ELSEVIER

European Journal of Operational Research 138 (2002) 260–273

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

www.elsevier.com/locate/dsw

Effectiveness evaluation of expert classification methods

Oleg Larichev *, Artyom Asanov, Yevgeny Naryzhny

Institute for Systems Analysis, Russian Academy of Sciences, Prospect 60 let Octjabrja 9, Moscow 117312, Russia

Abstract

A special class of multicriteria classification problems is under consideration in the paper: expert classification problem. It consists in the construction of an expert knowledge base in a classification task. The criterion of expert classification methods efficiency is the minimum number of questions needed for the construction of a complete classification. To arrive at an efficiency estimation system for different methods, a comparison with mathematical algorithms of monotone functions decoding is made. The two best algorithms optimal by Shannon are presented. A procedure of simulating different monotone functions is proposed. A new efficient method of expert classification, CYCLE, is proposed. The results of the simulation demonstrate that the method CYCLE has good evaluations of efficiency for arbitrary monotone functions.

Keywords: Multiple criteria analysis; Multicriteria classification; Expert knowledge; Complete and non-contradictory knowledge base; Monotone function decoding; Algorithm optimal by Shannon

1. Introduction

The problem of constructing a knowledge base imitating expert reasoning is one of the most important in artificial intelligence. The problem consists in creating a computer model that could behave like an expert in answering the user's questions. According to the famous Turing test [1] the construction of such a model could be a demonstration of artificial intellectual existence.

In many practical cases the problem of knowledge base construction could be represented as one of classification because expert knowledge consists in the assignment of objects to different decision classes. For example, in technical diagnostics an engineer analyses the troubles in a complex system and assigns them to one of the several classes. In the problems of medical diagnostics, a physician analyses the patient's state and refers it to one of the several possible diseases.

In a general case, a typical problem is to assign a decision class to an object with evaluations by several criteria. An expert demonstrates his/her knowledge while performing the classification. That is why such problems could be referred to as expert classification problems.

* Corresponding author.

E-mail addresses: oil@isa.ru (O. Larichev), art@isa.ru (A. Asanov), nev@isa.ru (Y. Naryzhny).

PII: S0377-2217(01)00245-4

There are different statements of multicriteria classification problems. For example, in the field of decision making there are two quite different problems:

- to construct a decision maker's rule for the classification of any set of multicriteria alternatives [2–4];
- to find a set of decision rules describing the given classification of multicriteria alternative subsets [5].

Let us call an expert decision concerning classification of one object (a combination of criteria estimates) as a production. In fact, the decision could be represented as a rule: if an object has the following evaluations by criteria then it belongs to a particular decision class. It was shown that a knowledge base could be represented as a set of productions [6]. The problem of expert knowledge base construction is a well-known problem of artificial intelligence.

There are different approaches to the problem of expert knowledge acquisition. The best known is a KADS approach for creating an efficient toolbox for a knowledge engineer [7].

In [8,9] a new statement of an expert classification problem has been proposed: the construction of a complete and non-contradictory knowledge base by a direct dialogue "expert-computer". The main ideas of the approach are as follows:

1. The structure of the classification task is put into computer as decision classes, criteria and estimation scales of the criteria. The Cartesian product of criteria scales defines the set of possible objects – possible descriptions of objects in terms of criteria.
2. The classification is complete if expert decisions allow one to assign all possible objects with estimates by multiple criteria to the decision classes. The classification is non-contradictory if there are no intransitive decisions in the assignment. The main purpose is to construct a complete and non-contradictory expert classification. A complete knowledge base provides insight into the structure of expert knowledge [9].
3. It is known that the major part of expert knowledge is unconscious. Therefore one cannot expect to obtain decision rules by simply asking an expert about them. But it is possible to present descriptions of objects in terms of estimates by criteria and ask an expert to classify them. This classification process is the usual activity of an expert.
4. An expert classifies directly some set of multicriteria objects. On the basis of some multicriteria space properties (see below), expert decisions are used for indirect classification of the remaining objects.
5. The analysis of expert decisions is carried out to check the transitivity property, find possible contradictions and provide the expert with some means for intransitivity exclusion.
6. On the basis of a complete classification, "boundaries" between the decision classes (sets of Pareto-optimal objects of a class) could be constructed. The "boundary" objects represent implicit expert decision rules [10].

On the basis of the proposed ideas, several methods of expert classification have been developed (see below). They have been applied to different practical problems: construction of medical expert systems [8], development of the tutoring system based on the expert knowledge [11] and some others.

The efficiency of the developed expert classification methods could be measured as a number of productions per hour that would be elicited by computer from an expert. For the methods based on the ideas presented above the number of productions classified directly and indirectly is 200–500 per hour [9,11]. For different classification problems the ratio of directly elicited productions varies from approximately 25% to 50%.

In spite of a relatively high efficiency of the developed methods, the question about their evaluation from the standpoint of maximum possible productivity remains.

The criterion of efficiency for an expert classification method is the minimal load on the expert: the minimal number of questions needed for the construction of a complete classification. Indeed, expert time is very valuable. A knowledge base may include hundreds and thousands of objects. The process of knowledge base construction could be very time consuming. That is why the criterion of a minimal number of questions needed for the construction of a complete classification goes well with the method of efficiency estimation.

When taking this criterion as the basis for the efficiency evaluation, it is necessary to note that in reality the analysis of possible expert contradictions requires additional questions. Therefore such an evaluation could give the absolute minimum of the required number of questions.

The aim of this paper is to compare the efficiency of different expert classification methods and obtain the evaluation of expert classification methods from the point of view of maximal possible efficiency. As the basis for the comparison mathematical methods of monotone functions decoding are chosen.

In a general case of expert classification non-monotonic functions may exist. Expert knowledge bases could be constructed in this case as well (see details in [5]). But our practical experience and psychological experiments demonstrate that the case of monotonicity occurs quite frequently in expert knowledge bases construction. That is why a comparison with mathematical decoding algorithms could be carried out.

The task of monotone functions decoding in algebra is one of the well-known problems of modern mathematics [12–14]. The methods of decoding have the same efficiency criterion. Formal efficiency estimates for different decoding methods are derived [12–14].

Below we present the problem statement for monotone functions decoding with an oracle. The ideas of two efficient algorithms by Shannon are given. We discuss the difference between the problem of expert classification and that of the monotone functions decoding. A procedure of efficiency evaluation is introduced. We briefly describe two existing expert classification methods. A new and efficient method of expert classification is presented. The comparison results allow evaluating the relative and absolute efficiency of different methods.

2. Monotone functions decoding problem in algebra of logic

The problem of monotone functions decoding can be briefly represented in the following way: Consider a set of N variables each taking values from some finite set $\{1,\dots,\omega_q\}$, $q \in \overline{1,N}$. A function f defined on this set and taking values from the set $\{1,\dots,m\}$ is called monotone if the condition $x_q \leqslant y_q$ for any $q \in \overline{1,N}$ implies that

$$f(x_1,x_2,\dots,x_N) \leqslant f(y_1,y_2,\dots,y_N).$$

Let us assume that we have a source of information called an oracle. The oracle means an objective source of information (for instance, a sensor) or a human being (a person who makes decisions, an expert).

Let us denote by M_N the class of all monotone binary $(m=2)$ functions of N variables. Any function $f \in M_N$ splits the set $Y = \{1,\dots,\omega_1\} \times \{1,\dots,\omega_2\} \times \cdots \times \{1,\dots,\omega_N\}$ into two disjoint subsets. We denote them by C_0 (the set of zeros) and C_1 (the set of ones).

Definition 1. The maximal elements of the set C_0 and the minimal elements of the set C_1 are called boundary objects.

Definition 2. The sets C_0 and C_1 are called decision classes.

The problem of decoding monotone functions with an oracle lies in the following: The oracle is supposed to correctly determine the value of a function f at a given point $\boldsymbol{y} \in Y$. In other words, the oracle decides to which class, C_0 or C_1, this point should be assigned. The oracle is asked until the values of f are found at all points of Y on the basis of its answers or using the monotonicity of f.

The design of algorithms for optimal decoding of monotone functions is one of the well-known problems of the Boolean algebra [12–14].

Let us introduce the notion of an algorithm optimal by Shannon. Let Ω be the set of all algorithms for decoding of an arbitrary, unknown in advance, function $f \in M_N$ by means of a certain number of questions to the operator B_f. This operator returns the value $f(\boldsymbol{y})$ for any vector $\boldsymbol{y} \in Y$. Let $\varphi(A,f,N)$ be the number of questions for algorithm $A \in \Omega$ to the operator B_f while decoding a function $f \in M_N$. Then, the Shannon function is

$$\tilde{\varphi}(N) = \min_{A\in\Omega} \max_{f\in M_N} \varphi(A, f, N).$$

The algorithm A^* is called optimal by Shannon if

$$\max_{f\in M_N} \varphi(A^*, f, N) = \tilde{\varphi}(N).$$

3. Expert classification problem

The problem of expert classification can be formulated in the following way:

Given:

- $K = \{K_1, K_2, \ldots, K_N\}$ – a set of criteria for estimation of an object.
- $K_q = \{k_1^q, k_2^q, \ldots, k_{\omega_q}^q\}$ – a set of estimates on the scale of qth criterion; ω_q – a number of estimates on the scale of qth criterion; the estimates are ordered from the best to the worst.
- $Y = K_1 \times K_2 \times \cdots \times K_N$ – the Cartesian product of the criteria scales that defines all possible alternatives (combinations of estimates) $\boldsymbol{y}^i \in Y$, $\boldsymbol{y}^i = (y_1^i, y_2^i, \ldots, y_N^i)$, where y_q^i is a gradation on the scale of qth criterion.
- $C = \{C_1, C_2, \ldots, C_m\}$ is the set of decision classes.
- The linear reflexive anti-symmetric transitive relation Q_C is given on the set C so $(C_i, C_j) \in Q_C$ if $i \leqslant j$. Also given is the linear anti-reflexive asymmetric transitive relation P_C so $(C_i, C_j) \in P_C$ if $(C_i, C_j) \in Q_C$ and $i \neq j$.
- The linear reflexive anti-symmetric transitive relation Q_q is given on each set K_q so $(k_i^q, k_j^q) \in Q_q$ if $i \leqslant j$. Also given is the linear anti-reflexive asymmetric relation P_q so $(k_i^q, k_j^q) \in P_q$ if $(k_i^q, k_j^q) \in Q_q$ and $i \neq j$.

Relations P_C and P_q reflect ordering correspondingly of the decision classes and scale gradations on the basis of expert knowledge. We introduce the reflexive anti-symmetric transitive *dominance relation*

$$Q = \left\{(\boldsymbol{v}, \boldsymbol{w}) \in Y \times Y \,|\, \forall q \in \overline{1, Q},\ (v_q, w_q) \in Q_q\right\}$$

on the set of all possible vectorial estimates Y as well as the anti-reflexive asymmetric transitive *strict dominance relation*

$$P = \Big\{(\boldsymbol{v}, \boldsymbol{w}) \in Y \times Y \,|\, \forall q \in \overline{1, Q},$$
$$(v_q, w_q) \in Q_q\ \exists \tilde{q} : (v_{\tilde{q}}, w_{\tilde{q}}) \in P_{\tilde{q}}\Big\}.$$

Needed: to build a reflection F: $Y \to \{Y_l\}$, $l = \overline{1, m}$ on the basis of expert knowledge (where Y_l is a set of vectors belonging to class C_l) so if $(\boldsymbol{v}, \boldsymbol{w}) \in Q$ and $\boldsymbol{v} \in Y_i$, then $\boldsymbol{w} \notin Y_j$ for any $j < i$.

None of the vectors from Y dominating the given one could be assigned to a less preferable class. We refer to the partition of the set Y as non-contradictory if this requirement is fulfilled. More formally, the partition of Y is non-contradictory if it meets the following condition:

$$\text{if } \boldsymbol{y}^i \in Y_k,\ \boldsymbol{y}^j \in \boldsymbol{Y}_l,\ (\boldsymbol{y}^i, \boldsymbol{y}^j) \in Q,\ \text{then } k \leqslant l. \quad (1)$$

Different methods for the solution of the expert classification problem are presented in [9,11].

4. Difference and similarity of two problems

The difference between the problems of monotone function decoding and the problem of expert classification is as follows:

1. The problem of expert classification has a more general nature. There can be non-monotone functions. The orderings of the estimates on criteria scales reflecting expert knowledge can be different with respect to different decision classes. Decision classes can intersect.
2. The expert and the oracle are not quite the same. The expert can make errors. That is why certain procedures for error search and elimination are incorporated into the expert classification methods.

In spite of such a difference, there is an important similar stage in the algorithms of monotone functions decoding and methods of expert classification.

One of the most important steps in the development of expert classification methods is the definition of an effective strategy for questioning an expert, i.e. defining the sequence of the vectors $\boldsymbol{y}^i$ being presented to an expert in the process of classification. This phase is intended to solve the same problem, as do the functions decoding algorithms. Assuming that a characteristic function, which defines the classification, is monotone, and an expert is unerring, both methods coincide. The criterion of non-contradictory classification construction through a minimal number of questions to an expert is perfectly suitable for expert classification problems.

This creates a possibility to use the evaluations of efficiency for decoding algorithms to estimate the efficiency of expert classification methods.

5. The approach to comparison

The formal evaluations of efficiency are given for a special kind of boundary between the decision classes: it is one with the maximum number of boundary objects. For the binary criteria scales, the minimax estimate for the number of questions to an oracle is determined as the Shannon function [13] in the form

$$\varphi(N) = C_N^{N/2} + C_N^{(N/2)+1}.$$

For the problem of expert classification, a kind of boundary is defined by expert knowledge in the classification task. That is why we are interested in efficiency evaluation for different kinds of boundaries between the decision classes. It is desirable that a method be close to the optimal one for various functions.

The analytical estimation of the method efficiency for various functions (boundaries between decision classes) is a very complicated problem. Therefore, it is reasonable to compare different algorithms and methods by using a simulation approach. For using such an approach we are to define which kinds of functions would be taken to simulate different boundaries between the decision classes.

The results of psychological research show the typical decision rules given by experts in classification tasks.

Psychological studies show that the expert decision rules determining the boundaries between the decision classes have a fairly definite structure [8]. The internal organisation of expert knowledge is determined by the characteristics of the human information processing system. During the expert's long-lasting practice, the expert (a physician, geologist or engineer) generates rules (subconscious and informal) for assigning objects with certain characteristics to decision classes. The number of such rules is relatively small. In a general case, each of the rules has the structure of a tree whose root contains combinations of values of r most important features. A certain number of less important features typical for the given class are added. Usually, minor features are "interchangeable", and the rule of adding them to the tree root has the form of the binomial coefficient C_k^t. Here, k is the total number of minor features ($N = r + k$) and t is the number of features that should be added to the basic features to make a decision. Let us note that an unconscious count is the typical operation for human information processing system [15].

Consequently, the functions to be decoded in simulation must be so chosen that they have the structure described above. The structure gives a possibility to also simulate the boundary with the maximum number of objects: in this particular case $r = 0$ and all features (criteria) are equally important.

This explanation is evident for the criteria with binary scales. But in a general case, the simulation could be done in the following way. For example, for two decision classes C_1, C_2 the decision rules for the class C_2 could be simulated. It is natural to suppose that an object belongs to the class C_1 if it does not belong to class C_2. Let us suppose that for each criterion the quality is equally distributed among the evaluations on the scale and the evaluations are more or less typical of class C_2 except the first one (most typical of class C_1). For the simulation we could use the following decision rule: the sum of numbers for evaluations by the most important criteria is equal to r, and sum of values of the other criteria is not less than t. Each decision rule depends on three parameters: subset of the most important criteria S, r and t.

Let us introduce special criteria for efficiency evaluation of different classification (decoding) methods.

A. *The average number of questions to an oracle*: $Q(r, t)$. Since the results of comparison are sensitive to the choice of a most important criteria set (S), the averaging is done on all possible variants of the choice of S (note that if $r = 0$, the decision rule does not depend on S; it depends only on t).

B. *The absolute efficiency of a method* $E_{\mathrm{abs}}(r, t)$. The index of absolute efficiency is taken as the relation of the sum of numbers of boundary objects for two neighbouring classes to $Q(r, t)$. E_{abs} is larger than 0 and not larger than 1. It is the redundancy measure of a number of questions posed to an oracle. If $E_{\mathrm{abs}} = 1$, the number of questions is optimal. If $E_{\mathrm{abs}} = 0.5$, the number of questions is two times larger than the optimal.

The general scheme for comparison can be represented as follows:

1. The functions separating the decision classes that determine the answer of the oracle are simulated.
2. Various methods of classification (decoding of functions) restore a particular function on the basis of the oracle answers.
3. The efficiency of each method is evaluated in terms of the special criteria given above.

6. The methods to be compared

6.1. Monotone functions decoding methods

Alexeev introduced in [14] a method for monotone functions decoding dealing with the functions defined on the set Y with arbitrary numbers of estimations on criteria scales. His method (hereinafter it will be referred to as Algorithm AL) extends an algorithm introduced in [12] based on the theorem about the existence of a catenary partition of a set Y. A chain here is an ordered sequence of vectors $r = \langle \boldsymbol{v}^1, \boldsymbol{v}^2, \ldots, \boldsymbol{v}^d \rangle$, where for any $i = \overline{1, d-1}$ the following applies: $(\boldsymbol{v}^i, \boldsymbol{v}^{i+1}) \in P$, $\boldsymbol{v}^i$ differs from $\boldsymbol{v}^{i+1}$ by the value of exactly one component j, and $v_j^{i+1} = v_j^i + 1$. The length d of the chain r is a number of vectors in the chain.

Algorithm AL constructs a catenary partition of the set Y. All the chains in the partition are ordered by their length. Then the following procedure is applied consequently starting from the shortest chains. A piece of a chain is defined, where the values of f are not yet defined. The middle element $\boldsymbol{u}$ is chosen on the selected piece. Then the value of f at $\boldsymbol{u}$ is requested from the B_f operator. The value $f(\boldsymbol{u})$ obtained is expanded by monotonicity on elements of Y. For that part of chain, where the values of f are still undefined, the bisection is repeated, etc. The algorithm stops after processing all the chains in the partition.

The following property of the algorithm AL is demonstrated in [14]:

$$\frac{\max_{f \in M_{Y,F}} \varphi(\mathrm{AL}, f, N)}{\tilde{\varphi}(N)} \leqslant \frac{1}{2}([\log_2 k] + 1),$$

where AL is Alexeev's algorithm and

$$k = \max_{q \in \overline{1,N}} \omega_q.$$

Thus the algorithm AL is optimal by Shannon in case of binary criteria scales ($k = 2$), i.e. for the longest borders between classes it poses to an expert the minimal number of questions.

Sokolov introduced in [13] a different method for decoding monotone Boolean functions in algebra of logic. This recursive algorithm uses decomposition of a problem into sub-problems. Unit N-dimensional hypercube E_N with even N is split into four subsets $E_{N-2}(0,0), E_{N-2}(0,1), E_{N-2}(1,0), E_{N-2}(1,1)$. Here $E_{N-j}(\beta_1, \ldots, \beta_j)$ denotes a subset of all α elements of the set Y such that $\alpha_1 = \beta_1, \alpha_2 = \beta_2, \ldots, \alpha_j = \beta_j$, $0 \leqslant j \leqslant n$. The values of f are first determined "in the middle" of unit hypercube E_N: separately at the points of the set $E_{N-2}(0,1)$ and $E_{N-2}(1,0)$; and then "at the ends": separately at the points of the set $E_{N-2}(0,0)$ and $E_{N-2}(1,1)$. Here, each of the mentioned subsets is, in turn, split into four subsets $E_{N-4}(\beta_1, \beta_2, 0, 0)$, $E_{N-4}(\beta_1, \beta_2, 0, 1)$, $E_{N-4}(\beta_1, \beta_2, 1, 0)$, $E_{N-4}(\beta_1, \beta_2, 1, 1)$, etc.; $\beta_1, \beta_2 \in \{0, 1\}$. The set E_N for the odd N is split into two subsets $E_{N-1}(0)$ and $E_{N-1}(1)$. Then the decoding problem is solved separately for each subset following the scheme for the even N presented above. Due to the absence of a catenary partition, the Sokolov algorithm is less computationally complex and demands less working memory.

The paper [13] includes a proof of the Shannon optimality for the Sokolov algorithm.

6.2. Expert classification methods

6.2.1. CLASS method

CLASS algorithm [9] uses a maximin procedure to choose an object for presentation to an expert. For each not yet classified object a minimum by all possible DM's answers of a number of indirectly classified objects is calculated using the dominance expansion algorithm. Then an object for which this figure is maximal is selected. The weak point of "CLASS" algorithm is its computational complexity which amounts to $O(|Y|^2)$ with a fixed number of classes.

6.2.2. ORCLASS method

At each step of the ORCLASS algorithm [3] one should calculate G^i – a set of class numbers possible for $\boldsymbol{y}^i$ according to (1). Before the classification all $G^i = \{1, 2, \ldots, m\}$, because there is not yet information about the expert's preferences. When the classification is finished, all G^i consist of exactly one element, i.e. $|G^i| = 1$.

Then among vectors $\boldsymbol{y}^i$ the algorithm chooses the one whose direct classification by an expert allows indirectly classifying by dominance relation as many objects as possible. Usually, the probability of attributing a vector to some class is taken into account. An index p_{ik} (estimating the probability of attributing a vector $\boldsymbol{y}^i$ to class C_k) is associated with the proximity of the vector to the specimens of the given class. For p_{ik} computation a normalized distance between $\boldsymbol{y}^i$ and C_k class centre was suggested to be used. Then it is possible to build a single quantitative index of "information density" for each not yet classified vector $\boldsymbol{y}^i$:

$$\Phi(\boldsymbol{y}^i) = \sum_{k \in G^i} p_{ik} g_{ik}, \tag{2}$$

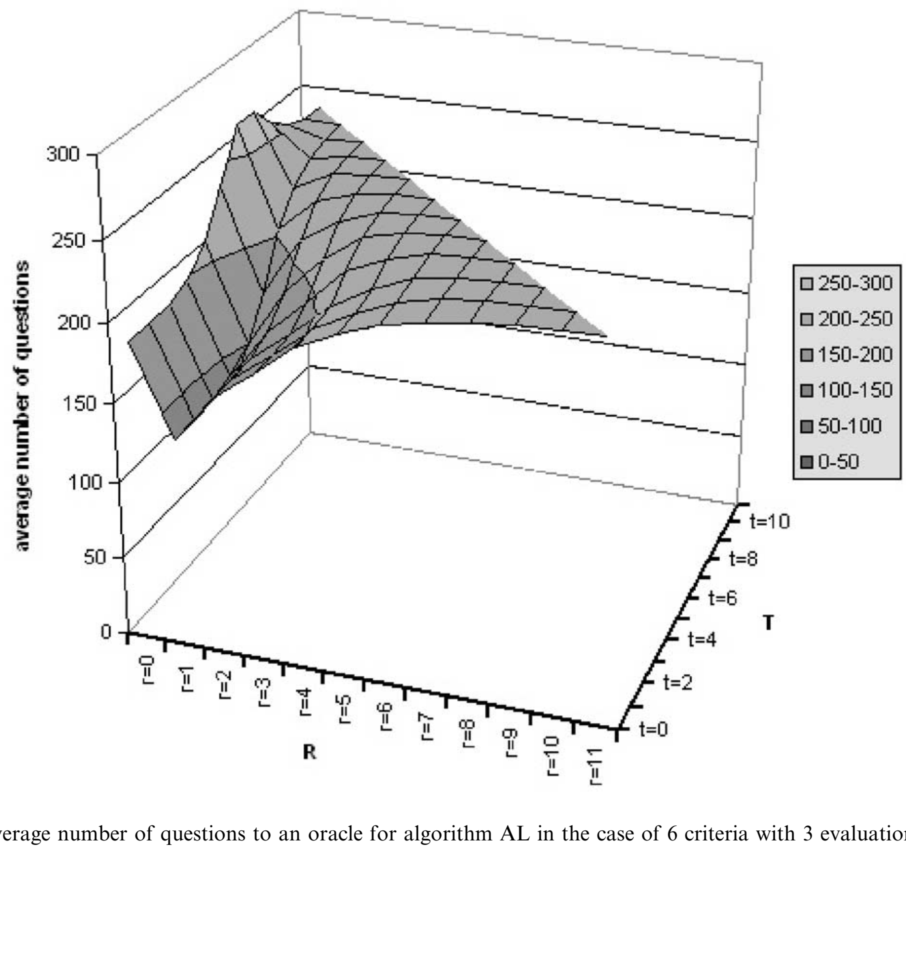

Fig. 1. The average number of questions to an oracle for algorithm AL in the case of 6 criteria with 3 evaluations on the scale.

where g_{ik} is the number of vectors from Y whose classification becomes known when an expert assigns vector $\boldsymbol{y}^i$ to class C_k. At each step of the algorithm the expert is presented vector $\boldsymbol{y}^i$ for direct classification, so the index (2) reaches a maximum.

There are also certain procedures for consistency control of the expert's answers, i.e. a check that condition (1) holds. A drawback of the ORCLASS algorithm is its high computational complexity which amounts to $\mathrm{O}(|Y|^2 \cdot m)$.

6.2.3. CYCLE method

CYCLE algorithm is a new method for expert classification. The Russian name of the method is an abbreviation of the "Catenary Interactive Classification". The CYCLE algorithm is an evolutionary derivative of the DIFCLASS algorithm [11]. DIFCLASS was designed to solve expert classification problems with two decision classes and binary criteria scales. CYCLE extends the DIFCLASS system to the case of an arbitrary number of decision classes and ordinal gradations on criteria scales. The results of our computing experiment demonstrated (see below) that efficiency characteristics of these algorithms are practically the same when criteria have binary scales. Therefore, in what follows we will not cite the results concerning DIFCLASS algorithm.

As in what precedes, we will not distinguish between an object described by estimates by criteria and its vector representation.

Consider a metrics $\rho(\boldsymbol{x}, \boldsymbol{y})$ in the discrete space Y:

$$\rho(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{N} |x_i - y_i|.$$

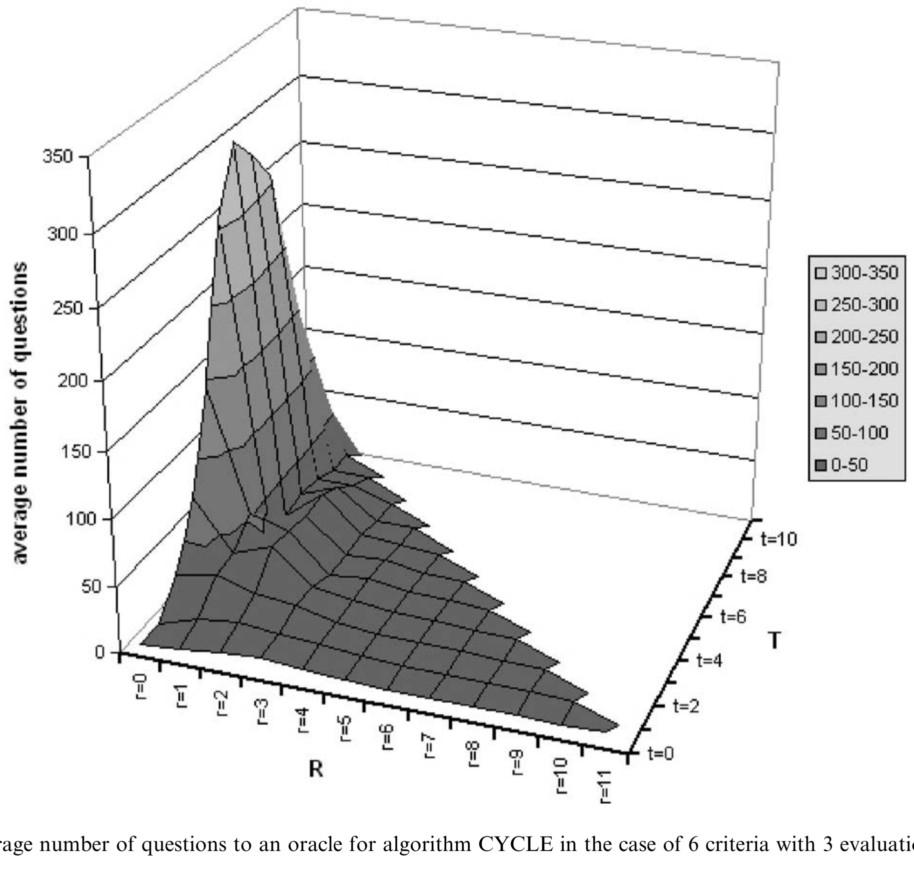

Fig. 2. The average number of questions to an oracle for algorithm CYCLE in the case of 6 criteria with 3 evaluations on the scale.

Let the index of a vector $\boldsymbol{x} \in Y$ (denoted as $\|\boldsymbol{x}\|$) be the number $\rho(\vec{0}, \boldsymbol{x})$, i.e. the sum of all its components.

The chain $\Re$ is an ordered sequence of vectors from Y:

$$\Re = \langle \boldsymbol{v}^1, \boldsymbol{v}^2, \ldots, \boldsymbol{v}^d \rangle$$

so $(\boldsymbol{v}^i, \boldsymbol{v}^{i+1}) \in P$ and $\rho(\boldsymbol{v}^i, \boldsymbol{v}^{i+1}) = 1$ for any $i = \overline{1, d-1}$. The length of a chain is a number of vectors in it and is denoted as $|\Re|$.

For the vectors $\boldsymbol{x}, \boldsymbol{y} \in Y$ so $(\boldsymbol{x}, \boldsymbol{y}) \in P$ consider the set

$$\Lambda(\boldsymbol{x}, \boldsymbol{y}) = \{\boldsymbol{v} \in Y \,|\, (\boldsymbol{x}, \boldsymbol{v}) \in Q, (\boldsymbol{v}, \boldsymbol{y}) \in Q\},$$

i.e. the set of vectors dominating $\boldsymbol{y}$ and dominated by $\boldsymbol{x}$.

Having denoted

$$\boldsymbol{y}' = (0, 0, \ldots, 0),$$
$$\boldsymbol{y}'' = (n_1 - 1, n_2 - 1, \ldots, n_N - 1),$$

one can easily notice that $\Lambda(\boldsymbol{y}', \boldsymbol{y}'')$ coincides with the space Y.

Let us also consider the set

$$M(\boldsymbol{x}, \boldsymbol{y}) = \left\{ \boldsymbol{v} \in \Lambda(\boldsymbol{x}, \boldsymbol{y}) \,\middle|\, \|\boldsymbol{v}\| = \frac{\|\boldsymbol{x}\| + \|\boldsymbol{y}\|}{2} \right\},$$

i.e. the subset of vectors from $\Lambda(\boldsymbol{x}, \boldsymbol{y})$, equidistant from $\boldsymbol{x}$ and $\boldsymbol{y}$ (hereinafter the division is accomplished without remainder).

Let us define numerical functions $C^{\mathrm{U}}(\boldsymbol{x})$ and $C^{\mathrm{L}}(\boldsymbol{x})$ on Y. The values of the functions correspond to the maximal and minimal number of a class that could be assigned to the vector $\boldsymbol{x}$ without violating the consistency condition (1). Vector $\boldsymbol{x}$ is assigned to the class C_k when

$$C^{\mathrm{U}}(\boldsymbol{x}) = C^{\mathrm{L}}(\boldsymbol{x}) = k.$$

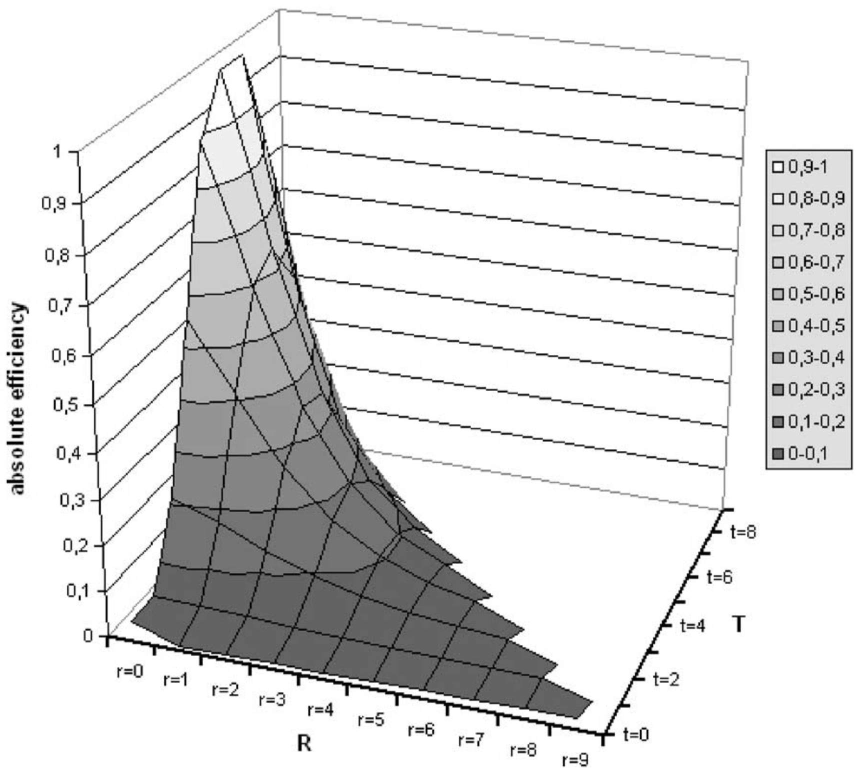

Fig. 3. Absolute efficiency of AL algorithm in the case of 10 criteria with 2 estimates on the scale.

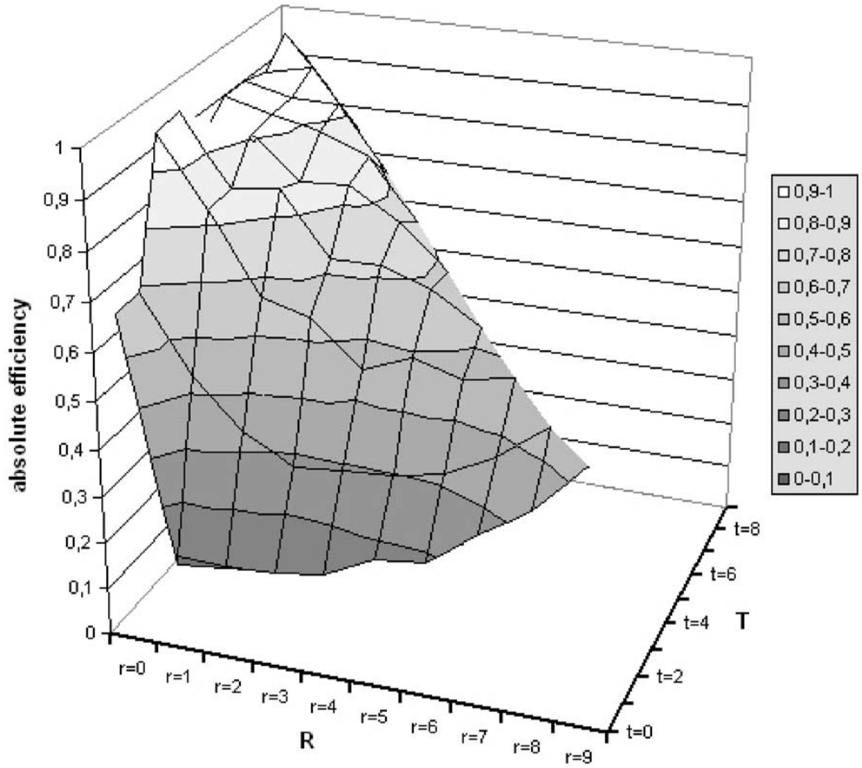

Fig. 4. Absolute efficiency of CYCLE algorithm in the case of 10 criteria with 2 estimates on the scale.

Let us define a procedure $S(\boldsymbol{x})$ (expansion by dominance). Assuming that the decision class for $\boldsymbol{x}$ is known: $\boldsymbol{x} \in Y_k$ (which means $C^{\mathrm{U}}(\boldsymbol{x}) = C^{\mathrm{L}}(\boldsymbol{x}) = k$), for each $\boldsymbol{y} \in Y$ so $(\boldsymbol{x}, \boldsymbol{y}) \in P$ and $C^{\mathrm{L}}(\boldsymbol{y}) < k$ the function $C^{\mathrm{L}}(\boldsymbol{y})$ is redefined: $C^{\mathrm{L}}(\boldsymbol{y}) = k$. In the same manner, for each $\boldsymbol{z} \in Y$ such that $(\boldsymbol{z}, \boldsymbol{x}) \in P$ and $C^{\mathrm{U}}(\boldsymbol{z}) > k$ the function $C^{\mathrm{U}}(\boldsymbol{z})$ is redefined: $C^{\mathrm{U}}(\boldsymbol{z}) = k$.

The main algorithm of CYCLE method could be presented as follows. Let us denote $D(\boldsymbol{a}, \boldsymbol{b})$ – the procedure of classification on the set $\Lambda(\boldsymbol{a}, \boldsymbol{b})$, which uses the idea of dynamic construction of the chains connecting vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. It is assumed that $(\boldsymbol{a}, \boldsymbol{b}) \in P$ and the classification of $\boldsymbol{a}$ and $\boldsymbol{b}$ is known: $\boldsymbol{a} \in Y_k$, $\boldsymbol{b} \in Y_l$.

For each $\boldsymbol{x} \in M(\boldsymbol{a}, \boldsymbol{b})$ the following steps are to be done:

1. If a decision class for $\boldsymbol{x}$ is unknown, then object $\boldsymbol{x}$ is presented to an expert for direct classification. [1] Let $\boldsymbol{x} \in Y_r$. The expansion by the dominance $S(\boldsymbol{x})$ is accomplished. The consistency condition (1) is checked. If it is violated, classification is revised (see next paragraph).
2. If $r > k$ and $(\boldsymbol{a}, \boldsymbol{x}) \in P$, then do $D(\boldsymbol{a}, \boldsymbol{x})$.
3. If $r < l$ and $(\boldsymbol{x}, \boldsymbol{b}) \in P$, then do $D(\boldsymbol{x}, \boldsymbol{b})$.

While classifying a vector $\boldsymbol{x}$ an expert can make an error and a pair of vectors $\boldsymbol{x}, \boldsymbol{y} \in Y$ violating the consistency condition (1) will appear. While such a pair exists, it is presented to an expert with the request to change the decision class of one or both vectors. After this, functions C^{U} are C^{L} redefined to their initial state and ex-

[1] It means that an object description in terms of criteria estimates is presented to an expert along with a set of decision classes. The expert is asked to assign the object to one of the decision classes.

pansion by dominance $S(\boldsymbol{v})$ is accomplished for each vector $\boldsymbol{v}$ that was directly classified by an expert.

For the selection of the vector $\boldsymbol{x}$ at step 1 the following heuristics is suggested: it is necessary to find among all not yet classified vectors of the set $M(\boldsymbol{a}, \boldsymbol{b})$ the one that directly dominates the maximum number of unclassified vectors; i.e. the vector $\boldsymbol{x}^*$ is chosen so

$$\boldsymbol{x}^* = \arg \max_{x \in M(\boldsymbol{a},\boldsymbol{b})} \left| \left\{ \boldsymbol{y} \in Y \,\middle|\, (\boldsymbol{x},\boldsymbol{y}) \in P \text{ or } (\boldsymbol{y},\boldsymbol{x}) \in P, \rho(\boldsymbol{x},\boldsymbol{y}) = 1,\ C^{\mathrm{L}}(\boldsymbol{y}) < C^{\mathrm{U}}(\boldsymbol{y}) \right\} \right|.$$

7. Results

In Figs. 1 and 2 we present typical surfaces that illustrate comparison of AL and CYCLE algorithms using the first criterion: average number of questions (Q) to an oracle, depending on the parameters r and t (with six criteria and three evaluations on the scales). Here, the smaller the value of $Q(r,t)$, the better.

*For all figures: **R** is the sum of evaluation numbers by the most important criteria; **T** is the sum of evaluation numbers by the other (less important) criteria.*

Comparison using the second criterion (average efficiency: $E(r,t)$) revealed that the average efficiency of CYCLE method is much better than that of any other method being compared. Two algorithms for monotone functions decoding are very close in efficiency (the difference being no more than 0.06). The method CYCLE proved to be the most efficient of expert classification methods. Figs. 3–6 present absolute efficiency of AL and CYCLE for different numbers of criteria and estimates on the scales. Here, the larger the value of E, the better.

Fig. 3 demonstrates that the algorithm AL is optimal by Shannon in the case of binary criteria scales. Moreover, it remains optimal by Shannon for other criteria scales; its efficiency reaches 1 in the case of the most difficult boundaries (boundaries with the maximal number of objects) – see points ($r = 0$, $t = 5$) for binary scales at Fig. 1 and

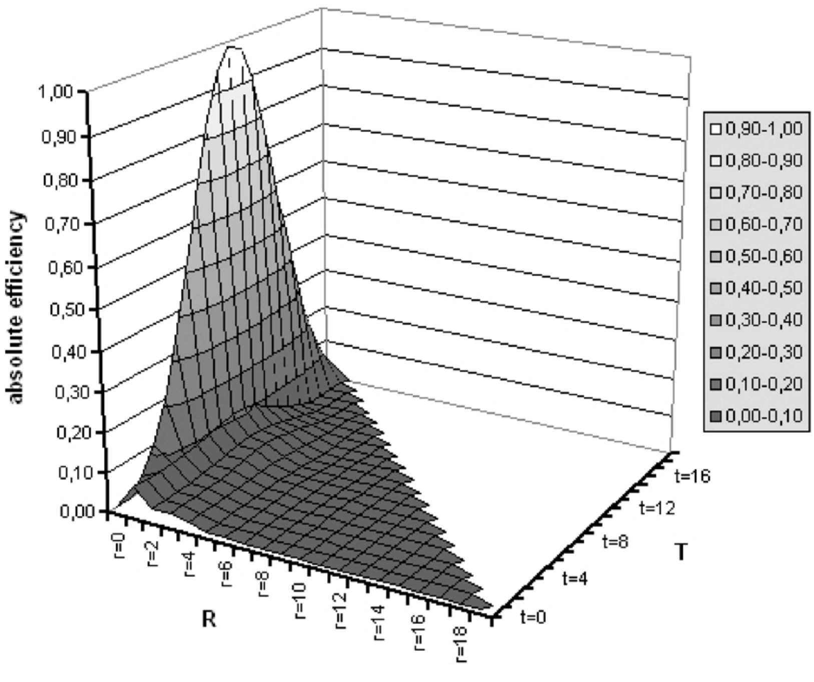

Fig. 5. Absolute efficiency of AL algorithm in the case of 4 criteria with 6 estimates on the scale.

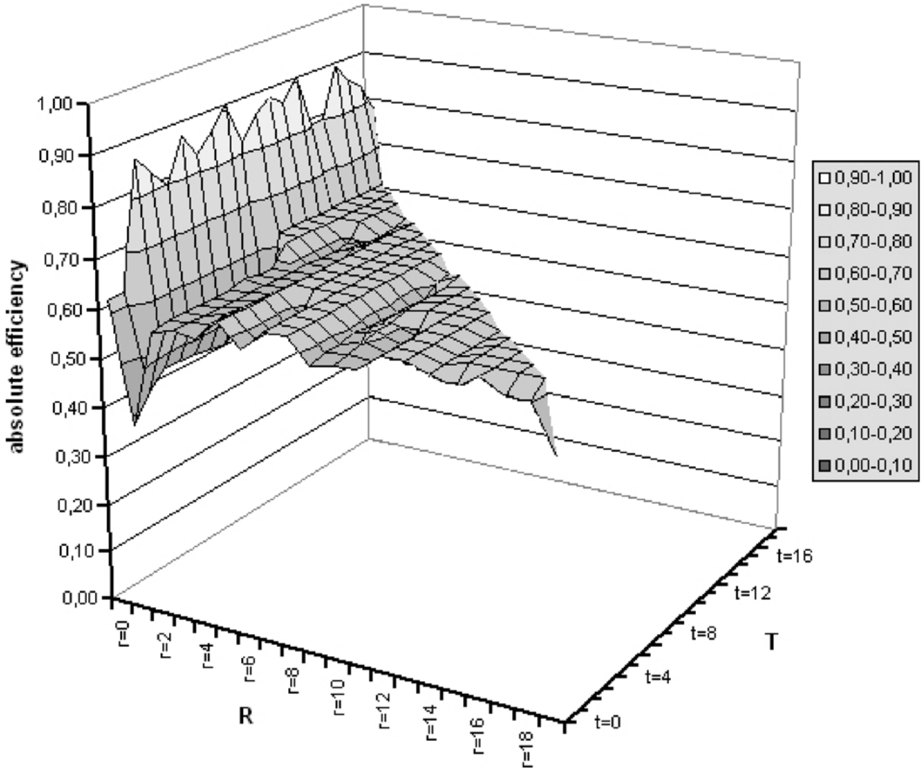

Fig. 6. Absolute efficiency of CYCLE algorithm in the case of 4 criteria with 6 estimates on the scale.

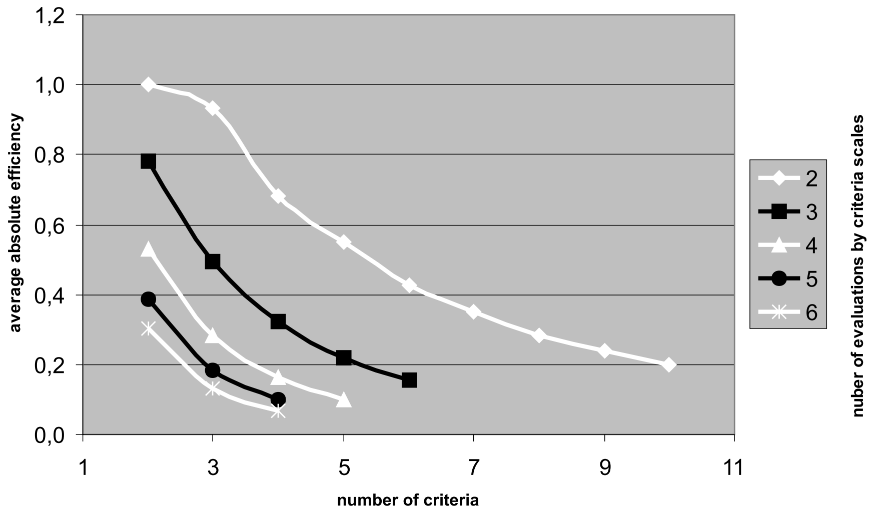

Fig. 7. Average absolute efficiency for AL algorithm.

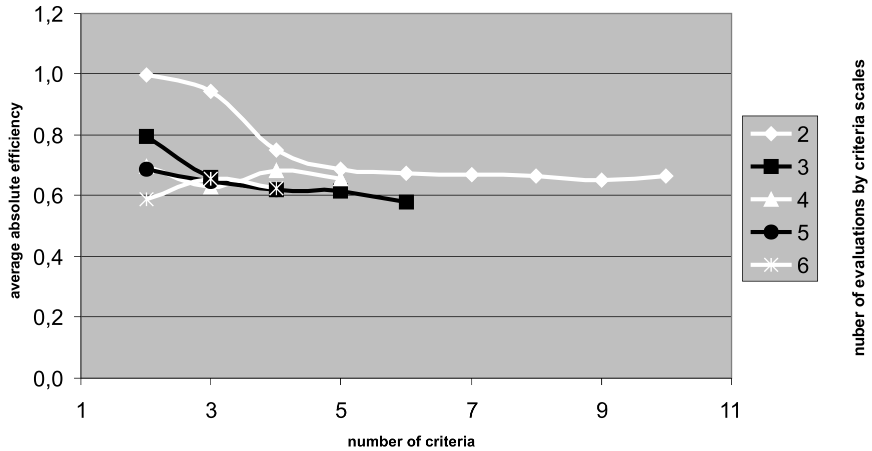

Fig. 8. Average absolute efficiency for CYCLE algorithm.

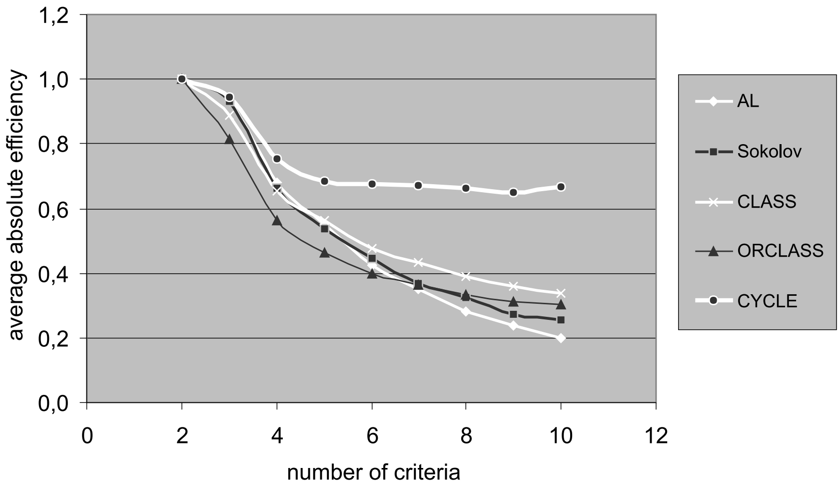

Fig. 9. Average absolute efficiency for all algorithms in the case of binary criteria scales.

$$\left(r = 0, t = \frac{(k-1)\cdot N}{2}\right)$$

for the scales with k estimates at Fig. 5.

The simulation demonstrates that for the most difficult boundaries algorithm AL is a little better than CYCLE. Absolute efficiency of AL is better by 3% with $k = 2$; by 5% with $k = 3$; by 10% with $k = 4$; by 15% with $k = 5$; by 13% with $k = 6$. But for other boundaries algorithm CYCLE is much better. For example, for $r = 3$, $t = 3$ absolute efficiency of CYCLE is 5–10 times larger.

For general comparison, we could find the average absolute efficiency for both algorithms taking the average value for all possible combinations of r and t (where $r = 0, \ldots, (k-1)\cdot N - 1$ and $t = 0, \ldots, (k-1)\cdot N - 1 - r$). Figs. 7 and 8 show the average absolute efficiency for a different number of estimates on criteria scales.

Fig. 9 summarizes the results of the comparison. It shows the average absolute efficiency for all algorithms.

Let us note that the standard deviation of the average absolute efficiency equals approx. 0.15–0.25.

8. Conclusion

The results demonstrate that on the average the CYCLE method of expert classification is much more effective in comparison with the algorithms of monotone functions decoding and the previous methods of expert classification.

The method CYCLE has very good adaptive properties. It uses the idea of dynamic chains construction during the classification. In the process of choice it adapts to a function dividing the decision classes by finding boundary elements with a minimal number of questions to the oracle. In comparison with decoding algorithms, the efficiency of CYCLE is less than that of the Alexeev's algorithm only by several percent in the most difficult case. But on the average it is from 3 to 10 times more effective.

Although the CYCLE method has good efficiency characteristics, there are some conditions and limitations on its applicability. First, a problem is to be structured in an appropriate way: there should be ordinal criteria scales and ordered decision classes. Second, there must be an expert – a person who has spent long time solving corresponding practical problems. A good expert has a deep understanding of a professional area and solves classification problems in a reliable manner. Third, the size of the problem (number of criteria, quality grades and decision classes) should correspond to the possibilities and limitations of the human information processing system. In practice the method has been used for construction of knowledge bases that include up to 20,000 productions.

The application of CYCLE to different practical tasks allows one to obtain a good imitation of expert knowledge and reasoning and save the expert's valuable time.

Acknowledgements

The research was partly supported by Russian Foundation for Basic Research, grant 98-00083 and by the programme "Intellectual Computer Systems", project 2.1. The authors express their gratitude to the unknown reviewers for the useful comments and suggestions.

References

[1] A.M. Turing, Computing machinery and intelligence, Mind 59 (1950) 433–460.
[2] J. de Montgolfier, P. Bertier, Approche multicritère des problèmes de décision, Éditions Hommes et Techniques, Paris, 1978 (in French).
[3] O. Larichev, H. Moshkovich, An approach to ordinal classification problems, International Journal of Operational Research 1 (1994) 375–385.
[4] B. Roy, D. Bouyssou, Aide multicritère à la décision: méthodes et cas, Economica, Paris, 1993.
[5] S. Greco, B. Matarazzo, R. Slowinski, The use of rough sets and fuzzy sets in MCDM, in: T. Gal, T.J. Stewart, T. Hanne (Eds.), Multicriteria Decision Making Advances in MCDM Models, Algorithms, Theory, and Applications.
[6] W.J. Clancey, Classification Problem Solving, Heuristic Programming Project, Memo HPP-84-7, 1984.
[7] B.J. Wielinga, A.Th. Schreber, J.A. Breuker, KADS: A modelling approach to knowledge engineering, Knowledge Acquisition 4 (1) (1992) 5–53.
[8] O. Larichev, A. Mechitov, V. Morgoev, H. Moshkovich, E. Furems, Exact duplicates of human judgements, in: K. Borcherding, O. Larichev, D. Messick (Eds.), Contemporary Issues in Decision Making, Elsevier, Amsterdam, 1990.
[9] O. Larichev, H. Moshkovich, E. Furems, A. Mechitov, V. Morgoev, Knowledge acquisition for the construction of the full and contradiction free knowledge bases, Proggamma, Groningen, 1991.
[10] O. Larichev, A study on the internal organization of expert knowledge, Pattern recognition and image analysis 5 (1) (1995) 57–63.
[11] O. Larichev, A. Bolotov, DIFCLASS system: construction of complete and noncontradictory expert knowledge bases in differential classification problems, Scientific and Technical Information, Series 2, Informational processes and systems 9 (1996) (in Russian).
[12] G. Hansel, Sur le nombre des fonctiones booléennes monotones de n variables, C. R. Acad. Sci. 262 (1966) 1088–1090 (in French).
[13] N. Sokolov, Optimal decoding of monotone Boolean functions, Journal of computational mathematics and mathematical physics 27 (12) (1987) 1878–1887 (in Russian).
[14] V. Alexeev, Decoding of some classes of monotone multivalued functions, Journal of computational mathematics and mathematical physics 16 (1) (1976) 189–198 (in Russian).
[15] C.X. Ling, M.A. Marinov, Symbolic model of the nonconscious acquisition of information, Cognitive Science 18 (1994) 595–621.