A Preliminary Study of Ordinal Metrics to Guide a Multi-Objective Evolutionary Algorithm

M. Cruz-Ramírez, C. Hervás-Martínez, J. Sánchez-Monedero and P.A. Gutiérrez
Department of Computer Science and Numerical Analysis
University of Córdoba, Spain
{*mcruz,chervas,jsanchezm,pagutierrez*}*@uco.es*

***Abstract*—There are many metrics available to measure the goodness of a classifier when working with ordinal datasets. These measures are divided into product-moment and association metrics. In this paper, the behavior of several metrics is studied in different situations. In addition, two new measures associated with an ordinal classifier are defined: the maximum and the minimum mean absolute error of all the classes. From the results of this comparison, a pair of metrics is selected (one associated to the overall error and another one to the error of the class with lowest level of classification) to guide the evolution of a multi-objective evolutionary algorithm, obtaining good results in generalization on ordinal datasets.**

***Keywords*-mean absolute error; multi-objective evolutionary algorithm; ordinal measures;**

I. Introduction

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. Nonetheless, the applications of ordinal regression frequently occur in domains where human-generated data plays an important role. Thereby, many of the performance measures that researchers use are ordinal. 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 analyze social class identification, scored responses as follows: "Respondents identifying 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 metric regression 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].

On a first consideration, various measures of ordinal association and product-moment correlation and regression seem to rely on very different foundations. This 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 association between ordinal measures as Goodman and Kruskal's γ [5], and Somers's d [6]. Each one of these four measures has several advantages and disadvantages in terms of interpretation, ease of computation, and suitability to the data under discussion. They seem to satisfy the necessity of measures of ordinal association in the bivariate case. For instance, in order to get r_S, we start by defining two rank vectors of length n, corresponding to the size of the dataset, which are associated with the variables C and C^*, corresponding respectively to the numbers representing the true and predicted classes.

More recently, various measures of association for ordered variables have been developed [7], and models useful for analyzing square tables with ordinal variables are diagonal-parameter models, quasi-uniform association models [8] and conditional symmetry models [9].

If methodology b) (product-moment system) is used, the most common considered measures in Machine Learning are the Mean Absolute Error (here denoted as MAE) [10], [11]; Root Mean Square Error ($RMSE$) [11], and Mean Zero-One Error (more frequently known as Error Rate) [11], being 1-CCR, where CCR is the Correct Classification Rate. However, these three measures are not very good when used to measure the performance of classifiers on ordinal unbalanced datasets because they assume a metric on the output space [10].

In this preliminary study, we define two new measures associated with an ordinal classifier. The first one is associated to the highest MAE value for of all the classes (maximum MAE) and the second one to the lowest MAE value (minimum MAE). Then, we study these ordinal metrics,

978-1-4577-1675-1 ©2011 IEEE

together with other measures, in order to determine the best pair of ordinal metrics, which will be used to guide a Multi-Objective Evolutionary Algorithm (MOEA) for the generalization improvement of classifiers.

The rest of paper is organized as follows: Section II shows a comparison of measures of ordinal classification; Section III describes the method used; Section IV describes the experimental design and the results obtained, while conclusions and future research are outlined in Section V.

II. Measures of association in ordinal classification

A ordinal classification problem with J classes and n patterns is considered with g as a classifier obtaining a $J \times J$ contingency or confusion matrix ($\mathbf{M} = M(g)$):

$$\mathbf{M} = M(g) = \{n_{ij};\ \sum_{i,j=1}^{J} n_{ij} = n\},$$

where n_{ij} represents the number of times the patterns are predicted by classifier g to be in class j when they really belong to class i and n is the total number of patterns.

There are many ordinal measures to determine the efficiency of the g classifier, but not all pairs formed by these metrics are valid to guide a MOEA. Before describing the ordinal metrics, accuracy and minimum sensitivity are presented, since they are conflicting [12]:

- CCR: The Correct Classification Rate or Accuracy is the percentage of correctly classified patterns:

$$CCR = \frac{1}{n}\sum_{j=1}^{J} n_{jj},$$

  where CCR values range from 0 to 1.
- MS: The Minimum Sensitivity is the lowest percentage of patterns correctly predicted as belonging to each class, with respect to the total number of examples in the corresponding class:

$$MS = \min\{S_j = \frac{n_{jj}}{n_j}; j = 1, ..., J\},$$

  where n_j is the number of patterns in the *j-th* class, S_j is the sensitivity of the *j-th* class and MS values range from 0 to 1.

On the other hand, there are other product-moment ordinal metrics used in ordinal classification:

- MAE: The Mean Absolute Error is the average deviation in absolute value of the predicted class from the true class [10]:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} e(x_i),$$

  where $e(x_i) = |C_i - C_i^*|$ is the distance between the true (C) and the predicted (C^*) rank, assuming that the J classes are integer numbers from 1 to J and, then, MAE values range from 0 to $J-1$.
- $MMAE$: The Maximum MAE value of all the classes. $MMAE$ is the MAE value of the class with higher distance from the true values to the predicted ones:

$$MMAE = \max\{MAE_j; j = 1, ..., J\},$$

  where MAE_j is the MAE value for the *j-th* class. $MMAE$ values range from 0 to $J-1$.
- $mMAE$: The minimum MAE value of all the classes. $mMAE$ is the MAE value of the class with smaller distance from the true values to the predicted ones:

$$mMAE = \min\{MAE_j; j = 1, ..., J\},$$

  where $mMAE$ values range from 0 to $J-1$.
- $AMAE$: The Average MAE is the mean of the classification errors across classes ($MAE^M \equiv AMAE$) [10]:

$$AMAE = \frac{1}{J}\sum_{j=1}^{J}\frac{1}{n_j}\sum_{i=1}^{n_j} e_j(x_i) = \frac{1}{J}\sum_{j=1}^{J} MAE_j,$$

  where $AMAE$ values range from 0 to $J-1$.

Finally association metrics are presented, which are also used in ordinal classification:

- τ_b: The Kendall's τ is a statistic used to measure the association between two measured quantities. Specifically, it is a measure of rank correlation [4]:

$$\tau_b = \frac{\sum c_{ij}^* c_{ij}}{\sqrt{\sum c_{ij}^{*2} \sum c_{ij}^2}},$$

  where c_{ij}^* is +1 if C_i^* is greater than C_j^*, 0 if C_i^* and C_j^* are the same, and -1 if C_i^* is less than C_j^* for $i, j = 1, ..., n$, and similar for C, and τ_b values range from -1 to 1.
- r_S: The Spearman's rank correlation coefficient is a non-parametric measure of statistical dependence between two variables [3]:

$$r_S = \frac{\sum (C_i - \overline{C})(C_i^* - \overline{C^*})}{\sqrt{\sum (C_i - \overline{C})^2 \sum (C_i^* - \overline{C^*})^2}},$$

  where $\overline{C}$ and $\overline{C^*}$ are the average of C and C^* respectively, $i = 1, ..., n$ and r_S values range from -1 to 1.
- r_{int}: The r_{int} is a metric to measure the performance of ordinal classifiers introduced by Pinto da Costa *et al.* [13]:

$$r_{int} = -1 + 2\frac{card(S_1 \bigcap S_2)}{\sqrt{card(S_1)card(S_2)}},$$

  where $card(S_1)$ and $card(S_2)$ are defined as follows:

$$card(S_1) = \sum_{i=1}^{J}\sum_{j=i}^{J} n_{i.}n_{j.} - n,$$

Table I: Confusion matrices for the study.

$$\mathbf{M}_1 = \begin{pmatrix} 10 & 0 & 0 & 0 \\ 20 & 0 & 0 & 0 \\ 30 & 0 & 0 & 0 \\ 40 & 0 & 0 & 0 \end{pmatrix} \quad \mathbf{M}_2 = \begin{pmatrix} 10 & 0 & 0 & 0 \\ 0 & 20 & 0 & 0 \\ 0 & 0 & 30 & 0 \\ 0 & 0 & 0 & 40 \end{pmatrix}$$

$$\mathbf{M}_3 = \begin{pmatrix} 0 & 0 & 0 & 10 \\ 0 & 0 & 20 & 0 \\ 0 & 30 & 0 & 0 \\ 40 & 0 & 0 & 0 \end{pmatrix} \quad \mathbf{M}_4 = \begin{pmatrix} 0 & 0 & 0 & 10 \\ 0 & 0 & 0 & 20 \\ 30 & 0 & 0 & 0 \\ 40 & 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{M}_5 = \begin{pmatrix} 0 & 10 & 0 & 0 \\ 20 & 0 & 0 & 0 \\ 0 & 0 & 0 & 30 \\ 0 & 0 & 40 & 0 \end{pmatrix} \quad \mathbf{M}_6 = \begin{pmatrix} 40 & 0 & 0 & 0 \\ 30 & 0 & 0 & 0 \\ 20 & 0 & 0 & 0 \\ 10 & 0 & 0 & 0 \end{pmatrix}$$

$$card(S_2) = \sum_{i=1}^{J} \sum_{j=i}^{J} n_{.i} n_{.j} - n,$$

where $n_{i.} = \sum_{j=1}^{J} n_{ij}$ and $n_{.j} = \sum_{i=1}^{J} n_{ij}$ for $i, j = 1, ..., J$. And $card(S_1 \bigcap S_2)$:

$$card(S_1 \bigcap S_2) = \sum_{i=1}^{J} \sum_{j=1}^{J} \sum_{i'=i}^{J} \sum_{j'=j}^{J} n_{ij} n_{i'j'} - n,$$

where r_{int} values range from -1 to 1.

- $WKappa$: The Weighted Kappa is a modified version of the Kappa statistic calculated to allow assigning different weights to different levels of aggregation between two variables [14]:

$$WKappa = \frac{p_{o(w)} - p_{e(w)}}{1 - p_{e(w)}},$$

where

$$p_{o(w)} = \frac{1}{n} \sum_{i=1}^{J} \sum_{j=1}^{J} w_{ij} n_{ij},$$

and

$$p_{e(w)} = \frac{1}{n^2} \sum_{i=1}^{J} \sum_{j=1}^{J} w_{ij} n_{i.} n_{.j},$$

where the weight w_{ij} quantifies the degree of discrepancy between the true (C_i) and the predicted (C_j^*) categories, and $WKappa$ values range from -1 to 1.

While the τ_b and r_S measures are dependent on the values chosen for the ranks that representing the classes, the MAE, $MMAE$, $mMAE$ and $AMAE$ depend on the distance between ranking of two consecutive classes.

A. Comparison of the Ordinal Metric

The purpose of analyzing these metrics is to determine which ones could be suitable as fitness functions in a MOEA. To study the performance of these metrics, we will use the 6 confusion matrices shown in Table I. These matrices are designed to cover all possible situations: all patterns classified in the same class ($\mathbf{M}_1$ and $\mathbf{M}_6$), perfect classification ($\mathbf{M}_2$), inverse association ($\mathbf{M}_3$) and other problematic classifications ($\mathbf{M}_4$ and $\mathbf{M}_5$). Matrices $\mathbf{M}_1 - \mathbf{M}_5$ follow the same distribution patterns per class. The distribution pattern of the matrix $\mathbf{M}_6$ is different to compare with the matrix $\mathbf{M}_1$.

Table II: Results obtained by the different metrics.

	$\mathbf{M}_1$	$\mathbf{M}_2$	$\mathbf{M}_3$	$\mathbf{M}_4$	$\mathbf{M}_5$	$\mathbf{M}_6$
CCR	0.1	1.0	0.0	0.0	0.0	0.4
MS	0.0	1.0	0.0	0.0	0.0	0.0
MAE	2.0	0.0	2.0	2.5	1.0	1.0
$MMAE$	3.0	0.0	3.0	3.0	1.0	3.0
$mMAE$	0.0	0.0	1.0	2.0	1.0	0.0
$AMAE$	1.5	0.0	2.0	2.5	1.0	1.5
τ_b	0.0	1.0	-0.9354	-0.7200	0.1972	0.0
r_S	0.0	1.0	-1.0	-0.8728	0.5570	0.0
r_{int}	0.6080	1.0	-0.0937	0.2171	0.5625	0.6080
$WKappa$	0.0	1.0	-0.4084	-0.4705	0.1228	0.0

Table II shows the results obtained after applying all the metrics on the confusion matrices. These results show that the matrix $\mathbf{M}_2$ obtains optimal values. For the other matrices, the values of the metrics are between their expected ranges. The worst results are obtained for $\mathbf{M}_3$ and $\mathbf{M}_4$. Matrices $\mathbf{M}_1$ and $\mathbf{M}_6$ get the same results on all metrics less in CC and MAE. This indicates that all metrics (except CCR and MAE) are not dependent on the distribution of patterns per class.

In [15], the CCR and MAE measures (and other metrics) are analyzed and compared. In our case, and preliminary manner, we analyze the pair formed by the $MMAE$ and $AMAE$ measures from a multi-objective point of view. These metrics are conflicting, since $AMAE$ seeks to minimize the overall error rate of the classifier, while $MMAE$ intends that all classes have an acceptable value of MAE.

B. Study of $AMAE$-$MMAE$ pair

To analyze the relationship between the two metrics we propose the following procedure.

Proposition: Let us consider a J-class classification problem. Let $AMAE$ and $MMAE$ be respectively the two measures associated with an ordinal classifier g, then:

$$\frac{MMAE}{J} \leq AMAE \leq MMAE. \quad (1)$$

Proof: We begin by proving the upper bound. The class with $MMAE$ is denoted without loss of generality by $j = J$, and the class with $mMAE$ is denoted by $j = 1$. In general $0 \leq MAE_j \leq MMAE \leq J - 1$, so:

$$AMAE = \frac{1}{J} \sum_{j=1}^{J} MAE_j \leq \frac{1}{J} \sum_{j=1}^{J} MMAE = MMAE.$$

Thus, $AMAE \leq MMAE$.

On the other hand, the lower bound can be obtained:

$$AMAE = \frac{1}{J} \sum_{j=1}^{J} \frac{1}{n_j} \sum_{i=1}^{n_j} e_j(x_i) =$$

$$= \frac{1}{J}\sum_{j=1}^{J-1}\frac{1}{n_j}\sum_{i=1}^{n_j} e_j(x_i) + \frac{1}{J}\frac{1}{n_J}\sum_{i=1}^{n_J} e_J(x_i),$$

but

$$\frac{1}{J}\sum_{j=1}^{J-1}\frac{1}{n_j}\sum_{i=1}^{n_j} e_j(x_i) \geq \frac{(J-1)mMAE}{J},$$

and

$$\frac{1}{J}\frac{1}{n_j}\sum_{i=1}^{n_J} e_J(x_i) = \frac{MMAE}{J},$$

then

$$AMAE = \frac{1}{J}\sum_{j=1}^{J}\frac{1}{n_j}\sum_{i=1}^{n_j} e_j(x_i) \geq$$

$$\geq mMAE + \frac{MMAE - mMAE}{J} \geq \frac{MMAE}{J},$$

for all $mMAE$. But since $0 \leq mMAE \leq MAE_j \leq J-1$, we have $AMAE \geq \frac{MMAE}{J}$. ■

1) Graphical Representation of $AMAE$-$MMAE$: The $AMAE$-$MMAE$ point of view allows us to represent the performance of a classifier in a two dimensional space, but in this pair of measures the total number of classes in the problem is considered. Concretely, the $MMAE \equiv M$ is represented on the horizontal axis and the $AMAE \equiv A$ on the vertical axis. One point in (M, A) space dominates another if it is below and to the left, i.e. it has less M and less A. Therefore, from the inequalities in (1), each classifier will be represented as a point in the white region in Figure 1. Several points in (M, A) space are important to note. The upper right point $(J-1, J-1)$ represents the worst classifier and the optimum classifier is located at the $(0, 0)$ point. Furthermore, the point $(J-1, \frac{J-1}{J})$ corresponds with a classifier that has, at least, one class with no patterns correctly classified. When the number of classes increases, the value of A tends to one. Note that it is possible to find among them classifiers with a low level of A, but with a higher level of M, specially when the number of classes is high. Thus, minimizing these two error functions simultaneously produces models which are a trade-off between average results that are acceptable to all classes and the lowest ranked class, i.e., the class that has patterns farthest from the corresponding class in the ordinal ranking.

From the concrete shape of the region, the following comments can be made. First of all, observe that a decrease in A does not imply a decrease in M. Reciprocally, a decrease in M does not mean a decrease in A. On the other hand, it should be noted that for a fixed value of A, a classifier will be better when it corresponds to a point closer to the diagonal of the $(J-1) \times (J-1)$ square.

It is important to analyze if M and A are not cooperative in general, especially at certain high levels. At the beginning of a learning process, M and A could be cooperative, however after some generations, objectives become competitive and a decrease of one objective usually involves an increase in the other one.

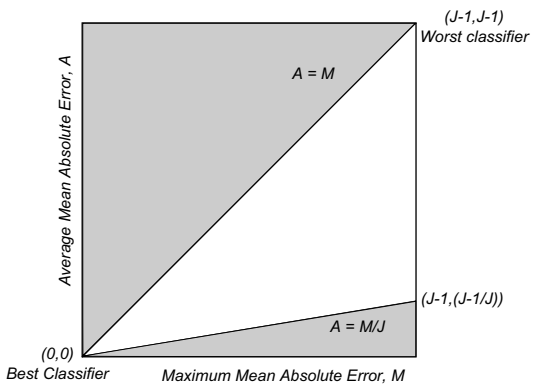

Figure 1: Unfeasible region in the two-dimensional (M, A) space for a given classification problem.

III. Method

To see how the selected metrics behave, this paper uses the MOEA described in [16] for training artificial neural networks with sigmoid basis functions. For more details about the base classifier framework, see [16]. The algorithm used is the MPDENN (Memetic Pareto Differential Evolution Neural Network) algorithm developed by R. Storn and K. Price in [17] and modified by H. Abbass to train neural networks [18]. MPDENN is adapted by analyzing the trade-off between the CCR and MS in [12], [19]. The fundamental bases of this algorithm are Differential Evolution and the concept of Pareto dominance.

For the study presented in this paper the local optimization process has been eliminated for two reasons: 1) the objective is to study the performance and relationship of the metrics and 2) the selected metrics are not continuous functions, which makes convergence more difficult or even impossible in local optimization. Because of this reason, we call the algorithm PDENN (without the local search step, non Memetic).

IV. Experimental Study

To verify the efficiency of the selected ordinal metric pair, we will use four ordinal datasets[1]. The results obtained by using the $AMAE$-$MMAE$ pair as fitness functions when evaluating individuals obtained from the MOEA will be compared with those obtained when the algorithm is guided by two nominal metrics, CCR and MS. The process to obtain the results with nominal metrics can be seen in [12]. The ordinal metrics are used as fitness functions without requiring any change in the evolution process of MOEA. The purpose of this comparison is to see if using ordinal metrics improves the results obtained with the nominal metrics.

[1]Datasets are available in http://weka.wikispaces.com/Datasets

Table III: Results obtained in generalization for different metrics pairs.

Dataset	Metrics pair	Method	CCR_G Mean±SD	MS_G Mean±SD	MAE_G Mean±SD	$MMAE_G$ Mean±SD	$AMAE_G$ Mean±SD
Balance-scale 3 classes	CCR vs. MS	PDENN-CCR	0.8754±0.0304	0.3205±0.2467	0.1512±0.0446	0.6795±0.2467	0.2955±0.0821
		PDENN-MS	0.8586±0.0512	0.7268±0.1238	0.1609±0.0591	0.2771±0.1201	0.1884±0.0688
	$AMAE$ vs. $MMAE$	PDENN-AMAE	**0.8786±0.0360**	**0.7660±0.1331**	**0.1348±0.0420**	**0.2408±0.1290**	**0.1547±0.0668**
		PDENN-MMAE	0.8771±0.0418	0.7490±0.1398	0.1363±0.0503	0.2570±0.1353	0.1621±0.0729
Squash-stored 3 classes	CCR vs. MS	PDENN-CCR	0.5359±0.1301	0.2300±0.2175	0.5026±0.1561	0.8756±0.2998	0.5504±0.1620
		PDENN-MS	0.5154±0.1091	0.2356±0.2249	0.5308±0.1462	0.9211±0.4006	0.5785±0.1938
	$AMAE$ vs. $MMAE$	PDENN-AMAE	**0.5641±0.1475**	**0.2744±0.2457**	**0.4718±0.1804**	**0.8411±0.3774**	**0.5007±0.2270**
		PDENN-MMAE	0.5615±0.1487	**0.2744±0.2457**	0.4744±0.1808	**0.8411±0.3774**	0.5030±0.2258
SWD 4 classes	CCR vs. MS	PDENN-CCR	0.4317±0.0476	0.0266±0.0600	0.6789±0.0912	1.3803±0.2785	0.8452±0.1060
		PDENN-MS	0.4471±0.0420	**0.2824±0.1191**	0.6768±0.0712	1.0065±0.2735	0.7453±0.1103
	$AMAE$ vs. $MMAE$	PDENN-AMAE	0.4845±0.0457	0.2538±0.1320	0.5849±0.0493	**0.8738±0.1520**	**0.6245±0.0681**
		PDENN-MMAE	**0.4860±0.0413**	0.2696±0.1273	**0.5835±0.0560**	0.9050±0.1961	0.6460±0.0762
Tae 3 classes	CCR vs. MS	PDENN-CCR	0.3684±0.0688	0.0545±0.0867	0.8289±0.1634	1.3964±0.4507	0.8308±0.1625
		PDENN-MS	0.4211±0.0909	0.2915±0.1005	0.7737±0.1216	1.0376±0.2017	0.7752±0.1207
	$AMAE$ vs. $MMAE$	PDENN-AMAE	0.4851±0.0727	0.3355±0.0973	**0.6114±0.1013**	**0.7889±0.1348**	**0.6129±0.1005**
		PDENN-MMAE	**0.4868±0.0823**	**0.3451±0.0952**	0.6184±0.1086	0.8073±0.1538	0.6200±0.1085

The best result is in **bold** face.

Due to the MOEA used is non-deterministic, we perform 10 times a holdout validation and 3 repetitions for each holdout (obtaining a total of $10 \times 3 = 30$ different results). Each holdout is a stratified random division of the data, where approximately 75% of the instances are used for the training set and the remaining 25% for the generalization set (maintaining the original distribution of classes).

In all the experiments, the population size is established at 25 and the number of generations is 150. The crossover probability is 0.8 and the mutation probability is 0.1. The number of neurons in the hidden layer ranges from 1 to 20, establishing the optimum value during the evolutionary process.

A. Results

Table III shows the results obtained after guiding the MOEA with the two pairs of selected metrics. The results presented correspond to average values in generalization for the 30 extreme models of the Pareto fronts generated in 30 runs in training (one Pareto front for each run). Thus, the PDENN-AMAE method shows the mean values in generalization of the 30 individuals with the best value of $AMAE$ in training. The same procedure is followed for the other methods.

Table III does not show the values of all the metrics described in Section II because our aim is to obtain conclusions about the metrics we are working with.

From a descriptive point of view, the $AMAE$-$MMAE$ pair obtain the best results in all datasets for all metrics, except in the SWD dataset, in which PDENN-MS method get the best value in MS. These results show that the MOEA has better performance on ordinals datasets when evolution is guided by the pair of ordinal metrics $AMAE$-$MMAE$.

In Figure 2, we can see the graphical results obtained for the PDENN algorithm with $AMAE$-$MMAE$ pairs for the dataset SWD. For the $AMAE$-$MMAE$ space, we select the Pareto front for one specific run of the 30 ones performed for the dataset, concretely the execution that presents the best individual on MAE for training data. On the generalization graphic, we show the $MMAE$ and $AMAE$ values over the generalization set for the individuals who are reflected in the training graphic. Observe that the $AMAE$-$MMAE$ values do not form Pareto fronts in generalization, and the individuals that in the training graphic were in the first Pareto front, now can be located within space in a worst region. In general the structure of a Pareto front in training is not maintained in generalization.

V. Conclusion

In this paper, the Maximum and the Minimum MAE are presented and used in a comparative study together with other ordinal measures in order to determine the pair of metrics to better guide a Multi-Objective Evolutionary learning Algorithm. A good performance for ordinal datasets is obtained when the algorithm is guided by the Average MAE and Maximum MAE pair. These results are better than those obtained by using a pair of nominal metrics.

As future work, the Average MAE and Maximum MAE metrics could be used to guide a specific evolutionary algorithm for ordinal classification, which use ordinal information during the evolutionary process.

Acknowledgment

This work has been partially subsidized by the TIN 2008-06681-C06-03 project of the Spanish Ministerial Commission of Science and Technology (MICYT), FEDER funds and the P08-TIC-3745 project of the "Junta de Andalucía" (Spain). Manuel Cruz-Ramírez's research has been subsidized by the FPU Predoctoral Program (Spanish Ministry of Education and Science), grant reference AP2009-0487.

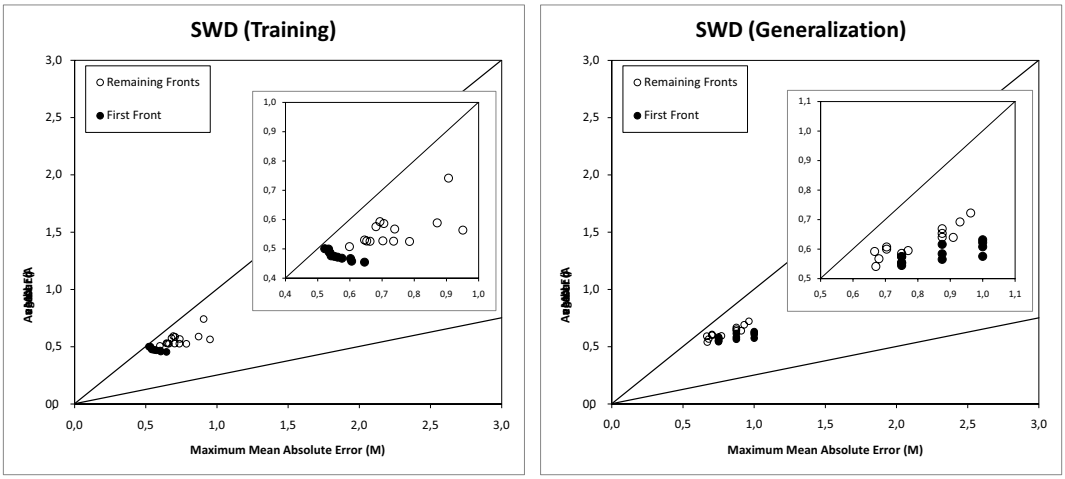

Figure 2: Pareto front in training and associated values in generalization.

Javier Sánchez-Monedero's research has been funded by the "Junta de Andalucía" Ph. D. Student Program.

REFERENCES

[1] R. W. Hodge and J. T. Donald, "Class identification in the united states." *American Journal of Sociology*, vol. 73, pp. 535–547, 1968.

[2] W. Chu and S. Keerthi, "New approaches to support vector ordinal regression," in *Proc. the 22nd International Conference on Machine Learning (ICML 2005), Bonn, Germany*, 2005, pp. 145–152.

[3] C. Spearman, "The proof and measurement of association between two things," *American Journal of Psychology*, vol. 15, pp. 72–101, 1904.

[4] M. G. Kendall, *Rank Correlation Methods*, 3rd ed. New York: Hafner Press, 1962.

[5] L. Goodman and W. Kruskal, "Measures of association for cross classifications," *Journal of the American Statistical Association*, vol. 49, pp. 732–764, 1954.

[6] R. H. Somers, "The rank analogue of product-moment partial correlation and regression with application to manifold, ordered contingency tables," *Biometrika*, vol. 46, pp. 241–246, 1955.

[7] A. Agresti, *Analysis of ordinal categorical data*. New York: Wiley, 1984.

[8] L. Goodman, "Simple models for the analysis of association in cross-classifications having ordered categories," *J. of the American Statistical Association*, vol. 74, pp. 537–552, 1979.

[9] Y. Bishop, S. Fienberg, and P. Holland, *Discrete multivariate analysis: theory and practice*. MIT Press, Cambridge, 1975.

[10] S. Baccianella, A. Esuli, and F. Sebastiani, "Evaluation measures for ordinal regression," in *Intelligent Systems Design and Applications, 2009. ISDA '09. Ninth International Conference on*, 2009, pp. 283 –287.

[11] K. Dembczyński, W. Kotlowski, and R. Slowiński, "Ordinal classification with decision rules," in *Proceedings of the ECML/PKDD'07 workshop on Mining Complex Data, Warsaw, PL*, 2007, pp. 169–181.

[12] J. C. Fernández-Caballero, F. J. Martínez-Estudillo, C. Hervás-Martínez, and P. A. Gutiérrez, "Sensitivity versus accuracy in multiclass problems using memetic Pareto evolutionary neural networks," *IEEE Transactions on Neural Networks*, vol. 21, no. 5, pp. 750 –770, may 2010.

[13] J. F. Pinto da Costa, H. Alonso, and J. S. Cardoso, "The unimodal model for the classification of ordinal data," *Neural Netw.*, vol. 21, pp. 78–91, 2008.

[14] J. L. Fleiss, J. Cohen, and B. S. Everitt, "Large sample standard errors of kappa and weighted kappa," *Psychological Bulletin*, vol. 72, no. 5, pp. 323–327, 1969.

[15] W. Waegeman and B. Baets, "A Survey on ROC-based Ordinal Regression," in *Preference Learning*. Springer, 2011, pp. 127–154.

[16] M. Cruz-Ramírez, J. Sánchez-Monedero, F. Fernández-Navarro, J. Fernández, and C. Hervás-Martínez, "Memetic Pareto differential evolutionary artificial neural networks to determine growth multi-classes in predictive microbiology," *Evolutionary Intelligence*, vol. 3, no. 3-4, pp. 187–199, 2010.

[17] R. Storn and K. Price, "Differential evolution. A fast and efficient heuristic for global optimization over continuous spaces," *J. Global Optimization*, vol. 11, pp. 341–359, 1997.

[18] H. A. Abbass, R. Sarker, and C. Newton, "PDE: a Pareto-frontier differential evolution approach for multi-objective optimization problems," in *Proc. of the 2001 Congress on Evolutionary Computation*, vol. 2, Seoul, South Korea, 2001.

[19] J. C. Fernández, C. Hervás, F. J. Martínez, P. A. Gutiérrez, and M. Cruz, "Memetic Pareto differential evolution for designing artificial neural networks in multiclassification problems using cross-entropy versus sensitivity," in *Hybrid Artificial Intelligence Systems*, vol. 5572. Springer Berlin / Heidelberg, 2009, pp. 433–441.