Label Ranking with Abstention

Predicting Partial Orders by Thresholding Probability Distributions

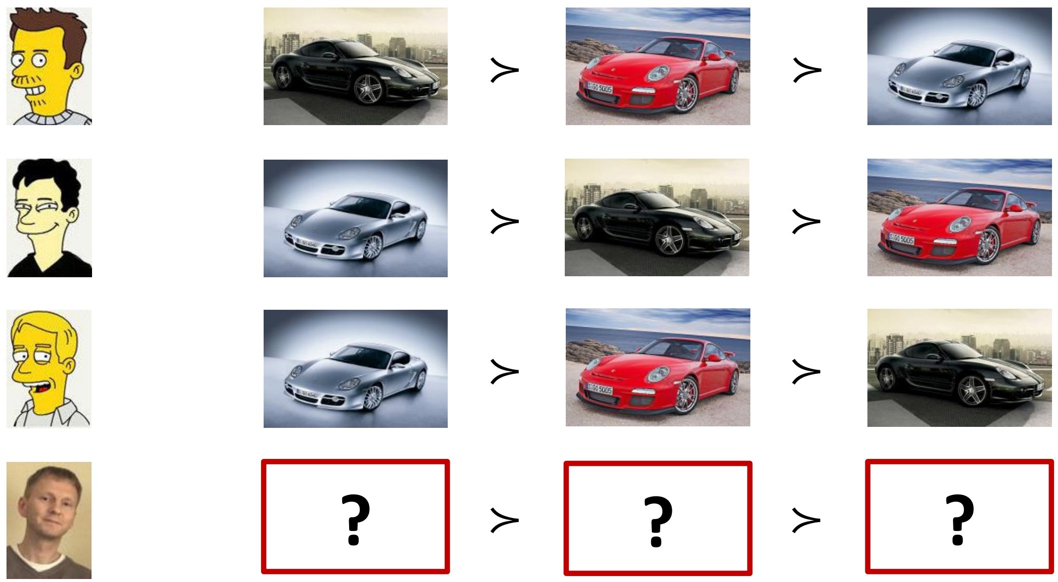
Weiwei Cheng **Eyke Hüllermeier**

Knowledge Engineering and Bioinformatics Lab
Department of Mathematics and Computer Science
Marburg University, Germany

Philipps Universität Marburg

Choice Models and Preference Learning, NIPS 2011

Label Ranking – An Example



where the customers could be described by feature vectors, e.g., (gender, age, place of birth, has children, …)

Label Ranking – An Example

			
	1	2	3
	2	3	1
	3	2	1
	**?**	**?**	**?**

$\pi(i)$ = position of the i-th label in the ranking

1:  2:  3: 

Label Ranking

Given:

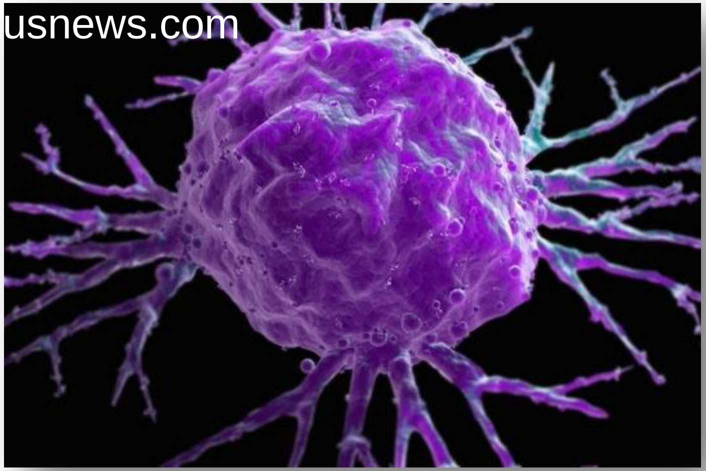
- a set of training instances $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\} \subseteq \mathrm{X}$
- a set of labels $\mathrm{Y} = \{y_1, \ldots, y_m\}$
- for each training instance $\boldsymbol{x}_k$: a set of *pairwise preferences* of the form $y_i \succ_{\boldsymbol{x}_k} y_j$ (for some of the labels)

Find:

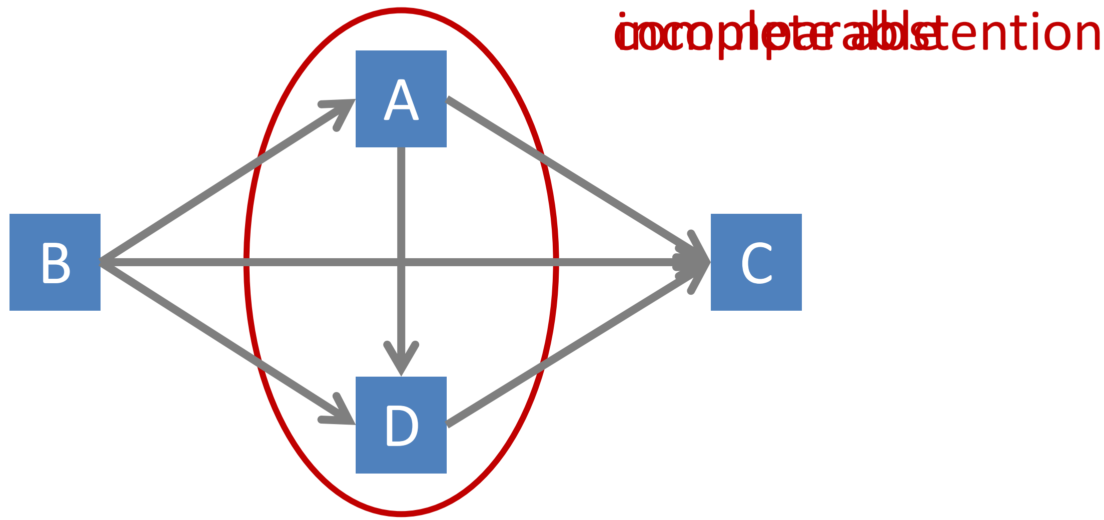
- A ranking function ($\mathrm{X} \rightarrow \Omega$ mapping) that maps each $\boldsymbol{x} \in \mathrm{X}$ to a ranking $\succ_{\boldsymbol{x}}$ of Y (permutation $\pi_{\boldsymbol{x}}$) and generalizes well in terms of a loss function on rankings

Learning with Reject Option

To train a learner that is able to say
“I don’t know”.



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 .

Partial Orders from Pairwise Comparisons

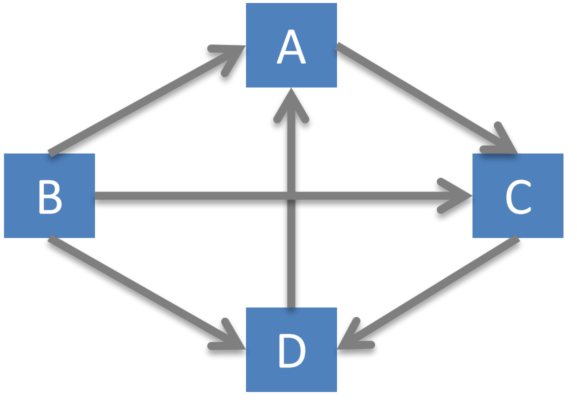
only rely on most confident comparisons → thresholding the relation

	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	

$P(A, D)$

thresholding at 0.5 →

	A	B	C	D
A		0	1	0
B	1		1	1
C	0	0		1
D	1	0	0	



Inconsistent*!*

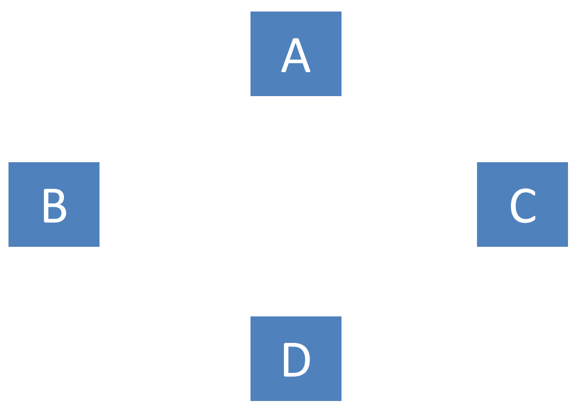
Partial Orders from Pairwise Comparisons

only rely on most confident comparisons → thresholding the relation

	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 1 →

	A	B	C	D
A		0	0	0
B	0		0	0
C	0	0		0
D	0	0	0	



complete abstention

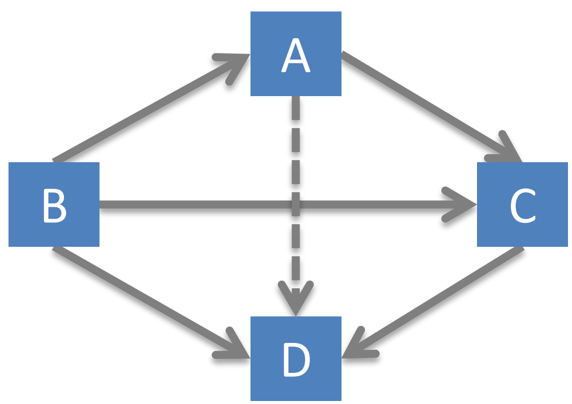
Partial Orders from Pairwise Comparisons

only rely on most confident comparisons → thresholding the relation

	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	



Consistent, but not a partial order!

Partial Orders from Pairwise Comparisons

- **Problem:** Given a (valued) relation P, find the smallest threshold q such that the transitive closure of P_q defines a proper partial order.

  → **maximally informative and consistent prediction**

- There is an $O(m^3)$ algorithm for this problem, with m the number of labels [Cheng et al., ECMLPKDD2010].

Our Ideas & Results

Can we restrict $\mathrm{P}(\cdot,\cdot)$ 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.

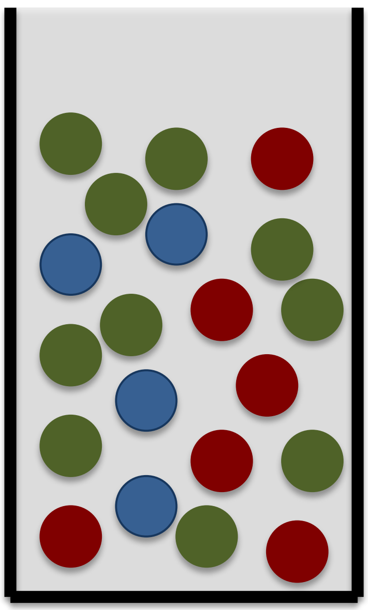
The Plackett-Luce Model

… is a **multistage** model specified by a vector $\boldsymbol{v} = (v_1, \dots, v_m) \in \mathrm{R}_+^m$:

$$\mathbf{P}(\pi \mid \boldsymbol{v}) = \prod_{i=1}^{m} \frac{v_{\pi(i)}}{v_{\pi(i)} + v_{\pi(i+1)} + \cdots + v_{\pi(m)}}$$

A ranking is produced by choosing labels one by one, with a probability proportional to their respective "skills".

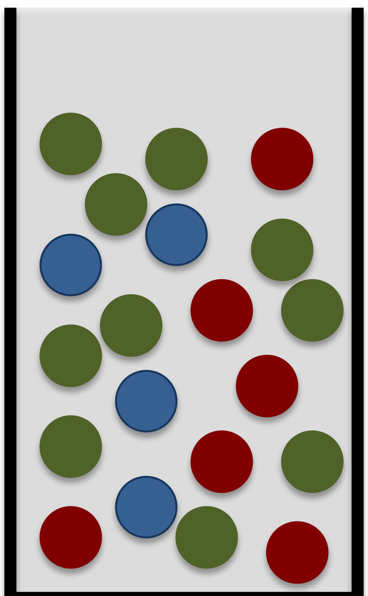
The Plackett-Luce Model



$v = 10, \quad v = 6, \quad v = 4$

$\mathbf{P}(\quad)$

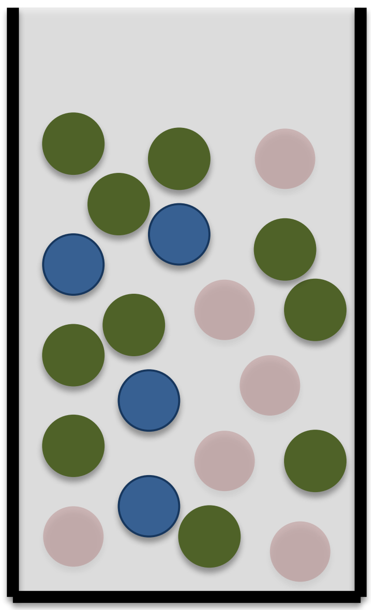
The Plackett-Luce Model



$v_{\bullet} = 10, \quad v_{\bullet} = 6, \quad v_{\bullet} = 4$

$$\mathbf{P}(\bullet \; \bullet \; \bullet) = \frac{6}{20}$$

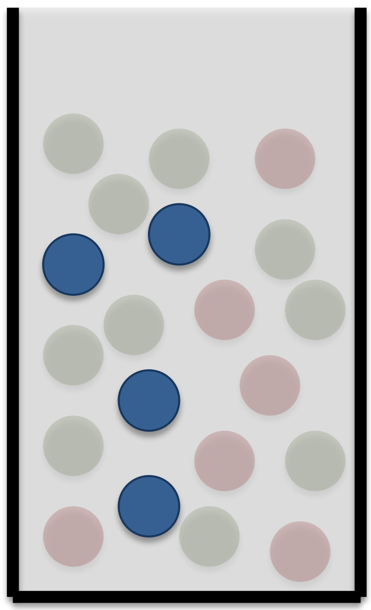
The Plackett-Luce Model



$$v = 10, \qquad v = 6, \qquad v = 4$$

$$\mathbf{P}(\quad) = \frac{6}{20} \times \frac{10}{14}$$

The Plackett-Luce Model



$$v_{\color{green}\bullet} = 10, \qquad v_{\color{red}\bullet} = 6, \qquad v_{\color{blue}\bullet} = 4$$

$$\mathbf{P}({\color{red}\bullet}\ {\color{green}\bullet}\ {\color{blue}\bullet}) = \frac{6}{20} \times \frac{10}{14} \times \frac{4}{4}$$

$$= \frac{3}{14}$$

The Mallows Model

… is a **distance-based** model from the exponential family:

$$\mathbf{P}(\pi \mid \pi_0, \theta) = \frac{\exp(-\theta\, \Delta(\pi, \pi_0))}{\phi(\theta)}$$

center ranking spread normalization constant

where $\Delta(\cdot,\cdot)$ is a (right-invariant) metric on rankings.

The probability of a ranking is higher if it is close to the mode, i.e., the center ranking of the distribution.

Some Common Choices of Δ

- Kendall's tau

$$T(\pi,\sigma) = \sum_{i<j} [\![\big(\pi(i)-\pi(j)\big)\cdot\big(\sigma(i)-\sigma(j)\big) < 0]\!]$$

- Spearman's rho

$$R(\pi,\sigma) = \sqrt{\sum_i \big(\pi(i)-\sigma(j)\big)^2}$$

- Spearman's footrule

$$F(\pi,\sigma) = \sum_i |\pi(i)-\sigma(j)|$$

- Hamming

$$H(\pi,\sigma) = \sum_i [\![\pi(i) \neq \sigma(i)]\!]$$

For example:

$$\pi = (1\ 2\ 3\ 4), \sigma = (1\ 4\ 2\ 3)$$

$$T(\pi,\sigma) = 2$$
$$R(\pi,\sigma) = 2.45$$
$$F(\pi,\sigma) = 4$$
$$H(\pi,\sigma) = 3$$

Transposition Property

Definition A distance Δ is said to have the *transposition property* iff

$$\Delta(\pi, \sigma) \leq \Delta(\pi', \sigma)$$

for any $\pi, \sigma \in \Omega$ and i, j such that

$$\pi(i) < \pi(j) \text{ and } \sigma(i) < \sigma(j).$$

Here π' is a ranking identical to π, except for a transposition of i and j.

Kendall's tau 

Spearman's rho 

Spearman's footrule 

Hamming 

Remarks on $\mathbf{P}(y_i \succ y_j)$

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

$\longrightarrow$ linear extensions of $y_i \succ y_j$

e.g., for $\mathrm{Y} = \{y_1, y_2, y_3\}$, $\mathrm{E}(y_1, y_2) = \left\{ \begin{array}{c} y_1 \succ y_2 \succ y_3, \\ y_1 \succ y_3 \succ y_2, \\ y_3 \succ y_1 \succ y_2 \end{array} \right\}$.

Model	$\mathbf{P}(y_i \succ y_j)$
Plackett-Luce	$\dfrac{v_i}{v_i + v_j}$
Mallows with Spearman's rho	$\dfrac{1}{1 + \exp\left(-2\theta \cdot (\pi_0(j) - \pi_0(i))\right)}$
Mallows with Kendall's tau	$\dfrac{\exp(\theta \cdot [\![\pi_0(j) - \pi_0(i) > 0]\!])}{1 + \exp \theta}$

Our Main Result

Let the preference relation P be given by a probability distribution $\mathbf{P}$ on Ω, that is $\mathrm{P}(y_i, y_j) = \mathbf{P}(y_i \succ y_j)$.

Theorem Let $\mathbf{P}$ be

(1) the Plackett-Luce model or

(2) the Mallows model with a distance Δ having the transposition property.

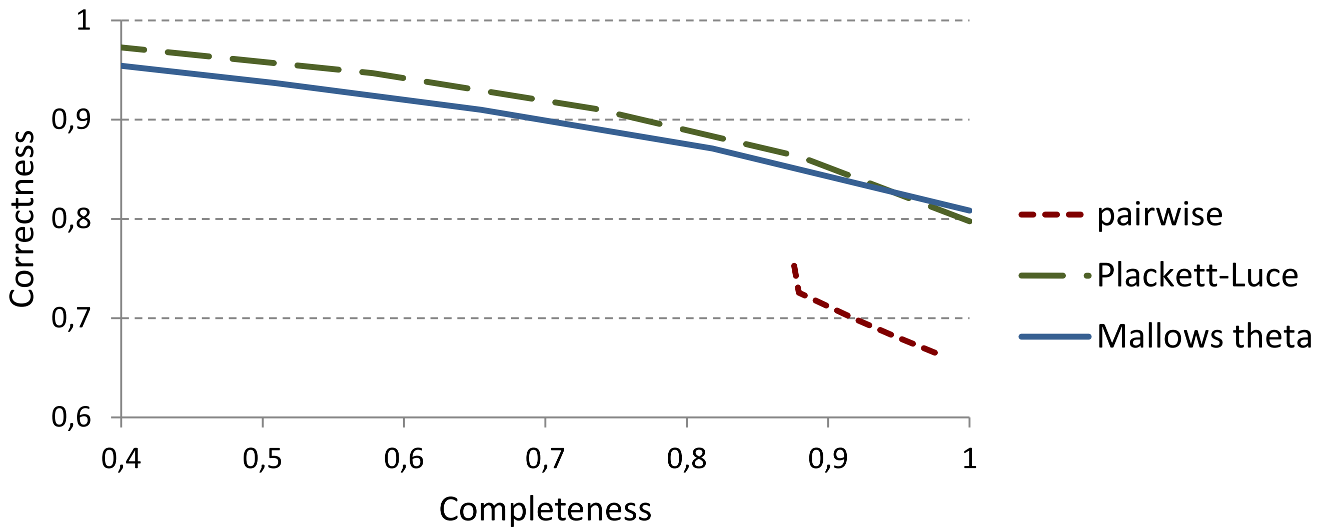
Moreover, let Q be the thresholded relation

$$\mathrm{Q}(y_i, y_j) = 1 \text{ if } \mathrm{P}(y_i, y_j) > q \text{ and}$$

$$\mathrm{Q}(y_i, y_j) = 0 \text{ otherwise.}$$

Then Q defines a proper partial order relation for all $q \in [1/2, 1)$.

Experimental Results



- Results on the UCI benchmark data set VOWEL;
- Correctness (measured by gamma rank correlation):

$$\frac{|\text{concordant pairs}|-|\text{discordant pairs}|}{|\text{concordant pairs}|+|\text{discordant pairs}|}$$

- Completeness: 1 − the relative number of pairwise comparisons on which the model abstains.

Take-Away Messages

- A natural way to derive partial orders is via **thresholding** a (valued) binary **preference relation**.
- While this 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.

Label Ranking with Abstention

Predicting Partial Orders by Thresholding Probability Distributions

Weiwei Cheng **Eyke Hüllermeier**

Knowledge Engineering and Bioinformatics Lab
Department of Mathematics and Computer Science
Marburg University, Germany

The statisticians, like the artists, have a bad habit of falling in love with their models.