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
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).

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.

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