Learning Conditional Lexicographic Preference Trees

Michael Bräuning and Eyke Hüllermeier

Abstract We introduce a generalization of lexicographic orders and argue that this generalization constitutes an interesting model class for preference learning in general and ranking in particular. We propose a learning algorithm for inducing a so-called conditional lexicographic preference tree from a given set of training data in the form of pairwise comparisons between objects. Experimentally, we validate our algorithm in the setting of multipartite ranking.

1 Introduction

Preference learning is an emerging subfield of machine learning that has received increasing attention in recent years (Fürnkranz and Hüllermeier, 2011). A specific though important special case of preference learning is "learning to rank", that is, the learning of models that can be used to predict preferences in the form of rankings of a set of alternatives (Cohen et al, 1999; Dekel et al, 2003). Ranking problems are often reduced to problems of a simpler type, such as learning a value function that assigns scores to alternatives (with better

Michael Bräuning
Philipps-University Marburg,
✉ braeunim@mathematik.uni-marburg.de

Eyke Hüllermeier
University of Paderborn,
✉ eyke@upb.de

ARCHIVES OF DATA SCIENCE, SERIES A
KIT SCIENTIFIC PUBLISHING
Vol. 1, No. 1, S. 41–55, 2016

DOI 10.5445/KSP/1000058747/03
ISSN 2363-9881
(cc) BY-SA

alternatives having higher scores) or learning a binary predicate that compares pairs of alternatives (Hüllermeier et al, 2008). While the former approach is close to regression, the latter is in the realm of classification learning.

Another approach to learning ranking functions is to proceed from specific model assumptions, that is, assumptions about the structure of the sought preference relations. This approach is less generic than the previous one, as it strongly depends on the concrete assumptions made. On the other hand, it typically offers the advantage of being more easily understandable and interpretable. As an example, let us mention CP-networks, that is, the representation of conditional dependence and independence of preference statements under a ceteris paribus (all else being equal) interpretation (Boutilier et al, 2004). Those preferences are encoded as a graph, in which each node is annotated with a preference table. Another example is lexicographic orders that are widely accepted as a plausible representation of (human) preferences (Schmitt and Martignon, 2006), especially in complex decision making domains (Ahlert, 2008). Here, the assumption is that the target ranking of a set of alternatives, each one described in terms of multiple attributes, can be represented as a lexicographic order.

From a machine learning point of view, assumptions of the above type can be seen as an *inductive bias* restricting the hypothesis space. Provided the bias is correct, this is clearly an advantage, as it may simplify the learning problem. On the other hand, an overly strong bias may prevent the learner from approximating the target ranking sufficiently well. For example, while being plausible in some situations, the assumption of a lexicographic order will be too restrictive for many applications.

In this paper, we therefore present a method for learning generalized lexicographic orders. While still being simple and easy to understand, the model class we consider relaxes some of the assumptions of a proper lexicographic order. More specifically, we increase flexibility thanks to two extensions of conventional lexicographic orders:

- First, we allow for *conditioning* (Booth et al, 2009, 2010): The importance of attributes as well as the preferences for the values of an attribute may depend on the values of other variables preceding that one in the underlying variable order.
- Second, we allow for *grouping* (Wilson, 2009): Several (one-dimensional) variables can be grouped into a single high-dimensional variable, and preferences can be specified on the Cartesian product of the corresponding domains.

The remainder of this paper is organized as follows. In the next section, we give a brief overview of related work. In Sect. 3, we introduce generalized lexicographic orders and the notion of conditional lexicographic preference trees. In Sect. 4, we present an algorithm for learning such preference models from data. An experimental study is presented in Sect. 5, prior to concluding the paper in Sect. 6.

2 Related Work

The use of lexicographic orders in preference modeling has already been considered in the seventies of the last century (Fishburn, 1974), whereas in machine learning, this type of structure has attracted attention only recently. Flach and Matsubara developed a lexicographic ranker called LexRank , using a linear preference ordering on attributes derived by the odds ratio (Flach and Matsubara, 2007, 2008). Experimentally, they show that LexRank is competitive to decision trees and naive Bayes in terms of ranking performance.

Further work on learning lexicographic orders was done by Schmitt and Martignon (2006), Dombi et al (2007), and Yaman et al (2008). However, these works are based on rather simplistic assumptions. More general models were studied by Booth et al (2009, 2010), and in fact, important parts of our approach (such as conditional importance of attributes and conditional preferences on attribute values) are inspired by these models. Their work remains rather theoretical, however, without a practical realization in terms of an implementation of algorithms or an experimental study with real data.

3 Generalized Lexicographic Orders

Formally, we proceed from an attribute-value representation of decision alternatives or objects, i.e., an object is represented as a vector

$$o \in \mathscr{O} = \mathscr{D}(V) = \mathscr{D}(A_1) \times ... \times \mathscr{D}(A_n),$$

where $V = \{A_1, ..., A_n\}$ is the set of attributes (variables) and $\mathscr{D}(A_i)$ is the domain of attribute A_i. For a subset $A = \{A_{i_1}, \ldots, A_{i_k}\} \subset V$ of attributes we define $\mathscr{D}(A) = \mathscr{D}(A_{i_1}) \times ... \times \mathscr{D}(A_{i_k})$.

An *assignment* or *instantiation* of a subset $A \subseteq V$ of attributes is an element $a \in \mathscr{D}(A)$; an assignment is called *complete* if $A = V$, otherwise it is called *partial*. For an object $o \in \mathscr{O}$ and a subset $A \subset V$, we denote by $o[A]$ the projection of o from $\mathscr{D}(V)$ to $\mathscr{D}(A)$; if $A = \{A_k\}$ is a single attribute, we also write $o[k]$ instead of $o[\{A_k\}]$.

A lexicographic order on $\mathscr{O}$ is a total order $\succ$ defined in terms of

- a total order $\sqsupset$ on V, i.e., a ranking of the attributes,
- a total order $\sqsupset_i$ on each attribute domain $\mathscr{D}(A_i)$.

More specifically, $o^* \succ o$ (suggesting that o^* is preferred to o) if and only if there exists a $k \in \{1,\ldots,n\}$ such that

$$\big(o^*[k] \sqsupset_k o[k]\big) \wedge \Big((A_i \sqsupset A_k) \Rightarrow \big(o^*[i] = o[i]\big)\Big)$$

for all $i \in \{1,\ldots,n\}$. The relations $\sqsupset_i$ indicate preference on individual attributes: $a \sqsupset_i b$ means that, for $a,b \in \mathscr{D}(A_k)$, a is preferred to b as a value for attribute A_i. Moreover, the relation $\sqsupset$ reflects the importance of attributes: $A_i \sqsupset A_j$ means that attribute A_i is more important than A_j, whence the former is considered prior to the latter. Without loss of generality, we shall subsequently assume that $A_1 \sqsupset A_2 \sqsupset \cdots \sqsupset A_n$ (unless otherwise stated).

3.1 Conditional preferences on attribute values

Conventional lexicographic orders assume that preferences $\sqsupset_k$ on attribute domains are independent of each other. Needless to say, this assumption is often violated in practice. For example, although it is possible that a person prefers red wine to white wine *in general*, it is also plausible that her preference for wine may depend on the main dish: red is preferred to white in the case of meat, whereas white is preferred to red in the case of fish.

In order to capture attribute dependencies of that type, the preference relations $\sqsupset_k$ can be conditioned on the values of the attributes A_j preceding A_k in the order $\sqsupset$ (Booth et al, 2009, 2010). That is, $\sqsupset_k$ is now replaced by a set of strict orders

$$\Big\{ \sqsupset_k^{(a_1,\ldots,a_{k-1})} \mid (a_1,\ldots,a_{k-1}) \in \mathscr{D}(\{A_1,\ldots,A_{k-1}\}) \Big\}$$

Moreover, the order relation $\succ$ on $\mathscr{O}$ is then defined as follows: $o^* \succ o$ for $o^* = (a_1^*,\ldots,a_n^*)$ and $o = (a_1,\ldots,a_n)$ if and only if there exists a $k \in \{1,\ldots,n\}$ such that

$$\Big(\forall i \in \{1,\ldots,k-1\} : a_i^* = a_i\Big) \wedge \big(a_k^* \sqsupset_k^{(a_1,\ldots,a_{k-1})} a_k\big).$$

3.2 Conditional attribute importance

Going one step further, one may assume that the values of the first attributes in the attribute order $\sqsupset$ do not only influence the preferences on the values of the attributes that follow, but also the importance of the attributes themselves (Booth et al, 2009, 2010). Thus, we are no longer dealing with a lexicographic order in the sense that $\sqsupset$ defines a *sequence* of the attributes V according to their importance. Instead, we are dealing with a *tree-like* structure. This structure is defined by the following (choice) function:

$$A = C\Big((A_{i_1}, A_{i_2}, \ldots, A_{i_k}), (a_{i_1}, a_{i_2}, \ldots, a_{i_k})\Big),$$

where $(A_{i_1}, A_{i_2}, \ldots, A_{i_k}) \in V^k$ is a sequence of attributes (such that $A_{i_j} \neq A_{i_k}$ for $j \neq k$) and $a_{i_j} \in \mathscr{D}(A_{i_j})$ for all $j \in \{1,\ldots,k\}$. Moreover, $A \in V \setminus \{A_{i_1}, \ldots, A_{i_k}\}$ is the most important attribute given that $A_{i_j} = a_{i_j}$ for all $j \in \{1,\ldots,k\}$.

3.3 Variable grouping

Another extension consists of grouping several variables, that is, to allow the expression of preferences on *attribute tuples* instead of single attributes only (Wilson, 2009). Formally, this means selecting an index set $\mathscr{I} \subseteq \{1,\ldots,n\}$ and defining a total order relation $\sqsupset_{\mathscr{I}}$ on the Cartesian product $\mathscr{D}(V_{\mathscr{I}})$ of the domains $\mathscr{D}(A_i)$, $i \in \mathscr{I}$.

Note that the possibility of variable grouping significantly increases the expressivity of the model class. In particular, by taking $\mathscr{I} = \{1,\ldots,n\}$, it is possible to define every order on $\mathscr{D}(V)$, that is, to sort the set of alternatives in any way. Since this level of expressivity is normally not desirable, it is reasonable to restrict to variable grouping of order g_{max}, meaning to impose the constraint $|\mathscr{I}| \leq g_{max}$ for a fixed $g_{max} \leq n$.

3.4 Conditional lexicographic preference trees

Combining the generalizations discussed above, we end up with what we call a Conditional Lexicographic Preference Tree (CLPT). Graphically, this is a tree structure in which

- every node is labeled with a subset of attributes $V_{\mathscr{I}}$ and a total order on the Cartesian product $\mathscr{D}(V_{\mathscr{I}})$ of the corresponding attribute domains $\mathscr{D}(A_i)$, $i \in \mathscr{I}$;
- there is one outgoing edge (descendant node) for each value $o[V_{\mathscr{I}}] \in \mathscr{D}(V_{\mathscr{I}})$;
- every attribute $A_i \in V$ occurs at most once on each branch from the root of the tree to a leaf node (i.e., the index sets $\mathscr{I}$ along a branch are disjoint).

We call a CLPT *complete* if every attribute $A_i \in V$ occurs exactly once on each branch from the root of the tree to a leaf node (i.e., the index sets $\mathscr{I}$ along a branch form a partition of $\{1,\ldots,n\}$).

A (complete) CLPT can be thought of as defining an order relation on $\mathscr{O}$ through recursive refinement of a weak order $\succeq$, that is, by refining an order relation with tie groups in a recursive manner (in the following, $\sim$ and $\succ$ denote, respectively, the symmetric and asymmetric part of $\succeq$):

- One starts with a single equivalence class (tie group), i.e., $o^* \sim o$ for all $o^*, o \in \mathscr{O}$.
- Let the root of the CLPT be labeled with the attribute set $V_{\mathscr{I}}$, and let $\sqsupset_{\mathscr{I}}$ denote the corresponding order on $\mathscr{D}(V_{\mathscr{I}})$. The current order $\succeq$ is then refined by letting $o^* \succ o$ whenever $o^*[V_{\mathscr{I}}] \sqsupset_{\mathscr{I}} o[V_{\mathscr{I}}]$; otherwise, if $o^*[V_{\mathscr{I}}] = o[V_{\mathscr{I}}]$, then o^* and o remain tied.
- Thus, a linear order of tie groups (equivalence classes) is produced.
- Each equivalence class (represented by a value $a \in \mathscr{D}(V_{\mathscr{I}})$) is then recursively refined by the subtree the objects of this equivalence class are passed to.

Note that, if the CLPT is complete, the order relation $\succeq$ eventually produced is a total order $\succ$.

4 Learning CLPTs

In this section, we outline a method for inducing a CLPT from training data

$$\mathscr{T} = \left\{(o_i^*, o_i)\right\}_{i=1}^{N} \tag{1}$$

that consists of a set of object pairs $(o_i^*, o_i) \in \mathscr{O}^2$, suggesting that o_i^* is preferred to o_i. Roughly speaking, this means finding a CLPT whose induced order relation $\succeq$ on $\mathscr{O}$ is as much as possible in agreement with the pairwise preferences in $\mathscr{T}$ (without overfitting the training data). The induced order relation $\succeq$ is a total order $\succ$ if the CLPT is complete.

4.1 Performance and evaluation measures

In order to evaluate the predictive performance of a CLPT, there is a need to compare the order relation $\succeq$ (with asymmetric part $\succ$) induced by this model with a ground truth order $\succ^*$. As will be seen below, the same measures can be used to fit a CLPT to a given set of training data (1) during the training phase. In this case, the "ground truth" is not a total order but a set of pairwise comparisons between objects. Since a total order $\succ^*$ can be decomposed into (a quadratic number of) such comparisons, too, we can assume (without loss of generality) that we compare $\succeq$ with a set $\mathscr{T}$ of pairs $(o^*, o) \in \mathscr{O}^2$, suggesting that o^* should be ranked higher than o.

Inspired by the corresponding notions introduced in Cheng et al (2010), we define two performance measures of *correctness* and *completeness*, respectively, as follows:

$$\mathrm{CR}(\succeq, \mathscr{T}) = \frac{|C| - |D|}{|C| + |D|}, \tag{2}$$

$$\mathrm{CP}(\succeq, \mathscr{T}) = \frac{|C| + |D|}{|\mathscr{T}|}, \tag{3}$$

where

$$C = \left\{(o^*, o) \in \mathscr{T} \,|\, o^* \succ o\right\},$$
$$D = \left\{(o^*, o) \in \mathscr{T} \,|\, o \succ o^*\right\}.$$

Note that $\mathrm{CR}(\succeq, \mathscr{T})$ assumes values between -1 (complete disagreement) and $+1$ (complete agreement), while $\mathrm{CP}(\succeq, \mathscr{T})$ ranges between 0 (no comparisons) and 1 (full comparison).

4.2 A greedy learning procedure

We implement an algorithm for learning a CLPT as a (greedy) search in the space of tree structures based on the greedy algorithms presented by Schmitt and Martignon (2006) as well as Booth et al (2009, 2010). This is done by constructing the tree from the root to the leaves in a recursive manner. In each step of the recursion, a new node is created with an associated subset $V_{\mathscr{J}}$ of attributes, where $|V_{\mathscr{J}}| \leq g_{max}$, and a total order $\sqsupset_{\mathscr{J}}$ on $\mathscr{D}(V_{\mathscr{J}})$.

4.2.1 Creating a node

The problem to be solved in each recursion is the following: Given a set of pairwise comparisons $\mathscr{T}$ and a set $V' \subseteq V$ of attributes still available, select the most suitable subset $V_{\mathscr{J}} \subseteq V'$ and an order $\sqsupset_{\mathscr{J}}$. Following a greedy strategy, we choose $(V_{\mathscr{J}}, \sqsupset_{\mathscr{J}})$ so as to maximize correctness (2), using completeness (3) as a second criterion to break ties. In the (unlikely) event of both correctness and completeness having ties, the first subset $V_{\mathscr{J}}$ and order $\sqsupset_{\mathscr{J}}$ identified are selected.

The selection of an attribute subset $V_{\mathscr{J}}$ can be done through exhaustive search if its size is sufficiently limited, i.e., if the upper bound g_{max} is small. Otherwise, a complete enumeration of all possibilities may become too expensive. Moreover, for each candidate subset $V_{\mathscr{J}}$, a total order $\sqsupset_{\mathscr{J}}$ needs to be determined. Again, all such orders can be tried if $\mathscr{D}(V_{\mathscr{J}})$ is not too large. Otherwise, heuristic ranking procedures such as a Borda count can be used (counting the number of "wins" and "losses" of each value $a \in \mathscr{D}(V_{\mathscr{J}})$ in the training data $\mathscr{T}$ and sorting according to the difference).

4.2.2 Limiting the number of candidate subsets

In order to avoid a complete enumeration of all candidate subsets $V_{\mathscr{J}}$ of size $\leq g_{max}$, we combine a greedy search with a kind of lookahead procedure: We provisionally create a node by selecting a single attribute instead of a subset, i.e., we tentatively set g_{max} to 1; apart from that, exactly the same selection procedure (as outlined above) is applied. This step is repeated g_{max} times, thereby producing a subtree of depth g_{max}. Let $V^* \subseteq V$ denote the subset of attributes that occur in this subtree, i.e., that are chosen in at least one of the

nodes. Then, as candidate subsets $V_{\mathscr{J}}$, we only try subsets V^*, i.e., subsets $V_{\mathscr{J}} \subseteq V^*$ such that $|V_{\mathscr{J}}| \leq g_{max}$. Obviously, the underlying assumption is that an attribute that has not been chosen in any of the g_{max} steps is not important at this point.

4.2.3 Recursion

Once an optimal subset $V_{\mathscr{J}}$ has been chosen, the training examples (o^*, o) with $o^*[V_{\mathscr{J}}] \neq o[V_{\mathscr{J}}]$ are removed from $\mathscr{T}$ (since they are sorted at this node). Moreover, for each value $a \in \mathscr{D}(V_{\mathscr{J}})$, a data set

$$\mathscr{T}_a = \left\{ (o^*, o) \in \mathscr{T} \,|\, o^*[V_{\mathscr{J}}] = o[V_{\mathscr{J}}] = a \right\}$$

is created and passed to the corresponding successor node (together with $V' \setminus V_{\mathscr{J}}$ as the attributes that have not been used so far). The same recursive procedure is then applied to each of these successor nodes.

4.2.4 Initialization and termination

The learning procedure is called with the original training set $\mathscr{T}$ and the full set V of attributes as candidates. The recursion terminates if no attribute is left ($V' = \emptyset$) or if the set of training examples is empty ($\mathscr{T} = \emptyset$). A description of the basic algorithm in the form of pseudocode is provided in Algorithm 1.[1]

4.2.5 CLeRa

We call the algorithm outlined above *CLeRa*, which is short for Conditional Lexicographic Ranker . The CLPT induced by CLeRa can be used to compare new object pairs $\{o^*, o\} \subset \mathscr{O}$. To this end, the tuple is submitted to the root and propagated through the tree until either a leaf node is reached or a node at which $o^*[V_{\mathscr{J}}] \neq o[V_{\mathscr{J}}]$; in this case, $o^* \succ o$ is decided if $o^* \sqsupset_{\mathscr{J}} o$ and $o \succ o^*$ if $o \sqsupset_{\mathscr{J}} o^*$. Otherwise, if $o^*[V_{\mathscr{J}}] = o[V_{\mathscr{J}}]$ in all nodes traversed by the two objects, then $o^* \sim o$.

Given not only a pair but a complete set of objects to be ranked, the pairwise comparison realized by the CLPT can be embedded in any standard sorting

[1] The pseudocode does not consider the lookahead procedure.

Algorithm 1: CLeRa

Input : training data $\mathscr{T}$, set of attributes V, maximal grouping size g_{max}
Output : CLPT ct

$ct \leftarrow \emptyset, V' \leftarrow V, \mathscr{I}' \leftarrow \{1,\ldots,n\}$
if $\mathscr{T} \neq \emptyset$ && $V' \neq \emptyset$ **then**
  $I' \leftarrow \emptyset, CR \leftarrow 0, CP \leftarrow 0$
  for $\mathscr{I} \subseteq \mathscr{I}'$, $|\mathscr{I}| \leq g_{max}$ **do**
    determine $\sqsupset_{\mathscr{I}}$ on $\mathscr{D}(V_{\mathscr{I}})$ maximally consistent with $\mathscr{T}$
    compute $\mathrm{CR}(\sqsupset_{\mathscr{I}}, \mathscr{T})$ and $\mathrm{CP}(\sqsupset_{\mathscr{I}}, \mathscr{T})$
    if $\mathrm{CR}(\sqsupset_{\mathscr{I}}, \mathscr{T}) = \mathrm{CR}$ && $\mathrm{CP} < \mathrm{CP}(\sqsupset_{\mathscr{I}}, \mathscr{T})$ **then**
      $\mathrm{CP} \leftarrow \mathrm{CP}(\sqsupset_{\mathscr{I}}, \mathscr{T})$
      $I' \leftarrow \mathscr{I}$
    else if $\mathrm{CR}(\sqsupset_{\mathscr{I}}, \mathscr{T}) > \mathrm{CR}$ **then**
      $\mathrm{CR} \leftarrow \mathrm{CR}(\sqsupset_{\mathscr{I}}, \mathscr{T})$
      $\mathrm{CP} \leftarrow \mathrm{CP}(\sqsupset_{\mathscr{I}}, \mathscr{T})$
      $I' \leftarrow \mathscr{I}$
  $\mathscr{I}' \leftarrow \mathscr{I}' \backslash I'$
  $V' \leftarrow V' \backslash V_{I'}$
  remove every $(o, o') \in \mathscr{T}$ decided by $\sqsupset_{I'}$
  add node $(V_{I'}, \sqsupset_{I'})$ to ct
  for $a \in \mathscr{D}(V_{\mathscr{I}})$ **do**
    $\mathscr{T}_a = \left\{ (o^*, o) \in \mathscr{T} \,|\, o^*[V_{\mathscr{I}}] = o[V_{\mathscr{I}}] = a \right\}$
    return $\mathrm{CLeRa}[\mathscr{T}_a, V', g_{max}]$
return ct

algorithm, such as insertion sort. Note that, since $o^* \sim o$ is possible in a pairwise comparison, the result of the sorting procedure will in general only be a weak order $\succeq$.

5 Experimental Results

We evaluate our approach on 15 benchmark data sets from the Statlog and the UCI repository (Asuncion and Newman, 2007). These data sets, which define binary or ordinal classification problems, were pre-processed as follows: numerical attributes and attributes with more than five values were discretized into four values using equal frequency binning. Moreover, instances with missing values were neglected.

The learning problem we consider is multipartite ranking (Fürnkranz et al, 2009): Given a set of test instances $X \subset \mathscr{O}$, the goal is to predict a ranking $\succeq$ that agrees with the (ordered) class labels of these instances. Formally, this

agreement is measured in terms of the so-called C-index, which can be seen as an extension of the area under the ROC curve (AUC):

$$C = \frac{1}{\sum_{i<j} n_i n_j} \sum_{1 \le i < j \le m} \sum_{(o,o^*) \in X_i \times X_j} \mathbb{I}(o^* \succ o) + \frac{1}{2}\mathbb{I}(o^* \sim o),$$

where $X_i \subseteq X$ denotes the set of instances with class labels y_i, and these class labels are assumed to have the order $y_1 < y_2 < \cdots < y_m$. $\mathbb{I}(\cdot)$ is the indicator function mapping false predicates to 0 and true predicates to 1. The training data consists of a set of labeled instances, just like in classification. Since CLeRa is learning from pairwise comparisons of the form (o^*, o), it first extracts such comparisons from the original data by looking at the class information: A preference (o^*, o) is generated for each pair (o^*, y_j) and (o, y_i) of labeled instances in the (original) training data such that $y_i < y_j$.

The ranking performance of CLeRa (with maximum grouping size of $g_{max} = 2$) is compared with LexRank, which was implemented as proposed by (Flach and Matsubara, 2007, 2008); therefore, this method was only applied to binary (two-class) problems but not to problems with more than two classes.[2] We applied naive Bayes (NB) and decision tree (J48) learning as additional baselines, using the standard implementations[3] in the Weka machine learning toolbox Hall et al (2009) and sorting instances according to the estimated probability of the positive class; note that these methods are not applicable to the multi-class case either.

The results of a 10-fold cross-validation are given in Table 1. Since CLeRa produced a completeness of 1 or extremely close to 1 throughout, these values are not reported here. Overall, the performance of the methods is quite comparable but slightly in favour of NB. In particular, CLeRa and LexRank produce quite similar results on many data sets (Asuncion and Newman, 2007). In some cases, however, the results are strongly in favor of CLeRa:

- Census Income: The census data provides information about whether an income exceeds 50,000 USD over a year. The root node of the CLPT is labeled with a single attribute (capital-loss) as well as the descendant node. The preferences on attribute values of the descendant nodes at the third stage depend on the values of the node following the root node. This is also true

[2] The red wine data actually has a target attribute with values between 1 and 10; it was binarized by thresholding at the median.

[3] Trees are not pruned.

Table 1 Average performance in terms of C-index based on a 10-fold cross-validation (best results per data set highlighted in bold font).

Dataset	CLeRa	LexRank	J48	NB
Red Wine	0.7827 ± 0.0479	0.8011 ± 0.0475	0.7378 ± 0.0272	**0.8110** ± 0.0225
Census Income	0.7952 ± 0.0523	0.5776 ± 0.0256	0.7401 ± 0.0356	**0.8607** ± 0.0192
Credit Approval	0.9201 ± 0.0298	**0.9229** ± 0.0389	0.8517 ± 0.0480	0.9061 ± 0.0377
Mammographic Mass	0.8831 ± 0.0289	0.8960 ± 0.0327	0.8524 ± 0.0430	**0.8999** ± 0.0307
Mushroom	**1.0000** ± 0.0000	0.9865 ± 0.0021	**1.0000** ± 0.0000	0.9484 ± 0.0164
SPECT Heart	0.6740 ± 0.0767	0.6590 ± 0.1430	0.5106 ± 0.0961	**0.7409** ± 0.0957
Ionosphere	**0.9198** ± 0.0494	0.5748 ± 0.0740	0.8059 ± 0.1290	0.9061 ± 0.0805
MAGIC Gamma Telescope	0.8218 ± 0.0302	0.7263 ± 0.0517	0.7841 ± 0.0304	**0.8241** ± 0.0329
Breast Cancer Wisconsin	0.9837 ± 0.0171	0.9901 ± 0,0093	0.9793 ± 0.0392	**0.9909** ± 0.0091
German Credit	0.6285 ± 0.0880	0.4523 ± 0.1092	0.6251 ± 0.0902	**0.7835** ± 0.0647
Car Evaluation	**0.9198** ± 0.0185	n/a	n/a	n/a
Nursery	**0.9052** ± 0.0288	n/a	n/a	n/a
Tic-Tac-Toe Endgame	**0.7728** ± 0.0389	n/a	n/a	n/a
Vehicle	**0.7554** ± 0.0459	n/a	n/a	n/a
Cardiocraphic	**0.9551** ± 0.0138	n/a	n/a	n/a

for the importance of the attributes at this stage. One level below, the CLPT also contains nodes that are labeled with grouped attributes.

- Ionosphere: The radar data contains information about whether radar returns are "good" or "bad".[4] With regard to the conditional dependencies and the grouping, the basic structure of the CLPT is very similar to the aforementioned case.
- MAGIC Gamma Telescope: The gamma telescope data contains information about the registration of gamma particles. The basic structure of the CLPT differs from the aforementioned CLPTs with respect to the occurrence of conditional dependencies. Already the first descendant nodes exhibit conditional dependencies on the attribute values of the root node.
- German Credit: In the credit data, customers are classified as good or bad. The respective CLPT makes even stronger use of the proposed extensions compared to the CLPT for the MAGIC Gamma Telescope data set. The first descendant nodes are labeled with grouped attributes.

Overall, these results indicate that the bias imposed by the assumption of a standard lexicographic order is inadequate for these data sets, and hence

[4] Good returns show evidence of some type of structure in the ionosphere.

our extensions (conditional attribute importance, conditional value preferences, variable grouping) clearly pay off.

6 Conclusion and Future Work

Lexicographic orders constitute an interesting model class for preference learning, which allows for representing rankings of a set of objects in a very compact and comprehensible way. Yet, as we have argued in this paper, this model class may not be flexible enough for many real-world applications. Therefore, we have proposed to weaken the assumptions underlying a lexicographic order in various directions, allowing for conditional attribute importance, conditional preferences on attribute values, and variable grouping. Moreover, we have proposed an algorithm called CLeRa, which learns preference models in the form of conditional lexicographic preference trees from training data in the form of pairwise comparisons between objects.

First experimental results in the setting of multipartite ranking are quite promising and show CLeRa to be competitive with other methods. In a direct comparison with an existing lexicographic ranker, the benefit of our extensions are becoming quite obvious.

Important topics of future work can be found both on the theoretical and practical side. In particular, we are currently studying formal properties of our generalized model class, such as its expressiveness and means for regularization and complexity control. Practically, there is certainly scope for improving our current algorithm, for example by devising a suitable procedure for estimating an optimal value g_{max} for the oder of variable grouping. Moreover, improving the computational efficiency of CLeRa would be desirable, too. Last but not least, we are of course interested in real applications for which (generalized) lexicographic models appear to be an adequate representation.

References

Ahlert M (2008) Aggregation of lexicographic orderings. Homo Oeconomicus 25(3):301–317

Asuncion A, Newman DJ (2007) UCI machine learning repository. URL `http://archive.ics.uci.edu/ml/`

Booth R, Chevaleyre Y, Lang J, Mengin J, Sombattheera C (2009) Learning various classes of models of lexicographic orderings. In: Hüllermeier E, Fürnkranz J (eds) Preference Learning, Springer, Berlin, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, pp 1–16

Booth R, Chevaleyre Y, Lang J, Mengin J, Sombattheera C (2010) Learning conditionally lexicographic preference relations. In: Proc. ECAI 2010, IOS Press, Amsterdam, The Netherlands, pp 269–274

Boutilier C, Brafman RI, Domshlak C, Hoos HH, Poole D (2004) CP-nets: A tool for representing and reasoning with conditional ceteris paribus preference statements. Journal of Artificial Intelligence 21:135–191

Cheng W, Rademaker M, De Baets B, Hüllermeier E (2010) Predicting partial orders: Ranking with abstention. In: Balcázar J, Bonchi F, Gionis A, Sebag M (eds) Machine Learning and Knowledge Discovery in Databases, Lecture Notes in Computer Science, vol 6321, Springer, Berlin, pp 215–230, DOI 10.1007/978-3-642-15880-3_20

Cohen W, Schapire R, Singer Y (1999) Learning to order things. Journal of Artificial Intelligence Research 10:243–270, DOI 10.1613/jair.587

Dekel O, Manning CD, Singer Y (2003) Log-linear models for label ranking. In: Thrun S, Saul LK, Schölkopf B (eds) Advances in Neural Information Processing Systems, MIT, 16, pp 497–504

Dombi J, Imreh C, Vincze N (2007) Learning lexicographic orders. European Journal of Operational Research 183(2):748–756, DOI 10.1016/j.ejor.2006.10.029

Fishburn PC (1974) Lexicographic orders, utilities and decision rules: A survey. Management Science 20(11):1442–1471, DOI 10.1287/mnsc.20.11.1442

Flach P, Matsubara E (2008) On classification, ranking, and probability estimation. In: de Raedt L, Dietterich T, Getoor L, Kersting K, Muggleton SH (eds) Probabilistic, Logical and Relational Learning - A Further Synthesis, Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Dagstuhl, Germany, no. 07161 in Dagstuhl Seminar Proceedings

Flach PA, Matsubara ET (2007) A simple lexicographic ranker and probability estimator. In: Proceedings of the 18th European Conference on Machine Learning, Springer, Berlin, Heidelberg, ECML '07, pp 575–582, DOI 10.1007/978-3-540-74958-5_55

Fürnkranz J, Hüllermeier E (2011) Preference Learning. Springer-Verlag, Berlin, Heidelberg, DOI 10.1007/978-3-642-14125-6

Fürnkranz J, Hüllermeier E, Vanderlooy S (2009) Binary decomposition methods for multipartite ranking. In: Buntine W, Grobelnik M, Mladenić D, Shawe-Taylor J (eds) Machine Learning and Knowledge Discovery in Databases, Lecture Notes in Computer Science, vol 5781, Springer, Berlin, pp 359–374, DOI 10.1007/978-3-642-04180-8_41

Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA data mining software: An update. SIGKDD Explorations 11(1):10–18, DOI 10.1145/1656274.1656278

Hüllermeier E, Fürnkranz J, Cheng W, Brinker K (2008) Label ranking by learning pairwise preferences. Artificial Intelligence 172(16–17):1897–1916, DOI 10.1016/j.artint.2008.08.002

Schmitt M, Martignon L (2006) On the complexity of learning lexicographic strategies. Journal of Machine Learning Research 7:55–83

Wilson N (2009) Efficient inference for expressive comparative preference languages. In: Proceedings of the 21st International Joint Conference on Artificial Intelligence, Morgan Kaufmann, San Francisco, IJCAI'09, pp 961–966

Yaman F, Walsh TJ, Littman ML, desJardins M (2008) Democratic approximation of lexicographic preference models. In: Proc. ICML-08, Helsinki, Finland, pp 1200–1207