10th IMEKO TC10 Conference on
Technical Diagnostics
Budapest, HUNGARY, 2005, June 09-10

IDENTIFICATION OF RELEVANT SIGNAL FEATURES

Wojciech Cholewa

Faculty of Mechanical Engineering, Silesian University of Technology, Gliwice, Poland

Abstract - The paper deals with design of diagnostic classifiers. The main goal is to illustrate an original way of identification of useful signal features on the basis of learning data prepared as a set of examples. One indicated some possibilities of application of a criterion based on the expectation that results of unsupervised clustering in a new limited space should be compatible with results of classification of learning data.

Keywords: diagnostic classifier, feature selection

1. INTRODUCTION

The goal of technical diagnostics is to recognize the state of an investigated object. Operation of numerous technical objects, particularly these ones whose elements can undergo wear, are accompanied by different residual processes (e.g. noise or vibration). These processes are sources of useful signals, which carry information about the state of an object. Accessible contemporary measuring and signal processing techniques cause that possibilities of construction of monitoring and diagnosing sets with the use of expert systems are repeatedly encountered. Taking into account the essence of their operation, expert systems, which aid monitoring and diagnosing technical objects, can be considered as classifiers of diagnostic signals. The basis for recognition of the state of an investigated technical object is to determine a class, to which observed signals belong. In comparison to simple classifiers, expert systems are characterized by additional possibilities to expound and interpret results of their operation.

2. DIAGNOSTIC MODELS

It is required that when a diagnostic expert system is built some proper knowledge related to the operation of an investigated object should be accessible. Such knowledge enables to construct a classifier of diagnostic signals. Diagnostic knowledge can be acquired from different sources including:

- active diagnostic experiments, where during the experiment the state of an object is allowed to be changed,
- passive diagnostic experiments, which consist in the observation of an object, whose state varies independently of the experiment,
- simulation research, which reduce a need for application of real objects.

Recently, the rapid development of methods of numerical modelling of technical objects and processes related to the objects, can be observed [8]. Results of such modelling are cause and effect models. Assuming a particular technical state, determined structural features and conditions of operation of an object, one can state that these models enable us to determine results of operation of the object (e.g. vibrations). Important techniques for accurate tuning of numerical models have been developed [5]. They consist in fitting these models into results of observations of real object operation. Appropriate quality of the tuning makes it possible to apply the models as sources of learning data in a process of diagnostic knowledge acquisition [7]. Methods, which allow inverting numerical models of objects were developed, too [1]. Results of these procedures are inverse models, which of course are not causal. They are so-called diagnostic models and enable us to indicate a class of the technical state as probable causes of observed symptoms that are results of object operation.

3. MULTI-LAYER DIAGNOSTIC MODELS

An advantage resulting from the application of tuned numerical models of objects is the possibility to generate numerous sets of training data, which make it possible to acquire diagnostic knowledge effectively. Unfortunately, the same models are sources of inconveniences related to the facts that:

- excessive number of features of diagnostic signals can be estimated,
- excessive number of state features, which are the basis of identification of a state class, can be determined.

Too high number of considered features, i.e. an increase of dimensionality of an undertaken task, forces us to redesign a lot of commonly applied algorithms. It also makes it difficult to verify correctness of obtained results. In order to limit enumerated inconveniences, decomposition of one selected global model into a set of local models may be carried out. The goal of this operation is to reduce a number of input and output features taken into account. The operation cannot be considered only as a spatial decomposition, which consists in division of an object into subassemblies and elements. Significant reduction of a number of considered features is possible only if decomposition of a global model into local models consists in uncoupling of local descriptions making them independent. Convenient way of disconnection of local models is to introduce additional features playing a role of instrumental variables. The variables can be features of virtual or hidden interactions. Their reference to observed real interactions is not required to be assumed.

Some additional features of a local model can be estimated by a method of trails and errors, e.g. by means of methods of multidimensional scaling applied to features related to the other local models. However, search for such the features is a difficult task, which does not guarantee satisfactory results. It manifests itself when proper domain knowledge and particularly knowledge related to object operation are not used in the search process. Local models, which are results of decomposition of a global model become elements of a mixed multi-stage model. The first stage of this model consists of local models and their inputs and outputs. Inputs include features of diagnostic signals, whereas outputs represent additional features. These features exemplify the second stage inputs of a mixed model.

4. USEFUL FEATURES OF SIGNALS

Side-effect of diagnostic knowledge acquisition on the basis of simulation research carried out with the use of numerical models is necessity to estimate very huge sets of signal features. Expensive modification of measuring devices and repetition of experiments is not required in order to consider a new feature. The only operation to be performed is to complete a program code, which estimates features on the basis of their recorded courses. Additionally it is possible to show that in general, excessive increase of a number of considered features does not ensure quality improvement of a classifier estimated on the basis of these features. It can lead to quality deterioration. This fact, indicated sometimes as contradictory to intuition, can be justified by disadvantageous influence of useless features, i.e. features that are not related to considered classes. Such the features are treated as noise.

It entails that reduction of a number of estimated or ready for estimation features is required to be performed. Selection is applied to these features, which can be recognized as useful in order to determine the state of an object. These features are also called relevant ones. The problem defined in such the way has been well known for many years. Numerous publications indicating different methods of estimations of useful features are accessible. Commonly applied methods for such limitation of a number of considered features are methods of multidimensional scaling [2]. The method enable us:

- to select a given number of features from a set of existing values, or
- to estimate a given number of new features as a result of transformation (most often linear one) of existing features.

Properties of the limited set of features, which are regarded as useful ones, are dependent from application of a selection criterion. Basic methods (e.g. principle components method), are based on the assumption that data dispersion and its main directions should be preserved. Similarly, (e.g. method of the main factors) relationships appearing between values of features should be retained.

These methods are perfectly fit for limitation of a number of features taken into account in grouping tasks carried out without a teacher. In these task, assumed division into classes is unknown, thus requirements of a given preservation of data distribution is justified. However, in the considered tasks dealing with building a local classifier some different requirements appear. Pattern division of learning data into classes is unknown. This fact should be taken into account by optimization criteria for features selection. It is possible to apply the following criteria:

- different criteria basing on minimized mean values of measures of internal distribution in classes, and maximized mean values of measures of exterior distribution between classes,
- criteria basing on resulting effectiveness of classifier estimated for a limited (as a result of the selection) set of features,
- criteria basing on compatibility between results of clustering (without a teacher) in a new space determined by means of limited set of features and results of pattern classification of learning data.

In the case of a group of criteria considering internal and extended distribution of classes, numerous effective algorithms determining their applications are known. These criteria are particularly recommended when some classes in the form of coherent concentrations of data are expected. Other criteria require iterative search for a solution.

It seems that a sensible compromise between expected quality of selection and limitations related to time of searching for a solution is the application of so-called deepening with turns back procedures. It can be carried out in repeating series in such the way that at first m consecutive features, which candidate to a selection is searched, and then n consecutive features to be omitted, where $n<m$, is estimated. These criteria can be also effectively applied in evolutionary optimization methods applied to a set of considered features.

Particular attention should be paid for the last enumerated group of criteria. It should be stressed that this method assumes that selection of a set of features, which makes it possible to reconstruct a division of data according to the pattern learning division is possible to be performed. Determination of such features allows assuming that a classifier estimated on the basis of the features is characterized by proper generalizing properties.

5. SELECTION OF RELEVANT FEATURES

The issue of selection of relevant signal features may be discussed on the basis of an object, whose scheme was shown in Fig. 1. The problem arises during identification of diagnostic models on the basis of simulation research. A state of the object is described by a vector s of values of a state features, whereas, its operation conditions are characterized by a vector x of parameters. A result of object operation are interactions y_i, which are dependent on the state s and conditions x.

Basing on the acquired knowledge related to the object, one can define a transformations v_i, which convert operation conditions x and states s into interactions y_i. It is assumed that operation conditions x are observable and states s cannot be observed. Additionally, one also assumes the ability to define an inverse transformation V^{1} (Fig. 2), which makes it possible to reconstruct the states s on the basis of known interactions y_i and operation conditions x.

Then, we assume that interactions y_i can be directly observed neither during research nor object maintenance. Their identification is possible in simulation experiments only. Considering real conditions, information on these interactions can be indirectly acquired only on the basis of features fea(z_i) of observed signals z_i. Transformations w_i, which convert interactions y_i into signals z_i are known only as approximate operations and their possible inversion does not ensure required accuracy. It makes identification of a state s on the basis of observed features fea(z_i) difficult.

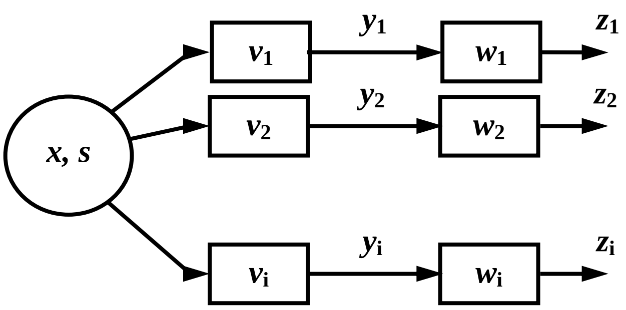

Fig. 1. Transformations v_i and w_i

However, such the operation is possible be means of inverse models N. The models realize searched transformations. They can be identified as a result of training with the use of learning data. A problem to be solved is proper selection of features of signals fea(z_i), which should be taken into account. On the basis of numerous experiments, one can state that direct attempts to optimize a set of such features during training of a model N results in high calculation costs. In most cases, discussed optimization requires restricted, narrow scope.

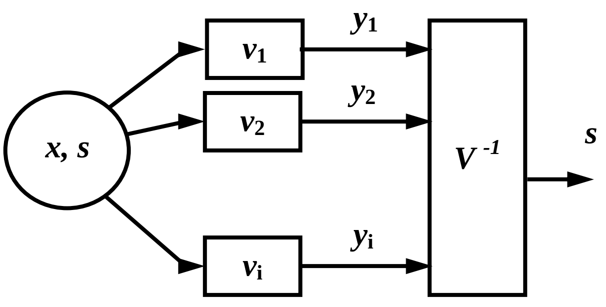

Fig. 2. Inverse transformation V^{-1}

A need of determination of an effective method for searching of relevant features, which provide a proper quality of inference process, appears. The selection of such features can be preformed on the basis of learning data z_i as well as on y_i. A basis of the suggested solution is an assumption that considered representations of z_i and y_i should be equivalent. It means that clustering of values of features fea(y_i) as well as clustering of values of features fea(z_i) will lead to similar or equivalent classes.

Selection of relevant features can be done be means of an appropriate searching. Depending on a task dimension one can apply different methods. Examples are exhaustive searching and evolutionary algorithms. The most important element of all considered searching algorithms is a measure of structural similarity between sets of values of features fea(y_i) and fea(z_i), used for estimation of the quality of a solution.

As criterion functions, one can apply similarity measures between dendrograms, estimated independently for data fea(y_i) and fea(z_i). An example of such the measure is a norm *ccc* (Cophenetic Correlation Coefficient) [6], [9]. The *ccc* measures make it possible to obtain satisfactory results of searching in the case of significantly separated classes appearing in sets of values of fea(y_i) and fea(z_i).

An another way of defining of a criterion function is to consider the assumption that the transformation w_i should preserve an order of distances between elements after converting space fea(y_i) into fea(z_i). It entails that elements placed near to each other in the set fea(y_i) should be also neighbors in the set fea(z_i). Analogous property should be valid for far-distant elements, too. Measures of such ordering can be rank correlations as well as a measure *kk* estimated as a result of the following operations:

- all distances between elements in sets y_i and z_i are generated,
- values of distances are replaced with their ranks,
- a distance *kk* between sets of ranks is estimated.

The result of described selection is a limited set of relevant signal features, which can be basis for the model N.

6. EXAMPLE

Let us consider a limited set of states of a turbo generator rotor supported by means of seven bearings.

The state of the rotor is described by horizontal and vertical thermal and assembly displacements of bearing housings in comparison to nominal placements. For seven bearings this state is described by 7*2=14 features (displacements). A numerical model of vibration of a set is accessible [4].

The results of operation of the rotor are among other things residual processes appearing e.g. in the form of vibration in a bearing housing. A change of any enumerated displacements of bearings influences on a change of the shape of vibration of the whole set.

A mixed multi-layer diagnostic model was built for the considered set. The first layer of this model consists of seven local models. Decomposition of the global model was performed under the assumption that local models represent consecutive bearings.

In order to disconnect local models, state features, which are displacements of supports, are replaced with additional features. The roles of additional features are played by values of horizontal and vertical forces in bearings. These forces are not observed directly in the case of a real object. They can be estimated by means of an accessible numerical model [4]. It must be noticed that the magnitude of forces in a selected bearing depends on displacement of all bearings, thus it is a direct carrier of information about all displacements.

Inputs to local models are values of features of diagnostic signals. Selection of considered features was performed on the basis of the criterion of compatibility between results of grouping in a new selected space and results of pattern classification. Each of seven local models is a classifier estimated on the basis of learning data. The kind of a classifier may be always a subject of separated optimization.

Local diagnostic models in the general case do not enable us to estimate state features accurately. The reason of observed identical values of symptoms can be often different states of the object. The models allow only indicating of possible state classes, where considered classes are the classes representing values of reactions in supports (e.g. very large forces directed to the right, medium forces directed upwards). Outputs of local models in the form of classes representing reactions connected with seven bearings are inputs to the second layer of the mixed model.

A task consisting in defining the second layer as a linear model transforming reactions of bearing housing into their displacements by means of given matrices of influence numbers was undertaken. However, non-linearity appearing in the considered object causes that the application of such the linear model is reduced. As a solution to this problem one introduces belief nets [3] describing relationships between classes of forces (reactions) and displacements.

7. RESULTS

Some satisfactory results of application of the described method of identification of useful features were obtained. Indicated criteria may be expected to ensure the possibility of effective reconstruction of state classes on the basis of values of useful features of signals. It must be stressed that discussed criteria are only suggested as criteria useful during identification of relevant features. Criteria, which are the basis of construction of classifiers exemplifying local models, should be selected independently. The obtained results make it possible to recommend presented criterion and the method to be applied to practical problems.

REFERENCES

[1] W. Cholewa, M. F. White, "Inverse modeling in rotordynamics for identification of unbalance distribution", Machine Vibration, Vol. 2, pp. 157-167, (1993).

[2] R. O. Duda, P. E. Hart, D. G. Stork, "Pattern Classification", John Wiley & Sons, 2001.

[3] V. J. Jensen, "Bayesian Networks and Decision Graphs", Springer-Verlag, (2002).

[4] J. Kiciński (Ed.), "Modeling and diagnostics of mechanical, aerodynamical and magnetic interactions" (in Polish), Wyd. IMP PAN, Gdańsk, (2005).

[5] J. Korbicz, J. M. Kościelny, Z. Kowalczuk, W. Cholewa (Eds.), "Fault Diagnosis. Models, Artificial Intelligence, Applications", Springer-Verlag, (2004).

[6] F.-J. Lapointe, P. Legendre, "Comparison Tests for Dendrograms: A Comparative Evaluation", Journal of Classification, 1, 265-282 (1995).

[7] W. Moczulski, "Problems of declarative and procedural knowledge acquisition for machinery diagnostics", In: Methods of Artificial Intelligence (T. Burczyński, W. Cholewa - Eds.) - Computer Assisted Mechanics and Engineering Sciences, Vol. 9, No. 1, pp. 71-86 (2002).

[8] H. G. Natke, C. Cempel, "Model-Aided Diagnosis of Mechanical Systems: Fundamentals, Detection, Localization, Assessment", Springer Verlag, (1997).

[9] J. Podani, "Simulation of Random Dendrograms and Comparison Tests: Some Comments", Journal of Classification, 17, 123-142 (2000).

AUTHOR: Wojciech CHOLEWA, Faculty of Mechanical Engineering, Silesian University of Technology, Konarskiego 18a, 44-100 Gliwice, Poland, E-mail: wch@polsl.pl