Neurocomputing 126 (2014) 29–35

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Optimal selection of ensemble classifiers using measures of competence and diversity of base classifiers

Rafal Lysiak*, Marek Kurzynski, Tomasz Woloszynski

Wroclaw University of Technology, Department of Systems and Computer Networks, Wyb. Wyspianskiego 27, 50-370 Wroclaw, Poland

ARTICLE INFO

Article history:
Received 9 May 2012
Received in revised form
30 November 2012
Accepted 10 January 2013
Available online 9 August 2013

Keywords:
Dynamic ensemble selection
Classifier competence
Diversity measure
Simulated annealing

ABSTRACT

In this paper, a new probabilistic model using measures of classifier competence and diversity is proposed. The multiple classifier system (MCS) based on the dynamic ensemble selection scheme was constructed using both developed measures. Two different optimization problems of ensemble selection are defined and a solution based on the simulated annealing algorithm is presented. The influence of minimum value of competence and diversity in the ensemble on classification performance was investigated. The effectiveness of the proposed dynamic selection methods and the influence of both measures were tested using seven databases taken from the UCI Machine Learning Repository and the StatLib statistical dataset. Two types of ensembles were used: homogeneous or heterogeneous. The results show that the use of diversity positively affects the quality of classification. In addition, cases have been identified in which the use of this measure has the greatest impact on quality.

1. Introduction

At present, in identification and classification, the Multiple Classification Systems (MCS) are very strongly developed, mostly because of the fact that committee, also known as an ensemble, can outperform its members [1]. It is well known that one of the most important steps in the design of MCS is the ensemble selection and the other is combining their answers. Currently, MCS which are using Dynamic Ensemble Selection (DES) schemes are becoming increasingly popular. The DES method is based on dynamic selection of classifiers for a classifying object due to its feature vector. In other words, the MCS each time select the new ensemble (called *dynamic way*) for each recognition object depending on the characteristics describing the object. Most DES schemes use the concept of classifier competence on a defined neighbourhood or region [2], such as the local accuracy estimation [3–5], Bayes confidence measure [6], multiple classifier behavior [7] or probabilistic model [8], among others.

Note that even the best MCS will not be able to outperform its members if classifiers in the team are identical. The ideal situation is when classifiers in the ensemble are the most competent and where the probability of correct classification for the recognition object is the greatest, but are possibly different from each other at the same time. It is popular to use the diversity measure to select such a committee. In the literature, there are many approaches to defining and determining diversification [9]. In this paper, the authors tried to create such a model which will select the best classifiers (most competent) while trying to differentiate their wrong answers. There are examples which show that the use of measure of diversification positively affects the performance of the whole recognition process [10].

In this paper, a novel model has been presented which uses both competence and diversity. In this way, we obtained a hybrid architecture [11] which uses two independent measures. Furthermore, two types of optimization problems were considered. Problem of classifiers selection, because of the criteria and constraints, is solved using simulated annealing [12]. Methods for calculating classifier competence and diversity using a probabilistic model are based on the original concept of a randomized reference classifier (RRC) [8], which – on average – acts like the evaluated classifier. The competence of a classifier is calculated as the probability of correct classification of the respective RRC, and the class-dependent error probabilities of RRC are used for determining the diversity measure, which evaluates the difference of incorrect outputs of classifiers [13,14]. The proposed methods are novel because they take under consideration the competence and diversity measures at the same time during the selection process.

The motivation of our work on the development of the algorithm described in this paper were the results of previous research [15]. It was the first time that both measures were combined with each other, and the results were promising. It should be noted that previously used algorithms, selecting subsets of classifiers, which are involved in the recognition process, were

* Corresponding author.
E-mail addresses: rafal.lysiak@pwr.wroc.pl (R. Lysiak), marek.kurzynski@pwr.wroc.pl (M. Kurzynski), tomasz.woloszynski@pwr.wroc.pl (T. Woloszynski).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.01.052

intuitive. In the following work, we used the simulated annealing algorithm, which gives better results both in terms of classification efficiency and the time required for the recognition process. It is also a generally known and popular heuristic algorithm because of the large number of possibilities of parameterization. It should also be noted that the problem of classifiers selection due to two independent measurements is complex as described in Section 3.

The paper is organized as follows. In Section 2, the randomized reference classifier (RRC) is presented and measures of base classifier competence and ensemble diversity are developed. The constructed multiple classifier systems which use both measures are presented in Section 3. There are also two optimization problems defined and a solution is proposed. The conducted experiments and the results with discussion are presented in Section 4. Section 5 concludes the paper.

2. Theoretical framework

2.1. Preliminaries

Consider a classification problem with a set $\mathcal{M} = \{1, 2, \ldots, M\}$ of class labels and a feature space $\mathcal{X} \subseteq \mathcal{R}^n$. Let a pool of classifiers, i.e. a set of trained classifiers $\Psi = \{\psi_1, \psi_2, \ldots, \psi_L\}$, be given. Let

$$\psi_l : \mathcal{X} \to \mathcal{M} \tag{1}$$

be a classifier that produces a vector of discriminant functions $[d_{l1}(x), d_{l2}(x), \ldots, d_{lM}(x)]$ for an object described by a feature vector $x \in \mathcal{X}$. The value of $d_{lj}(x)$, $j \in \mathcal{M}$ represents a support given by the classifier ψ_l for the fact that the object x belongs to the j-th class. Assume without loss of generality that $d_{lj}(x) \geq 0$ and $\sum_j d_{lj}(x) = 1$. Classification is made according to the maximum rule

$$\psi_l(x) = i \Leftrightarrow d_{li}(x) = \max_{j \in \mathcal{M}} d_{lj}(x). \tag{2}$$

Now, our purpose is to determine the following characteristics, which will be the basis for dynamic selection of classifiers from the pool:

(1) A competence measure $C(\psi_l|x)$ of each base classifier ($l = 1, 2, \ldots, L$), which evaluates the competence of classifier ψ_l, i.e. its capability to correct activity (correct classification) at a point $x \in \mathcal{X}$.
(2) A diversity measure $D(\Psi_E|x)$ of any ensemble of base classifiers Ψ_E, considered as the independency of the errors made by the member classifiers at a point $x \in \mathcal{X}$.

In this paper trainable competence and diversity functions are proposed using a probabilistic model. It is assumed that a learning set

$$\mathcal{S} = \{(x_1, j_1), (x_2, j_2), \ldots, (x_N, j_N)\}; \quad x_k \in \mathcal{X}, \; j_k \in \mathcal{M} \tag{3}$$

is available for the training of competence and diversity measures.

In the next section, the original concept of a reference classifier will be presented, which – using a probabilistic model – will state the convenient and effective tool for determining both competence and diversity measures.

2.2. Randomized reference classifier – RRC

A classifier[1] ψ from the pool Ψ is modeled by a randomized reference classifier (RRC) [8] which takes decisions in a random manner. A randomized decision rule (classifier) is, for each $x \in \mathcal{X}$, a probability distribution on a decision space [14] or – for the classification problem (2) – on the product $[0,\ 1]^M$, i.e. the space of vectors of discriminant functions (supports).

[1] Throughout this subsection, the index l of classifier ψ_l and class supports $d_{lj}(x)$ is omitted for clarity.

The RRC classifies object $x \in \mathcal{X}$ according to the maximum rule (2) and it is constructed using a vector of class supports $[\delta_1(x), \delta_2(x), \ldots, \delta_M(x)]$, which are observed values of random variables $[\Delta_1(x), \Delta_2(x), \ldots, \Delta_M(x)]$. Probability distributions of the random variables satisfy the following conditions:

(1) $\Delta_j(x) \in [0, 1]$;
(2) $E[\Delta_j(x)] = d_j(x)$, $j = 1, 2, \ldots, M$;
(3) $\sum_{j=1,2,\ldots,M} \Delta_j(x) = 1$,

where E is the expected value operator. In other words, class supports produced by the modeled classifier ψ are equal to the expected values of class supports produced by the RRC.

Since the RRC performs classification in a stochastic manner, it is possible to calculate the probability of classifying an object x to the i-th class:

$$P^{(RRC)}(i|x) = Pr[\forall_{k=1,\ldots,M,\ k \neq i}\ \Delta_i(x) > \Delta_k(x)]. \tag{4}$$

In particular, if the object x belongs to the i-th class, from (4) we simply get the conditional probability of correct classification $Pc^{(RRC)}(x)$.

The key element in the modeling presented above is the choice of probability distributions for the rv's $\Delta_j(x)$, $j \in \mathcal{M}$ so that the conditions 1–3 are satisfied. In this paper beta probability distributions are used with the parameters $\alpha_j(x)$ and $\beta_j(x)$ ($j \in \mathcal{M}$). The justification of the choice of the beta distribution can be found in [8] and furthermore the MATLAB code for calculating probabilities (4) was developed and it is freely available for download [16].

Applying the RRC to a learning point x_k and putting in (4) $i = j_k$, we get the probability of correct classification of RRC at a point $x_k \in \mathcal{S}$, namely

$$Pc^{(RRC)}(x_k) = P^{(RRC)}(j_k|x_k), \quad x_k \in \mathcal{S}. \tag{5}$$

Similarly, putting in (4) a class $j \neq j_k$ we get the class-dependent error probability at a point $x_k \in \mathcal{S}$:

$$Pe^{(RRC)}(j|x_k) = P^{(RRC)}(j|x_k), \quad x_k \in \mathcal{S}, \; j(\neq j_k) \in \mathcal{M}. \tag{6}$$

In the next sections probabilities of correct classification (5) and conditional probabilities of error (6) for learning objects will be utilized for determining the competence and diversity functions of base classifiers.

2.3. Measure of classifier competence

Since the RRC can be considered equivalent to the modeled base classifier $\psi_l \in \Psi$, it is justified to use the probability (5) as the competence of the classifier ψ_l at the learning point $x_k \in \mathcal{S}$, i.e.:

$$C(\psi_l|x_k) = Pc^{(RRC)}(x_k). \tag{7}$$

The competence values for the validation objects $x_k \in \mathcal{S}$ can be then extended to the entire feature space $\mathcal{X}$. To this purpose the following normalized Gaussian potential function model was used [8]:

$$C(\psi_l|x) = \frac{\sum_{x_k \in \mathcal{S}} C(\psi_l|x_k) \exp(-dist(x, x_k)^2)}{\sum_{x_k \in \mathcal{S}} \exp(-dist(x, x_k)^2)}, \tag{8}$$

where $dist(x, y)$ is the Euclidean distance between the objects x and y.

2.4. Measure of diversity of classifiers ensemble

As it was mentioned previously, the diversity of a classifier ensemble Ψ_E is considered as an independency of the errors made by the member classifiers. Hence the method in which the diversity measure is calculated as a variety of class-dependent error probabilities is fully justified.

Similarly, as in competence measure, we assume that at a learning point $x_k \in \mathcal{S}$ the conditional error probability for the class $j \neq j_k$ of the base classifier ψ_l is equal to the appropriate probability of the equivalent RRC, namely:

$$Pe^{(\psi_l)}(j|x_k) = Pe^{(RRC)}(j|x_k). \tag{9}$$

Next, these probabilities can be extended to the entire feature space $\mathcal{X}$ using Gaussian potential function (8):

$$Pe^{(\psi_l)}(j|x) = \frac{\sum_{x_k \in \mathcal{S}, j_k \neq j} Pe^{(\psi_l)}(j|x_k)\exp(-dist(x,x_k)^2)}{\sum_{x_k \in \mathcal{S}, j_k \neq j} \exp(-dist(x,x_k)^2)}. \tag{10}$$

According to the presented concept, using probabilities (10), first we calculate pairwise diversity at the point $x \in \mathcal{X}$ for all pairs of base classifiers ψ_l and ψ_k from the pool Ψ:

$$D(\psi_l, \psi_k|x) = \frac{1}{M} \sum_{j \in \mathcal{M}} |Pe^{(\psi_l)}(j|x) - Pe^{(\psi_k)}(j|x)|, \tag{11}$$

and finally, we get the diversity of the ensemble of n $(n \leq L)$ base classifiers $\Psi_E(n)$ at a point $x \in \mathcal{X}$ as a mean value of pairwise diversities (11) for all pairs of member classifiers, namely

$$D(\Psi_E(n)|x) = \frac{2}{n \cdot (n-1)} \sum_{\psi_l, \psi_k \in \Psi_E(n); l \neq k} D(\psi_l, \psi_k|x). \tag{12}$$

It should be noted that two different possibilities to optimize the problem of selecting the classifier ensemble have been proposed below. Due to the differences in the defined objectives and constraints, we use the non-pairwise diversity measure (12) for Problem 1 and the pairwise one (11) for Problem 2 [10].

3. Dynamic ensemble selection systems

The design of DES system may be formulated as an optimization problem in which we look for such value of decision variable for which the objective function takes an extreme value, subject to constraints imposed on decision. In the considered problem, the decision answers the question of which base classifiers should be selected as member classifiers of an ensemble of size n $(n \leq L)$ $\Psi_E(n)$ for classification of a test point $x \in \mathcal{X}$.[2]

Two DES systems can be formulated depending on the role which competence and diversity measures play in optimization problem.

In the procedure of DES-CD$_{\text{d-opt}}$ system design, the diversity measure (12) of an ensemble makes the objective function, whereas competence (8) of member classifiers are included in constraints. In other words, the DES-CD$_{\text{d-opt}}$ system maximizes the diversity of the ensemble and simultaneously keeps competence of member classifiers on an acceptable level.

In the procedure of DES-CD$_{\text{c-opt}}$ system design, the role of both measures is exactly reversed: the total competence of member classifiers creates the objective function and the diversity of the ensemble is a constraint in optimization problem. It means, that the DES-CD$_{\text{c-opt}}$ system maximizes the sum of competence of member classifiers and simultaneously keeps the ensemble relatively diverse.

The next two subsections describe both DES systems in detail.

[2] Formally, the decision variable has the form of binary sequence of size L in which 1 (0) at the l-th position $(l = 1, 2, \ldots, L)$ denotes that base classifier ψ_l has been selected (has not been selected) as a member classifier of an ensemble $\Psi_E(n)$.

3.1. DES-CD$_{d\text{-}opt}$ system

This system is constructed as follows:

(1) For a given test pattern $x \in \mathcal{X}$ the competence (8) are calculated for each base classifier and pairwise diversities (11) are calculated for all pairs of base classifiers from the pool Ψ;

(2) The ensemble $\Psi_E^*(n)$ is found as a solution of the following optimization problem (Problem 1):

$$D(\Psi_E^*(n)|x) = \max_{\Psi_E(n)} D(\Psi_E(n)|x), \tag{13}$$

subject to constraints

$$C(\psi_l|x) \geq \alpha \quad \text{for} \quad \psi_l \in \Psi_E^*(n), \tag{14}$$

where α $(0 \leq \alpha \leq 1)$ is a given competence threshold value.

(3) The supports of member classifiers of the ensemble $\Psi_E^*(n)$ are combined by the weighted sum method:

$$d_j^{(d-opt)}(x) = \sum_{\psi_l \in \Psi_E^*(n)} C(\psi_l|x)\, d_{jl}(x) \tag{15}$$

and finally, the DES-CD$_{\text{d-opt}}$ system classifies x according to the maximum rule:

$$\psi^{d-opt}(x) = i \Leftrightarrow d_i^{(d-opt)}(x) = \max_{j \in \mathcal{M}} d_j^{(d-opt)}(x). \tag{16}$$

3.2. DES-CD$_{c\text{-}opt}$ system

This system is the same as the DES-CD$_{\text{d-opt}}$ system except for step 2. Now, the ensemble $\Psi_E^*(n)$ is found as a solution of the following optimization problem (Problem 2):

$$\sum_{\psi_i \in \Psi_E^*(n)} C(\psi_i|x) = \max_{\Psi_E(n)} \sum_{\psi_i \in \Psi_E(n)} C(\psi_i|x), \tag{17}$$

subject to constraint

$$D(\psi_l, \psi_k|x) \geq \beta, \tag{18}$$

where β $(0 \leq \beta \leq 1)$ is a given diversity threshold value.

3.3. Solution of optimization problems

Problems 1 and 2 are combinatorial optimization problems in which we have to choose the best solution from a finite number of solutions. It is obvious that for both problems the number of feasible solutions is equal to $\binom{L}{n}$. For example, if the size of pool of base classifiers is $L = 50$ and if we want to obtain the ensemble containing $n = 10$ classifiers, then the set of possible solutions is equal to $50!/10!40! = 10,272,278,170$. This means that, even for typical sizes of DES system, the exhaustive enumeration method for the solution of optimization problems (13) and (17) is completely ineffective. In order to solve these problems we propose to apply the simulated annealing (SA) algorithm, which has demonstrated to be an effective method for different optimization problems [17–20]. The main reason why SA was chosen in this paper was the speed of its operation. In the pretests, it turned out to be faster than other heuristic algorithms, such as tabu search or genetic algorithms. The proposed classification algorithms based on RRC have a high computational complexity, and therefore a fast optimization algorithm selection was crucial. In addition, the SA algorithm gives a lot of possibilities for parameterization of the optimization process. SA is a random-search technique which exploits an analogy between the way in which a metal cools and freezes into a minimum energy structure and the search for minimal value of the objective function [12,21]. In this

Table 1
Pseudocode of the solution of Problem 1.

Input data:
$\mathcal{S}$ – learning set;
Ψ_L – the pool of classifiers;
n – the size of enesemble;
$x \in \mathcal{X}$ – the testing point;
α – the threshold of competence
T – current temperature; initial value of T is defined as algorithm input parameter
T_{min} – minimum temperature
1. For each $\psi_l \in \Psi_L$ calculate competence $C(\psi_l|x)$ at the point x
2. Create temporal set of competent classifiers at the point x $\Psi(x) = \{\psi_l \in \Psi_L : C(\psi_l|x) \geq \alpha\}$
3. $\Psi_E^*(n) = \{\psi_{(1)}, \psi_{(2)}, \ldots, \psi_{(n)}\}$ and $\Psi(x) = \Psi(x) - \{\psi_{(1)}, \psi_{(2),\ldots,\psi_{(n)}}\}$ where $\{\psi_{(1)}, \psi_{(2)}, \ldots, \psi_{(n)}\}$is the randomly selected subset
4. until $T > T_{min}$
   (a) Randomly change a random classifier from $\Psi_E^*(n)$ to one from $\Psi(x)$ and store a new set as $\Psi_E^{**}(n)$
   (b) If diversity $D(\Psi_E^{**}(n)|x)$ is better than the best solution so far; store the $\Psi_E^{**}(n)$ as the best solution
   (c) If $rv(0,1) < e^{(-((D(\Psi_E^{**}(n)|x) - D(\Psi_E^{*}(n)|x))/T))}$ accept the change - $\Psi_E^*(n) = \Psi_E^{**}(n)$;
   ($rv(0,1)$ is a random value uniformly distributed on [0, 1])
   (d) $T = 0.95$ T

Table 2
Pseudocode of the solution of the Problem 2.

Input data:
$\mathcal{S}$ – learning set;
Ψ_L – the pool of classifiers;
n – the size of ensemble;
$x \in \mathcal{X}$ – the testing point;
β – the threshold of diversity
T – current temperature; initial value of T is defined as algorithm input parameter
T_{min} – minimum temperature
1. For each pair of classifiers ψ_l and $\psi_k \in \Psi_L$ calculate pairwise diversity $D(\psi_l, \psi_k|x)$ at the point x
2. Create temporal set of diversed classifiers at the point x where $\Psi(x) = \{\psi_l, \psi_k \in \Psi_L : D(\psi_l, \psi_k|x) \geq \beta\}$ where $l \neq k$
3. $\Psi_E^*(n) = \{\psi_{(1)}, \psi_{(2)}, \ldots, \psi_{(n)}\}$ and $\Psi(x) = \Psi(x) - \{\psi_{(1)}, \psi_{(2),\ldots,\psi_{(n)}}\}$where $\{\psi_{(1)}, \psi_{(2)}, \ldots, \psi_{(n)}\}$ is the randomly selected subset
4. until $T > T_{min}$
   (a) Randomly change a random classifier from $\Psi_E^*(n)$ to one from $\Psi(x)$ and store a new set as $\Psi_E^{**}(n)$
   (b) If $\sum_{\psi_i \in \Psi_E^{**}(n)} C(\psi_i|x)$ is greater than the best solution so far; store the $\Psi_E^{**}(n)$ as the best solution
   (c) If $rv(0,\ 1) < e^{(-((\sum_{\psi_i \in \Psi_E^{**}(n)} C(\psi_i|x) - \sum_{\psi_i \in \Psi_E^{*}(n)} C(\psi_i|x))/T))}$accept the change $\Psi_E^*(n) = \Psi_E^{**}(n)$;
   ($rv(0,1)$ is a random value uniformly distributed on [0,1])
   (d) $T = 0.95$ T

method, the following elements must be determined: (1) a representation of possible solutions, (2) a procedure of random changes in solutions, (3) a method of evaluating the objective function and (4) an annealing schedule, i.e. an initial temperature and rules for lowering it s the search procedure progresses.

Application of the SA algorithm in the described optimization problems allows us to create new methods, pseudocodes of which are presented in Tables 1 and 2.

4. Experiments

In order to study the performance of the developed DES systems two computer experiments were made using 7 benchmark databases. In the first experiment, the two constructed systems were evaluated for different threshold values in the constraints (14) and (18) of optimization problems and the values that showed the best performance of DES systems were identified. In the second experiment, the DES systems with the best values of thresholds were compared against other multiple classifier systems (MCSs).

4.1. Databases and experimental setup

The benchmark databases used in the experiments were taken from the UCI Machine Learning Repository and StatLib statistical datasets. A brief description of the databases is given in Table 3.

The experiments were conducted in MATLAB using PRTools, which automatically normalizes feature vectors for zero mean and unit standard deviation and, for a given $x \in \mathcal{X}$, produces classifying functions (supports) for all base classifiers according to the paradigms of their activity [22]. The training and testing datasets were extracted from each database using two-fold cross-validation. The base classifiers and both competence and diversity measures were trained using the same training dataset.

Two types of classifier ensembles were used in the experiments: homogeneous and heterogeneous. The homogeneous ensemble consisted of 20 pruned decision tree classifiers with Gini splitting criterion. To prevent overlearning and obtaining diversity between classifiers, each classifier was trained using randomly selected 70% of objects from the training dataset. The proposed percentage has been determined experimentally.

The pool of heterogeneous base classifiers used in the experiments consisted of the following nine classifiers [23]: (1–2) linear (quadratic) discriminant classifier based on normal distributions with the same (different) covariance matrix for each class; (3) nearest mean classifier; (4–6) k-NN - k-nearest neighbours classifiers with $k=1$, 5, 15; (7–8) Parzen classifier with the Gaussian kernel and the optimal smoothing parameter h_{opt} (and the smoothing parameter $h_{opt}/2$); (9) pruned decision tree classifier with Gini splitting criterion.

4.2. Experiment 1

In this experiment the influence of values of thresholds α and β on the classification quality of DES systems was examined.

For competence threshold α five levels were applied in experiments: $\alpha \in \{1/M,\ 1/M + \alpha',\ 1/M + 2\alpha',\ 1/M + 3\alpha',\ 1/M + 4\alpha'\}$, where $\alpha' = (0.9 - 1/M)/4$ and M denotes the number of classes. Such a choice evenly covers the competence interval from the value $1/M$, which refers to competence of random-guessing classifier, to the value 0.9, which was accepted as the maximal practical threshold of competence.

In order to define values of diversity threshold β, first, some pretests were conducted which enabled the maximum value of diversity D_{max} to be calculated for each database and for given size of ensemble. Next, for diversity threshold β, four levels were defined: $\beta \in \{0.2D_{max}, 0.4D_{max}, 0.6D_{max}$ and $0.8D_{max}\}$.

Table 3
The databases used in the experiments.

Data set	Source	# Objects	# Features	# Classes
Breast C. W.	UCI	699	9	2
Biomed	StatLib	194	5	2
Glass	UCI	214	9	4
Iris	UCI	150	4	3
Sonar	UCI	3823	64	10
Ionosphere	UCI	351	34	2
CNAE-9	UCI	1080	856	9

Half of the number of base classifiers fulfiling constraints in optimization problems was adopted as the ensemble size n (but no less than 2), i.e. $n = \max\{1/2|\Psi(x)|,\ 2\}$.

4.3. Experiment 2

In this experiment the DES-CD$_{\text{d-opt}}$ and DES-CD$_{\text{c-opt}}$ systems with the best competence/diversity thresholds identified in the previous experiment were compared against three multiclassifier systems:

(1) SB (the single best) system – this system selects the single best classifier in the pool [2];
(2) MV (majority voting) system – this system is based on majority voting of all classifiers in the pool [2];
(3) DES-SC system – this system defines the competence of a base classifier ψ for a test object x according to (8) and next the ensemble of competent (better-than-random) classifiers is selected – the final decision is made as in (16).

4.4. Results and discussion

Experiment 1. Classification accuracies (i.e. the percentage of correctly classified objects) averaged over 20 runs (10 replications of two-fold cross validation) for experiment 1 are shown in Tables 4–7.

Values of thresholds α and β significantly affect quality of DES systems. For the parameter α and for heterogeneous (homogeneous) classifiers the maximum difference of classification accuracy ranges from 0.87% (Iris) to 6.8% (Sonar) (from 1.17% (Iris) to 5.05% (Sonar)). The corresponding ranges for the parameter β are as follows: for heterogeneous classifiers – from 0.64% (Iris) to 10.46% (Ionosphere); for homogeneous classifiers – from 1.13% (Breast C.W.) to 11.76% (Sonar).

For heterogeneous classifiers the best classification accuracies were obtained for smaller values of the threshold α (for $\alpha = 1/M$, $1/M + \alpha'$). For homogeneous classifiers the best classification accuracies were obtained for the middle value of the threshold $\alpha = 1/M + 2\alpha'$.

Table 4
Dependence of classification accuracies % of the DES-CD$_{\text{d-opt}}$ using heterogeneous ensembles from α threshold.

Benchmark database name	$\frac{1}{M}$	$\frac{1}{M}+\alpha'$	$\frac{1}{M}+2\alpha'$	$\frac{1}{M}+3\alpha'$	$\frac{1}{M}+4\alpha'$
Breast C.W.	98.27	98.65	98.01	95.37	95.42
Biomed	90.53	90.92	87.63	87.52	87.61
Glass	74.09	74.11	69.54	69.51	69.39
Iris	97.71	97.68	96.81	96.83	96.79
Sonar	83.33	76.48	76.52	76.81	76.63
Ionosphere	90.49	90.51	88.91	86.72	86.61
CNAE-9	88.25	88.36	86.67	85.86	85.21

Table 5
Dependence of classification accuracies % of the DES-CD $_{\text{d-opt}}$ using homogeneous ensembles from α threshold.

Benchmark database name	$\frac{1}{M}$	$\frac{1}{M}+\alpha'$	$\frac{1}{M}+2\alpha'$	$\frac{1}{M}+3\alpha'$	$\frac{1}{M}+4\alpha'$
Breast C.W.	96.39	96.46	96.53	95.45	95.24
Biomed	87.26	87.88	88.31	87.31	86.91
Glass	73.10	74.41	75.22	73.86	72.91
Iris	91.89	91.87	92.01	91.65	90.84
Sonar	78.50	79.36	81.26	79.16	76.21
Ionosphere	90.43	90.53	90.48	90.21	89.36
CNAE-9	88.39	88.67	85.98	86.29	85.04

Table 6
Dependence of classification accuracies % of the DES-CD $_{\text{c-opt}}$ using heterogeneous ensembles from γ threshold.

Benchmark database name	$0.2D_{max}$	$0.4D_{max}$	$0.6D_{max}$	$0.8D_{max}$
Breast C.W.	97.67	98.01	97.26	95.43
Biomed	89.93	89.27	84.59	83.23
Glass	73.86	75.21	69.83	65.91
Iris	96.81	97.21	96.98	96.57
Sonar	81.03	79.96	74.69	71.59
Ionosphere	87.29	89.97	84.98	79.51
CNAE-9	86.51	87.21	87.55	86.95

Table 7
Dependence of classification accuracies % of the DES-CD$_{\text{c-opt}}$ using homogeneous ensembles from γ threshold.

Benchmark database name	$0.2D_{max}$	$0.4D_{max}$	$0.6D_{max}$	$0.8D_{max}$
Breast C.W.	96.02	95.69	95.23	94.89
Biomed	86.31	86.38	85.27	83.04
Glass	73.08	72.19	67.59	62.67
Iris	90.86	90.37	90.29	88.29
Sonar	74.98	77.05	67.98	65.29
Ionosphere	89.25	89.88	86.27	80.53
CNAE-9	87.05	87.86	85.22	83.23

The best classification accuracies for both homogeneous and heterogeneous ensembles were achieved for smaller values of the threshold $\beta = 0.2D_{max}$, $0.4D_{max}$.

Experiment 2. The results obtained for the MCSs using heterogeneous and homogeneous ensembles are shown in Table 8. For each database and for the DES systems, the mean sizes of classifier ensembles are given under the classification accuracy. The row “Average” contains results averaged over all datasets.

Statistical differences between the performances of the DES-CD systems and the three MCSs were evaluated using Student's *t*-test [24]. The level of $p < 0.05$ was considered as statistically significant. In Table 8, statistically significant differences are given as indices of the systems evaluated, e.g. for the Biomed database and the heterogeneous ensemble the DES-CD$_{\text{d-opt}}$ system produced statistically different classification accuracies from the SB and MV systems.

These results imply the following conclusions:

(1) The DES-CD$_{\text{d-opt}}$ system outperformed the SB, MV, DES-CS, DES-CD$_{\text{c-opt}}$ classifiers by 7.32%, 3.80%, 0.35% and 0.72% for heterogeneous ensemble and by 7.83%, 2.41%, 1.21% and 1.62% for homogeneous ensemble, respectively;
(2) The DES-CD$_{\text{d-opt}}$ system achieved the highest classification accuracy for 6 datasets for heterogeneous and 7 for homogeneous ensembles; it produced statistically significant higher scores in 27 out of 56 cases;
(3) There is a statistically significant difference between the classification accuracies of the DES-CS and the DES-CD$_{\text{d-opt}}$ systems in one database for heterogeneous ensembles and in one database for homogeneous ensemble;
(4) The relative difference between the mean ensemble sizes for the DES-CS and the DES-CD$_{\text{d-opt}}$ systems is on average equal to 49.21% and 50.79% for heterogeneous and homogeneous ensembles, respectively;
(5) The relative difference between the mean ensemble sizes for the DES-CS and the DES-CD$_{\text{c-opt}}$ systems is on average equal to 36.68% and 55.5% for heterogeneous and homogeneous ensembles, respectively;

Table 8
Classification accuracies of the MCSs using heterogeneous/homogeneous ensembles. The mean sizes of classifier ensembles and statistically significant differences are given under the classification accuracies. The best result for each database is bolded.

Database	SB (1)	MV (2)	DES-CS (3)	DES-CD	
				d-opt (4)	c-opt (5)
Breast C.W.	95.51/94.86	96.25/95.98	98.06/96.19 8.51/19.28 1,2/1,2	**98.65/96.53** 4.7/9.61 1,2/1,2	98.01/96.02 5.89/8.25 1,2/1,2
Biomed	83.90/83.30	87.50/86.73	90.32/87.48 8.03/18.03 1,2/1,2	**90.92/88.31** 4.38/8.98 1,2/1,2	89.93/86.38 5.35/8.18 1,2/1
Glass	71.41/61.43	69.99/71.07	**75.95**/72.65 8.2/19.35 1,2,4,5/1,2,5	74.11/**75.22** 4.79/9.29 1,2/1,2,5	75.21/73.08 5.12/8.47 1,2/1,2
Iris	95.93/91.41	97.07/90.61	96.41/91.41 7.26/19.79 /2	**97.71/92.01** 4.43/9.13 1,3/2	97.21/90.86 4.59/9.03 1/
Sonar	73.60/69.96	76.54/76.61	82.59/78.47 8.58/18.97 1,2/1,2	**83.33/81.26** 4.68/9.89 1,2/1,2,3	81.03/77.05 4.87/9.02 1,2/1
Ionosphere	84.78/88.50	86.14/90.03	90.47/90.44 8.33/19.04 1,2/1	**90.51/90.53** 4.54/9.28 1,2/1	89.97/89.88 5.01/8.11 1,2/
CNAE-9	67.29/68.24	83.54/84.63	87.89/87.38 7.55/18.93 1,2/1	**88.36/88.67** 4.26/9.47 1,2/1	87.21/87.86 4.93/8.28 1,2/
Avarage	81.77/79.67	85.29/85.09	88.74/86.29 8.07/19.06	**89.09/87.50** 4.54/9.38	88.37/85.88 5.11/8.48

Based on these experiments, it can be concluded that DES-CD$_{d\text{-}opt}$ has obtained the best results thanks to a hybrid approach to the problem. Therefore, the ensemble of classifiers selected by the proposed method consist of only competent classifiers. At the same time, those classifiers commit various errors. This is the reason why the DES-CD$_{d\text{-}opt}$ algorithm was able to increase the quality of recognition. The second proposed algorithm DES-CD$_{c\text{-}opt}$ has acquired worse because choosing a highly diversified classifiers created the possibility of rejecting competent ones.

5. Conclusion

In this study a novel method for dynamic ensemble selection has been proposed using probabilistic measures of competence and diversity of member classifiers. These measures are calculated on the basis of the original concept of the randomized reference classifier (RRC). RRC acts – on average – as an evaluated classifier and hence its probability of correct classification can be considered as the competence of that classifier and the probability of misclassification can be used for the construction of measuring ensemble diversity.

Results of the experimental investigations indicate that the proposed method can eliminate weak classifiers and keep the ensemble maximally diverse. This approach leads to the DES system for which classification accuracy (for 7 benchmark datasets regardless of the ensemble type used) is better than the classification accuracy of the DES system using only the competence measure or, on average, is very close to this accuracy but achieved by means of a smaller number of classifiers in the ensemble.

To the best of the authors' knowledge, the proposed approach to the DES system construction is the first method that simultaneously uses the measure of competence of base classifiers and the diversity measure of an ensemble.

References

[1] J. Kittler, M. Hatef, R. Duin, J. Matas, On combining classifiers, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (2006) 226–239.
[2] I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms, Wiley-Interscience, West Florida, 2004.
[3] K. Woods, W. Kegelmeyer, W. Bowyer, Combination of multiple classifiers using local accuracy estimates, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (1997) 405–410.
[4] L. Didaci, G. Giacinto, F. Roli, G. Marcialis, A study on the performances of dynamic classifier selection based on local accuracy estimation, Pattern Recognition 38 (2005) 2188–2191.
[5] P. Smits, Multiple classifier systems for supervised remote sensing image classification based on dynamic classifier selection, IEEE Transactions on Geoscience and Remote Sensing 40 (2002) 717–725.
[6] F. Huenupan, N. Yoma, Confidence based multiple classifier fusion in speaker verification, Pattern Recognition Letters 29 (2008) 957–966.
[7] G. Giacinto, F. Roli, Dynamic classifier selection based on multiple classifier behaviour, Pattern Recognition 34 (2001) 1879–1881.
[8] T. Woloszynski, M. Kurzynski, A probabilistic model of classifier competence for dynamic ensemble selection, Pattern Recognition 44 (2011) 2656–2668.
[9] M. Aksela, J. Laaksonen, Using diversity of errors for selecting members of a committee classifier, Pattern Recognition 39 (2006) 608–623.
[10] L. Kuncheva, C. Whitaker, Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy, Machine Learning 51 (2003) 181–207.
[11] E. Corchado, A. Abraham, A. de Carvalho, Editorial: hybrid intelligent algorithms and applications, Information Sciences: an International Journal 180 (2010) 2633–2634.
[12] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, Optimization by simulated annealing, Science 220 (1983) 621–630.
[13] M. Aksela, Comparison of classifier selection methods for improving committee performance, in: Proceedings of the Multiple Classifier Systems, pp. 84–93.
[14] J. Berger, Statistical Decision Theory and Bayesian Analysis, Springer Verlag, New York, 1987.
[15] R. Lysiak, M. Kurzynski, T. Woloszynski, Probabilistic approach to the dynamic ensemble selection using measures of competence and diversity of base classifiers, in: Proceedings of the Hybrid Artificial Intelligent Systems. Part II. Lecture Notes in Computer Science, vol. 6678, 2011, pp. 345–351.
[16] T. Woloszynski, Matlab Central File Enchange, ⟨http://www.mathwork.com/~matlabcentral/~fileenchange/~28391-classifier-competence-based-on-probabilistic-modeling⟩, 2010.
[17] J. Liu, Algorithm of QOS multicast routing based on genetic simulated annealing algorithm, Computer Application and System Modeling (ICCASM) 5 (2010), pp. V5–220–V5–223.

[18] A. Aly, Y. Hegazy, M. Alsharkawy, A simulated annealing algorithm for multi-objective distributed generation planning, Power and Energy Society General Meeting (2010) 1–7.
[19] C. Queirolo, L. Silva, O. Bellon, M. Segundo, 3d face recognition using simulated annealing and the surface interpenetration measure, Pattern Analysis and Machine Intelligence 32 (2010) 206–219.
[20] L. Zhong, J. Sheng, M. Jing, Z. Yu, X. Zeng, D. Zhou, An optimized mapping algorithm based on simulated annealing for regular NOC architecture, in: 9th International Conference on ASIC, 2011, pp. 389–392.
[21] D. Bertsimas, J. Tsitsiklis, Simulated annealing, Statistical Science 8 (1993) 10–15.
[22] R. Duin, P. Juszczak, P. Paclik, PRTools4. A Matlab Toolbox for Pattern Recognition, Delft University of Technology, Delft, 2007.
[23] R. Duda, P. Hart, G. Stork, Pattern Classification, John Wiley and Sons, New York, 2000.
[24] T. Dietterich, Approximate statistical tests for comparing supervised classification learning algorithms, Neural Computation 10 (1998) 1895–1923.

Rafal Lysiak in 2009 graduated five years daily studies at Wroclaw University of Technology. He Defended his thesis “Review of methods and analysis of the allocation algorithms in mesh structures, with particular approach of self-learning algorithms”. He graduated with a very good result and obtained a Masters degree in the field of designing networks. Then he decided to continue his education in Ph.D. programs in the Department of Systems and Networks. He is currently in his third year. In September 2010 he participated in summer school, “The International Summer School on Pattern Recognition” in Plymouth, UK.

Marek Kurzynski is an M.Sc. in Automatic Control, Wroclaw University of Technology, Faculty of Electronics – 1972, Ph.D. in Computer Science, Wroclaw University of Technology, Institute of Engineering Cybernetics, 1974, D.Sc. in Computer Science, Silesian Technical University, Faculty of Automation and Computer Science, 1987 and Professor's Scientific Tittle, 1998.

Tomasz Woloszynski presently works as a Graduate Research Assistant at the University of Western Australia.