Artificial Intelligence 175 (2011) 1223–1250

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

On the ERA ranking representability of pairwise bipartite ranking functions

Willem Waegeman *, Bernard De Baets

KERMIT, Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Coupure links 653, B-9000 Ghent, Belgium

ARTICLE INFO

Article history:
Received 31 January 2009
Received in revised form 19 June 2010
Accepted 19 June 2010
Available online 2 December 2010

Keywords:
Pairwise bipartite ranking
Reciprocal preference relation
Cycle transitivity
Receiver operating characteristics (ROC) analysis
Graph theory
Multi-class classification
Decision theory
Machine learning

ABSTRACT

In domains like decision theory and social choice theory it is known for a long time that stochastic transitivity properties yield necessary and sufficient conditions for the ranking or utility representability of reciprocal preference relations. In this article we extend these results for reciprocal preference relations originating from the pairwise comparison of random vectors in a machine learning context. More specifically, the expected ranking accuracy (ERA) is such a reciprocal relation that occurs in multi-class classification problems, when ranking or utility functions are fitted to the data in a pairwise manner. We establish necessary and sufficient conditions for which these pairwise bipartite ranking functions can be simplified to a single ranking function such that the pairwise expected ranking accuracies of both models coincide. Similarly as for more common reciprocal preference relations, cycle transitivity plays a crucial role in this new setting. We first consider the finite sample case, for which expected ranking accuracy can be estimated by means of the area under the ROC curve (AUC), and subsequently, we further generalize these results to the underlying distributions. It turns out that the ranking representability of pairwisely compared random vectors can be expressed elegantly in a distribution-independent way by means of a specific type of cycle transitivity, defined by a conjunctor that is closely related to the algebraic product.

1. Introduction

Multi-class classification and ordinal regression can be seen as two closely related machine learning settings that share many properties. Multi-class classification refers to the supervised learning problem of inferring a predictive model capable of classifying data into a finite number of classes. This simply means that the model predicts for new data instances an output (also called label or response variable) that takes values in a finite unordered set (for example, class labels red, green, blue). Ordinal regression considers a slightly different setting. Labels here come from a finite ordered set, in which the order naturally follows from the semantics of the classes (for example, class labels bad, moderate, good). As a specific case of preference learning, ordinal regression problems typically arise in situations where humans are involved in the data generation process, like human experts or internet users expressing preferences on objects w.r.t. characteristics such as quality, beauty, appropriateness, etc.

So, the different semantics of the data respectively result in the absence or presence of an order relation on the classes in multi-class classification or ordinal regression. Owing to this important interpretation of the classes, substantially different methods have been proposed in the past for the two types of learning problems. Briefly summarized, the absence or presence of an order relation leads to two main differences in assumptions:

* Corresponding author. Tel.: +329 264 6018, fax: +329 264 6220.
E-mail address: Willem.Waegeman@UGent.be (W. Waegeman).

0004-3702/$ – see front matter
doi:10.1016/j.artint.2010.11.006

(1) Firstly, both models typically differ in the type of performance measure they optimize. If an order relation on the classes can be assumed, then a performance measure that takes this order into account must be utilized, both for optimization and evaluation. For example, in ordinal regression, misclassifying an object of class “bad” into class “good” must typically lead to a higher loss than misclassifying the same object into class “moderate”.
(2) Secondly, the absence or presence of an order relation on the classes gives rise to a different model structure for the two types of problems. The model structure of multi-class classification methods typically consists of an ensemble of binary classifiers, such as one-versus-one [26,30] and one-versus-all [41] ensembles, while typically only one global model is considered in ordinal regression. Moreover, this global model always consists of an underlying latent variable that reflects the order on the classes. Let $\mathcal{X}$ denote the set of data objects, then this latent variable serves as a ranking function $f:\mathcal{X}\to\mathbb{R}$ that defines a total order on the data objects. The final decision rule is then in the end obtained by placing a number of thresholds on the ranking function. This is for example the case in traditional statistical ordinal regression algorithms [2,38] and kernel-based methods [6,42].

Several authors [27,35,44] empirically analyzed in recent work the relationship between multi-class classification and ordinal regression, in which they primarily aim to improve ordinal regression algorithms by using ideas from multi-class classification, without considering an underlying ranking function. Conversely, the motivation of this article is to improve multi-class classification algorithms by using techniques from ordinal regression. Moreover, we will mainly focus on the theoretical connections between both problem settings, and to establish such a connection, we will take the ranking function that characterizes ordinal regression models as starting point. In this context, expected ranking accuracy (ERA) is a ranking-based performance measure that has recently been introduced for bipartite ranking [1] and further extended to ordinal regression [47]. Expected ranking accuracy can be easily considered too in multi-class classification, especially for one-versus-one ensembles, where the ensemble contains a set of pairwise bipartite ranking functions (i.e. one bipartite ranking function for each pair of classes). By using concepts from receiver operator characteristics (ROC) analysis, graph theory, decision theory and preference modeling, we will show that transitivity properties of the reciprocal relation generated by expected ranking accuracy result in a connection between multi-class classification and ordinal regression models.

Roughly speaking, we will investigate the conditions for which a one-versus-one ensemble, containing a set of bipartite ranking functions, can be reduced to an ordinal regression model with only one underlying ranking function, such that both models obtain an identical performance in terms of expected ranking accuracy. We will further refer to this property as ERA ranking representability of a one-versus-one ensemble. ERA ranking representability can be interpreted as a natural extension to the infinite sample case of AUC ranking representability, as previously introduced in [46]. It is well known that the area under the ROC curve (AUC) forms an unbiased estimator of the expected ranking accuracy on a finite dataset. Let us as an introductory example in a multi-class classification setting consider the following hypothetical three-class dataset that contains six objects of each class:

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
y_i	$\mathcal{C}_1$	$\mathcal{C}_1$	$\mathcal{C}_1$	$\mathcal{C}_1$	$\mathcal{C}_1$	$\mathcal{C}_1$	$\mathcal{C}_2$	$\mathcal{C}_2$	$\mathcal{C}_2$	$\mathcal{C}_2$	$\mathcal{C}_2$	$\mathcal{C}_2$	$\mathcal{C}_3$	$\mathcal{C}_3$	$\mathcal{C}_3$	$\mathcal{C}_3$	$\mathcal{C}_3$	$\mathcal{C}_3$

We have for simplicity assigned the indices in such a way that pairwise AUCs can be computed easily for a given ranking. Remark that the AUC simply computes the fraction of (lower class, higher class) couples that are correctly ranked by the classifier. Let us suppose that the following triplet of bipartite ranking functions is statistically inferred by a one-versus-one ensemble for this small toy problem:

	i											
ranking for f_{12}	7	8	1	2	9	3	4	5	6	10	11	12
ranking for f_{23}	13	7	14	8	9	10	11	12	15	16	17	18
ranking for f_{13}	13	1	2	3	14	15	16	17	18	4	5	6

So, from left to right, the numbers represent the ranking of the indices of the data objects, respectively obtained with the ranking functions f_{12}, f_{23} and f_{13}. For the pairwise AUCs we find:

$$\widehat{A}_{12}(f_{12},D)=20/36,\qquad \widehat{A}_{23}(f_{23},D)=25/36,\qquad \widehat{A}_{13}(f_{13},D)=15/36. \tag{1}$$

In other words, one finds for instance that 20 of the 36 couples are correctly ranked by the ranking function f_{12}: object number 1 is ranked before four objects of class $\mathcal{C}_2$, as well as object number 2, object number 3 is ranked before three objects of class $\mathcal{C}_2$, and so on. A more formal definition of the AUC will be given in Section 2.

In this example, the triplet of bipartite rankings can still be replaced in different ways by a single ranking of the whole data set such that the same pairwise AUCs are measured, for example

	i																	
ranking for global f	13	1	2	3	7	8	9	10	11	14	15	16	17	18	4	5	12	6

is such a ranking that results in the same pairwise AUCs. Verification of AUC ranking representability is much more difficult for larger datasets, since enumerating all global rankings is then computationally infeasible. However, in [46] we have shown