misfiring in cylinder 4 (100 objects).
4. Insufficient cooling (50 objects).
5. Insufficient cooling & misfiring in cylinder 4 (50 objects).
6. Misfiring in cylinder 4 (100 objects).

When using rosetta as a fault classification tool for rotating machinery, a number of challenges have emerged. One of them is the problem of dealing with the large amount of data collected from different measurement points on the engine. This paper summarise an approach to reduce the amount of data in consideration, using data collected from the large diesel engine as an example. The approach addresses both the problem of how much data is needed to induce a model and the problem of how to handle a situation where different amounts of data exist for different faults.

METHODOLOGY

The methodology used is depicted in Figure 3. The decision table already described is imported into rosetta, and divided into a training set and a test set. The idea behind the method is to use the training set to induce rules, and then use the test set to validate these rules. In general, non-categorical attributes need to be discretized in order to induce effective rules. The cut-off points generated during the discretization step, are produced using the training set. The test set is then discretized using these cuts. Rules are then generated on the basis of the discrete training table using reducts. Reducts are minimal sets of attributes capable of discerning objects with different decision. The resulting confusion matrix from applying the generated rules on the discrete test set, indicates how successfully the rules classified the objects in the test set. The measuring unit used is *accuracy*. Accuracy equals the number of objects correctly classified divided by the number of objects classified altogether.

A number of algorithms are available both for discretization and reduct computation. Since the decision table used in this paper totally consists of continuous attributes, discretization has to be performed. Two different algorithms were used subsequently:

a) Boolean reasoning: This algorithm reduces the search for appropriate cut-off points to finding minimal Boolean expressions.

b) Equal frequency binning: This algorithm dicretize the attribute in question into a given number of intervals such that each interval contains approximately the same number of objects.

Two algorithms for reduct computation where used and compared:

i) Johnson algorithm: This algorithm uses a simple greedy algorithm to compute single reducts only.

ii) Genetic algorithm: This is an implementation of a genetic algorithm for computing minimal hitting sets.

In general, the genetic algorithm computes more reducts than the Johnson algorithm, thus rosetta generates a larger number of rules when using the genetic algorithm compared with using the Johnson algorithm.

More details about the rosetta system and the different algorithms available can be found in Øhrn (1999).

As mentioned, the maximum load and its corresponding speed gives a point called the optimal operating point. In a 30% range above and below this point experts assume linearity (see Figure 2). Thus one way of reducing the amount of data one has to consider is by using objects only from this area. The results from comparing classifiers using those objects, versus using objects from the whole range of speeds, are outlined later.

When using *accuracy* as a measure of quality for the classifier, the distribution of the amount of objects from each decision class should as far as possible reflect the distribution from the real world. However, in our data set the *normal* - class would totally overshadow the *fault* - classes, since an engine can spend years operating normally. With respect to the example data in this paper, each fault is considered to be equally important. Thus an equal amount of objects will be used from each decision class. When too few objects are available from a given decision class, randomly selected objects from this class will be copied until enough objects are obtained. In the same manner objects will be selected randomly from decision classes when too many objects are available. Note, however, that the copying and discharging of objects only will be performed in the training set, never in the test set. Thus this has to be done after the decision table has been divided into a training set and a test set.

RESULTS

Figure 4 shows the results from comparing classifiers based on objects from the area around the optimal operation point, versus using objects from the whole range of speeds. Accuracy is measured for six different sets of objects, each set containing an equal amount of objects from each decision class.

The results shown in Figure 4 emerged from two different algorithms for reduct computation. Each experiment was repeated four times, each time using a different split. Generally speaking, the algorithm which produces a large set of reducts (Genetic alg.) did best using training sets with few objects (5, 10, 20), while the algorithm only finding single reducts (Johnson alg) actually did almost as good for training sets using a large number of objects (30, 40, 50). Figure 3 also indicates that using 30 objects from each decision class seems to be the optimal solution given this experiment.

In addition to the results shown in Figure 4, another experiment was done using objects from the whole range of speeds in the test set and only objects from the area around the optimal operating point in the training set. Only training sets with 30 objects from each decision where considered. Again using four splits, an accuracy of 0.62 using the Johnson algorithm and 0.76 using the genetic algorithm was obtained. This indicates that even when classifying objects from the whole range of speeds, inducing rules using objects from the area around the optimal operating point still gives a rather good result.

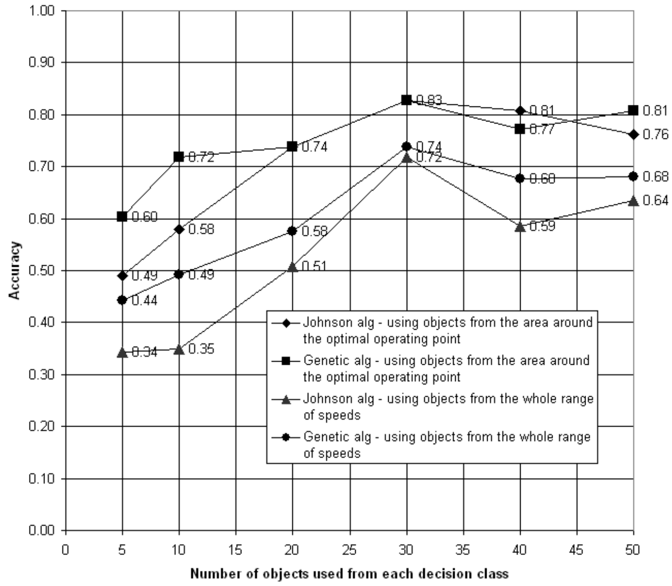

Figure 4: The results from comparing classifiers using objects from the area around the optimal operating point, versus using objects from the whole range of speeds.

Last, one should also note that it is possible to get a significantly better result by not using an equal amount of objects from each decision class. This can be justified noting the fact that some faults are harder to classify than others. Using more objects from these decision classes will make it easier to classify them, and thus give a better classification altogether. By doing so it is possible to classify with an accuracy around 90% for the given data.

ANALYSIS

Given the quality of the data used, experts on the field of fault diagnosis in rotating machinery consider an accuracy of 67% (2/3) as a rather good result. Thus the results outlined in Figure 2 are quite acceptable. As expected, classifiers based on objects from the area around the optimal operating point did considerably better than those using objects from the whole range of speeds. Actually what gave the best results was a training set based on 30 objects from each decision class (a total of 180 objects). This could be an interesting result, showing that using many objects does not necessarily give a good result. However, for the classifiers using objects from the area around the optimal operating point, this could also just indicate that copying the same objects does not give the classifier any new information. Anyway, the results clearly shows that 30 objects from each decision class is enough to get a good result.

Another interesting result concerns the algorithms used. In both cases the Johnson algorithm did nearly as good as the genetic algorithm, using 30 or more objects from each decision class. Since the Johnson algorithm generates only a fraction of the rules the genetic algorithm does, Johnson seems to be the obvious choice when a data mining approach is taken.

FUTURE RESEARCH

There are several considerations which have to be taken into account before any definite conclusion can be drawn. A more thorough testing should be made, using a larger number of different test and training sets. Experiments should also be done using different approaches than copying existing objects when too few objects are available.

In this paper, classifiers have been measured according to accuracy. Another much used measure is area under the ROC (receiver operating characteristic) curve (AUC) (Øhrn, 1999). However, AUC can only directly be used to evaluate binary classifiers. Thus to use AUC with the data considered in this paper, one would have to build a pipeline of seven binary classifiers.

Looking into the future, one can imagine the rough set method, and in particular the rosetta system, being used to design online expert systems for fault diagnosis of rotating machinery partly replacing the role of the expert.

ACKNOWLEDGEMENT

Many thanks to Alexander Øhrn for answering many questions about the rosetta system, and for always taking the time to do so.

REFERENCES

E. Cholewa; M. F. White, 1993, “Inverse modelling in rotordynamics for identification of unbalance distribution, Machine Vibration, Springer Verlag, Vol. 2, No. 3.

J. Komorowski; A. Skowron; A. Øhrn, 1999, “Rosetta and Rough Set Software Systems for Data Mining and Knowledge Discovery”, to appear in *Handbook of Data Mining and Knowledge Discovery*, (W. Klösgen, J. Zytkow, Eds.), Oxford University Press.

A. Øhrn; J. Komorowski; A. Skowron; P. Synak, 1998, “The Design and Implementation of a Knowledge Discovery Toolkit Based on Rough Sets - The Rosetta System”, in: *Rough Sets in Knowledge Discovery*, (A. Skowron, L. Polkowski, Eds.), Springer Verlag.

A. Øhrn, 1999, “ROSETTA, Technical Reference Manual”, http://www.idi.ntnu.no/∼aleks/rosetta.

C. Steinebach; M. F. White, 1990, “VIBEX (Part1) - An intelligent knowledge based system for fault diagnosis in turbo-machinery”, NATO ASI Series, Vol. 174, Vibration and Wear in High Speed Rotating Machinery, Ed. J.M. Montalvao e Silva & F.A. Pina da Silva, Kluwer Academic Publishers, 759 - 772.

M. F. White, 1991, “Expert Systems for Fault Diagnosis of Machinery”, Measurement, Journal of the International Measurement Confederation, Vol9, No4.