Contents lists available at ScienceDirect

Neural Networks

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

Corrigendum

Corrigendum to "The unimodal model for the classification of ordinal data" [Neural Netw. 21 (2008) 78–79]

Joaquim Pinto da Costa [a,*], Hugo Alonso [b], Jaime S. Cardoso [c]

[a] FCUP-DM/CMUP, University of Porto, Portugal
[b] University of Aveiro, Portugal
[c] Faculdade de Engenharia da Universidade do Porto/INESC Porto, Portugal

ARTICLE INFO

Article history:
Received 8 May 2014
Received in revised form 27 May 2014
Accepted 1 June 2014
Available online xxxx

In 2008 (Pinto da Costa, Alonso, & Cardoso, 2008) we proposed the unimodal paradigm for ordinal data classification and also a new coefficient to measure the performance of ordinal data classifiers; that is, classifiers for the important and often neglected case where there is an order relation between the classes. Many real life settings exist which involve classifying examples (instances) into classes which have a natural ordering, like for instance econometric modelling; information retrieval and collaborative filtering; stock trading support systems, where one wants to predict, for instance, whether to buy, keep or sell a stock. In fact in a very large number of applications the classes are ordered, although that was rarely taken into account. Recently, the subject has attracted a growing interest (Cruz-Ramírez, Hervás-Martínez, Sánchez-Monedero, & Gutiérrez, 2011; Hu, Guo, Yu, & Liu, 2010; Pinto da Costa, Sousa, & Cardoso, 2011; Seah, Tsang, & Ong, 2012).

We also proposed in Pinto da Costa et al. (2008), apart from the unimodal paradigm, a new coefficient to measure the performance of ordinal data classifiers. Since then a growing number of works have also addressed this issue (Baccianella, Esuli, & Sebastiani, 2009; Cardoso & Sousa, 2011; Cruz-Ramírez, Hervás-Martínez, Sánchez-Monedero, & Gutiérrez, 2014; Lee & Liu, 2002; Sánchez-Monedero, Gutiérrez, Tino, & Hervás-Martínez, 2013; Waegeman, De Baets & Boullart, 2006; Vanbelle & Albert, 2009). The coefficient r_{int} that we have proposed in Pinto da Costa et al. (2008) has a minor error in its final formulation which must be corrected otherwise its values make no sense. In order to correct its formula we will first describe briefly the way the coefficient was constructed.

In supervised classification problems with ordered classes, it is common to assess the performance of the classifier using measures which are not really appropriate. Very often, every misclassification is considered equally costly and the Misclassification Error Rate (MER) is used. Two other measures that are also usually used are the Mean Square Error (MSE) and the Mean Absolute Deviation (MAD), which must assign numbers to each class. However, this assignment is arbitrary and the numbers chosen to represent the existing classes will evidently influence the performance measurement given by MSE or MAD. 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. With that in mind, the use of Spearman's rank correlation coefficient, r_S (Press, Flannery, Teukolsky, & Vetterling, 1992; Spearman, 1904) and specially Kendall's τ_b (Press et al., 1992), considered in some works, is progress, although Spearman's coefficient is still dependent on the values chosen for the ranks representing the classes. Kendal's coefficient is in our view better than the Spearman's one to measure the performance of ordinal data classifiers, but still has some problems that we explained in Pinto da Costa et al. (2008).

Let us designate by C the true value of the class of a test instance and by $\hat{C}$ the value predicted by some model. The variables C and $\hat{C}$ are two special ordinal variables because, as there are usually very few classes compared to the number of observations, these variables will take many tied values (most of them, in fact). The coefficient introduced by us in Pinto da Costa et al. (2008), r_{int}, takes this into account by defining a suitable order relation associated to each variable, and it is not sensitive to the values that are chosen to represent the ordinal classes.

DOI of original article: http://dx.doi.org/10.1016/j.neunet.2007.10.003.
* Corresponding author. Tel.: +351 220402270.
E-mail addresses: jpcosta@fc.up.pt (J. Pinto da Costa), hugo.alonso@ua.pt (H. Alonso), jaime.cardoso@inescporto.pt (J.S. Cardoso).

http://dx.doi.org/10.1016/j.neunet.2014.06.003

Table 1
Values of v_1 and v_2 in $\mathcal{O}$.

$\mathcal{O}$	v_1	v_2
o_1	3	1
o_2	2	2
o_3	1	3
o_4	4	4

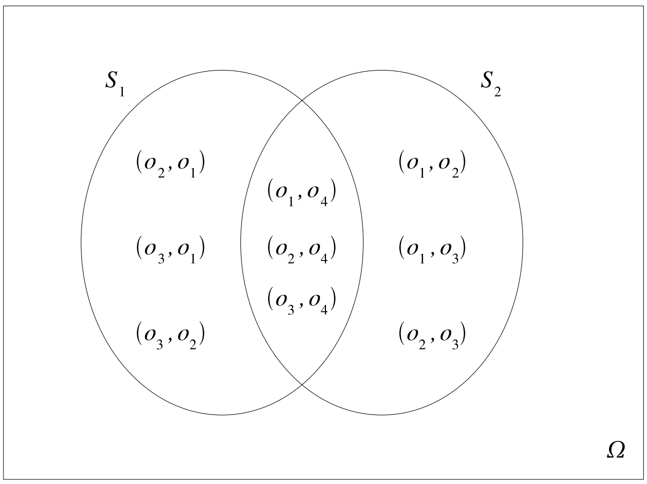

Fig. 1. $S_1 \cap S_2$.

Let $\mathcal{O} = \{o_1, o_2, \ldots, o_n\}$ represent the dataset where the performance is to be measured and take $\Omega = \mathcal{O} \times \mathcal{O} - \{(o_1, o_1), (o_2, o_2), \ldots, (o_n, o_n)\}$ to represent the set of pairs corresponding to different observations. Now, define the order relation R_v as $o_i R_v o_j$ if and only if $v(o_i) \leq v(o_j)$, where v represents a qualitative ordinal variable on $\mathcal{O}$. Let v_1 and v_2 be two qualitative ordinal variables on $\mathcal{O}$, whose exact values are of no importance since the only information of interest to us is the order relation between those values. Each of the variables v_1 and v_2 can thus be fully represented by a subset of Ω. As a motivating example, suppose that $\mathcal{O} = \{o_1, o_2, o_3, o_4\}$ and consider the following data in Table 1:

According to v_1, the subset of Ω whose elements verify the relation R_{v_1} is $S_1 = \{(o_1, o_4), (o_2, o_1), (o_2, o_4), (o_3, o_1), (o_3, o_2), (o_3, o_4)\}$. Similarly, v_2 is represented by $S_2 = \{(o_1, o_2), (o_1, o_3), (o_1, o_4), (o_2, o_3), (o_2, o_4), (o_3, o_4)\}$. The intersection $S_1 \cap S_2$, which in this case has three elements, is the key point for the comparison of the two variables v_1 and v_2 (see Fig. 1).

We defined our measure of association between the two ordinal variables v_1 and v_2 to be

$$r_{\text{int}} = A + B \frac{\text{card}(S_1 \cap S_2)}{\sqrt{\text{card}(S_1)\,\text{card}(S_2)}},$$

where the denominator is considered to normalise the coefficient, and the constants A and B are such that r_{int} takes values in the range $[-1, 1]$: 1 when the two variables are identical ($S_1 = S_2$), and -1 when they are completely opposite ($S_1 \cap S_2 = \emptyset$). These two conditions imply that $A + B = 1$ and $A = -1$, respectively; thus, $A = -1$, $B = 2$ and therefore

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

This coefficient allows us to compare any two ordinal variables. Our purpose now is to apply it to the performance measurement of a classifier. As above, we will do that by comparing the two variables C and $\hat{C}$, corresponding to the true and predicted classes. As these two variables take values in the set $\{1, 2, \ldots, K\}$, there will be many observations with the same value in each variable. This fact led us to define a much quicker way of computing r_{int}, which is based on the contingency table crossing C with $\hat{C}$:

C \ $\hat{C}$	1	2	...	K	TOTAL
1	n_{11}	n_{12}	...	n_{1K}	$n_{1\bullet}$
2	n_{21}	n_{22}	...	n_{2K}	$n_{2\bullet}$
⋮	⋮	⋮	⋮	⋮	⋮
K	n_{K1}	n_{K2}	...	n_{KK}	$n_{K\bullet}$
TOTAL	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet K}$	n

In this table, n_{ij} represents the number of observations whose true class is $\mathcal{C}_i$ and whose predicted class is $\mathcal{C}_j$. The total number of observations whose true class is $\mathcal{C}_i$ is given by the sum of row i, $n_{i\bullet}$. The total number of observations whose predicted class is $\mathcal{C}_j$ is given by the sum of column j, $n_{\bullet j}$. Hence,

$$\text{card}(S_1) = \sum_{i=1}^{K} n_{i\bullet} \left(\sum_{j=i}^{K} n_{j\bullet} - 1 \right) = \sum_{i=1}^{K} \sum_{j=i}^{K} n_{i\bullet} n_{j\bullet} - n,$$

$$\text{card}(S_2) = \sum_{i=1}^{K} n_{\bullet i} \left(\sum_{j=i}^{K} n_{\bullet j} - 1 \right) = \sum_{i=1}^{K} \sum_{j=i}^{K} n_{\bullet i} n_{\bullet j} - n,$$

and

$$\begin{aligned} \text{card}(S_1 \cap S_2) &= \sum_{i=1}^{K} \sum_{j=1}^{K} n_{ij} \left(\sum_{i'=i}^{K} \sum_{j'=j}^{K} n_{i'j'} - 1 \right) \\ &= \sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{i'=i}^{K} \sum_{j'=j}^{K} n_{ij} n_{i'j'} - n, \end{aligned}$$

which can now be considered in the definition of r_{int}. We note that in Pinto da Costa et al. (2008) there was an error in $\text{card}(S_1) = \sum_{i=1}^{K} \sum_{j=1}^{K} n_{i\bullet} n_{j\bullet} - n$ and similarly in $\text{card}(S_2)$. The computation of r_{int} involves terms of order K^4, but as K is usually very small, this shall present no problem. Finally, we get the expression:

$$r_{\text{int}} = -1 + 2 \frac{\sum_{i=1}^{K} \sum_{j=1}^{K} \sum_{i'=i}^{K} \sum_{j'=j}^{K} n_{ij} n_{i'j'} - n}{\sqrt{\left(\sum_{i=1}^{K} \sum_{j=i}^{K} n_{i\bullet} n_{j\bullet} - n \right) \left(\sum_{i=1}^{K} \sum_{j=i}^{K} n_{\bullet i} n_{\bullet j} - n \right)}}.$$

Acknowledgments

The first author was partially funded by the European Regional Development Fund through the program COMPETE and by the Portuguese Government through the FCT - Fundação para a Ciência e a Tecnologia under the project PEst-C/MAT/UI0144/201.

References

Baccianella, Stefano, Esuli, Andrea, & Sebastiani, Fabrizio (2009). Evaluation measures for ordinal regression. In *Proceedings of the ninth international conference on intelligent systems design and applications* (pp. 283–287).

Cardoso, Jaime S., & Sousa, Ricardo (2011). Measuring the performance of ordinal classification. *International Journal of Pattern Recognition and Artificial Intelligence*, *25*(8), 1173–1195.

Cruz-Ramírez, M., Hervás-Martínez, C., Sánchez-Monedero, J., & Gutiérrez, P. A. (2011). A preliminary study of ordinal metrics to guide a multi-objective evolutionary algorithm. In *2011 11th international conference on proceedings of the intelligent systems design and applications (ISDA)* (pp. 1176–1181). IEEE.

Cruz-Ramírez, M., Hervás-Martínez, C., Sánchez-Monedero, J., & Gutiérrez, P. A. (2014). Metrics to guide a multi-objective evolutionary algorithm for ordinal classification. *Neurocomputing*, *135*(0), 21–31. Advances in Learning Schemes for Function Approximation Selected papers from the 11th International Conference on Intelligent Systems Design and Applications (ISDA 2011).

Hu, QingHua, Guo, MaoZu, Yu, DaRen, & Liu, JinFu (2010). Information entropy for ordinal classification. *Science China Information Sciences*, *53*, 1188–1200.

Lee, J.W.T., & Liu, Da-Zhong (2002). Induction of ordinal decision trees. In *Proceedings of the international conference on machine learning and cybernetics*, Vol. 4 (pp. 2220–2224).

Pinto da Costa, J. F., Alonso, Hugo, & Cardoso, Jaime S. (2008). The unimodal model for the classification of ordinal data. *Neural Networks*, *21*, 78–91.

Pinto da Costa, Joaquim F., Sousa, Ricardo, & Cardoso, Jaime S. (2011). An all-at-once unimodal svm approach for ordinal classification. In *Proceedings of the 2010 ninth international conference on machine learning and applications* (pp. 59–64). IEEE.

Press, William, Flannery, Brian, Teukolsky, Saul, & Vetterling, William (1992). *Numerical recipes in C: the art of scientific computing*. Cambridge University Press.

Sánchez-Monedero, J., Gutiérrez, Pedro A., Tiño, Peter, & Hervás-Martínez, C. (2013). Exploitation of pairwise class distances for ordinal classification. *Neural Computation*, *25*, 2450–2485.

Seah, Chun-Wei, Tsang, Ivor W., & Ong, Yew-Soon (2012). Transductive ordinal regression. *IEEE Transactions on Neural Networks and Learning Systems*, *23*, 1074–1086.

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, *15*, 72–101.

Waegeman, Willem, De Baets, Bernard, & Boullart, Luc (2006). A comparison of different roc measures for ordinal regression. In *Proceedings of the CML 2006 workshop on ROC Analysis in Machine Learning*.

Vanbelle, S., & Albert, A. (2009). A note on the linearly weighted kappa coefficient for ordinal scales. *Statistical Methodology*, *6*(2), 157–163.