Cautious label-wise ranking with constraint satisfactions

Sébastien Destercke and Yonatan Carranza-Alarcon[1]

Abstract. Ranking problems are usually difficult to solve, due to their combinatorial nature. One way to circumvent this issue is to adopt a decomposition scheme, in which the initial difficult problem is split into a set of simpler problems. The predictions obtained from these simplified settings must then be combined into one single output, possibly resolving observed inconsistencies between the outputs. In this paper, we consider such an approach for the label ranking problem, where in addition we allow the predictive model to produce cautious inferences in the form of sets of rankings when it lacks information to produce reliable, precise predictions. More particularly, we propose to combine a rank-wise decomposition, in which every sub-problem becomes an ordinal classification one, with a constraint satisfaction problem (CSP) approach to verify the overall consistency of the predictions.

1 Introduction

In recent years, machine learning problems with structured outputs received an increasing interest. These problems appear in a variety of fields such as biology [12] or image analysis [7], ...

In this paper, we deal with the problem of *label ranking*, where one has to learn a mapping from instances to rankings (complete orders) defined over a finite, usually limited number of labels. Most solutions to this problem reduce the initial complexity of the problem, e.g., through a decomposition scheme. For example, ranking by pairwise comparison (RPC) [9] transforms the problem of label ranking into binary problems. Other label-wise approaches learn one model per label, trying to predict its relative importance/rank [2].

In ranking problem and in preference learning in general, it may also be interesting [1] to predict partial rather than complete rankings, abstaining to make a precise prediction in presence of too little information. Such cautious predictions can prevent harmful decisions based on incorrect predictions, and have been applied for different decomposition schemes [3, 5], always producing cautious predictions in the form of partial order relations.

In this paper, we propose a label-wise decomposition where each sub-problem intends to predict a set of rank. More precisely, we propose to learn for each label an imprecise ordinal regression model of its rank [6], and use these models to infer a set of possible ranks. We then use CSP techniques on the set of resulting predictions to check whether the outputs are consistent with a global ranking (i.e., that each label can be assigned a different rank)

Section 2 explains the main principles of our method, while Section 3 provides some elements indicating why we think that a rank-wise decomposition can be useful to consider.

[1] Sorbonne Université, UMR CNRS 7253 Heudiasyc, Université de Technologie de Compiègne CS 60319 - 60203 Compiègne cedex, France.

2 Method description

In label ranking, an instance $\mathbf{x}$ is associated to an order relation $\succ_{\mathbf{x}}$ over $\Lambda \times \Lambda$, or equivalently to a complete ranking over the labels in Λ, where $\Lambda = \{\lambda_1, \ldots, \lambda_k\}$ is the set of possible labels. Since learning a predictive function directly from the full rankings is usually difficult (as the size of the output space is $k!$), we propose first to consider a rank-wise aggregation technique, splitting the initial data sets into several ordinal classification data sets (each label is then associated to its ranks in the different instances), as shown in Figure 1.

For each data set $\mathbb{D}_i$, we then have to solve a problem known as ordinal classification or regression [10]. For each of this problem, we propose to learn a convex set Π of probabilities, which size reflect our lack of knowledge about the rank of λ_i. To transform Π into a set of possible rank, we combine a model p-boxes, already used to solve cautious ordinal problems [6], and a L_1 loss function, whose interest is that the resulting set of predicted rank is an interval [4], in the sense that $\{1, 3\}$ cannot be a prediction but $\{1, 2, 3\}$ can.

As for the RPC approach (and its cautious versions [1, 3]), the label-wise decomposition requires to aggregate all decomposed models into a single (partial) prediction. Indeed, nothing forbids to predict the same rank for multiple labels.

Checking whether a solution exists where all labels have a different rank and assessing the number of possible completions can be done by solving a CSP with a *alldifferent* constraint, whose resolution can be achieved by assignment and flow problems [8]. Also, in the case where possible ranks are intervals, we can speed up computations by using appropriate algorithms [11].

Example 1. *To illustrate the issue, let us consider the case where we have four labels* $\lambda_1, \lambda_2, \lambda_3, \lambda_4$*. Then the following predictions*

$$\hat{R}_1 = \{1, 3\}, \hat{R}_2 = \{1, 3\}, \hat{R}_3 = \{1, 3\}, \hat{R}_4 = \{2, 4\}$$

are inconsistent, simply because labels $\lambda_1, \lambda_2, \lambda_3$ *cannot be given simultaneously a different rank (note that pairwisely, they are not conflicting). On the other hand, the following predictions*

$$\hat{R}_1 = \{1, 2\}, \hat{R}_2 = \{1, 2, 3\}, \hat{R}_3 = \{2\}, \hat{R}_4 = \{1, 2, 3, 4\}$$

are consistent, but could be reduced to the unique ranking

$$\hat{R}'_1 = \{1\}, \hat{R}'_2 = \{3\}, \hat{R}'_3 = \{2\}, \hat{R}'_4 = \{4\}$$

as the strong constraint $\hat{R}_3 = \{2\}$ *propagates to all other constraints.*

3 Why rank-wise decomposition?

Most existing methods that propose to make set-valued or cautious predictions in ranking problems consider partial orders as their final predictions, that is pairwise relations $\succ_{\mathbf{x}}$ that are transitive and asymmetric, but no longer necessarily complete. However, rank-wise decomposition techniques also have some interest, that we briefly recall here.

$\mathbb{D}$

X_1	X_2	X_3	X_4	Y
107.1	25	*Blue*	60	$\lambda_1 \succ \lambda_3 \succ \lambda_2$
−50	10	*Red*	40	$\lambda_2 \succ \lambda_3 \succ \lambda_1$
200.6	30	*Blue*	58	$\lambda_2 \succ \lambda_1 \succ \lambda_3$
107.1	5	*Green*	33	$\lambda_1 \succ \lambda_2 \succ \lambda_3$
...	...	...	...	...

$\mathbb{D}_1$

X_1	X_4	Y
107.1	60	1
−50	40	3
200.6	58	2
107.1	33	1
...	...	...

$\mathbb{D}_2$

X_1	X_4	Y
107.1	60	3
−50	40	1
200.6	58	1
107.1	33	2
...	...	...

$\mathbb{D}_3$

X_1	X_4	Y
107.1	60	2
−50	40	2
200.6	58	3
107.1	33	3
...	...	...

Figure 1. Label-wise decomposition of rankings

3.1 Rank-wise decomposition in learning

Compared to pair-wise decomposition, in which $K(K-1)/2$ different sub-problems must be solved, which increases quadratically in the number of labels, a rank-wise decomposition only involves K different sub-problems, a linear increase in the number of labels.

One possible problem of the rank-wise approach is that partial observations of rankings [2] cannot always be reduced to sets of ranks (a prototypical example are comparisons of pairs). Nevertheless, some partial observations lends themselves well to rank-wise decomposition, such as top-k lists, which are quite common.

3.2 Rank-wise imprecise predictions

While a complete order can be equivalently described by the relation $\succ_{\mathbf{x}}$ or by the rank associated to each label, this is no longer true when one considers partial predictions. Indeed, consider for instance the case where the set of rankings over three labels $\{\lambda_1, \lambda_2, \lambda_3\}$ we would like to predict is $S = \{\lambda_1 \succ \lambda_2 \succ \lambda_3, \lambda_1 \prec \lambda_2 \prec \lambda_3\}$, which could correspond to an instance where λ_2 is a good compromise, and where the population is quite divided about λ_1 and λ_3 that represents more extreme options.

While the set S can be efficiently and exactly represented by providing sets of ranks for each items, none of the information it contains can be retained in a partial order. Indeed, the prediction $\hat{R}_1 = \{1,3\}, \hat{R}_2 = \{2\}, \hat{R}_3 = \{1,3\}$ perfectly represents S, while representing it by a partial order would result in the empty relation (since for all pairs i,j, we have $\lambda_i \succ \lambda_j$ and $\lambda_j \succ \lambda_i$ in the set S). We could, of course, find an example that would disadvantage a rank-wise cautious prediction over one using partial orders, as one representation is not more general than the other. Yet, our small example shows that considering both approaches make sense, as one cannot encapsulate the other, and vice-versa.

4 Work to do

In the next weeks, we will perform experiments on several benchmarks, and will develop simple metrics to measure the quality of our predictions, including whether or not we can reach a good balance between incompleteness and accuracy of the predictions.

REFERENCES

[1] W. Cheng, M. Rademaker, B. De Baets, and E. Hüllermeier, 'Predicting partial orders: ranking with abstention', *Machine Learning and Knowledge Discovery in Databases*, 215–230, (2010).

[2] Weiwei Cheng and Eyke Hüllermeier, 'A Nearest Neighbor Approach to Label Ranking based on Generalized Labelwise Loss Minimization', in *M-PREF'13, 7th Multidisciplinary Workshop on Preference Handling*, Beijing, (2013).

[3] Weiwei Cheng, Eyke Hüllermeier, Willem Waegeman, and Volkmar Welker, 'Label ranking with partial abstention based on thresholded probabilistic models', in *Advances in neural information processing systems*, pp. 2501–2509, (2012).

[4] Sébastien Destercke, 'On the median in imprecise ordinal problems', *Annals of Operations Research*, **256**(2), 375–392, (2017).

[5] Sébastien Destercke, Marie-Hélène Masson, and Michael Poss, 'Cautious label ranking with label-wise decomposition', *European Journal of Operational Research*, **246**(3), 927–935, (2015).

[6] Sébastien Destercke and Gen Yang, 'Cautious ordinal classification by binary decomposition', in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 323–337. Springer, (2014).

[7] X. Geng, 'Multilabel ranking with inconsistent rankers', in *Proceedings of CVPR 2014*, (2014).

[8] Ian P Gent, Ian Miguel, and Peter Nightingale, 'Generalised arc consistency for the alldifferent constraint: An empirical survey', *Artificial Intelligence*, **172**(18), 1973–2000, (2008).

[9] E. Hüllermeier, J. Furnkranz, W. Cheng, and K. Brinker, 'Label ranking by learning pairwise preferences', *Artificial Intelligence*, **172**, 1897–1916, (2008).

[10] Ling Li and Hsuan-Tien Lin, 'Ordinal regression by extended binary classification', in *Advances in neural information processing systems*, pp. 865–872, (2007).

[11] Alejandro López-Ortiz, Claude-Guy Quimper, John Tromp, and Peter Van Beek, 'A fast and simple algorithm for bounds consistency of the alldifferent constraint', in *IJCAI*, volume 3, pp. 245–250, (2003).

[12] N. Weskamp, E. Hullermeier, D. Kuhn, and G. Klebe, 'Multiple graph alignment for the structural analysis of protein active sites', *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, **4**(2), 310–320, (2007).