Available online at www.sciencedirect.com

ScienceDirect

Pattern Recognition 40 (2007) 2198–2210

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/pr

Pairwise fusion matrix for combining classifiers

Albert H.R. Ko[a,*], Robert Sabourin[a], Alceu de Souza Britto Jr.[b], Luiz Oliveira[b]

[a]*LIVIA, École de Technologie Supérieure, University of Quebec, 1100 Notre-Dame West Street, Montreal, Que., Canada H3C 1K3*
[b]*PPGIA, Pontifical Catholic University of Parana, Rua Imaculada Conceicao, 1155, PR 80215-901, Curitiba, Brazil*

Received 6 July 2006; received in revised form 13 October 2006; accepted 26 January 2007

Abstract

Various fusion functions for classifier combination have been designed to optimize the results of ensembles of classifiers (EoC). We propose a pairwise fusion matrix (PFM) transformation, which produces reliable probabilities for the use of classifier combination and can be amalgamated with most existent fusion functions for combining classifiers. The PFM requires only crisp class label outputs from classifiers, and is suitable for high-class problems or problems with few training samples. Experimental results suggest that the performance of a PFM can be a notch above that of the simple majority voting rule (MAJ), and a PFM can work on problems where a behavior–knowledge space (BKS) might not be applicable.

Keywords: Fusion function; Combining classifiers; Confusion matrix; Pattern recognition; Majority voting; Ensemble of learning machines

1. Introduction

Different classifiers usually make different errors on different samples, which means that we can arrive at an ensemble that makes more accurate decisions by combining classifiers [1–9]. For this purpose, diverse classifiers are grouped together into what is known as an ensemble of classifiers (EoC). There are two problems in optimizing the performance of an EoC: first, how classifiers are selected, given a pool of different classifiers, to construct the best ensemble; and second, given all the selected classifiers, choosing the best rule to combine their outputs. These problems are fundamentally different, and should be solved separately to reduce the complexity involved in optimizing EoCs; the former focuses on ensemble selection [3,6,10–14] and the latter on ensemble combination, i.e. the choice of fusion functions [2,5,9,14,15]. Various fusion functions for classifier combination have been designed to facilitate a consensus decision from the outputs of each individual classifier. Through experimentation, some fusion functions have been shown to perform better than the single best classifier. But, we have no adequate understanding of the reasons why some classifier combination schemes are better than others [2,7,14,16,17].

An important consideration in classifier combination is that much better results can be achieved if diverse classifiers, rather than similar classifiers, are combined. There are several methods for creating diverse classifiers, among them are Random Subspaces [18], Bagging and Boosting [19–21]. The Random Subspaces method creates various classifiers by using different subsets of features to train them. Bagging generates diverse classifiers by randomly selecting subsets of samples to train classifiers. Boosting also uses parts of samples to train classifiers, but not randomly; in this case, difficult samples have a greater probability of being selected and easier samples have less chance of being used for training. To summarize, diverse classifiers are needed to optimize the performance of an EoC, as well as an adequate fusion function for classifier combination. A number of different combination schemes have been suggested [2,5–7,9,11,14,15,22,23]. In general, two kinds of fusion functions are available: (a) fusion functions of label outputs, such as majority voting, behavior–knowledge space (BKS), naive Bayes (NB) methods, etc. and (b) fusion functions of continuous-value outputs, which require the class probabilities

* Corresponding author.
E-mail addresses: albert@livia.etsmtl.ca (A.H.R. Ko), robert.sabourin@etsmtl.ca (R. Sabourin), alceu@ppgia.pucpr.br (A. Britto), soares@ppgia.pucpr.br (L. Oliveira).

0031-3203/$30.00 © 2007 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2007.01.031

outputs from classifiers. Different from the continuous-valued fusion functions, the label outputs fusion functions could not apply *a posteriori* probabilities of classes provided by each individual classifier. In the case where only class labels are offered as outputs by each individual classifier, then the simple majority vote rule (MAJ) is suggested.

To improve the performance of the fusion functions of label outputs, the BKS [11] has been proposed as an interesting fusion function that takes into account the interaction of classifiers. The method does not require any *a posteriori* probabilities of classes provided by each individual classifier. By contrast, it estimates the probability of each possible class label by constructing a table with $L+1$ dimensions for an ensemble of L classifiers, each dimension corresponds to the output of each classifier, and the additional dimension is for the true labels of concerned samples. By this means, with only the class label outputs of each classifier the BKS can estimate the likelihood of a given sample belonging to a class. The problem of the BKS is that it can apply only on low-dimensional problems. Moreover, in order to have an accurate probability estimation, it requires a large number of samples for the training.

On the other hand, the continuous-valued fusion functions require *a posteriori* probabilities of classes provided by each individual classifier and thus can use simple probability combination functions, such as sum, product, maximum and minimum. Moreover, they can also be more sophisticated classifier combination schemes than label outputs fusion functions, such as decision templates (DTs), Dempster–Shafer combination (DSC), fuzzy integral, or multilayer perceptrons (MLP) [6,11,22,23]. While it is true that these functions deal with the problem of combining classifiers as a problem of pattern recognition and take into account the interactions from classifiers, most of them do need further training. As insufficient training data usually lead to imperfect training, these sophisticated fusion functions might perform worse than the simple fusion functions [24]. It has, in fact, been suggested that, given insufficient training samples, simple fusion functions may outperform some trained fusion functions [24].

Herein lies the dilemma of EoCs. Given a limited number of samples, we need to take into account the interaction among classifiers. When the number of samples is too small, most trained fusion functions will not work well. For classifiers with crisp label outputs, this is especially a serious problem, because the number of fusion functions for label outputs is limited, and the BKS is suited neither to high-dimensional class problem nor to ensembles with a large number of classifiers. Therefore, we note three constraints for classifier combination: (a) classifiers without *a posteriori* probabilities of classes as outputs cannot use continuous-valued fusion functions; (b) trainable fusion functions need a number of samples for training, otherwise they will not perform well; (c) in most cases the independence of each classifier is the basic assumption. This assumption is, however, usually not true. Here are the key questions that need to be addressed:

(1) Can label outputs classifiers apply continuous-valued fusion functions?
(2) Can a trainable fusion function perform well without a large training data set?
(3) Can we take the interaction among classifiers into account in combining classifiers?

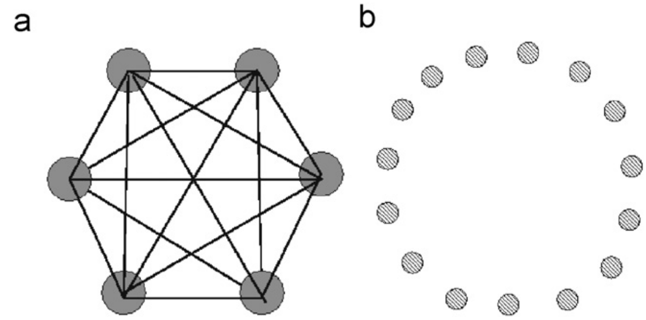

Fig. 1. An example of pairwise confusion matrices transformation in a six-classifier ensemble. (a) The original ensemble with six classifiers and (b) the transformation yields to 15 classifier pairs, each classifier pair is equal to the link between two classifiers in (a).

Given the challenge of combining classifiers, we suggest that the methods for combining classifiers can be improved by a simple transformation of an EoC into an ensemble of classifier pairs. We propose a pairwise fusion matrix (PFM) for classifier combination. A PFM is actually a three-dimensional confusion matrix consisting of the label outputs of any two classifiers and the real labels of samples. It is a method for transforming EoCs (Fig. 1) by which an ensemble of L classifiers is transformed into another ensemble of $L \times (L-1)/2$ classifier pairs.

With the prospect of using classifier pairs, it becomes possible to transform the crisp class label outputs into class probability outputs and thus allow the use of other fusion functions of continuous-valued outputs. At the same time we do take into account the interaction between classifiers in a pairwise manner. Moreover, the construction of PFM does not require as many samples needed for ensemble training as the BKS.

It is important to note that the classifier combination problem is very complex, and there are still a great many issues associated with it that we do not fully understand. It is difficult to say whether or not a method is better if we have an insufficient theoretical framework with which to assess it. The analysis and the method in this paper constitute only a small step towards a considerably improved understanding of classifier combination.

The paper is organized as follows. In Section 2, we introduce label outputs fusion functions for classifier combination. The proposed pairwise confusion matrices are presented in Section 3, and we discuss its relationship with BKS in Section 4. Experimental results are compared in Section 5. Discussion and our conclusion are presented in the remaining sections.

2. Fusion functions for label outputs classifier combination

Several fusion functions of label outputs for combining classifiers have been proposed [2,7,16,17]. These directly compare the outputs from all individual classifiers in an ensemble. Some related theoretical studies are presented in Refs. [2,7,17]. As