Information Sciences 247 (2013) 1–20

Contents lists available at SciVerse ScienceDirect

Information Sciences

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

Tackling the problem of classification with noisy data using Multiple Classifier Systems: Analysis of the performance and robustness

José A. Sáez [a,*], Mikel Galar [b], Julián Luengo [c], Francisco Herrera [a]

[a] Department of Computer Science and Artificial Intelligence, University of Granada, CITIC-UGR, Granada 18071, Spain
[b] Department of Automática y Computación, Universidad Pública de Navarra, Pamplona 31006, Spain
[c] Department of Civil Engineering, LSI, University of Burgos, Burgos 09006, Spain

ARTICLE INFO

Article history:
Received 25 May 2012
Received in revised form 13 May 2013
Accepted 2 June 2013
Available online 13 June 2013

Keywords:
Noisy data
Class noise
Attribute noise
Multiple Classifier System
Classification

ABSTRACT

Traditional classifier learning algorithms build a unique classifier from the training data. Noisy data may deteriorate the performance of this classifier depending on the degree of sensitiveness to data corruptions of the learning method. In the literature, it is widely claimed that building several classifiers from noisy training data and combining their predictions is an interesting method of overcoming the individual problems produced by noise in each classifier. This statement is usually not supported by thorough empirical studies considering problems with different types and levels of noise. Furthermore, in noisy environments, the noise robustness of the methods can be more important than the performance results themselves and, therefore, it must be carefully studied. This paper aims to reach conclusions on such aspects focusing on the analysis of the behavior, in terms of performance and robustness, of several Multiple Classifier Systems against their individual classifiers when these are trained with noisy data. In order to accomplish this study, several classification algorithms, of varying noise robustness, will be chosen and compared with respect to their combination on a large collection of noisy datasets. The results obtained show that the success of the Multiple Classifier Systems trained with noisy data depends on the individual classifiers chosen, the decisions combination method and the type and level of noise present in the dataset, but also on the way of creating diversity to build the final system. In most of the cases, they are able to outperform all their single classification algorithms in terms of global performance, even though their robustness results will depend on the way of introducing diversity into the Multiple Classifier System.

1. Introduction

Classifier learning algorithms aim to extract the knowledge from a problem from the available set of labeled examples (training set) in order to predict the class for new, previously unobserved, examples [8]. Classic learning algorithms [36,4] build a unique model, called a classifier, which attempts to generalize the peculiarities of the training set. Therefore, the success of these methods, that is, their ability to classify new examples, highly depends on the usage of a concrete feature descriptor and a particular inference procedure, and directly on the training data.

* Corresponding author. Tel.: +34 958 240598; fax: +34 958 243317.
E-mail addresses: smja@decsai.ugr.es (J.A. Sáez), mikel.galar@unavarra.es (M. Galar), jluengo@ubu.es (J. Luengo), herrera@decsai.ugr.es (F. Herrera).

0020-0255/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2013.06.002

Real-world data, which is the input of the classifier learning algorithms, are affected by several components [42,52,37]; among them, the presence of noise is a key factor. Noise is an unavoidable problem, which affects the data collection and data preparation processes in Data Mining applications, where errors commonly occur [48,50]. The performance of the classifiers built under such circumstances will heavily depend on the quality of the training data, but also on the robustness against noise of the classifier itself. Hence, classification problems containing noise are complex problems and accurate solutions are often difficult to achieve with a unique classifier system – particularly if this classifier is noise-sensitive.

Several works have claimed that simultaneously using classifiers of different types, complementing each other, improves classification performance on difficult problems, such as satellite image classification [27], fingerprint recognition [30] and foreign exchange market prediction [33]. Multiple Classifier Systems (MCSs) [15,14,32,45] are presented as a powerful solution to these difficult classification problems, because they build several classifiers from the same training data and therefore allow the simultaneous usage of several feature descriptors and inference procedures. An important issue when using MCSs is the way of creating *diversity* among the classifiers [24], which is necessary to create discrepancies among their decisions and hence, to take advantage from their combination.

MCSs have been traditionally associated with the capability of working accurately with problems involving noisy data [15]. The main reason supporting this hypothesis could be the same as one of the main motivations for combining classifiers: the improvement of the generalization capability (due to the complementarity of each classifier), which is a key question in noisy environments, since it might allow one to avoid the overfitting of the new characteristics introduced by the noisy examples [39]. Most of the works studying MCSs and noisy data are focused on techniques like bagging and boosting [7,25,20], which introduce diversity considering different samples of the set of training examples and use only one baseline classifier. For example, in [7] the suitability of randomization, bagging and boosting to improve the performance of C4.5 was studied. The authors reached the conclusion that with a low noise level, boosting is usually more accurate than bagging and randomization. However, bagging outperforms the other methods when the noise level increases. Similar conclusions were obtained in the paper of Maclin and Opitz [25]. Other works [20] compare the performance of boosting and bagging techniques dealing with imbalanced and noisy data, reaching also the conclusion that bagging methods generally outperforms boosting ones. Nevertheless, explicit studies about the adequacy of MCSs (different from bagging and boosting, that is, those introducing diversity using different base classifiers) to deal with noisy data have not been carried out yet. Furthermore, most of the existing works are focused on a concrete type of noise and on a concrete combination rule. On the other hand, when data are suffering from noise, a proper study on how the robustness of each single method influences the robustness of the MCS is necessary, but this fact is usually overlooked in the literature.

This paper aims to develop a thorough analysis of the behavior of several MCSs with noisy training data with respect to their individual components (classifiers), studying to what extent the behavior of these MCSs depends on that of the individual classifiers. The classic hypothesis about the good behavior of MCSs with noisy data will be checked in detail and the conditions under which the MCSs studied work well with noisy data will be analyzed. In order to reach meaningful conclusions based on the characteristics of the noise, a large collection of real-world datasets will be considered and different types of noise, present in real-world data, and several noise levels will be introduced into them, since these are usually unknown in real-world data. Two different types of noise, class and attribute noise, and four different schemes to introduce them will be considered. The experimentation will consist of a total of 1640 datasets. The results taken from these datasets will be analyzed taking into account two different factors: (i) the *performance*, and (ii) the *robustness*, i.e., the capability of the classifier to be insensitive to the increments in the noise level, of each method in each noisy dataset. The results obtained will also be contrasted using the proper statistical tests, as recommended in the specialized literature [16,17,10,11].

The choice of the single classification algorithms used to build the MCSs is based on their behavior with noisy data. In such a way, one is able to extract meaningful conclusions from the point of view of noise. Three algorithms have been selected, among the top ten algorithms in Data Mining [47], each one belonging to a different learning paradigm, and having a well-known differentiated robustness to noise: a *Support Vector Machine* (SVM) [5], C4.5 [36] and *k-Nearest Neighbors* (*k*-NN) [29].

All these classification algorithms will be combined using different decisions combination methods [28,38,41,18] to create MCSs of different sizes and characteristics. Two forms to create diversity will be considered: (i) considering different individual classifiers trained with the whole training data (heterogeneous base classifiers) and (ii) considering only one baseline classifier trained with different random samples of the training data of equal size as the original training data (using bagging). In this way, whether the performance and noise-robustness of the individual components is related to that of the corresponding MCS will be verified. All the conclusions and lessons learned from the analysis of the empirical results will be included in a specific section at the end of this paper.

A web-page with all the complementary material associated with this paper is available at http://www.sci2s.ugr.es/mcs_noise, including the basic information of this paper, all the datasets created and the complete results obtained for each classification algorithm, in such a way that this work becomes easily reproducible by other researchers.

The rest of this paper is organized as follows. Section 2 presents an introduction to classification with noisy data. Section 3 gives the motivations for the usage of MCSs. Next, Section 4 describes the experimental framework. Section 5 includes the experimental results and their analysis. Section 6 studies the results of different decisions combination methods. Section 7 presents the lessons learned and, finally, Section 8 presents some concluding remarks.