COST-SENSITIVE FEATURE SELECTION*

Krzysztof CIUPKE

Silesian University of Technology, Faculty of Mechanical Engineering
Konarskiego Street 18A, 44-100 Gliwice, Poland, e-mail: kciupke@polsl.pl

Summary

The paper concerns the selection of features in the technical diagnostics domain. The author focused his attention on a wrapper approach. In this approach an application of the ant algorithm as a search engine is proposed. The proposed method of so-called ant wrapper approach is presented. The method takes advantage of cost of features, where the cost is connected with the cost of sensors. The algorithm as a pseudo-code and some results of a verification experiment are shown. The verification was carried out on data derived from an active diagnostic experiment concerning a rotating machine. The obtained results show, that the proposed method could allow to reduce the number of used sensors.

Keywords: feature selection, ant algorithm, machine learning, artificial intelligence, technical diagnostics.

SELEKCJA CECH Z UWZGLĘDNIENIEM KOSZTU ICH POZYSKANIA

Streszczenie

W artykule opisano metodę selekcji cech z zastosowaniem algorytmu mrówkowego. Metoda pozwala także na uwzględnienie kosztu atrybutu, przy czym jego koszt związany jest z kosztem pozyskanie sygnału diagnostycznego. W przypadku gdy sygnał ten jest już wykorzystywany uznaje się, że koszt wyznaczenia danej cechy jest pomijalnie mały. Metodę przedstawiono w postaci pseudo-kodu i zweryfikowano dla danych pochodzących z czynnego eksperymentu diagnostycznego. Uzyskane wyniki pokazują, że istnieje możliwość ograniczenia liczby stosowanych czujników.

Słowa kluczowe: selekcja cech, algorytm mrówkowy, uczenie maszynowe, sztuczna inteligencja, diagnostyka techniczna.

1. INTRODUCTION

The main problem in a process of constructing expert systems is knowledge acquisition. One of the sources of knowledge could be databases. The author focused his attention on problems of knowledge acquisition from databases using *machine learning* methods. The collected data should be pre-processed. The pre-processing includes, among others, *feature subset selection*.

The data collected in the domain of technical diagnostics is usually result of observations and measurements carried out on investigated objects. Contemporary monitoring systems of machinery allow to observe plenty of diagnostic signals simultaneously. For each signal many features can be estimated. So that the total number of features describing a state of a machine can be very high. Not all of the features are important from the point of view of reasoning about a state of the machine. Besides, as shown in [5], quality of the acquired knowledge could deteriorate while using or adding an improper feature. That is why the selection of a set of considered features becomes a crucial task.

2. THE PROBLEM OF COST OF FEATURES IN THE FEATURE SELECTION TASK

In many problems, for each feature a cost of its acquisition can be specified, and every feature can have its own cost. But there are problems, such as in technical diagnostics, that the cost of features can be defined in different way. As was mentioned above, plenty of signals can be observed and for each signal many features can be estimated. And the main cost which should be taken into consideration is the cost of measuring the signal (among others the cost of sensor, its mounting on the machinery, transmitting the signal) and not the cost of estimation its features.

Let *Attr* be the set of considered features $a \in Attr$, n be the number of considered features and let the cost of obtaining the features is $C=\{C_1, \ldots, C_n\}$. Besides let as assume, that the number of observed signals in m, so the set of features could be grouped

* The research was partially financed by the Ministry of Science and Higher Education (grant No. 4 T07B 018 27)

into m groups $G=\{G_1, \ldots, G_m\}$. Thus the groups can be defined as:

$$G_j = \{a \in Attr \mid sensor(a) = j\} \quad (1)$$

where *sensor*(.) specifies the sensor (signal) number for which the feature a was estimated.

Next, we could make an assumption, that the cost of features in the group is the same, i.e. if the cost of feature $a_l \in G_j$ is C_l then:

$$C_i = C_l, \ \forall a_i \in G_j \quad (2)$$

On the basis of cost of features defined in such a way becomes an idea of its application into the feature selection problem. The idea could be formulated as follows:

if the feature $a \in G_j$ is chosen by the selection algorithm as a relevant one, then the cost of the other features $b \in G_j$ can be omitted.

Assuming that the cost of features estimation from the given signal is negligible the above mentioned rule becomes true.

3. SELECTION OF FEATURES

In the research the wrapper approach [7] was applied to selection of the set of relevant features. There are many possibilities while defining an search engine using this approach. In the paper an application of ant algorithm is presented.

3.1. Ant algorithm

Ant algorithms are inspired by real ants behaviour and are one of the most successful examples of swarm intelligence systems. The problem representation is suitably mapped on a graph. To each arc (i, j) of the graph is associated a variable τ_{ij} called *pheromone trail*. Pheromone trails are read and written by ants. The utility of the arc (i, j) to build good solutions is proportional to the amount of pheromone τ_{ij} [4].

An ant k located in node i uses the pheromone trials to choose the next node to move from the set of one-step neighbours N_i with some probability. The probability, in the simple ant colony optimisation algorithm, is computed as follows [4]:

$$p_{ij}^k = \begin{cases} \tau_{ij} & j \in N_i \\ 0 & j \notin N_i \end{cases} \quad (3)$$

While moving from node i to j each ant deposit some pheromone $\Delta\tau_{ij}$ on the arc (i, j). In the cycle ant colony optimisation algorithm ants deposit a constant amount of pheromone after reaching the goal state. So each ant will change the value of τ_{ij} in time t:

$$\tau_{ij}(t) = \tau_{ij}(t-1) + \Delta\tau \quad (4)$$

Similarly to real pheromone trails, artificial pheromone trails „evaporate". The evaporation is carried out in an exponential way:

$$\tau(t) = (1-\rho)\tau(t-1), \quad \rho \in (0,1] \quad (5)$$

at each iteration of the algorithm. So finally, the amount of pheromone on the arc (i, j) could be calculated as follows:

$$\tau_{ij}(t) \leftarrow (1-\rho)\tau_{ij}(t-1) + \sum_{k=1}^{K} \Delta\tau_{ij}^k , \quad (6)$$

where: K - number of ants, $\Delta\tau_{ij}^k = 1/L_k$ if the ant k chosen arc (i, j) otherwise $\Delta\tau_{ij}^k = 0$, L_k - length of the k-th ant route.

The probability (3) could also be defined using not only the information about pheromone trials, but also heuristic information η_{ij} [1,6]:

$$p_{ij}^k = \frac{[\tau_{ij}]^\alpha [\eta_{ij}]^\beta}{\sum_{l \in N_i} [\tau_{il}]^\alpha [\eta_{il}]^\beta}, \quad (7)$$

where α and β are parameters that control the relative weight of pheromone trail and heuristic information.

In the research as heuristic information the cost of features was used.

3.2. Ant wrapper approach

Using the ant algorithm as a search engine in the wrapper approach some assumption should be made. Let *Attr* be the set of input attributes, and let K denotes cardinality of the set *Attr*. Each node represents an attribute, and each arc (i, j) corresponds to choosing the j-th attribute as the next attribute after the i-th one.

Moreover, let the length of each arc be the same and be equal one. Then, the length of the k-th ant route L_k is equal the number of visited nodes, i.e. the number of chosen attributes.

Assuming this, the procedure of attribute selection using wrapper approach with the ant algorithm as a search engine, here called *ant wrapper approach* [3], could be formulated. The algorithm as a pseudo-code is presented in Fig. 1.

The procedure *AntWA* begins computations with the whole set of attributes *Attr*. At the first step (line 1) the set *Attr* is evaluated, e.g. the quality *q_0* of classification using all the attributes is calculated. Then, ants are located in nodes (line 2) and arcs are initiated with some small amount of pheromone (line 3). Also the cost is initiated (line 4). For each ant k the set of neighbour nodes N (i.e. not visited yet) is calculated (line 9) and the probability P of choosing each neighbour node is computed (line 10). The next node to move to is selected by the *NextAttr()* procedure at line 11. If as the next attribute an attribute a is chosen then – according to

```
procedure AntWA(Attr)
1   q_0 = SetOfAttrEvaluation(Attr);
2   M = AntsInit();
3   A = PheromonInit();
4   K = NumberOfAttr(Attr);
5   C = CostOfAttrInit(K);
6   while(terminat. crit. not satisf.)
7     for(k=1 to K)
8       while(M(k) ≠ Attr)
9         N = Attr \ M(k);
10        P = NextAttrProb(A,N,C(k));
11        a = NextAttr(P);
12        C(k) = ChangeCosts(C(k),a);
13        M(k) = AddAttr(M(k),a);
14        q = SubsetOfAttrEval(M(k));
15        if(CloseEnough(q, q_0))
16          A = UpdatePheromone(M(k));
17          DeleteAnt(k);
18          NewAnt(k);
19          C(k) = CostOfAttrInit();
20        end
21      end
22    end
23    PheromoneEvaporation(A);
24  end
end procedure
```

Fig. 1. Ant wrapper approach pseudo-code

the idea described above – the costs of attributes from the group G_j where $a \in G_j$ are set to one (line 12). Each ant builds its own solution. So that, the costs should be defined for each ant separately ($C(k)$ at line 12). A list of nodes (attributes) visited by the ant is stored in the memory matrix M. Next, at line 14, the subset of attributes is evaluated (the same procedure as in line 1 is applied), and if the quality q is close enough to q_0 (e.g. $q=q\_0$) then the pheromone trial is updated (line 16), the ant is removed and new ant begins its route with initial costs of attributes. The evaporation process is done by the procedure at line 23, and the computation is repeated till termination criteria (specified at line 6) are not satisfied.

3. VERIFICATION OF THE PROPOSED METHOD

To confirm the usefulness of the considered method for the diagnostic knowledge acquisition process, verification research was required. The research was carried out for a data set derived from an active diagnostic experiment.

3.1. The investigated object and the data set

As an object a model of a rotor machine called *RotorKit* [2] was used (see Fig. 2). The object allows to observe a rotor-bearings system during its operation at different rotating speeds and allows to introduce a few malfunctions. Two discs mounted on the rotor allow to produce the system imbalance and additional equipment allows to produce other malfunctions as e.g. overload.

Four sensors were used to observe relational vibrations in the points A and B (see Fig. 2). The

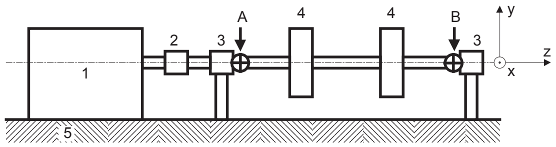

Fig 2. *RotorKit* - a rotor machine model. 1 - DC motor,2 - coupling, 3 - bearings, 4 - discs, 5 - foundation, A, B - measuring points of relative vibrations

vibrations were observed in two directions: horizontal (x) and vertical (y).

Five technical states (classes) of the model were considered:

- one-plane imbalance (IMB1),
- two-plane imbalance (IMB2),
- overload (OVL),
- rub (RUB),
- whirl (WRL).

The number of examples in each class is presented in Table 1.

Table 1. Number of measurements per class

Class	IMB1	IMB2	OVL	RUB	WRL	Total
Number	40	98	95	60	48	371

On the ground of signals measurements a set of data was obtained. The initial set of diagnostic signal features was specified on the basis of literature. Displacement amplitudes of points A and B in x and y directions for different frequency components (e.g. 1X, 1.5X,…) were mainly used. Finally a set of 38 features was applied in the research.

The features were grouped according to the signal for which they were estimated. Four signals were used, so the features were grouped to four groups. Two features: phase difference for the components 1X observed in points A and B in x or y directions were grouped into successive two groups.

3.2. Obtained results

The ant wrapper approach was applied to the data set described above. As the machine learning method the See5 system was applied [8]. The procedure had only one termination criterion: the number of steps which was fixed to 300.

The results describing minimal and maximal length of obtained subsets of attributes are presented in Fig. 3. In Fig. 4 the number of subsets obtained in every step of the procedure is shown as well as the average length of he subsets.

As one could see (Fig. 3a) the smallest subsets were obtained for the parameters α=1, β=0. The subsets were obtained very quickly and as shown in Fig. 4a there are many subsets of relevant features. By increasing the importance of the cost of features ants were "forced" to change their "natural" route. Much fewer subsets were found (see Fig. 4b) and the cardinality of the smallest subsets varied from 3 to 5 (see Fig. 3b).

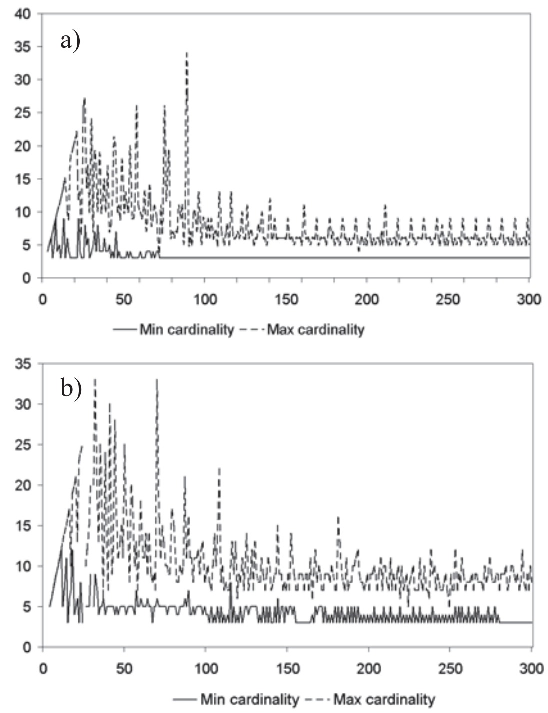

Fig. 3. Minimal and maximal cardinalities of subsets of features for a) α=1, β=0 and b) α=1, β=1

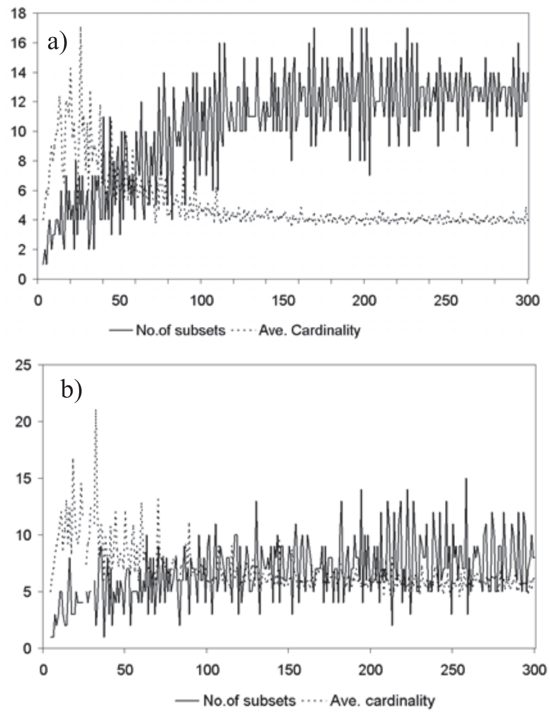

Fig. 4. Average cardinalities and number of subsets of features for a) α=1, β=0 and b) α=1, β=1

The most important is the answer to the question whether the increasing the value of the parameter β produced desired results considering groups of chosen features. The answer is positive, i.e. using the parameters α=1, β=0 the most frequently and the smallest subsets of features contains features from three groups (three sensors are needed), while in other cases (α=1, β>0) mainly features from two groups were chosen, so that only two sensors are needed to identification of the technical state of the considered object. The features allowing the state identification are: maximal relative displacement amplitude of the shaft in the point B in y direction for the rotating frequency range 0.2-0.5 and displacement amplitude of the shaft for frequency 3X in A and B points x direction (where X denotes the rotating frequency of the shaft). The efficiency of the classification using ten-fold cross-validation technique is 71.5%.

3. SUMMARY

In the article a new method of cost-sensitive feature selection is presented. The cost of features is strictly connected with the cost of acquiring the diagnostic signal for which the feature is evaluated. The verification of the method was carried out for the data derived from the active diagnostic experiment. The obtained results show that the method allows not only to reduce the number of features but also the number of signals, and simultaneously, the number of used sensors.

REFERENCES

[1] Boryczka U.: *Co-operation through stigmergy in ant colony optimization for TSP*, Proc. of the 7th Intern. Conf. on Soft Computing Mendel 2001, Brno, Czech Republic, pp. 191-196.
[2] Ciupke K. A*pplication of ant algorithm to selection of diagnostic information.*, (in Polish). Pomiary, Automatyka, Kontrola, 2005, No. 9bis, pp. 134-135.
[3] Dorigo M., Di Caro G., Gambardella L.: Ant algorithms for discrete optimization. Atrificial Life, 1999, Vol. 5, No. 3, pp. 137-172.
[4] Hand D. J.: Discrimination and classification. Chichester, J. Wiley and Sons, 1981
[5] Jensen R., Shen Q.: Finding Rough Set Reducts with Ant Colony Optimization. Proceedings of the 2003 UK Workshop on Computational Intelligence, pp. 15-22, 2003.
[6] Kohavi R., John G.: The wrapper approach. In: H. Liu and H. Motoda: Feature extraction, construction and selection: A data mining perspective. Kluwer Academic Publisher, 1998.
[7] C5/See5 Home page, url: www.rulequest.com.

Krzysztof CIUPKE - Assistant Professor at the Department of Fundamentals of Machinery Design, Silesian University of Technology at Gliwice. His research activities include application of methods of artificial intelligence in technical diagnostics and machinery designing.