Neurocomputing 135 (2014) 21–31

Contents lists available at ScienceDirect

Neurocomputing

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

Metrics to guide a multi-objective evolutionary algorithm for ordinal classification ☆

M. Cruz-Ramírez *,1, C. Hervás-Martínez, J. Sánchez-Monedero, P.A. Gutiérrez

Department of Computer Science and Numerical Analysis, University of Córdoba, Rabanales Campus, Albert Einstein Building 3rd Floor, 14071 Córdoba, Spain

ARTICLE INFO

Article history:
Received 22 October 2012
Received in revised form
23 March 2013
Accepted 30 May 2013
Available online 8 January 2014

Keywords:
Mean absolute error
Multi-objective evolutionary algorithm
Ordinal measures
Ordinal classification
Ordinal regression
Proportional odds model

ABSTRACT

Ordinal classification or ordinal regression is a classification problem in which the labels have an ordered arrangement between them. Due to this order, alternative performance evaluation metrics are need to be used in order to consider the magnitude of errors. This paper presents a study of the use of a multi-objective optimization approach in the context of ordinal classification. We contribute a study of ordinal classification performance metrics, and propose a new performance metric, the maximum mean absolute error (*MMAE*). *MMAE* considers per-class distribution of patterns and the magnitude of the errors, both issues being crucial for ordinal regression problems. In addition, we empirically show that some of the performance metrics are competitive objectives, which justify the use of multi-objective optimization strategies. In our case, a multi-objective evolutionary algorithm optimizes an artificial neural network ordinal model with different pairs of metric combinations, and we conclude that the pair of the mean absolute error (*MAE*) and the proposed *MMAE* is the most favourable. A study of the relationship between the metrics of this proposal is performed, and the graphical representation in the two-dimensional space where the search of the evolutionary algorithm takes place is analysed. The results obtained show a good classification performance, opening new lines of research in the evaluation and model selection of ordinal classifiers.

1. Introduction

Ordinal classification or ordinal regression is a supervised learning problem of predicting categories that have an ordered arrangement. Although classification and regression metric problems have been thoroughly investigated in the literature, the ordinal regression problems have not received as much attention as nominal (binary or multiclass) classification. For example, people can be classified by considering whether they are high, medium, or low on some attribute or in a set of categories varying from strong agreement to strong disagreement with respect to some attitude item. Hodge and Treiman [1], to analyse social class identification, scored responses as follows: "Respondents identified with the lower, working, middle, upper middle, and upper class were assigned the scores 1, 2, 3, 4, and 5, respectively". Though sequential numbers may be assigned to such categories, the numbers assigned serve only to identify the ordering of the categories. In contrast to regression metric problems, these ranks are finite types and the metric distances between the ranks are not defined; in general, in contrast to classification problems, these ranks are also different from the labels of multiple classes due to the existence of the ordering information [2].

In the previous example, it is straightforward to think that predicting class *lower* when the real class is *upper middle* should be considered as a more severe error than the one associated to a *working* prediction. Thereby, ordinal classification problems should be evaluated with specific metrics. In the first consideration, various measures of ordinal association and product-moment correlation and regression seem to rely on very different foundations. That is, the ordinal measures are developed from (a) the notion of comparing pairs of cases, or (b) the product-moment system, which is considered in terms of measures of individual cases.

If methodology (a) is used, and there is an ordering of the categories but the absolute distances among them are unknown, an ordinal categorical variable is obtained. In that respect, in order to avoid the influence of the numbers chosen to represent the classes on the performance assessment, we should only look at the order relation between "true" and "predicted" class numbers. The use of Spearman's rank correlation coefficient r_S [3] and specially Kendall's τ_b [4] is a step forward in that direction. Moreover, other coefficients are frequently used to describe the association between ordinal measures as Goodman and Kruskal's γ [5], and Somers's d [6].

If methodology (b) (product-moment system) is used, the most commonly considered measures in machine learning are the mean

☆This paper is a significant extension of the work "A preliminary study of ordinal metrics to guide a multi-objective evolutionary algorithm" appearing in the 11th International Conference on Intelligent Systems Design and Applications (ISDA2011).

* Corresponding author. Tel.: +34 957 218 349; fax: +34 957 218 630.
E-mail addresses: mcruz@uco.es (M. Cruz-Ramírez), chervas@uco.es (C. Hervás-Martínez), jsanchezm@uco.es (J. Sánchez-Monedero), pagutierrez@uco.es (P.A. Gutiérrez).

0925-2312/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.05.058

absolute error (here denoted as *MAE*) [7,8], root mean square error (*RMSE*) [8], and mean zero-one error (*MZE*, more frequently known as error rate) [8], with $MZE = 1 - CCR$, where *CCR* is the correct classification rate. However, these three measures are not suitable when used to evaluate the performance of classifiers on ordinal unbalanced datasets [7]. The first contribution of this work is a newly proposed metric associated to an ordinal classifier that is the highest *MAE* value from *MAE*s measured independently for each class (maximum *MAE* or *MMAE*). This metric evaluates the performance on the worst classified class. The second contribution of this work is the analysis of the state-of-the-art performance metrics. Finally, we empirically show that some of the metric pairs can be non-cooperative, and consequently justify the use of a multi-objective framework to address the classifier optimization problem.

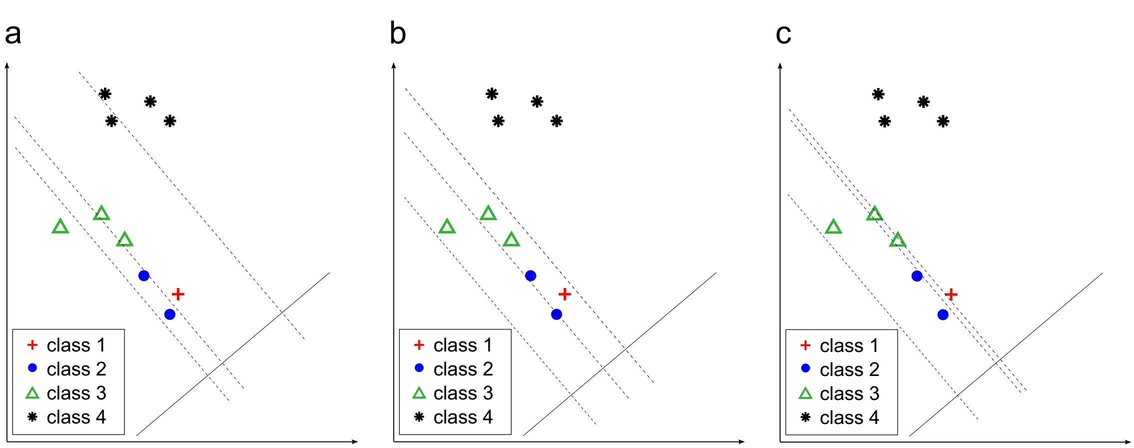
Fig. 1 presents a motivational example for the present work depicting three classifiers on a fourth class ordinal classification problem. This figure illustrates how different variations of decision thresholds can affect the classification performance specially influenced by patterns placed on the class boundaries. More specifically, this example raises two issues that will be studied in the current work. First, using a unique performance measure may not be enough to evaluate a classifier, specially in the field of ordinal regression. Second, some of the performance metrics can result in competitive objectives on a general optimization process since moving a threshold on a direction can produce an improvement in one metric, but a detrimental on a second one.

In the present paper, the aforementioned issues are studied under a multi-objective optimization approach. Multi-objective algorithms are algorithms that optimize simultaneously objectives that are non-cooperative. In many problems there are several conflicting objectives, such as execution speed or computational cost and kindness of the results. For example, in [9,10] the authors try to obtain optimal results in the shortest time and at the lowest cost. In other problems, the execution speed is not the most important and what is relevant is achieving good results in different conflicting error functions.

In the field of artificial neural networks (ANNs), classification performance and model simplicity are objectives that typically guide the training process of an evolutionary multi-objective algorithm (MOEA) [11], with the purpose of finding a trade-off between performance and model readability. Other works present the optimization of global performance versus the worst classified class in a Pareto based algorithm [12] or also by simplifying both objectives as a weighted linear combination of the functions [13].

In ordinal classification, it is common to use several error functions when some of the classes have a number of patterns much lower than the others, i.e. ordinal imbalanced datasets. Because of this reason, we proposed the *MMAE* metric measuring the performance in the worst classified class. One real world application where this problem can be found is in the extension of donor–recipient allocation in liver transplants [14], where the classifiers aim at predicting the survival of the organ (describing this survival in three different classes, class 1: lower than 15 days, class 2: between 15 days and 3 months, and class 3: higher than 3 months). The problem is that, in real cases, the number of patterns of class 1 is much lower than that of class 2 or 3. The hospital would be interested in classifiers able to correctly classify all classes equally, but the bad performance for class 1 can be hidden by the fact that the number of patterns of this class is very low (for example, a good *MAE* value can be obtained when class 1 is associated to a 5% of the patterns and the classifier never assigns a pattern to class 1). As can be seen, both objectives are conflicting (*MAE* and *MMAE*), because improving *MMAE* usually involves worsening *MAE* and vice versa. In [15] another ordinal problem is solved from a multi-objective perspective, where six different objectives are considered, including *MZE*, *MAE* and four different formulations for the expected ranking accuracy. In this work, several different ordinal measures that could be combined in the context of ordinal regression are analysed and combined in pairs for a MOEA.

The present work aims at identifying which pair of ordinal classification performance metrics can be more suitable to guide a MOEA to obtain classifiers with a good performance (considering both the order of the mis-classification errors and the worst classified class errors). The most common ordinal classification performance metrics are reviewed, and some of them are selected to evaluate the performance of four nominal and ordinal classifiers, including also the proposed metric. Then, a correlation study is done between all the metrics in order to find the less correlated ones. We hypothesize that the more uncorrelated metrics are the more suitable for acting as optimization objectives for the MOEA (given that all of them highlight positive aspects of the classifiers). The selected metrics are grouped into different pairs that will be simultaneously optimized by the MOEA. The base classifier considered is an ANN based on the proportional odds model (POM) [16] and it is evolved using a differential evolution MOEA [17,18].



$$\mathrm{M_A} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 \\ 1 & 0 & 2 & 0 \\ 0 & 0 & 1 & 3 \end{pmatrix}, \quad \begin{matrix} MAE = 0.5 \\ MMAE = 2.0 \\ CCR = 0.7 \end{matrix} \qquad \mathrm{M_B} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}, \quad \begin{matrix} MAE = 0.4 \\ MMAE = 2.0 \\ CCR = 0.7 \end{matrix} \qquad \mathrm{M_C} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 2 & 0 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}, \quad \begin{matrix} MAE = 0.4 \\ MMAE = 3.0 \\ CCR = 0.8 \end{matrix}$$

Fig. 1. An example of three classifiers decision boundaries for a four class ordinal classification problem. Decision thresholds vary from right to left leading to three different situations regarding performance evaluation metrics. Classifiers in subfigures (a) and (b) have the same *CCR* and *MMAE*, whereas *MAE* varies and confusion matrices M_A and M_B are different. A similar comment applies when comparing (b) and (c) situations, but in this case *MAE* is kept constant while *CCR* and *MMAE* vary. Finally, when comparing the classifiers (a) and (c), the *CCR* and *MAE* values improve and the value of *MMAE* worsens.