Predicting Partial Orders: Ranking with Abstention

Weiwei Cheng[1], Michael Rademaker[2], Bernard De Baets[2], Eyke Hüllermeier[1]

[1] Knowledge Engineering & Bioinformatics Lab
University of Marburg, Germany

[2] Research Unit Knowledge-Based Systems
Ghent University, Belgium

ECMLPKDD 2010, Barcelona, Sep.23.2010

Supervised Ranking Problems

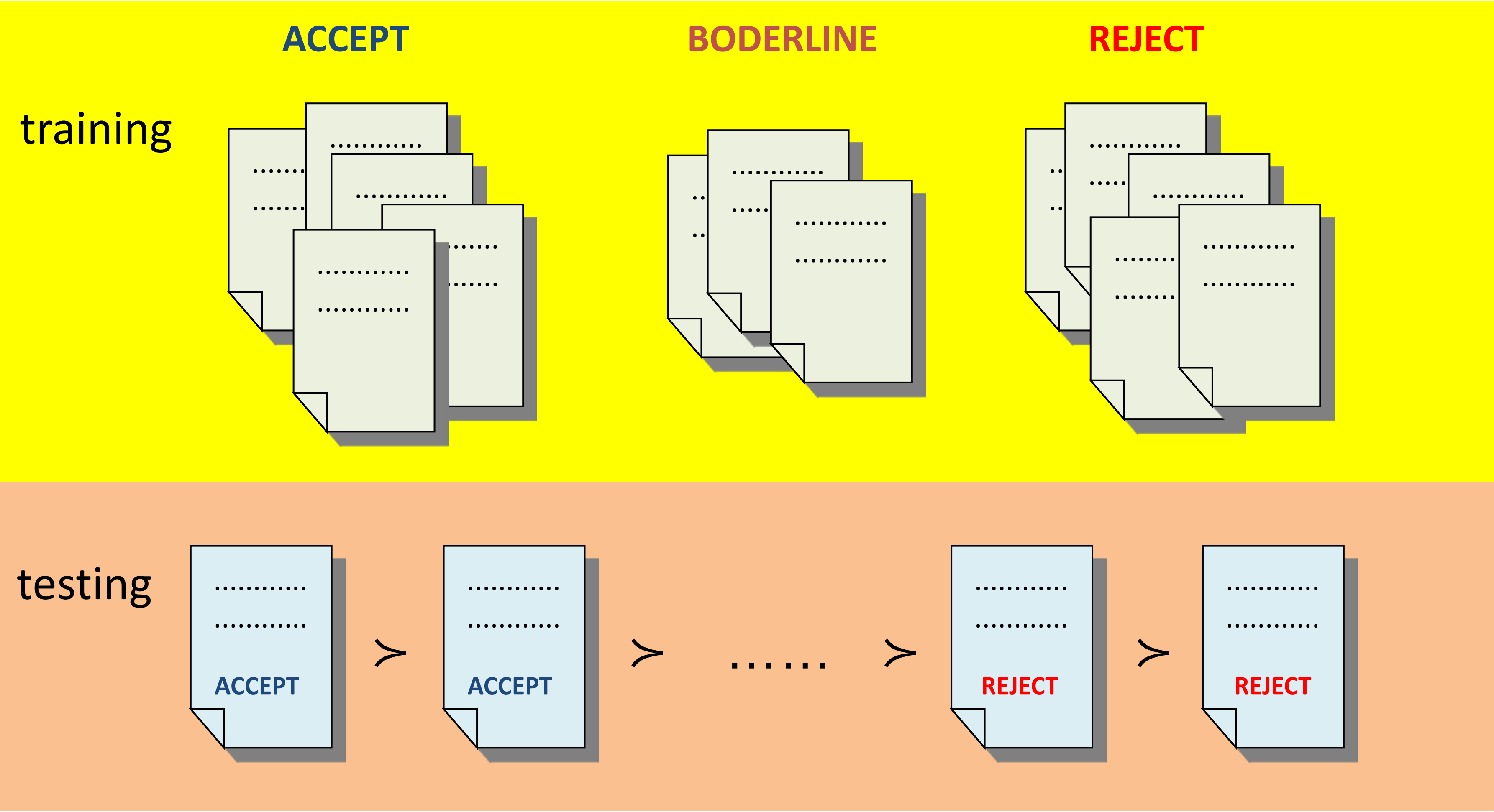
- Label ranking
- Object ranking
- Instance ranking

More in: J. Fürnkranz and E. Hüllermeier, *Preference Learning*, Springer, 2010

Output is a total order of alternatives.

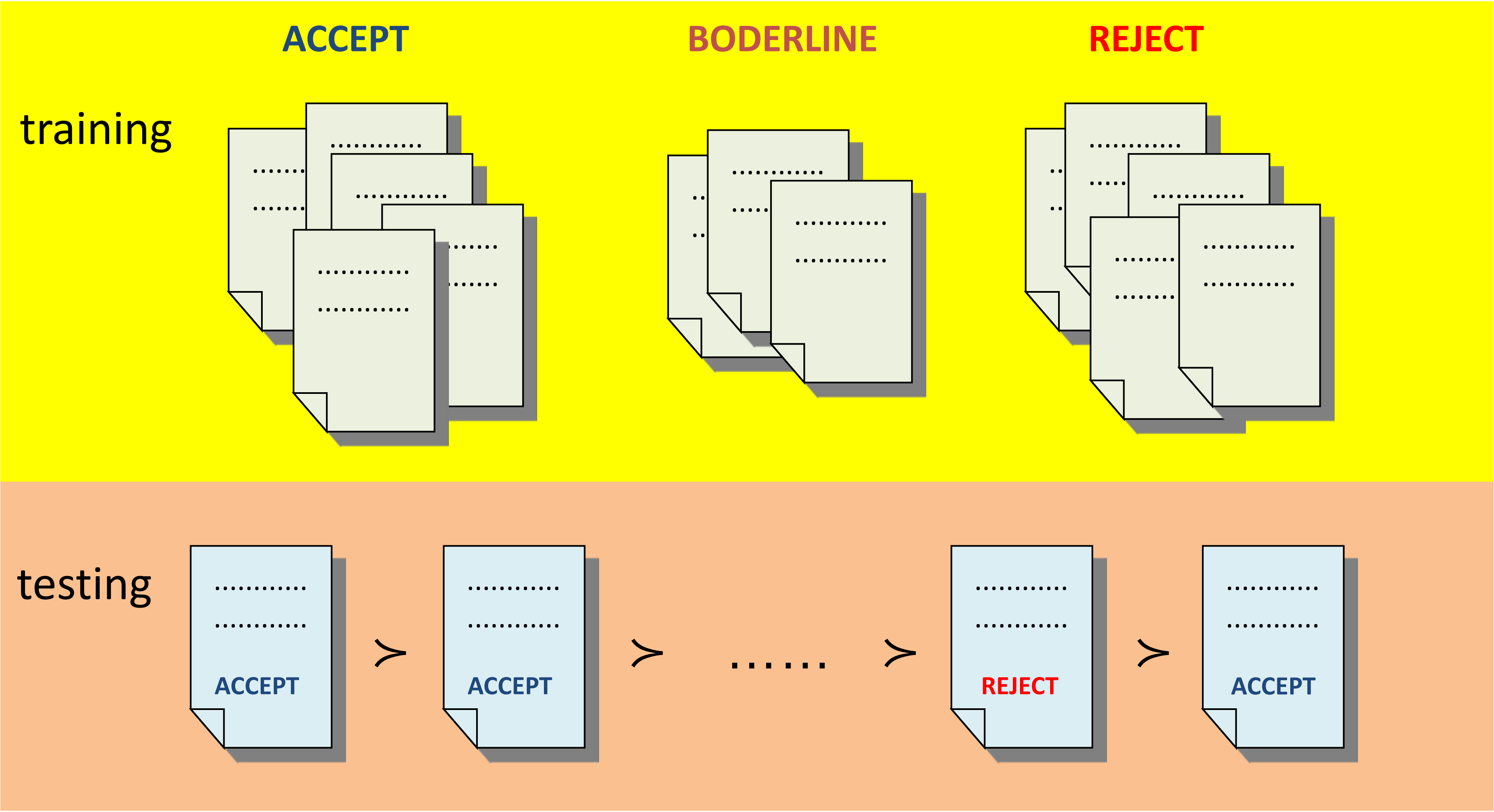
Instance Ranking – An Example

Learning reviewer's preferences on papers



Instance Ranking – An Example

Learning reviewer's preferences on papers



Instance Ranking

Given:

- a set of training instances $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subseteq \mathcal{X}$
- a set of labels $\mathcal{Y} = \{y_1, \ldots, y_k\}$ endowed with an order $y_1 < y_2 < \ldots < y_k$
- for each training instance $\mathbf{x}_l$ an associated label y_l

Find:

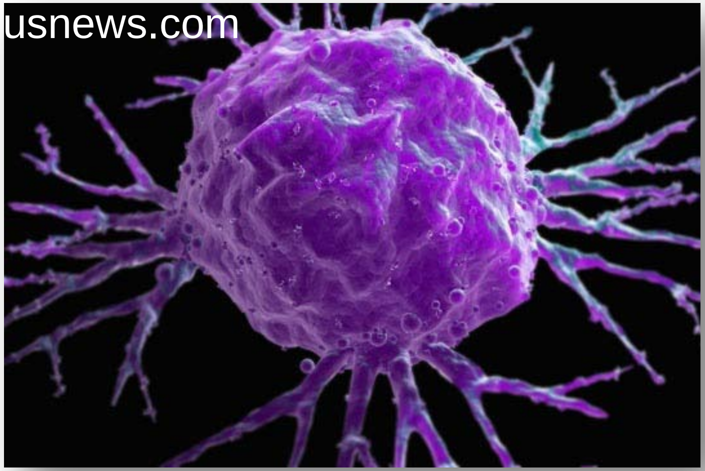
a ranking function that orders a new set of instances $\{\mathbf{x}'_j\}_{j=1}^{t}$ according to their (unknown) preference degrees

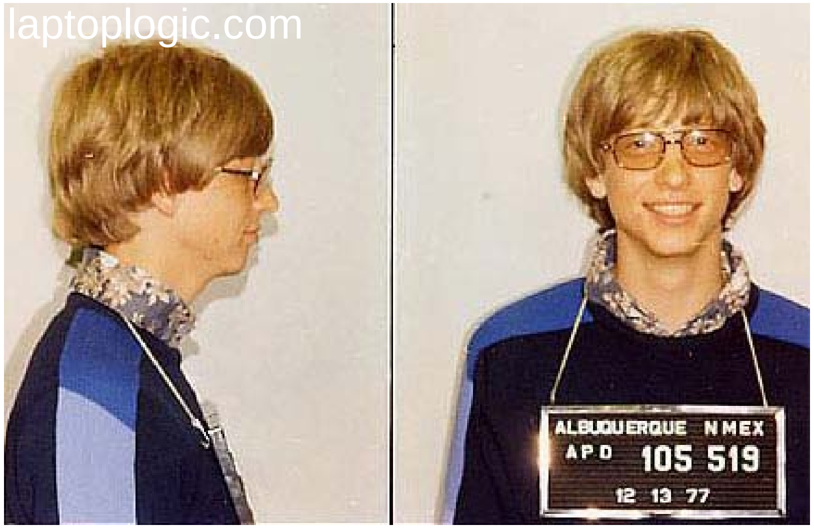
Performance measures:

- AUC in the dichotomous case (k = 2, i.e., bipartite ranking)
- C-index in the polytomous case (k > 2, i.e., k-partite ranking)

Learning with Reject Option

To train a learner that is able to say
"I don't know".





Ranking with Reject Option

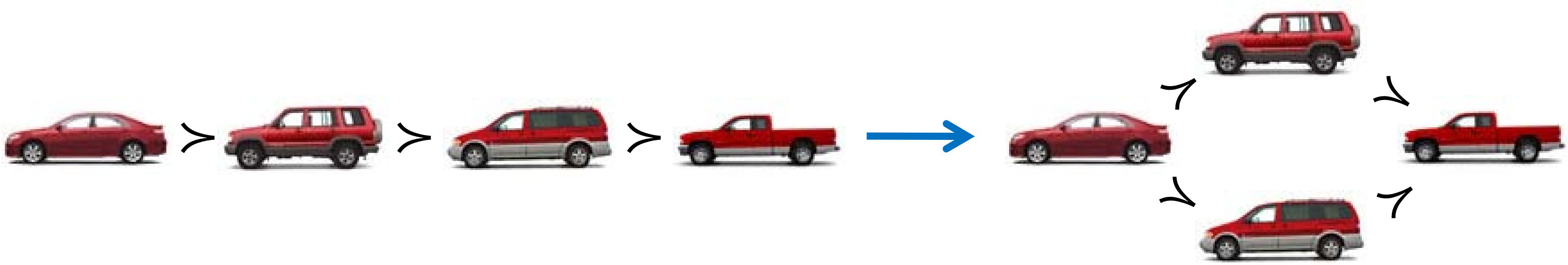
For **a pair of items** a and b to be ranked, the learner can

– predict $a \succ b$ or $b \succ a$, OR

– abstain from prediction (reject option).

The learner should be consistent (transitive & acyclic).

e.g. rank among *cars, vans*, *suvs*, and *trucks* according to a custom's preference



strict total orders ⟶ strict partial orders

The Roadmap of Our Approach

1. Predicting a binary preference relation *P* that specifies, for each pair of alternatives *a* and *b*, a degree of uncertainty regarding their relative comparison.

2. Deriving a (strict) partial order that is maximally compatible with the preference relation *P*.

Predicting a Preference Relation

A preference relation $P : \boldsymbol{A} \times \boldsymbol{A} \longrightarrow [0, 1]$ provides a measure of support for the pairwise preference $a \succ b$, with $P(a, b) = 1 - P(b, a)$ for all $a, b \in \boldsymbol{A}$.

We use a generic approach that can turn every ranker into a partial ranker via ensembling.

thefirstpost.co.uk

Predicting a Preference Relation

A preference relation $P : \boldsymbol{A} \times \boldsymbol{A} \longrightarrow [0, 1]$ provides a measure of support for the pairwise preference $a \succ b$, with $P(a, b) = 1 - P(b, a)$ for all $a, b \in \boldsymbol{A}$.

We use a generic approach that can turn every ranker into a partial ranker via ensembling.

1. With a ranker L, train k ranking models $M_1 \ldots M_k$ by resampling from the original data set, i.e., by k bootstrap samples. By querying these models, k rankings $\succ_1 \ldots \succ_k$ will be produced.
2. For each pair of alternatives a and b, we define the degree of preference $P(a,b) = \frac{1}{k} \left|\{i \mid a \succ_i b\}\right|$.

Predicting a Strict Partial Order

Based on P, we seek to induce a (partial) order relation
$\mathcal{R} : \mathbf{A} \times \mathbf{A} \to \{0,1\}$. $\mathcal{R}(a,b) = 1 \;\Leftrightarrow\; (a,b) \in \mathcal{R} \;\Leftrightarrow\; a\mathcal{R}b$

$$\mathcal{R}_\alpha = \{(a,b) \,|\, P(a,b) \geq \alpha\}$$

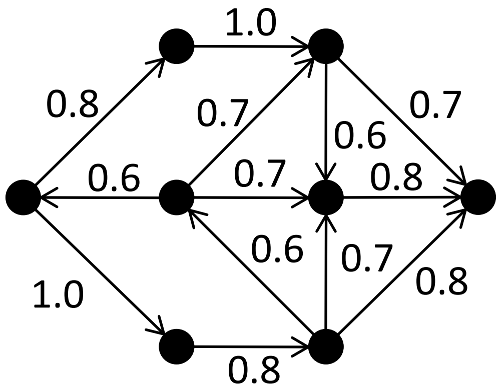
Two intuitive choices for α:

1. Consensus, i.e., $\alpha = 1$.
   Most items will be declared as incomparable.

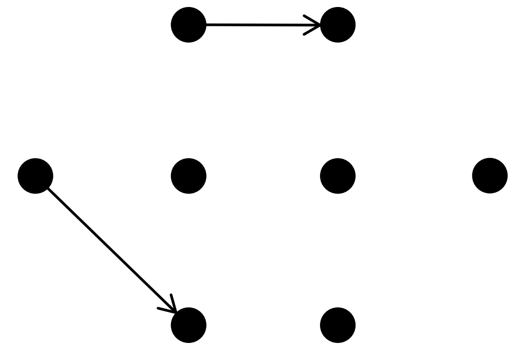
2. Majority (aka. Condorcet criterion), i.e., $\alpha = 0.5$.
   A cyclic relation can be produced, e.g.,

$$\begin{matrix} a \succ_1 b \succ_1 c \\ b \succ_2 c \succ_2 a \\ c \succ_3 a \succ_3 b \end{matrix} \quad \Longrightarrow \quad P(a,b) = P(b,c) = P(c,a) = 2/3$$

Predicting a Strict Partial Order

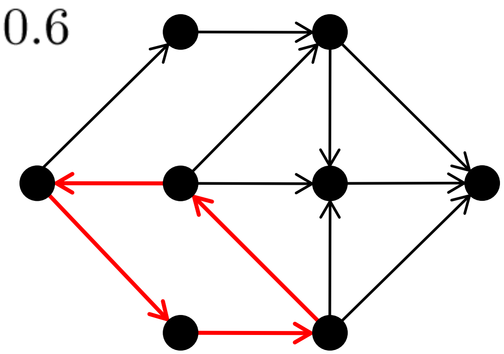


$\alpha = 1$



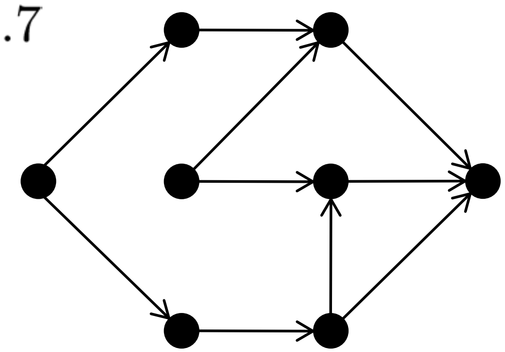
consistent partial order, but not very informative

$\alpha = 0.6$



cyclic relation due to choosing a too small threshold

$\alpha = 0.7$



consistent partial order

Searching the Minimal Threshold

Looking for a minimal α (denoted as α^*) such that the transitive closure of $\mathcal{R}_\alpha$ (denoted as $\overline{\mathcal{R}}_\alpha$) is a strict partial order relation.

- The domain of α can be restricted to $\{0, 1/k, 2/k, \ldots, 1\}$.
- If $\mathcal{R}_\alpha$ is cyclic, $\mathcal{R}_\beta$ is cyclic as well, unless $\beta > \alpha$.

Moreover, we can show that

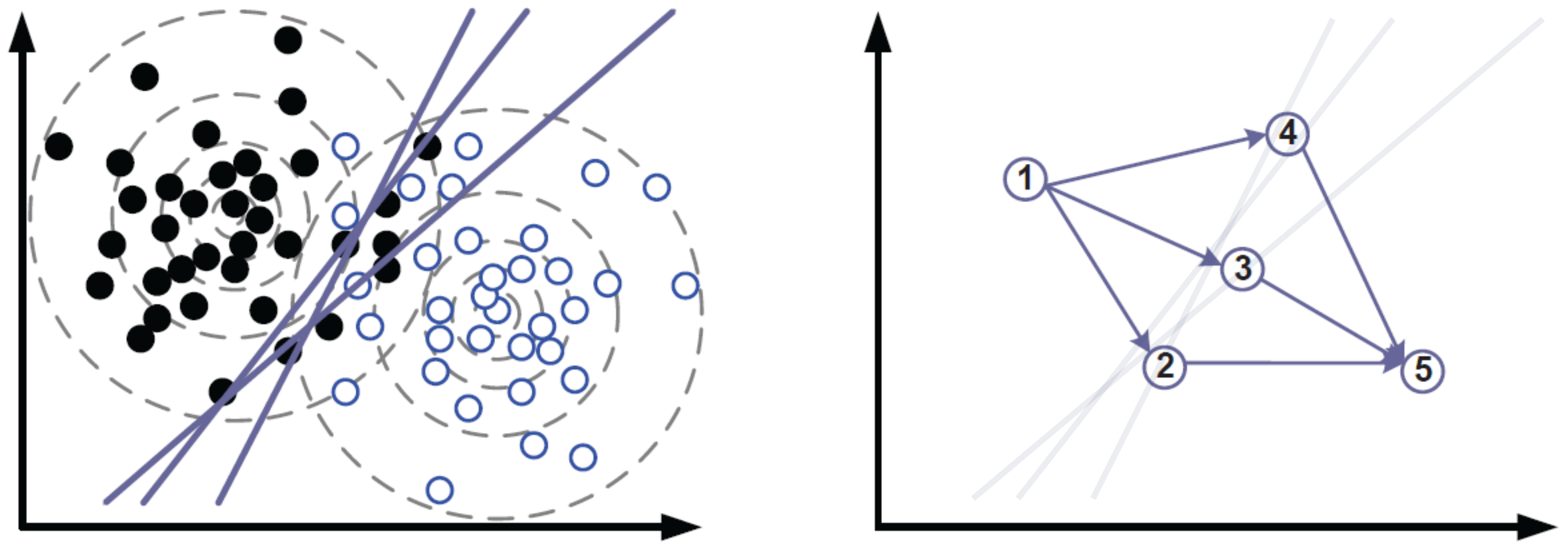
$$\alpha_u = 1 \geq \alpha^* \geq \alpha_l = \frac{1}{k} + \max_{a,b} \min(P(a,b), P(b,a)).$$

Repeat until $\alpha_u - \alpha_l < 1/k$

1) Set α to the middle point between α_u and α_l
2) Compute $\mathcal{R}_\alpha$
3) Compute $\overline{\mathcal{R}}_\alpha$ (e.g., with the Floyd-Warshall's algorithm)
4) If $\overline{\mathcal{R}}_\alpha$ is a partial order, set α_u to α
5) else set α_l to α

$\mathcal{O}(|\mathbf{A}|^3)$

An Illustrating Example



Left: training data and ensemble models
Right: partial order predicted for a set of five query instances

Evaluation Measures

As now the ranker has the ability to reject predictions, there is a trade-off between correctness and completeness.

	$a \sqsupset_* b$	$b \sqsupset_* a$	$a \perp_* b$
$a \sqsupset b$	C	D	$\times$
$b \sqsupset a$	D	C	$\times$
$a \perp b$	$\times$	$\times$	$\times$

C: concordant D: discordant

Evaluation Measures

As now the ranker has the ability to reject predictions, there is a trade-off between correctness and completeness.

- Correctness is measured by *gamma rank correlation*:

$$\mathrm{CR}(\sqsupset, \sqsupset_*) = \frac{|C| - |D|}{|C| + |D|}$$

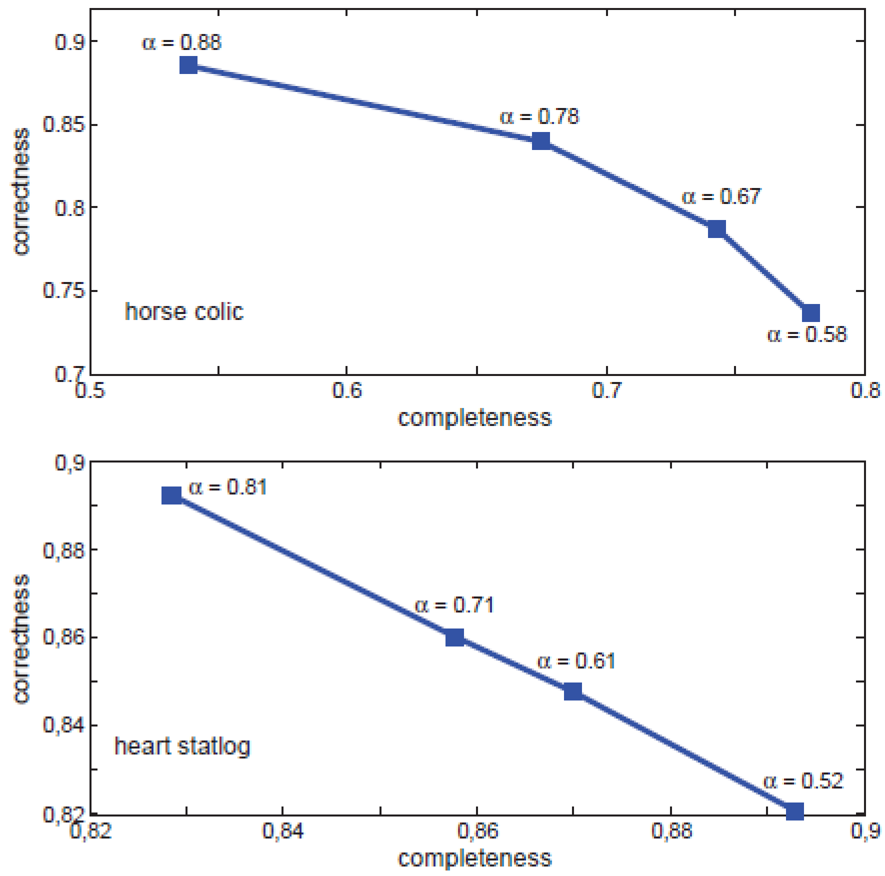
- Completeness measure punishes the abstention from comparisons that should actually be made.

$$\mathrm{CP}(\sqsupset, \sqsupset_*) = \frac{|C| + |D|}{|\sqsupset_*|}$$

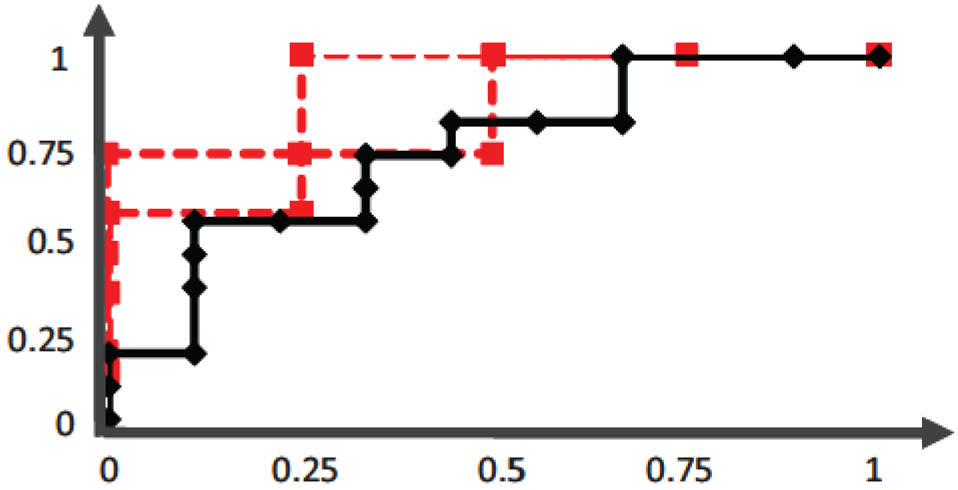
Experiments – Instance Ranking

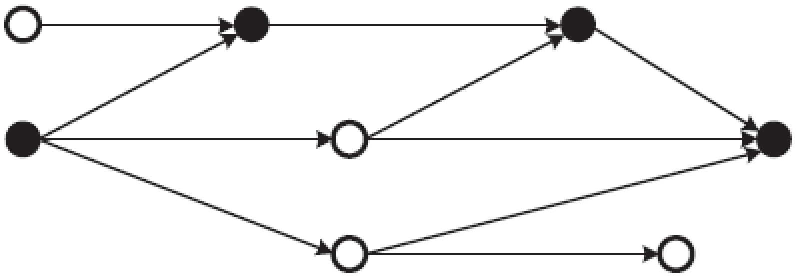
data set	#attr.	#inst.	correctness with abstention	correctness w/o abstention	completeness
breast	9	286	0.330±0.150	0.318±0.141	0.578±0.074
breast-w	9	699	0.988±0.014	0.987±0.015	0.982±0.015
horse colic	22	368	0.734±0.135	0.697±0.142	0.790±0.044
credit rating	15	690	0.858±0.062	0.827±0.065	0.888±0.038
credit german	20	1000	0.610±0.088	0.568±0.084	0.741±0.060
pima diabetes	8	768	0.684±0.084	0.666±0.086	0.819±0.047
heart statlog	13	270	0.811±0.102	0.797±0.101	0.890±0.060
hepatitis	19	155	0.709±0.292	0.697±0.271	0.797±0.084
ionosphere	34	351	0.771±0.174	0.722±0.190	0.814±0.098
kr-vs-kp	36	3196	0.992±0.006	0.980±0.007	0.991±0.006
labor	16	57	0.990±0.049	0.985±0.060	0.989±0.052
mushroom	22	8124	1.000±0.000	1.000±0.000	0.808±0.017
thyroid disease	29	3772	0.890±0.071	0.883±0.070	0.928±0.040
sonar	60	206	0.684±0.224	0.575±0.271	0.575±0.056
tic-tac-toe	9	958	0.253±0.127	0.221±0.120	0.908±0.013
vote	16	435	0.981±0.032	0.976±0.036	0.913±0.035

Experiments – Instance Ranking



Experiments – Instance Ranking





Conclusions

We have addressed the problem of "reliable" prediction in the context of learning to rank.

- A relaxation of the conventional setting where predictions are given in terms of **partial instead of total orders**.
- A **generic approach to predicting partial orders** that is applicable to different types of ranking problems.
- **Measures for evaluating the performance** of a ranker with (partial) reject option.
- Empirically, we have shown that our method is able to **trade off accuracy against completeness**.