Pattern Recognition Letters 34 (2013) 1623–1629

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

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

Classifier fusion with interval-valued weights

Robert Burduk *

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

ARTICLE INFO

Article history:
Available online 4 June 2013

Communicated by Manuel Graña

Keywords:
Classifier fusion
Interval information
Ensemble classifier
Multiple classifier system

ABSTRACT

The article presents a new approach of calculating the weight of base classifiers from a committee of classifiers. The obtained weights are interpreted in the context of the interval-valued sets. The work proposes four different ways of calculating weights which consider both the correctness and incorrectness of the classification. The proposed weights have been used in the algorithms which combine the outputs of base classifiers. In this work we use both the outputs, represented by rank and measure level. Research experiments have involved several bases available in the UCI repository and two data sets that have generated distributions. The performed experiments compare algorithms which are based on calculating the weights according to the resubstitution and algorithms proposed in the work. The ensemble of classifiers has also been compared with the base classifiers entering the committee.

1. Introduction

Creating a group of classifiers is one of the ways to improve the classification accuracy in the recognition process (Kuncheva, 2004; Kittler et al., 1998). In particular, the aim is to allow the ensemble of classifiers to obtain greater values of the classification correctness than in the case of a single classifier from the pool. Research regarding the problem of recognition has been in focus for more than fifteen years now. Due to the large number of different methods for combining classifiers and creating an ensemble of classifiers it is difficult to identify the best method for a particular recognition task (Alkoot and Kittler, 1999; Chen and Cheng, 2001; Kittler and Alkoot, 2003; Zhang and Duin, 2011). Problems involved in these areas are still evolving and there are new concepts associated with them (Cyganek, 2012; Woloszynski et al., 2011).

While analysing the outcomes of the base classifiers we can distinguish three primary situations (Xu et al., 1992). In the first one, the label description (crisp label) of the class of the recognized object is available. In the second, the labels obtained from a base classifier are ranked in a queue. In the third, each base classifier returns a posteriori probability of membership of the recognized object to each of the possible class labels. In recent years, many studies have presented different issues related to this recognition task. One of them defines the weights assigned to the given component classifiers. The selection of the appropriate system of weights has been widely discussed in the literature regarding the construction of the complex classifiers (Woods et al., 1997; Wozniak et al., 2009). This problem can also be formulated as a separate element of the process of learning in the classifiers committee (Kuncheva et al., 2000).

Recently, many papers describe the use of the interval information in pattern recognition (Bhadra et al., 2009; Kulczycki et al., 2011; Silva et al., 2006; Viertl, 1996). In particular, they refer to the inaccurate data description expressed by the interval information. The interval information is based on the interval analysis which belongs to the field of mathematics (Alefeld and Herzberger, 1983). Its advantage is modelling the uncertainty of the given value in the simplest way possible. The investigated value meets the dependency and thus can be regarded as the value in the range.

In the article, four ways of calculating weights of the classifiers entering the classifiers' committee will be suggested. These weights will be interpreted as the lower and upper values of the base classifiers' weights. The defined boundaries refer to the correctness or incorrectness of these classifiers.

The text is organized as follows: in Section 2 the definitions associated with the classifier fusion are presented. In particular, two approaches will be introduced that differ in the type of data received from the base classifiers' outcomes. In Section 3 the new methods of assigning weights of individual base classifiers are presented. Section 4 includes the description of research experiments comparing the suggested algorithms with others that are based on the same data received on the outcome of the of base classifiers. Finally, conclusions from the experiments are presented.

2. Classifier fusion

Let us assume that we possess K of different classifiers $\Psi_1, \Psi_2, \ldots, \Psi_K$. Such a set of classifiers, which is constructed on the basis of the same learning sample is called an ensemble classifier or a combining classifier. However, each of the Ψ_i classifiers is

* Tel.: +48 71 320 2877; fax: +48 71 320 2902.
E-mail address: robert.burduk@pwr.wroc.pl

0167-8655/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.05.022

described as a component or base classifier. As a rule K is assumed to be an odd number and each of Ψ_i classifiers makes an independent decision. As a result, of all the classifiers' action, their K responses are obtained. Having at the disposal a set of base classifiers one should determine the procedure of making the ultimate decision regarding the allocation of the object to the given class. It implies that the output information from all K component classifiers is applied to make the ultimate decision.

2.1. Fusion techniques for class labels

One of the possible types of information obtained from the component classifiers is a class label that is assigned to the given observation. Having at the disposal a set of K labels and in order to obtain the final decision different methods of connecting outputs of classifiers' sets are applied. The way of reaching the decision is based on counting the votes and is defined as the voting method.

One of the most common methods of connecting classifiers is the majority voting. The method implies that each of the component classifiers of the committee gives a rightful vote and the object is assigned to the class which gets the largest number of votes given by the base classifiers. The advantage of the method is its simplicity and lack of any calculation apart from counting votes of the individual classifiers. One of the drawbacks of the approach to counting scores is the draw situation, which denotes that the same number of classifiers points to more than one class. In the tasks of binary classification the solution of the problem can be using the odd number of base classifiers. The algorithm of making the ultimate decision by the set of classifiers in this approach is the following:

$$\Psi_{MV}(x) = \arg\max_{1 \leqslant i \leqslant M} \sum_{k=1}^{K} I(\Psi_k(x) = i), \tag{1}$$

where i denotes the set of class labels and $I(\cdot)$ is the indicator function.

Another method of combining the classifiers is the weighted voting. In this approach each of the classifiers has an allocated weight, which is taken into account when reaching the final decision of the group. Weights depend largely on the quality of their base classifiers. In the case when each classifier has one weight for all the possible classes or for the complete features space an adequate group classification formula is presented as follows:

$$\Psi_{wMV}(x) = \arg\max_{1 \leqslant i \leqslant M} \sum_{k=1}^{K} w_k * I(\Psi_k(x) = i), \tag{2}$$

where $w_k = 1 - Pe_{\Psi_k}$, and Pe_{Ψ_k} is the empirical error of Ψ_k classifier estimated on the testing set. In the case when the error is estimated on the learning set, we can talk about the estimation error based on the resubstitution method. Then w_k weight of each component classifier is calculated depending on the:

$$w_k = \frac{\sum_{n=1}^{N} I(\Psi_k(x_n) = i, j_n = i)}{N}. \tag{3}$$

The N value refers to the number of the learning set observations, which is used for estimating classifiers' weights, and j_n is the class number of the object with n index.

Another way of calculating weights is the approach of giving each of the classifiers as much weight as there are pre-defined classes in the recognition task. In this case, the classification rule is the following:

$$\Psi_{wcMV}(x) = \arg\max_{1 \leqslant i \leqslant M} \sum_{k=1}^{K} w_{ki} * I(\Psi_k(x) = i), \tag{4}$$

with $w_{ki} = \frac{\sum_{n=1}^{N} I(\Psi_k(x_n)=i, j_n=i)}{\sum_{n=1}^{N} I(j_n=i)}$.The obtained weights are normalised for each of the classes according to the formula:

$$\sum_{k=1}^{K} w_{ki} = 1, \tag{5}$$

which means that the sum of weights of all the base classifiers for the given class i is equal to unity.

2.2. Fusion techniques for a posteriori probability

Suppose that we have at our disposal the probability evaluation that the tested object belonging to each of the classes for all base classifiers. In particular, this evaluation is a posteriori probability, which can be obtained in the way of the parametric as well as non-parametric estimation. Let us denote a posteriori probability estimation by $\hat{p}_k(i|x),\ k = 1, 2, \ldots, K,\ i = 1, 2, \ldots, M$. In the literature several methods of denoting the scores of classifiers' groups have been proposed for this type of the problem. We can distinguish the sum, prod and mean methods. In the sum method the score of the group of classifiers is based on the application of the following sums:

$$s_i(x) = \sum_{k=1}^{K} \hat{p}_k(i|x), \quad i = 1, 2, \ldots, M, \tag{6}$$

however, following the prod method a posteriori probability prods are applied:

$$pr_i(x) = \prod_{k=1}^{K} \hat{p}_k(i|x), \quad i = 1, 2, \ldots, M. \tag{7}$$

The final decision of the group of classifiers is made following the maximum rule and is presented accordingly, depending on the sum method (6) or the prod method (7):

$$\Psi_{SUM}(x) = \arg\max_{i} s_i(x), \tag{8}$$

$$\Psi_{PROD}(x) = \arg\max_{i} pr_i(x). \tag{9}$$

In the presented methods (8) and (9) discrimination functions obtained from the individual classifiers take an equal part in building the combined classifier. Also, the weighted versions of these methods can be created effortlessly. In this case, similarly as in the case of classifiers (2) and (4) one needs to formulate firstly the way of calculating the individual weights of classifiers. It can be done according to the dependence (3), which means that each classifier is assigned weight depending on its classification error.

3. Interval-valued weights in classifier fusion

The methods of calculating weights, which were presented in the previous chapter take into account the quality of classification of the individual base classifiers. The calculation is done simultaneously for all the classifiers and is concluded by the normalisation process. The calculated weights do not therefore consider decisions made by other classifiers. We will now suggest a new way of calculating weights, which takes into account the result of all base classifiers entering the committee, for each of the learning objects. Two main cases will be discussed. In the first one the correctness of the classification will be taken into consideration. The second case, however, will refer to the incorrectness of the base classifiers. For each of the cases the upper and lower value of the base classifiers' weight will be suggested. The obtained range between the upper and lower values defines the uncertainty in estimating the quality