Computer Assisted Mechanics and Engineering Sciences, 9: 71–86, 2002.

Problems of declarative and procedural knowledge acquisition for machinery diagnostics

Wojciech Moczulski
Silesian University of Technology, Department of Fundamentals of Machinery Design
ul. Konarskiego 18a, 44-100 Gliwice, Poland

(Received September 7, 2001)

The paper deals with selected problems of knowledge acquisition for intelligent information systems that may be applied for aiding technical diagnostics of machinery and equipment. Two main kinds of knowledge are discussed, i.e. declarative and procedural knowledge. Some methods of declarative knowledge acquisition from domain experts and from databases are presented, the latter being divided into machine learning methods and knowledge discovery ones. Examples of declarative knowledge acquisition and discovery from databases are shown. Moreover, an example of procedural knowledge acquisition from a domain expert is presented. The paper concludes with new issues of knowledge acquisition methodology.

Keywords: intelligent information systems, knowledge base, procedural knowledge, declarative knowledge, knowledge acquisition, knowledge discovery

1. Introduction

Needs of modern society systematically grow. To satisfy these needs by limited resources accessible to the societies, even more and more complex technical means are being built. They have higher efficiency, greater power, and ensure better quality of their product. These means, and especially machinery and equipment, becomes even more complex. Thus, persons and organizations whose are involved in designing, production and exploitation of these technical means, are faced with growing requirements concerning operational reliability of machinery and equipment, its maintability, serviceability and user-friendliness. A person who is active in one of the above mentioned fields needs sufficient *knowledge* and *skill*, that are possessed by *domain experts*, usually very few. They acquire their knowledge and experience during long-term professional activity, by observations or by studies of professional literature.

It is worth drawing our attention to many important arguments, as:

1. **Knowledge is not unambiguous.** It can be also incomplete, or even contradictory. There exist divergent opinions of experts on the same subject matter. The same symptoms are related to different faults.

2. **Required promptness of operation.** A monitoring system of a critical machine or system (as a nuclear power station or chemical plant) has to react immediately on event of appearance of early symptoms of catastrophic failure.

3. **Accessibility of an expert *on-situ*.** An expert not always is accessible *on-situ*. Moreover, he/she usually is not going to share his/her knowledge and experience with other people.

4. **Knowledge is valuable.** Even more, if not transferred to other people or memory, it may be lost. J. Pokojski [27] draws our attention to the fact that organizations concerned on designing and exploitation of machinery start to treat intellectual values of employers as a very precious

property. Design and exploitation knowledge is collected by subsequent generations of employers and is often not acquired at all, therefore this knowledge is lost if the experienced workers leave the company. However, this knowledge and experience is necessary for efficient initiation of new workers.

5. **We are subject to stream of messages.** They arrive in the form of innumerable quantity of data. It is almost impossible for us to pick up from this stream such messages that carry essential information, or to identify important regularities, which may be considered as knowledge of both qualitative and quantitative form.

All arguments quoted above point out at the need for replacement of a human expert by specialized *intelligent information systems* as e.g. intelligent databases, expert systems, intelligent sensors and so on. Their basic elements are *knowledge bases*, in which such knowledge is stored that is needed for aiding activities in some (usually narrow) problem domain. Knowledge bases are directly connected with the subject matter of the paper, which concerns problems of *knowledge acquisition on technical diagnostics of machinery*. Such knowledge may be applied in intelligent information systems that support activities of exploitators and maintenance staff (including diagnosticians) of machinery and equipment.

This paper is based on the works [20, 25] and is composed as follows. In Section 2, some problems of knowledge in technical diagnostics are briefly addressed and two main knowledge kinds as declarative and procedural knowledge are introduced. The research problem itself is presented in Section 3. In Section 4, some more frequently used methods of declarative and procedural knowledge acquisition are described. Moreover, Section 5 contains examples of knowledge acquisition from databases and from experts. The paper concludes with future work and recapitulation.

2. Knowledge in technical diagnostics

Knowledge in relation to a human being is everything that some given person knows [5]. Knowledge in the given technical domain concerns [16] objects (as machinery, equipment, and also their parts) and classes of objects belonging to this domain, taxonomies of classes of objects, properties of objects and classes of objects, relations between objects and their classes. This knowledge includes also skills, understanding of general laws, procedures of proceeding etc.

Knowledge is acquired especially by learning of the given person. It is purposeful to consider also collected experience and skill. They are usually acquired by an individual himself/herself and are results of long-term activity of the expert in his/her domain. In the following the term "knowledge" will denote both the actual knowledge and practical skill of the expert.

Let's notice that knowledge understand in such a way concerns *facts* and *processes.* Taking this reason and others else into account, it is purposeful to distinguish *declarative knowledge* and *procedural knowledge.*

2.1. Declarative knowledge

Knowledge concerning facts, objects, relations between objects, classes of objects, features of objects etc. is represented in declarative form and in the following will be referred to as *declarative knowledge.* Most work done so far, also in our Department, concerned methods of representation and acquisition of declarative knowledge. More important sources of knowledge in technical diagnostics are shown in Fig. 1.

Declarative knowledge can be acquired either from experts or from databases. Experts may take part directly in the knowledge acquisition process, or they can be authors of publications, handbooks and other technical documentation that may be carriers of information to be next acquired in a separate process, in which their participation is unnecessary.

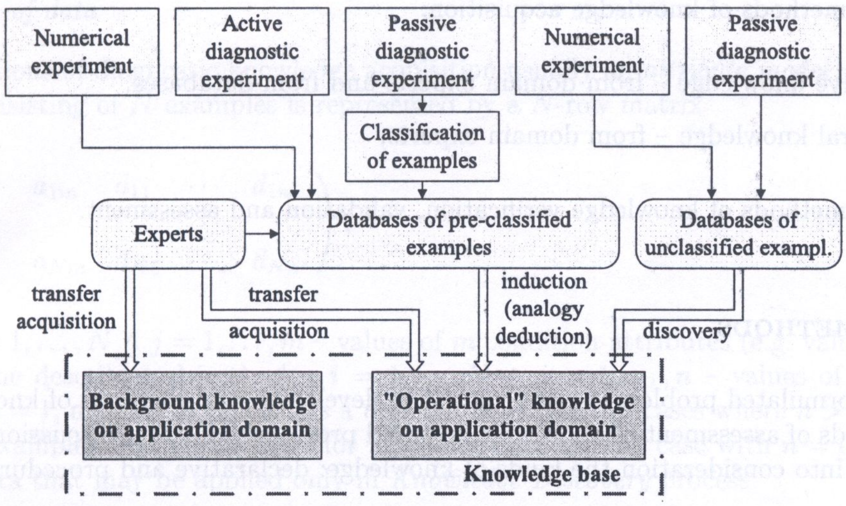

Fig. 1. Declarative diagnostic knowledge and its sources [19]

Databases are other important source of declarative knowledge. In this case knowledge may be acquired in automated way using either *Machine Learning* methods (if records in the database are pre-classified) or *methods of discovering* of qualitative and quantitative (functional) relationships.

2.2. Procedural knowledge

Declarative knowledge discussed so far is inadequate for aiding such processes as diagnostic examination of a machine or equipment. In this case *procedural knowledge* is required. It may be represented by procedures. In our research such knowledge is acquired from domain experts who take part in the knowledge acquisition process directly or can be authors of publications that are subject to analysis in order to extract procedural knowledge [33].

3. Problem description

A result of the process of knowledge acquisition is a *knowledge base* corresponding to some narrow application domain. From accessible publications concerning acquisition of diagnostic knowledge one can conclude that:

- the most common situation is that knowledge is acquired from domain experts and that an intermediary person as a *knowledge engineer* often takes part in this process;
- if knowledge is acquired using automated methods, *machine learning* applications are dominating;
- domain knowledge of an expert is used non-intensively;
- there is no generally acknowledged methodology of acquisition of diagnostic knowledge.

The author identified then the need to work out a *methodology of knowledge acquisition on technical diagnostics*, which would take into consideration all kinds of sources of diagnostic knowledge, both declarative and procedural one. This methodology includes:

1. selection of methods of data and knowledge representation;

2. selection of methods of knowledge acquisition:
   - declarative knowledge – from domain experts and from databases,
   - procedural knowledge – from domain experts;
3. selection of methods of knowledge verification, validation and assessment.

4. APPLIED METHODS

To solve such formulated problem it was required to develop several methods of knowledge acquisition and methods of assessment of knowledge acquired previously. Further discussion will be carried on with taking into consideration the kinds of knowledge: declarative and procedural one.

4.1. Methods of acquisition of declarative knowledge

Recalling main knowledge sources we will discuss:

- *methods of knowledge acquisition from domain experts*, and
- *methods of knowledge acquisition from databases.*

4.1.1. Methods of acquisition of declarative knowledge from experts

Experts are fundamental knowledge sources, therefore they usually cannot be omitted in the knowledge acquisition process, or their exclusion from this process is at least inadvisable. Experts provide knowledge bases with *background knowledge on the application domain.* Two kinds of knowledge acquisition methods are possible, namely with and without participation of knowledge engineer.

The method of knowledge acquisition from domain expert(s) with participation of a knowledge engineer was the most early used [1]. However, it has many disadvantages, one of them (very common) connected with problems with understanding the expert by the knowledge engineer, who is not an expert in the expert's domain. Other cause of problems is connected with the need of interpretation by the knowledge engineer the statements acquired from the expert, and then with representation of this new knowledge in the knowledge base.

In our Department we focused our attention on methods of knowledge acquisition from domain experts *without participation of a knowledge engineer* [16, 33]. Our attempt depends on making some special aiding means (usually software) available to the expert, in order to enable him/her to represent his/her knowledge in unaided manner. Hence, the knowledge engineer as intermediary in knowledge acquisition process may be eliminated at least from introductory stages of the process.

4.1.2. Metods of acquisition of declarative knowledge from databases

Acquisition of diagnostic knowledge from domain experts is usually less effective. For example, if rules are acquired, an expert is able to formulate only a dozen or few dozens of rules during one session of knowledge acquisition. For that reason, *methods of knowledge acquisition from databases* are even more frequently applied, which are far more efficient than methods of knowledge acquisition from human experts [15].

Representation of data

In real applications of diagnostic knowledge acquisition usually an *attribute model* is applied, where the dataset consisting of N examples is represented by a N-row matrix

$$\begin{pmatrix} a_{11} & \cdots & a_{1m} & d_{11} & \cdots & d_{1n} \\ & \cdots & & & \cdots & \\ a_{N1} & \cdots & a_{Nm} & d_{N1} & \cdots & d_{Nn} \end{pmatrix}, \tag{1}$$

where: a_{ij} , $i = 1, \ldots, N \wedge j = 1, \ldots, m$ – values of m condition attributes (e.g. values of symptoms observed for the described object), d_{ij} , $i = 1, \ldots, N \wedge j = 1, \ldots, n$ – values of $n \geq 0$ decision attributes. If $n = 1$, matrix (1) represents a *decision table* [26]. All cases where $n > 0$ correspond to pre-classified examples that are suitable for *machine learning*. The case with $n = 0$ corresponds to unclassified data that may be applied only in *Knowledge Discovery* process.

Data discretization and selection of attributes

Two different kinds of operations are required in order to better prepare data for knowledge extraction: *data discretization* and *selection of features*. Both of them are discussed by K. Ciupke [4].

The need of data discretization is triggered by fact that in practice values of condition attributes usually are quantitative ones. Decision attributes may also have such values. Most algorithms of Machine Learning require that data be represented qualitatively. Hence, a proper discretization of attributes is necessary, that is connected with a selection of q cutting points v_i , $i = 1, \ldots, q$. They may be determined either in a *supervised manner* if a domain expert takes part in their determination, or in an *unsupervised manner* if their selection is *data-driven* [6]. Moreover, cutting points may be determined as *absolute* or *relative* ones [19]. First solution is possible if some limit values of attributes are known, as permissible vibration level or allowable imbalance of a rotating part. If these limit values are unknown or undefined, cutting points may be determined with respect to some basic condition of the machine or equipment (as e.g. vibration levels of machine parts after commissioning the machine). It is worth to stress that by selection of values v_i we shall avoid excessive fit of these cutting points to data to be discretized. Selection of cutting points is a separate partial task, which may be a subject of optimization [18].

The problems of data discretization are very complex and are dealt with in [4, 6, 8, 10–12].

Other important task is connected with *selection of attributes*. Classical attempt to data analysis requires that attributes be uncorrelated, since correlated attributes carry the same information. Hence, one is going to select some subset of attributes (possibly not numerous) that allows correct classification of examples. Several algorithms of selection of attributes are discussed in [4]. The most interesting of them is that based on *rough sets theory*, which consist in determining of an intersection of all *minimal reducts* of a given decision table [26, 30].

In the author's opinion, however, the important case of dependent attributes is worth brief discussion [25]. Such database is in some sense redundant. This redundancy is often faced with in physical systems, where data are collected using multi-sensor measuring systems. Let's focus our attention on dependent attributes, that describe motion of selected points of the same shaft of a machine. Values of these attributes are estimated from vibration and displacement signals that may be partly coupled (by the same element and/or process). Because this coupling between vibrations need not to be linear, the signals need not to be correlated. Hence, information carried by signals acquired from different sensors is in some sense complementary, and may also be repeated. In most technical applications nature of this coupling may be non-linear. Since coupling between vibrations and, as result of that, between signals observed by individual sensors, is likely, some redundancy in the dataset may exist. However, redundancy in the data discussed so far may yield reliability and better quality of predictions [25].

Data sources

Data used in knowledge acquisition process may origin from different sources, as: *experts*, *observations* (done within the confines of generally understand experiments, both passive and active) and *numerical experiments*, while:

1. unclassified data may come in particular from observations and passive diagnostic experiments, as well as numerical ones,

2. classified data (examples) may be collected from domain experts, from observations, active and passive experiments and also numerical experiments.

Machine learning induction methods

Diagnostic knowledge is acquired inductively using *Machine Learning* (ML) methods [14, 15, 26, 28]. The author in his research applied *selective induction* of rules by generation of covers [15]. Moreover, an *induction of decision trees* was used, where optimal tree was selected with the use of the *criterion of maximum information gain* caused by application of new attributes [28, 29].

In near future we are going to apply *constructive induction* that depends on various transformations of representation space and creation of new attributes [15]. New attributes may be created in a supervised way (basing on experts' knowledge, who in their diagnostic reasoning apply *complex and multidimensional attributes* as simultaneous appearance of a specific group of attributes with their characteristic values, related between each other [7]). Constructive induction may also be applied in unsupervised manner, where some transformations, often of logical character, are applied automatically, as it is used in the program *AQ17-HCI* [15] or discovery system *49er/Bacon-3* [35].

Classifiers of states. Knowledge base content may be applied for classification of cases represented by values of attributes. Hence, this content determines a kind of classifier. It is suitable to distinguish two general types of classifiers: *binary classifiers* and *multi-class classifiers* [2]. The binary classifier is used for recognition of $K = 2$ states, while multi-class one for recognition of $K > 2$ states.

It may be interesting to find a criterion, which kind of classifier is suitable for a given diagnostic problem. If elementary states are taken into account and number of classes (corresponding to these elementary states) $K > 2$, a multi-class classifier is recommended. This classifier may also be applied in tasks where complex states are encountered. A set of classes may then be defined, each of them corresponding to considered complex state. Such a solution requires that new classes be added when the knowledge base will be extended in order to cover new (possibly complex) states. Usually this means that *renewed knowledge acquisition* should be carried out from the set of examples supplemented by new examples corresponding to added states. Other possibility is the application of an *incremental learning* [15].

Other possibility of diagnosing complex states, suggested by W. Cholewa[1], consists in application of a family of binary classifiers $\mathcal{C}_k$, $k = 1, \ldots, K$, each of them corresponding to single elementary state. Each classifier is trained on the same set of learning examples E partitioned into two subsets,

$$E = E_k^+ \cup E_k^- , \tag{2}$$

where E_k^+ denotes *positive* examples of the k-th state (i.e. such that the k-th fault is encountered) and E_k^- negative ones (the k-th fault is not perceived). Although such an attempt seems very elegant, no interesting results have been obtained in our introductory research [22].

Taking this into consideration, the author has introduced suitable hierarchical classifiers [16], which make possible sequential recognition of complex and elementary (sub)states that co-appear for the given complex state. Such hierarchical classifier is a special group of multi-class classifiers. Its structure may be represented by an acyclic graph or even tree, which corresponds to an expanded

[1] Private message

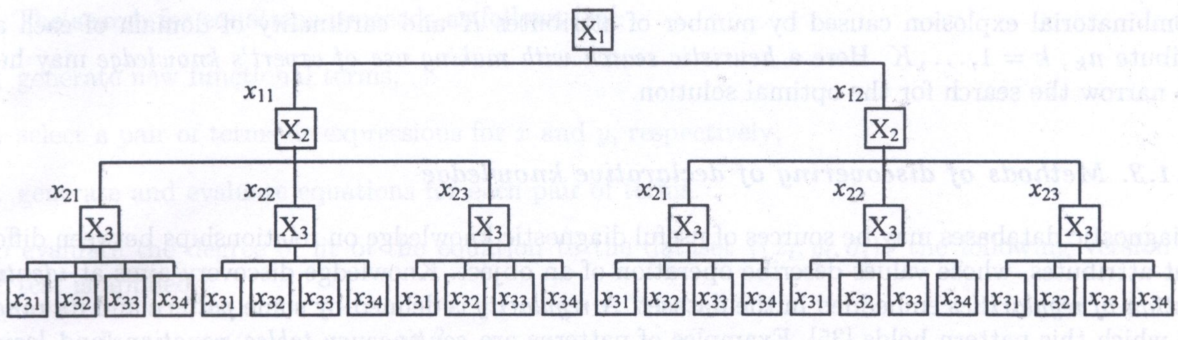

Fig. 2. Exemplary states tree (description in text) [17]

states tree as the one shown in Fig. 2. This tree corresponds to an example where an object is classified using values of its three attributes X_1, X_2, X_3 that may take 2, 3 and 4 values, respectively. A share of knowledge base (e.g. a set of rules) that makes possible recognition of the given state is attached to each node of the tree. Such classifiers are applied sequentially according to the sequence defined by the structure of the states tree (starting from its root).

Classifier's performance. Performance of a classifier is usually evaluated using *wrapper approach* that depends on application of the classifier under test to classification of *test examples*. Several techniques are applied, as *Leave-one-out*, *Random subsampling* or *k-fold* (they are often applied by the author's group in the research – see [4, 13, 16]). All these techniques depend on dividing the accessible set of examples $E = E^l \cup E^t$ into two subsets: *learning examples* E^l and *testing examples* E^t that are *unseen* (not used for learning of the classifier).

Performance η_{ov} of a classifier may be evaluated from the formula,

$$\eta_{\mathrm{ov}} = 1 - \epsilon_{\mathrm{ov}}\,, \tag{3}$$

where ϵ_{ov} is the *empirical overall error rate* defined as

$$\epsilon_{\mathrm{ov}} = \frac{n_{\mathrm{err}}}{\mathrm{card}(E^t)} \tag{4}$$

and n_{err} corresponds to number of test examples incorrectly classified. Only classifiers with high performance measure are accepted, the limit acceptance value being dependent from the number of classes K to be recognized. For strongly unbalanced distribution of set of examples among classes some *weighted error rates* may be applied [16].

Criteria of optimization of states tree structure. The states tree may represent diagnostic knowledge on the given class of objects. This knowledge may be acquired from domain experts or derived from examples. In the second case examples may be clustered with respect to similarity of attribute values. Next, results of clustering may be presented to the expert who may assign or define respective states for individual subsets (clusters) of examples.

There are numerous different structures of a states tree, hence optimal selection of one structure with respect to some established optimization criterion is needed. The most obvious selection criterion is the *minimum classification error rate*. This error rate is evaluated for each node of the tree. The overall error rate and partial error rates for a given branch of the tree may be estimated by consideration of dependent events and calculation of corresponding conditional probabilities. The best optimization criterion would be *minimum of risk of erroneous diagnostic decision* (e.g. with taking into consideration health and safety risk of human beings and environment), however we do not have experience of the right method of risk calculations until now.

A separate problem consists in application of respective search methods in order to find the optimal states tree. Cardinality of the set of all possible structures of states trees is subject to combinatorial explosion caused by number of attributes K and cardinality of domain of each attribute n_k , $k = 1, \ldots, K$. Here a ***heuristic search with making use of expert's knowledge*** may help to narrow the search for the optimal solution.

4.1.3. Methods of discovering of declarative knowledge

Diagnostic databases may be sources of useful diagnostic knowledge on relationships between different attributes, whose values describe operation of an object. Knowledge discovery aims at *identification of regularities* embedded in the dataset. A *regularity* is defined by some *pattern* and its *range*, in which this pattern holds [35]. Examples of patterns are *contingency tables*, *equations* and *logical equivalences*. The range of regularity is defined as a subset of data that satisfy complex condition, which is conjunction of elementary conditions as inequalities.

The research on application of *Knowledge Discovery in Databases* (KDD) methods to acquisition of diagnostic knowledge has been initiated by the author with J.M. Żytkow [24].

We stated in [24], that equations of general type

$$X = f(Y, U), \tag{5}$$

where: Y – diagnostic symptoms, X – values of attributes of technical state, and U – values of attributes that characterize operating conditions of the object, are the best tool for diagnostic reasoning. However, results of measurements and observations are corrupted by noise, are incomplete, approximate or fuzzy, so that datasets contained in diagnostic databases carry incomplete information on quantitative dependencies between attributes. Therefore, contingency tables are applied in the first stage of KDD process, when qualitative relationships are searched for [35].

Discovered patterns are evaluated by *significance measure* Q defined as probability of random generation of such a pattern for attributes that are mutually independent. It has been stated empirically [34] that very small values $Q < 10^{-5}$ should be applied, which means that there is very small probability of random arising of such patterns. The *prediction power* of the given contingency table is evaluated by means of Cramer's V measure,

$$V = \sqrt{\frac{\chi^2}{N \cdot \min\{(M_{\text{row}} - 1), (M_{\text{col}} - 1)\}}} \in [0, 1], \tag{6}$$

where values of the statistics χ^2 (for the empirical frequencies A_{ij} compared with frequencies E_{ij} expected if the null hypothesis of lack of any dependence between two considered attributes is satisfied) are defined as

$$\chi^2 = \sum_{i,j} \frac{(A_{ij} - E_{ij})^2}{E_{ij}}, \tag{7}$$

and N is number of records in the data table, M_{row}, M_{col} are numbers of rows and columns in the contingency table, respectively. The greater value of measure V, the more unique predictions can be obtained basing on this contingency table. An essential feature of the measure V is its independence from dimensions of the contingency table and from number of records. If $V > 0.9$, then the relationship represented by this contingency table may be considered as equality [34].

These contingency tables, which feature sufficient value of the measure V, are candidate regularities for searching for equations. However, is makes sense to search for equations if the given contingency table satisfies *functionality test*:

Given set of pairs

$$D = \{(x_i, y_i) \mid x_i \in X \land i = 1, \ldots, N\}.$$

y is a function of x iff for each $x_0 \in X$ there exists only one y_0 such, that $(x_0, y_0) \in D$.

The search for equations proceeds as follows [34]:

1. generate new functional terms;
2. select a pair of terms as expressions for x and y, respectively;
3. generate and evaluate equations for each pair of terms.

To evaluate the degree of fit of the equation to the dataset $\{(x_i, y_i, \sigma_i)\}$ the following version of χ^2 test is applied,

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{y_i - f(x_i, a_1, \ldots, a_q)}{\sigma_i} \right)^2, \tag{8}$$

where σ_i is standard deviation of the dependent variable y_i , $y = f(x, a_1, \ldots, a_q)$ is a mathematical model of the pattern (as e.g. $y = a_1 + a_2 x + a_3 x^2$), and $a_1, \ldots, a_q$ are selected in order to minimize the value of χ^2 defined in (8).

4.1.4. Methods of assessment of declarative knowledge

Many methods of assessment of declarative knowledge may be used, corresponding to the goal that may be [16]:

1. *detailed assessment* – when single "portion" of knowledge (e.g. single rule or decision tree) is being assessed and interrelationships occurring between this portion and remaining part of knowledge base are not taken into account;
2. *integral assessment* – when complete knowledge base is undergoing the assessment; such an assessment can be:
   (a) in *content-related respect* – when the whole content of knowledge base is being assessed, what can be accomplished:
      - *by an expert* – then it turns out to be very difficult task, since the expert must take into consideration the complete knowledge base,
      - *by means of test examples* that make possible automatic evaluation of completeness of the knowledge base,

   (b) in *formal respect* – if the complete knowledge base is being assessed, but the goal depends on identification of contradictions, absorption or loops; such tests are usually carried out using respective software.

4.2. Methods of acquisition of procedural knowledge

Acquisition of procedural knowledge is a separate and new task [32, 33]. Procedural knowledge may concern procedures of operation, but also procedures of reasoning and concluding. Engineers often represent procedures by means of *block diagrams*. In our research the basic source of procedural knowledge is an expert, who may take part in this process either directly, or indirectly as author of publications, which may then be searched for pieces of procedural knowledge that may be further interpreted by a knowledge engineer.

The applied method of procedural knowledge acquisition, as a general rule, aims at possible elimination of interventions of knowledge engineer [33]. It consists in making a *special editor of procedural knowledge* available to the expert. The editor not only aids the expert by representation of his/her knowledge, but also facilitates the expert to elicit his/her own knowledge. M. Wyleżoł proposed *top-down* approach for this purpose, which depends on step-wise increasing of level of detail contained in the block diagram. To make available respective means, he introduced a *multi-layer block diagram* [33]. Knowledge representation in this case consists in expanding of individual tasks and activities into ever more and more detailed sub-procedures (Fig. 3), where each subsequent layer corresponds to the increased level of detail of procedure representation.

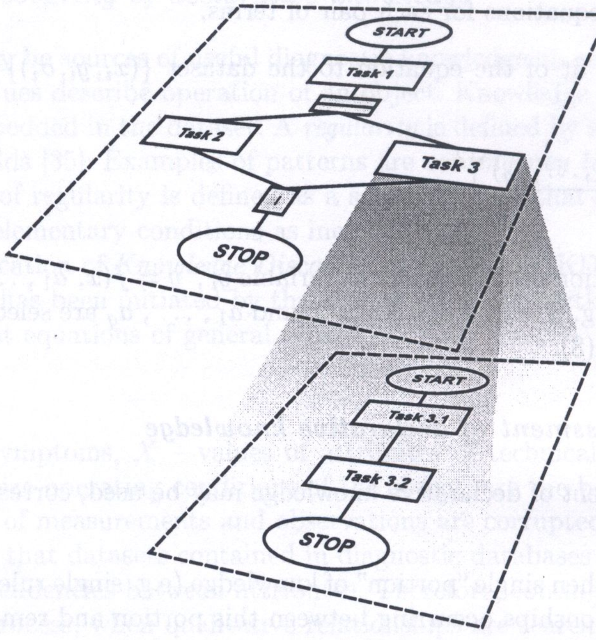

Fig. 3. Representation of procedural knowledge using a multi-layer block diagram [33]

5. Examples of application

In the following some examples of application of the described methods are presented, concerning:

- *acquisition of declarative knowledge* from a database collected within the confines of *active diagnostic experiment* (on the model of rotating machine called *Rotor Kit*) [16],
- *discovery of declarative knowledge* from a database obtained during *numerical experiment* on a model of the laboratory kit [16, 23, 25],
- *acquisition of procedural knowledge* [33].

5.1. Acquisition of declarative knowledge from active experiment

The active experiment concerned a special stand located in the laboratory of the Department. Several faults were considered, as: *differing states of imbalance* (rough dynamic balancing, moment imbalance and quasi-static imbalance), *local rub of the shaft against stationary elements* of the model, and *overload of bearing system*. P. Kostka [9] has carried out this experiment, invoking complex technical states of the object, which were combinations of elementary states mentioned above. The experiment plan is presented in Table 1. (abbreviation "Rub+ovld" denotes simultaneous occurrence of rub and overload).

Induction of decision trees was carried out on the collected dataset using the *C4.5* program [29]. However, very low performance of the classifier amounting to 58 per cent was obtained. The *Random subsampling* technique with learning to testing data ratio 70:30 was applied. Then the author has

Table 1. Numbers of observations carried out for different combinations of factors considered in the active experiment [16]

State of imbalance	Additional factors				Total
	None	Rubbing	Overload	Rub+ovld	
rough dynamic balancing	13		12		25
moment imbalance	7				7
quasi-static imbalance	16	98	23	9	146
Total	36	98	35	9	178

put forward the hypothesis that low performance of learned classifiers was caused by too complex structure of the set of technical states, evoked during the active experiment. Therefore, different *grades of detail of diagnosis* have been introduced together with the *top–down approach* consisting in gradual increasing of minuteness of detail of diagnosis.

An algorithm applied for optimization of a structure of the set of technical states is similar to the algorithm of search for optimal decision tree [29]. Three auxiliary decision attributes (elementary attributes of technical state) were introduced:

- attribute *state of imbalance* with values: *rough dynamic balancing*, *moment imbalance*, *quasi-static imbalance*;
- attribute *bearings overload* with values: *exists*, *doesn't exist*;
- attribute *rubbing* with values: *exists*, *doesn't exist*.

The optimization of the structure of the states tree was carried out with respect to the criterion of maximal classifier's performance. There were only seven different tree structures because of incomplete plan of the active experiment. The search was supported by domain knowledge of the author[2]. The optimal structure of the tree is shown in Fig. 4. Individual rectangles contain names of corresponding technical states, together with the overall error rate of the classifier assigned to this node of the tree. As we discussed in Section 4.1.2, the hierarchical classifier has been learned from data. Its performance was better than this obtained for ordinary multi-class classifier. For example, the performance of the classifier corresponding to the sub-tree starting from the node *quasi-static imbalance* is far better than in previous classifier.

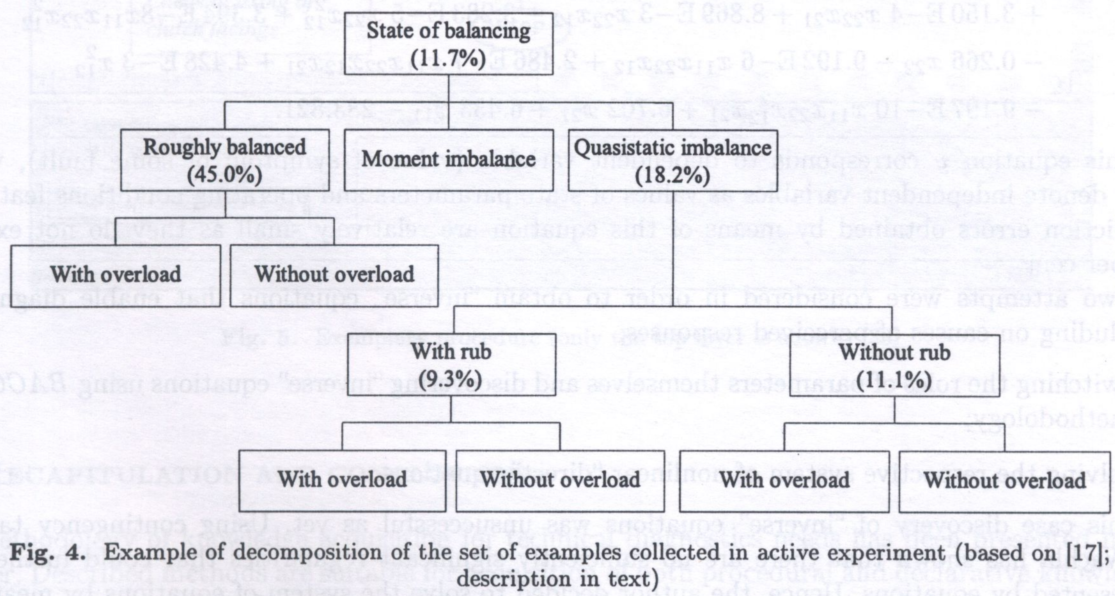

Fig. 4. Example of decomposition of the set of examples collected in active experiment (based on [17]; description in text)

[2]All calculations were carried out by P. Kostka.

5.2. Discovery of equations

We show some application of KDD methods to discovery of diagnostic knowledge in the database obtained during a numerical experiment. The experiment concerned different elementary states of balancing. Similar numbers of examples for each considered class have been taken into account. A detailed descriptions of this database and diagnostic problem can be found in [21, 25]. Further discussion requires some information on the database that is summarized in Table 2.

Table 2. Simulation data [23]

Property	Number
control attributes:	7
– attributes of operating condition	1
– attributes of imbalance distribution	6
dependent attributes	16
decision attribute	1
no. of records	5076
no. of classes	5

The fault diagnosis problem concerned the possibility to obtain accurate predictions concerning imbalance distribution along the shaft. More detailed description of search for equations and results obtained to-date is contained in [23, 25]. In the following only some more important issues connected with this research are given.

In the discussed example the $BACON-3$ [24] methodology was applied for the discovery of "causal" (i.e. direct) equations, which enable calculations of responses given inputs to the object. Very simple models as linear and quadratic ones were sufficient enough to achieve small prediction errors in considered subsets of records. However, application of simple equations of single variable can give complex nonlinear multi-variate equation, as the following one obtained in our research [23],

$$\begin{aligned} y_{11} = &-1.303\, x_{12} - 3.652\,\mathrm{E}{-5}\, x_{11}x_{12}x_{21} - 1.046\,\mathrm{E}{-5}\, x_{22}x_{12}x_{21} + 3.859\,\mathrm{E}{-8}\, x_{22}x_{12}^2x_{21} \\ &+ 1.240\,\mathrm{E}{-7}\, x_{11}x_{12}^2x_{21} - 7.464\,\mathrm{E}{-6}\, x_{11}x_{22}x_{21} - 7.839\,\mathrm{E}{-3}\, x_{11}x_{21} + 0.001\, x_{11}x_{12} \\ &- 4.564\,\mathrm{E}{-6}\, x_{11}x_{12}^2 + 2.760\,\mathrm{E}{-4}\, x_{11}x_{22} + 0.002\, x_{12}x_{21} - 5.198\,\mathrm{E}{-6}\, x_{12}^2x_{21} \\ &+ 3.150\,\mathrm{E}{-4}\, x_{22}x_{21} + 8.869\,\mathrm{E}{-3}\, x_{22}x_{12} - 3.283\,\mathrm{E}{-5}\, x_{22}x_{12}^2 + 3.394\,\mathrm{E}{-8} x_{11}x_{22}x_{12}^2 \\ &- 0.266\, x_{22} - 9.192\,\mathrm{E}{-6}\, x_{11}x_{22}x_{12} + 2.486\,\mathrm{E}{-7}\, x_{11}x_{22}x_{12}x_{21} + 4.428\,\mathrm{E}{-3}\, x_{12}^2 \\ &- 9.197\,\mathrm{E}{-10}\, x_{11}x_{22}x_{12}^2x_{21} + 6.702\, x_{21} + 6.433\, x_{11} - 283.821. \end{aligned}$$

In this equation y corresponds to dependent variable (value of symptom of some fault), while all x denote independent variables as values of state parameters and operating conditions features. Prediction errors obtained by means of this equation are relatively small as they do not exceed 6.5 per cent.

Two attempts were considered in order to obtain "inverse" equations that enable diagnostic concluding on causes of perceived responses:

- switching the roles of parameters themselves and discovering "inverse" equations using *BACON-3* methodology;
- solving the respective system of nonlinear "direct" equations.

In this case discovery of "inverse" equations was unsuccessful as yet. Using contingency tables, D. Wachla has shown that there are no sufficiently significant regularities that could further be represented by equations. Hence, the author decided to solve the system of equations by means of symbolic calculations. Although larger prediction error rates (up to 16 per cent) were noticed, they may still give some rough estimate of the imbalance distribution along the shaft [23].

5.3. Acquisition of procedural knowledge

The method described in Section 4.2 has been applied in the course of acquisition of relevant diagnostic knowledge on a passenger car [33]. An expert represented and recorded his procedural knowledge using a special software called *EMPREG* that was an editor of multi-layer block diagrams. An extensive description of this tool is contained in [33]. During his work the expert consulted individual procedures with some service stations. He also used accessible specialist publications.

Figure 5 shows an exemplary result of acquisition of procedural knowledge, in this case – procedure of check of technical state of the clutch. The procedure consists of three tasks, which are then expanded into sub-procedures. These tasks are represented in the next sub-layers of the represented block diagram (not visible at the screen shown in Fig. 5). Each of elementary tasks (at the lowest level of description) contains detailed instructions of carrying out the given check and limit values (as e.g. allowable clearances) that enable evaluation of checked part or assembly.

The acquired procedural knowledge has been presented to workers of the co-laboring service stations by means of the same editor of block diagrams. This attempt enabled them to verify and validate the procedural knowledge base [33].

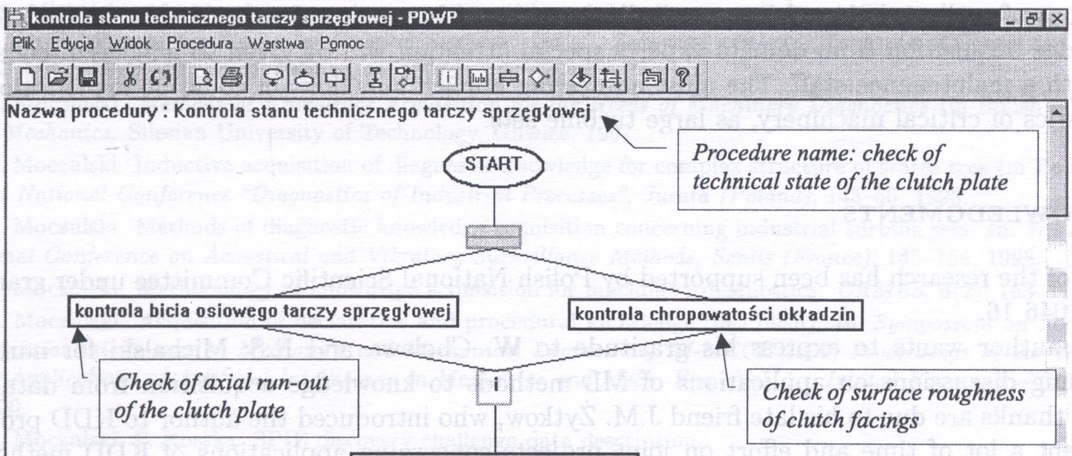

Fig. 5. Exemplary procedure (only the top layer is shown) [33]

6. Recapitulation and conclusions

A methodology of knowledge acquisition for technical diagnostics needs has been presented in the paper. Described methods are suitable for acquisition of both procedural and declarative knowledge. They conform to various knowledge sources, as domain experts or databases. Particular attention is paid to methods of knowledge acquisition from examples, which may be aided by computer systems. Two groups of methods are addressed: Machine Learning methods and Knowledge Discovery ones.

From the first group a new method suitable for acquisition of knowledge on complex technical states was briefly described, which consists in gradual decomposition of the set of learning examples into subsets, carried out with respect to structure of the states tree. Since there are usually many different trees, a search for optimal one is carried out with respect to an optimization criterion that is the overall performance of the hierarchical classifier determined by the given tree. Furthermore, some examples of application of all these methods in both declarative and procedural knowledge are given.

Future work will focus on the following issues:

- determination of structure of a set of examples basing on knowledge of domain experts and respective direction of search for optimal structure (in order to avoid combinatorial explosion);
- development of methods of knowledge acquisition from sequences of events and observations;
- development of methods of quantitative knowledge discovery, e.g. in the form of multi-variate equations.

This research will be carried out for more numerous sets of data, collected during diagnostic observations of really existing objects, as well as generated in numerical experiments. In the scope of knowledge acquisition from domain experts special attention will be given to long-term collaboration with a maintenance staff. The most interesting applications domain could be exploitation and diagnostics of critical machinery, as large turbine sets.

ACKNOWLEDGMENTS

A part of the research has been supported by Polish National Scientific Committee under grant No. 7T07B 046 16.

The author wants to express his gratitude to W. Cholewa and R.S. Michalski for numerous interesting discussions on applications of ML methods to knowledge acquisition from databases. Special thanks are due to his late friend J.M. Żytkow, who introduced the author to KDD problems and spent a lot of time and effort on joint projects concerning applications of KDD methods to acquisition of diagnostic knowledge from databases.

Particular results presented as examples were obtained by the author himself and his Ph.D. students. They were presented at several conferences and discussed with many participants of these conferences.

A part of the research has been carried out using software that was kindly made accessible to the author and his research group by J. Kiciński (system *MESWIR* [3]) and R.S. Michalski (systems *AQ15c* [31] and *AQ17-HCI* [15]).

REFERENCES

[1] B.G. Buchanan, et al. Constructing an expert system. In: F. Hayes-Roth, D.A. Waterman, eds., *Building Expert Systems*, 127–168. Addison-Wesley, Reading, MA, 1983.

[2] W. Cholewa, J. Kaźmierczak. *Data Processing and Reasoning in Technical Diagnostics.* WNT, Warszawa, 1995.

[3] W. Cholewa, J. Kiciński, eds. *Machinery Diagnostics. Inverted Diagnostic Models* (in Polish). Silesian University of Technology, Gliwice, 1997.

[4] K. Ciupke. *Selection and reduction of information in machinery diagnostics* (in Polish). Ph.D. thesis, Silesian University of Technology, Department of Fundamentals of Machinery Design, Gliwice, 2001.

[5] J. Dietrych. *System and Design* (in Polish). WNT, Warszawa, 1985.

[6] J. Dougherty, R. Kohavi, M. Sahami. Supervised and unsupervised discretization of continuous features. In: A. Prieditis, S. Russell, eds., *Machine Learning: Proceedings of the 12th International Conference.* Morgan Kaufman, San Francisco, CA, 1995.

[7] A. Giordana, L. Saitta, F. Bergadano, F. Brancadori, D. De Marchi. Enigma: A system that learns diagnostic knowledge. *IEEE Trans. on Knowledge and Data Engineering*, 5(1): 15–28, 1993.

[8] J.W. Grzymała-Busse. Managing uncertainty in machine learning from examples. In: M. Dąbrowski, M. Michalewicz, Z. Raś, eds., *Intelligent Information Systems: Proceedings of the Conference "Practical Aspects of Artificial Intelligence III", Wigry (Poland)*, 70–84. Insitute of Computer Science, Polish Academy of Sciences, Warszawa, 1994.

[9] P. Kostka. *Expert System for Application of Multi-Dimensional Signal Analysis for Investigations of Rotor Vibrations* (in Polish). Master's thesis, Silesian University of Technology, Department of Fundamentals of Machinery Design, 1997.

[10] P. Kostka. Examination of sensitivity of features of diagnostic signals (in Polish). In: *3rd National Conference "Diagnostics of Industrial Processes", Jurata (Poland)*, 85–90. Insitute of Computer Science, Polish Academy of Sciences, 1998.

[11] P. Kostka. Optimization of representation space for knowledge acquisition on rotating machinery using machine learning methods. In: *Intelligent Information Systems VII: Proceedings of the Workshop, Malbork (Poland)*, 235–238. Insitute of Computer Science, Polish Academy of Sciences, Warszawa, 1998.

[12] P. Kostka. Application of machine learning methods to the recognition of shaft bearing misalignment. In: M. Kłopotek, M. Michalewicz, S.T. Wierzchoń, eds., *Intelligent Information Systems IX, Proceedings of the Workshop, Bystra*, 51–55. Insitute of Computer Science, Polish Academy of Sciences, Warszawa, 2000.

[13] P. Maniak. *Examination of usefulness of inductive methods to acquisition of design knowledge* (in Polish). Ph.D. thesis, Silesian University of Technology, Department of Fundamentals of Machinery Design, Gliwice, 1999.

[14] R.S. Michalski. A theory and methodology of inductive learning. *Artificial Intelligence*, **20**: 111–161, 1983.

[15] R.S. Michalski. Machine learning, data mining and knowledge discovery. Principles and applications. Tutorial on the Workshop *"Intelligent Information Systems" IIS'97, Zakopane (Poland)*. Institute of Computer Science, Polish Academy of Sciences, Warszawa, 1997.

[16] W. Moczulski. *Methods of Knowledge Acquisition for the Needs of Machinery Diagnostics* (in Polish). Vol. 130 of *Mechanics*. Silesian University of Technology, Gliwice, 1997.

[17] W. Moczulski. Inductive acquisition of diagnostic knowledge for complex structure of states tree (in Polish). In: *3rd National Conference "Diagnostics of Industrial Processes", Jurata (Poland)*, 145–56, 1998.

[18] W. Moczulski. Methods of diagnostic knowledge acquisition concerning industrial turbine sets. In: *3rd International Conference on Acoustical and Vibratory Surveillance Methods, Senlis (France)*, 145–154, 1998.

[19] W. Moczulski. Methodology of knowledge acquisition for machinery diagnostics. *CAMES*, **6**(2): 163–175, 1999.

[20] W. Moczulski. Acquisition of declarative and procedural knowledge (in Polish). In: *Symposium on Methods of Artificial Intelligence in Mechanics and Mechanical Engineering AI-MECH'2000, Proceedings of the Workshop on Applications of Artificial Intelligence in Mechanics and Mech. Eng.*, Vol. 1 – *Invited Papers*, 55–73. Gliwice, 2000.

[21] W. Moczulski, P. Kostka. SPIE discovery challenge data description. `http://www.polsl.gliwice.pl/~moczulsk/spie_challenge`, 2000.

[22] W. Moczulski, P. Kostka, P. Maniak. Knowledge acquisition from results of simulation concerning a 200 MW turbine set (in Polish). Report DT6D390.01 on grant PBZ-038-06, Silesian University of Technology, Department of Fundamentals of Machinery Design, Gliwice, 1998.

[23] W. Moczulski, D. Wachla. Acquisition of diagnostic knowledge using discoveries in databases. In: *V National Conference "Diagnostics of Industrial Processes" DPP'01, Łagów (Poland)*, 2001 (in print).

[24] W. Moczulski, J.M. Żytkow. Automated search for knowledge on machinery diagnostics. In: M. Kłopotek, M. Michalewicz, Z. Raś, eds., *Intelligent Information Systems VI, Proceedings of the Workshop, Zakopane, 1997*, 194–203. Warszawa, 1997.

[25] W. Moczulski, J.M. Żytkow. Discovery of diagnostic knowledge from multi-sensor data. In: *AeroSense – SPIE's 15th International Symposium on Aerospace/Defense Sensing, Simulation, and Controls, "Data Mining and Knowledge Discovery: Theory, Tools, and Technology III"*, Vol. 4384, 104–115. Proceedings of SPIE, Orlando, FL, 2001.

[26] Z. Pawlak. *Rough Sets. Theoretical Aspects of Reasoning About Data*. Kluwer Academic Publishers, 1991.

[27] J. Pokojski, ed. *Intelligent support of the process of integration of an environment for computer-aided machine design* (in Polish). WNT, Warszawa, 2000.

[28] J.R. Quinlan. Induction of decision trees. *Machine Learning*, **1**: 81–106, 1986.

[29] J.R. Quinlan. *C4.5 Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA, 1993.

[30] A. Skowron. A synthesis of decision rules: Applications of discernibility matrix properties. In: M. Dąbrowski, M. Michalewicz, Z. Raś, eds., *Intelligent Information Systems: Proceedings of the Conference "Practical Aspects of Artificial Intelligence II", Augustów (Poland)*, 30–46. Insitute of Computer Science, Polish Academy of Sciences, Warszawa, 1993.

[31] J. Wnek, K. Kaufman, E. Bloedorn, R.S. Michalski. *Selective Induction Learning System AQ15c: The Method and User's Guide*. Tech. rep., George Mason University, Center for Machine Learning and Inference, Fairfax, VA, 1995.

[32] M. Wyleżoł. Electronic form–editor for acquisition of empirical rules from experts for the purpose of diagnostics. In: *Intelligent Information Systems: Proceedings of the Workshop "Intelligent Information Systems – IIS'98", Malbork (Poland)*, 231–234. Insitute of Computer Science, Polish Academy of Sciences, Warszawa, 1998.

[33] M. Wyleżoł. *Methods of acquisition of diagnostic procedures and relationships from machinery exploitation domain experts* (in Polish). Ph.D. thesis, Silesian University of Technology, Department of Fundamentals of Machinery Design, Gliwice, 2000.

[34] R. Zembowicz, J.M. Żytkow. From contingency tables to various forms of knowledge in databases. In: U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy, eds., *Advances in Knowledge Discovery and Data Mining*, 329–349. AAAI Press, 1996.

[35] J.M. Żytkow, R. Zembowicz. Database exploration in search of regularities. *Journal of Intelligent Information Systems*, 2(1): 39–81, 1993.