Contents lists available at ScienceDirect

Information Sciences

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

Empowering difficult classes with a similarity-based aggregation in multi-class classification problems

Mikel Galar [a,*], Alberto Fernández [b], Edurne Barrenechea [a], Francisco Herrera [c,d]

[a] *Departamento de Automática y Computación, Universidad Pública de Navarra, 31006 Pamplona, Spain*
[b] *Department of Computer Science, University of Jaén, 23071 Jaén, Spain*
[c] *Department of Computer Science and Artificial Intelligence, University of Granada, 18071 Granada, Spain*
[d] *Faculty of Computing and Information Technology – North Jeddah, King Abdulaziz University, 21589 Jeddah, Saudi Arabia*

ARTICLE INFO

Article history:
Received 1 June 2013
Received in revised form 15 October 2013
Accepted 29 December 2013
Available online 9 January 2014

Keywords:
Multi-class classification
Pairwise learning
One-vs-One
Decomposition strategies
Tuning
Difficult classes

ABSTRACT

One-vs-One strategy divides the original multi-class problem into as many binary classification problems as pairs of classes. Then, independent base classifiers are learned to face each problem, whose outputs are combined to predict a single class label. This way, the accuracy of the baseline classifiers without decomposition is usually enhanced, aside from enabling the usage of binary classifiers, i.e., Support Vector Machines, to solve multi-class problems. This paper analyzes the fact that existing aggregations favor easily recognizable classes; hence, the accuracy enhancement mainly comes from the higher correct classification rates over these classes. Using other evaluation criteria, the significant improvements of One-vs-One are diminished, showing a weakness due to the presence of difficult classes. Difficult classes can be defined as those obtaining a lower correct classification rate than that obtained by the other classes in the problem. After studying the problem of difficult classes in this framework and aiming to empower these classes, a novel similarity-based aggregation is presented, which generalizes the well-known weighted voting. The experimental analysis shows that the new methodology is able to increase the recognition of difficult classes, obtaining a more balanced performance over all classes, which is a desirable behavior. The methodology is tested within several Machine Learning paradigms and is compared with the state-of-the-art on aggregations for One-vs-One strategy. The results are contrasted by the proper statistical tests, as suggested in the literature.

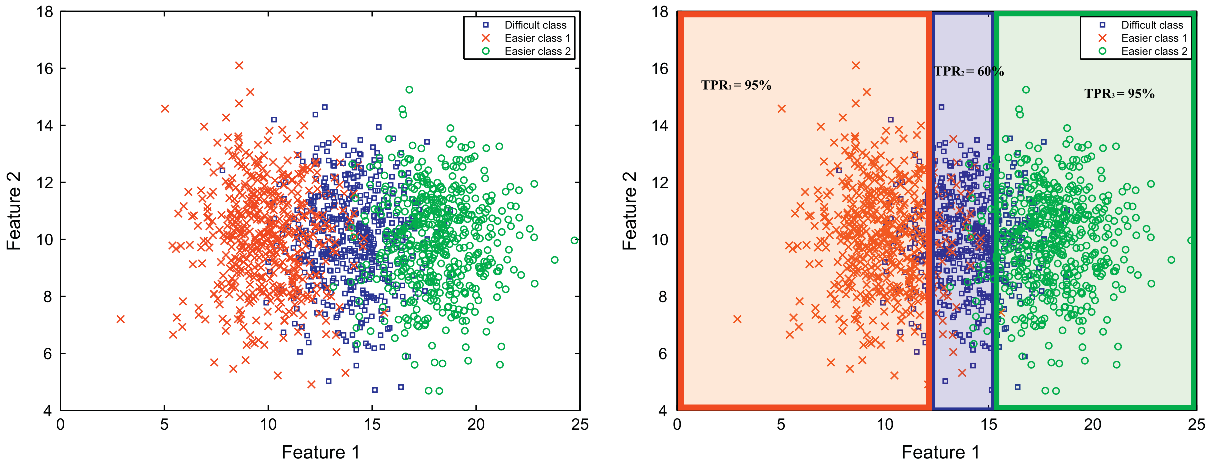
1. Introduction

Classification problems involving multiple categories are more general than their binary counterparts. When multiple classes are present, the complexity of finding the decision boundaries usually increases, making the construction of the classifiers more difficult. A number of real-world problems involve the classification of multiple classes, for instance, the classification of texts [44], microarrays [58] or textures [40].

Decomposition strategies [42] are commonly used to overcome these type of problems. In some cases, because the base classifier cannot deal with multiple classes by itself, whereas in others, because these strategies enhance the results of the baseline classifiers (without using decomposition) [24,56]. These strategies, also called binarization strategies, are based on divide and conquer paradigm, and most of them can be included within Error Correcting Output Codes (ECOC) [14,4]

[*] Corresponding author. Tel.: +34 948 166048; fax: +34 948 168924.
E-mail addresses: mikel.galar@unavarra.es (M. Galar), alberto.fernandez@ujaen.es (A. Fernández), edurne.barrenechea@unavarra.es (E. Barrenechea), herrera@decsai.ugr.es (F. Herrera).

0020-0255/$ - see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2013.12.053



(a) A three class problem with one difficult class.

(b) Classifier maximizing accuracy rate, even though a low TPR is achieved for the difficult class.

Fig. 1. An example of the difficult classes problem. The class in the center is more difficult to be correctly classified due to its overlapping with the other two classes.

framework. One of the most commonly used strategy is One-vs-One (OVO) [37], where the original problem is divided in as many pairs of classes as possible. When a new instance is presented to all the base classifiers, which were independently learned for each pair of classes, an aggregation is used to decide the final class label. In [24], an extensive review of different aggregations was carried out, concluding that the best aggregation depends on the base learner, but the weighted voting [35] and those based on probability estimates [62] were proven to be the most robust, even though no significant differences were often found.

In the specialized literature different studies analyzing the behavior of multi-class learning algorithms have been carried out, such as those learning class structures, which speed-up the test phase in problems with a huge amount of classes [9,64,45], those studying the consistency of multi-class classification [41,59] or works dealing with the calibration of probabilistic classifiers in multi-class problems [38,20].

Within a multi-class problem, the characteristics defining the classes are usually different: for example, the number of instances, the inter-class relations and the overlapping with other classes may vary. Depending on these characteristics, some of the classes might be easier to distinguish than others. *Difficult classes* can be defined as those obtaining a lower correct classification rate; that is, the number of correctly classified examples from the class divided by the total number of examples from that class (True Positive Rate, TPR[1]). The TPR of a class varies depending on both the mentioned characteristics and the classifier used, hence, this definition is subjective, since some classes might be easier or more difficult for a classifier than for another one. However, most of the classifiers are affected by the characteristics of the classes, being this part the most important one.

In this scenario, using the most commonly considered metric, i.e., accuracy rate (percentage of correctly classified examples) as an evaluation criterion, these classes might lose their importance, since it averages the results over all instances without taking into account the TPR over each class [36]. As a consequence, it becomes easier to increase the accuracy improving the classification of the easiest classes in exchange for misclassifying some of the instances from these difficult ones. This problem comprises the well-known class imbalance problem [31,25], where the difficult classes are those under-represented in the data-set; however, the problem of difficult classes is more general, since the hitch is not only caused by the skewed class distribution. For this reason, traditional solutions proposed for class imbalance, such as balancing the data-set, are not useful in this context and different approaches must be studied.

In this paper, we aim to undertake multi-class learning problems from a different perspective and centering on a completely distinct problem, i.e., the problem of difficult classes and its possible solutions from the point of view of decomposition strategies, and more specifically, paying attention to OVO strategy. We tackle the problem with a double study:

1. OVO strategy weakens when the enhancement is sought for the difficult classes. We aim to explain the reason why this occurs, which is mainly due to the way in which the decomposition is carried out and the aggregation used.
2. We introduce a new aggregation model based on similarity measures [10], which enables the modification of the decision boundaries of the base classifiers; in such a way, the classification of the difficult classes can be boosted without changing the underlying base classifiers. Hence, this methodology is independent of the base classifier, allowing the achievement of

[1] We refer to the true values of the TPR to show the difficult classes problem in Sections 2 and 3. Since these values cannot be obtained from data, the estimated TPR, by means of the results on the test sets in the cross-validation procedure, are considered to report the results along the experiments.