International Journal of Approximate Reasoning 76 (2016) 96–110

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Ordinal classification based on the sequential covering strategy

Juan Carlos Gámez [a,*], David García [b,1], Antonio González [b,1], Raúl Pérez [b,1]

[a] Dpto de Arquitectura de Computadores, Electrónica y Tecnología Electrónica, Universidad de Córdoba, Spain
[b] Dpto de Ciencias de la Computación e IA, ETSIIT, Universidad de Granada, Spain

ARTICLE INFO

Article history:
Received 15 December 2015
Received in revised form 17 May 2016
Accepted 19 May 2016
Available online 24 May 2016

Keywords:
Ordinal classification
Sequential covering strategy
Genetic algorithms
Fuzzy rules
NSLV
Supervised learning

ABSTRACT

Ordinal classification is a supervised learning problem. The distinctive feature of ordinal classification is that there is an order relationship among the categories to learn. In this paper, we present a fuzzy rule learning algorithm based on the sequential covering strategy applied to ordinal classification. This proposal modifies a nominal classification algorithm, called NSLV, to adapt it to this kind of problems. To take into account the order relationship among the categories, a new fitness function and a new concept of negative examples for a rule are proposed. Moreover, we introduce a new rule evaluation model for ordinal classification problems. Experimental results show that the proposed algorithm offers a better performance compared to other ordinal algorithms.

1. Introduction

Ordinal classification problems [1,2] are supervised learning problems that share some features with nominal classification [3] and regression-based [4] classification. In a similar way to nominal classification, the ordinal classification learns classes, but in this case there is an established order among them. From the regression point of view, the ordinal classification learns an ordered set of patterns, but in this case it is a set of finite labels, and the distance among the label values is not defined. The ordinal classification problems can be defined as prediction problems of an unknown value of an attribute $y = \{y_1, y_2, \ldots, y_Q\}$, being Q the number of classes. Moreover, the class labels have a predetermined order $y_1 \prec y_2 \prec \ldots \prec y_Q$. For example, in weather prediction, we can consider $y = \{VeryCold, Cold, Warm, Hot, VeryHot\}$ as a set of predicted values where there is a determined order between the classes $VeryCold \prec Cold \prec Warm \prec Hot \prec VeryHot$. In this example, considering the real class Hot, it is obvious that there is a more serious error by predicting the class $VeryCold$ than $Warm$. Because of that, different techniques are required to take advantage of the knowledge about the pattern distribution in the input [5].

Due to the similarities of ordinal classification with nominal classification and regression, classification and regression metrics are applied for the evaluation of this kind of algorithms, although there are specifics metrics which are also used in this kind of problems [6–8].

Ordinal classification is being used in multiple areas, such us economy [9], medicine [10,11], recommendation systems [12], movie industry [13] or wind speed reconstruction [14], among others.

* Corresponding author. Tel.: +34 957 212 039; fax: +34 957 218 315.
E-mail addresses: jcgamez@uco.es (J.C. Gámez), dgarcia@decsai.ugr.es (D. García), A.Gonzalez@decsai.ugr.es (A. González), Raul_Perez@decsai.ugr.es (R. Pérez).
[1] Tel.: +34 958 244019; fax: +34 958 243317.

http://dx.doi.org/10.1016/j.ijar.2016.05.002
0888-613X/

There are several approaches to tackle the ordinal classification. Several works deal with ordinal classification by converting the class value to a numeric quantity and applying a regression learner to the transformed data, and then they translate the output back to a discrete class value in a post-processing step. A disadvantage of this method is that it can only be applied in conjunction with a regression scheme. Another approach is to convert the learning of the ordinal classifier into learning of a set of two-class classifiers. The resulting two-class classifiers are trained by a modified well-known classification algorithm, and finally another conversion is needed to get the result as an ordinal value in any case. Our proposal approaches the ordinal classification with the adaptation of a multi-class classifier [15] in order to consider the features of ordinal problems.

In this paper, we present a proposal based on a sequential covering strategy for dealing with ordinal classification problems. This proposal is based on the use of a new fitness function together with a new definition of the negative examples for a rule. In this new definition we take into account the distances among the classes in order to take advantage of the specific features of the ordinal classification problems. The baseline algorithm of this proposal is NSLV [15,16], a fuzzy rule learning algorithm which has demonstrated a good behavior in nominal classification. The aforementioned algorithm uses the sequential covering strategy [17] where an iterative process extracts a rule in each step until no more rules can be extracted. Finally, this algorithm returns a set of weighted fuzzy rules.

In this work we try to test the performance of a new approach to tackle ordinal classification. In this proposal a set of fuzzy rules is obtained. This set of fuzzy rules can help us to understand the behavior of each particular ordinal problem.

This paper is organized as follows. Firstly, a brief summary of related works is presented in section 2. Section 3 presents the sequential covering strategy of NSLV. Several necessary modifications to adapt the algorithm to ordinal classification problems are proposed in section 4. Section 5 shows the experimental results and compares the proposal with several works of the literature. Finally, the conclusions and the future lines of work are presented in section 6.

2. Related works

In this section we present several approaches to ordinal classification. We show some ordinal classification algorithms as well as other proposals which are used to solve this kind of problems. These approaches can be classified in several ways.

Regarding the method or technique used to solve the problem, we can find several works which use SVM (Support Vector Machine) in some way, such as:

- In [18,19] the authors propose two new support vector approaches for ordinal regression. The first one takes into account only the adjacent ranks in order to set the thresholds. The authors introduces explicit constraints in the problem formulation by enforcing the inequalities on the thresholds. The second approach considers the training samples from all ranks to determine each threshold. The ordinal inequalities on the thresholds are automatically satisfied at the optimal solution, although there are no explicit constraints on these thresholds. An adaptation of the SMO (Sequential Minimal Optimization) algorithm [20,21] for SVMs is presented.
- Three methods for multi-class SVM are introduced in [22]. Although they are not exactly ordinal classifiers, they have been used to solve ordinal classification problems, so they are included in this section. The authors present the methods “one-against-all”, “one-against-one” and “direct acyclic graph SVM” (DAGSVM) to deal with ordinal classification problems. The experimental results show that “one-against-one” method may be the best suitable method for practical use. This method builds two-classes classifiers. Considering k classes, k(k-1)/2 classifiers are created for all possible combinations among classes.
- Paper [23] sets the base for C-RNK algorithm used in the experimentation section. This work, which is considered in others works [24–28], proposes a method for training rankSVMs efficiently. The approach allows the flexible use of various unconstrained optimization methods for training the rankSVM kernel efficiently by focusing on the loss function.
- In [29] the authors deal with ordinal regression by focusing the attention on the imbalanced classification problems. The authors propose methods to carry out an over-sampling process from a graph-based perspective. These methods consider the ordinal information in the over-sampling process. They use the SMOTE (Synthetic Minority Over-sampling TEchnique) algorithm [30] and a reformulation of the SVM paradigm for ordinal classification [19] in order to improve both the classification and the ordering of minority classes. Moreover, they propose a cost-sensitive extension of the ordinal SVM classifier where the minority class errors gain importance.

There are other proposals which use a classical statistical approach instead of SVM-based methods. For instance, a regression method which extends the KDL (Kernel Discriminant Learning) using a rank constraint is proposed in [31]. They extend the traditional LDA (Linear Discriminant Algorithm) and KDA (Kernel Discriminant Algorithm) algorithms in order to take advantage of the class rank information for ordinal regression. In [32] the authors propose to jointly learn the thresholds across samples and class centroids by seeking an optimal direction along which all the samples are distributed as in order as possible and maximally cater for nearest-centroid distributions.

Moreover, other authors approach the ordinal classification like a set of binary classifications:

- The algorithm proposed in [22], and previously described, which, in spite of they are not ordinal classifiers itself, can be included in this classification.