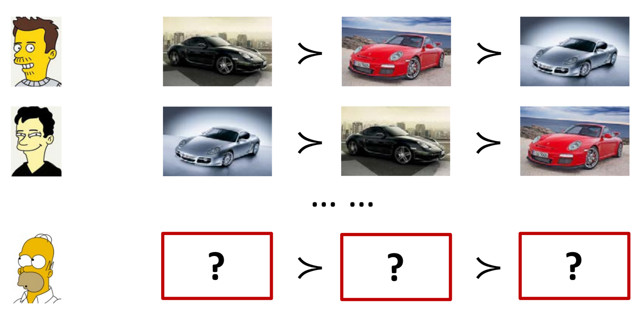
Label Ranking with Partial Abstention based on Thresholded Probabilistic Models

Weiwei Cheng[1], Eyke Hüllermeier[1], Willem Waegeman[2], Volkmar Welker[1]

[1]Mathematics and Computer Science, University of Marburg
[2]Mathematical Modeling, Statistics and Bioinformatics, Ghent University

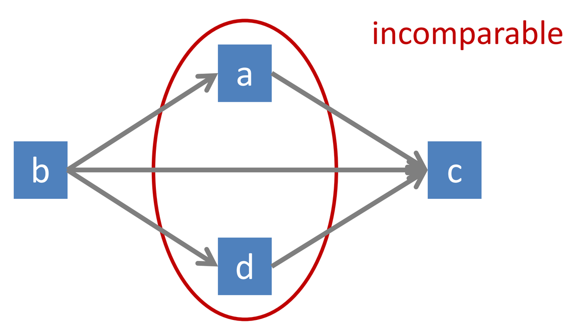
Background and Motivation

Label Ranking



Given training instances $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\} \subseteq \mathrm{X}$, labels $\mathrm{Y} = \{y_1, \dots, y_M\}$, and for each training instance $\boldsymbol{x}_k$ a set of *pairwise preferences* $y_i \succ_{\boldsymbol{x}_k} y_j$, find a ranking function ($\mathrm{X} \to \Omega$ mapping) that maps each $\boldsymbol{x} \in \mathrm{X}$ to a ranking $\succ_x$ of Y (permutation π_x).

From Total to Partial Order Relations



Partial abstention: The target is a total order, and a predicted partial order expresses incomplete knowledge about the target.

Predicting a partial order Q instead of a total order R: If $Q(i,j) = Q(j,i) = 0$, the ranker abstains on the label pair (y_i, y_j).

Partial Orders from Pairwise Comparisons

The idea is to rely on confident comparisons via thresholding:

$$Q(i,j) = [\![P(i,j) > q]\!] = \begin{cases} 1 & \text{if } P(i,j) > q \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

The probability of a pairwise preference can be obtained through marginalization:

$$P(i,j) = \mathbf{P}(y_i \succ y_j) = \sum_{\pi \in \mathrm{E}(i,j)} \mathbf{P}(\pi) \tag{2}$$

where $\mathrm{E}(i,j)$ is the set of linear extensions of $y_i \succ y_j$.

Previous Approach

	a	b	c	d
a		0.3	0.8	0.4
b	0.7		0.9	0.7
c	0.2	0.1		0.7
d	0.6	0.3	0.3	

thresholding at 0.6 →

	a	b	c	d
a		0	1	0
b	1		1	1
c	0	0		1
d	0	0	0	

In *Predicting Partial Orders: Ranking with Abstention* by Cheng et al. (ECML 2010), the pairwise probabilities $P(i,j)$ are learned independently of each other, leading to two problems:

- If q is not large enough, Q may have cycles.
- Even if Q is acyclic, it is not guaranteed to be transitive.

Our Ideas and Results

Can we restrict $P(\cdot,\cdot)$ to exclude the possibility of cycles and violations of transitivity from the very beginning?

- We make use of label ranking methods that produce probability distributions $\mathbf{P}$ over the ranking space Ω.
- We show that thresholding pairwise preferences induced by certain distributions yields partial order relations.
- Via variation of threshold, we are able to exploit the whole spectrum between a complete ranking and an empty relation.

Probability Models on Rankings

The Plackett-Luce model, a multi-stage model:

$$\mathbf{P}(\pi \mid \boldsymbol{v}) = \prod_{i=1}^{M} \frac{v_{\pi^{-1}(i)}}{v_{\pi^{-1}(i)} + v_{\pi^{-1}(i+1)} + \dots + v_{\pi^{-1}(M)}} \tag{3}$$

with $\boldsymbol{v} = (v_1, \dots, v_M) \in \mathrm{R}_+^M$, where $\pi^{-1}(i)$ is the index of the label ranked at position i.

The Mallows model, a distance-based model:

$$\mathbf{P}(\pi \mid \pi_0, \theta) = \frac{\exp(-\theta\, \mathrm{D}(\pi, \pi_0))}{\phi(\theta)} \tag{4}$$

where π_0 is the center ranking, θ is the spread, and D is a distance function on rankings, e.g., Kendall, Spearman's rho.

Theoretical Results

Thresholded Relations are Partial Orders

Definition 1. A distance D on Ω has the transposition property, if the following holds: Let π and π' be rankings and let (i,j) be an inversion, i.e., $i < j$ and $(\pi(i) - \pi(j))(\pi'(i) - \pi'(j)) < 0$. Let $\pi'' \in \Omega$ be constructed from π' by swapping y_i and y_j, that is, $\pi''(i) = \pi'(j), \pi''(j) = \pi'(i)$, and $\pi''(m) = \pi'(m)$ for all $m \in [M] \backslash \{i,j\}$. Then, $\mathrm{D}(\pi, \pi'') < \mathrm{D}(\pi, \pi')$.

Lemma 1. Let P be a reciprocal relation and let Q be given by (1). Then Q defines a strict partial order relation for all $q \in [0.5,1)$ if and only if P satisfies partial stochastic transitivity, i.e., $P(i,j) > 0.5$ and $P(j,k) > 0.5$ implies $P(i,k) \geq \min(P(i,j), P(j,k))$ for each triple $(i,j,k) \in [M]^3$.

Lemma 2. If y_i precedes y_j in the center ranking π_0 in (4), then $\mathbf{P}(y_i \succ y_j) \geq 0.5$. Moreover, if $\mathbf{P}(y_i \succ y_j) > q \geq 0.5$, then y_i precedes y_j in the center ranking π_0.

Lemma 3. If y_i precedes y_j and y_j precedes y_k in the center ranking π_0 in (4), then $\mathbf{P}(y_i \succ y_k) \geq \max(\mathbf{P}(y_i \succ y_j), \mathbf{P}(y_j \succ y_k))$.

Theorem 1. Let $\mathbf{P}$ in (2) be the PL model (3). Moreover, let Q be given by the threshold relation (1). Then Q defines a strict partial order relation for all $q \in [0.5,1)$.

Theorem 2. Let $\mathbf{P}$ in (2) be the Mallows model (4), with a distance D having the transposition property. Moreover, let Q be given by the threshold relation (1). Then Q defines a strict partial order relation for all $q \in [0.5,1)$.

Expressivity of the Model Classes

Lemma 5. For fixed $q \in (0.5,1)$ and a set A of subsets of $[M]$, the following are equivalent:

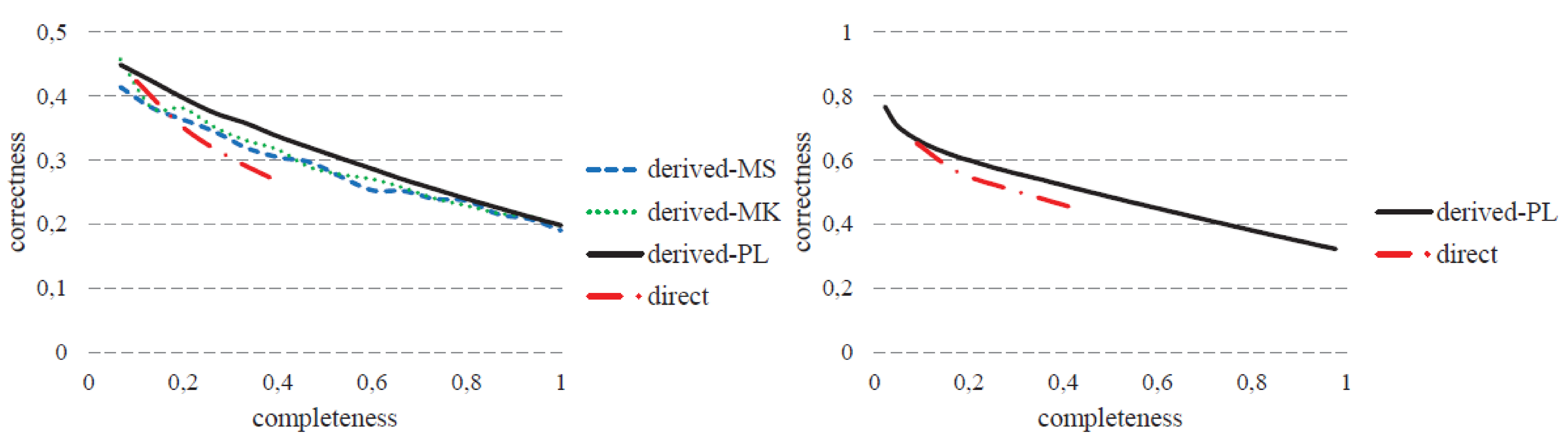
- The set A is the set of maximal antichains of a partial order induced by (2) on $[M]$ for some $v_1 > \dots > v_M > 0$.
- The set A is a set of mutually incomparable intervals that cover $[M]$.

Theorem 3. Let Γ_{M} denote the set of different partial orders (up to isomorphism) that can be represented as a threshold relation Q defined by (1), where P is derived according to (2) from the Mallows model (4) with D the Kendall distance. Then $|\Gamma_{\mathrm{M}}| = M$.

Theorem 4. Let Γ_{PL} denote the set of different partial orders (up to isomorphism) that can be represented as a threshold relation Q defined by (1), where P is derived according to (2) from the PL model (3). For any given threshold $q \in [0.5, 1)$, the cardinality of this set is given by the M^{th} Catalan number:

$$|\Gamma_{\mathrm{PL}}| = \frac{1}{M+1}\binom{2M}{M}.$$

Experiment and Conclusion



Tradeoff between completeness and correctness for the SUSHI label ranking data set: Existing pairwise method (direct) versus the probabilistic approach based on the PL model and Mallows model with Spearman's rho (MS) and Kendall (MK) as distance measure. The figure on the right corresponds to the original data set with rankings of size 10, and the figure on the left shows results for rankings of size 6.

- Thresholding a valued preference relation is a natural way of (partial) abstention in label ranking; the induced partial order can be interpreted as a confidence set (consisting of its linear extensions).
- While thresholding may yield inconsistencies in general, we have shown that proper partial orders are produced when restricting to preference relations induced by specific types of probability distributions on rankings.
- This approach is not only theoretically sound, but also performs well in experimental studies.
- While our focus was on label ranking, the results immediately apply to other ranking problems, too.

Neural Information Processing Systems Foundation

Poster presented at NIPS 2012, Lake Tahoe, Nevada, United States, December 2012