PREFERENCES IN ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Eyke Hüllermeier
Intelligent Systems Group
Department of Computer Science
University of Paderborn, Germany

eyke@upb.de

PFIA 2015, Rennes, France, 28-JUN-2015

PREFERENCES ARE UBIQUITOUS

Preferences play a key role in many applications of computer science and modern information technology:

COMPUTATIONAL ADVERTISING	RECOMMENDER SYSTEMS	COMPUTER GAMES
AUTONOMOUS AGENTS	ELECTRONIC COMMERCE	ADAPTIVE USER INTERFACES
PERSONALIZED MEDICINE	ADAPTIVE RETRIEVAL SYSTEMS	SERVICE-ORIENTED COMPUTING

PREFERENCES ARE UBIQUITOUS

Preferences play a key role in many applications of computer science and modern information technology:

COMPUTATIONAL ADVERTISING	RECOMMENDER SYSTEMS	COMPUTER GAMES
AUTONOMOUS AGENTS	ELECTRONIC COMMERCE	ADAPTIVE USER INTERFACES
PERSONALIZED MEDICINE	ADAPTIVE RETRIEVAL SYSTEMS	SERVICE-ORIENTED COMPUTING

medications or therapies specifically tailored for individual patients

COMMERCIAL INTEREST

INTELLIGENT SYSTEMS

Amazon files patent for "anticipatory" shipping

10 Comments / Shares / Tweets / Stumble / Email / More +

Amazon.com has filed for a patent for a shipping system that would anticipate what customers buy to decrease shipping time.

Amazon says the shipping system works by analyzing customer data like, purchasing history, product searches, wish lists and shopping cart contents, the Wall Street Journal reports. According to the patent filing, items would be moved from Amazon's fulfillment center to a shipping hub close to the customer in anticipation of an eventual purchase.

PREFERENCES IN AI

“**Early work in AI focused on the notion of a goal—an explicit target that must be achieved**—and this paradigm is still dominant in AI problem solving. But as application domains become more complex and realistic, it is apparent that **the dichotomic notion of a goal**, while adequate for certain puzzles, **is too crude in general**. The problem is that in many contemporary application domains ... **the user has little knowledge about the set of possible solutions or feasible items,** and what she typically seeks is the best that’s out there. But since the user does not know what is the best achievable plan or the best available document or product, she typically cannot characterize it or its properties specifically. **As a result, she will end up either asking for an unachievable goal, getting no solution in response, or asking for too little, obtaining a solution that can be substantially improved.**”

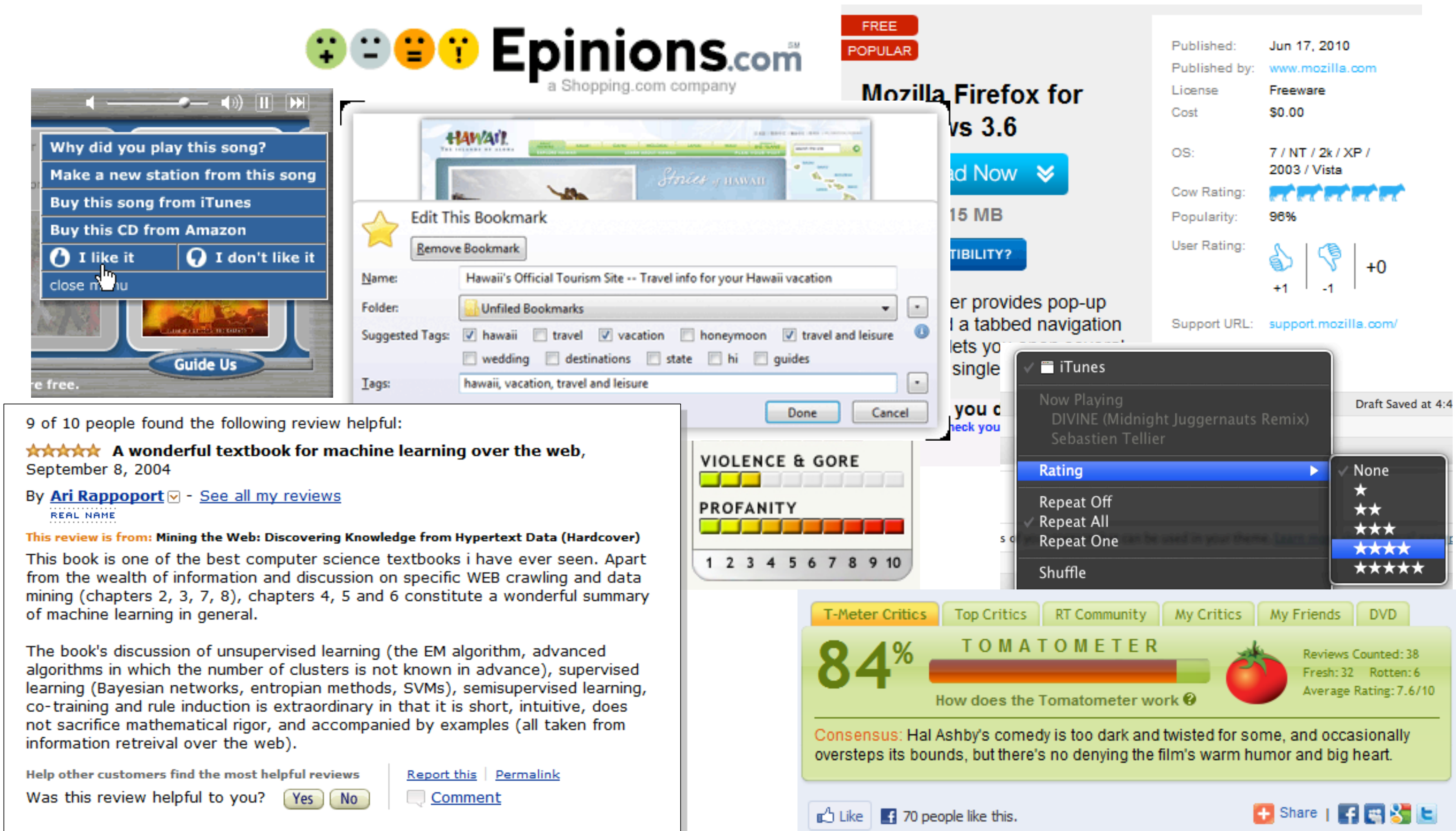
[Brafman & Domshlak, 2009]

*... compared with the dichotomic notion of a **goal**, preference formalisms significantly increase **flexibility** in knowledge representation and problem solving!*

PREFERENCES IN ARTIFICIAL INTELLIGENCE RESEARCH:

- **preference representation** (preference relations, CP nets, GAI networks, logical representations, fuzzy constraints, …)
- **preference handling and reasoning** with preferences (decision theory, constraint satisfaction, non-monotonic reasoning, …)
- **preference acquisition** (preference elicitation, **preference learning**, ...)

PREFERENCE INFORMATION



PREFERENCE INFORMATION

| Offizielle Homepage | **Daniel Baier** |
www.**daniel-baier**.com/
Willkommen auf der offiziellen Homepage von Fussballprofi **Daniel Baier** - TSV 1860 München.

NOT CLICKED ON

Prof. Dr. **Daniel Baier** - Brandenburgische Technische Universität ...
www.tu-cottbus.de/fakultaet3/de/.../team/.../prof-dr-**daniel-baier**.html
Vökler, Sascha; Krausche, **Daniel**; **Baier**, Daniel: Product Design Optimization Using Ant Colony And Bee Algorithms: A Comparison, erscheint in: Studies in ...

CLICKED ON

Daniel Baier
www.weltfussball.de/spieler_profil/**daniel-baier**/
Daniel Baier - FC Augsburg, VfL Wolfsburg, VfL Wolfsburg II, TSV 1860 München.

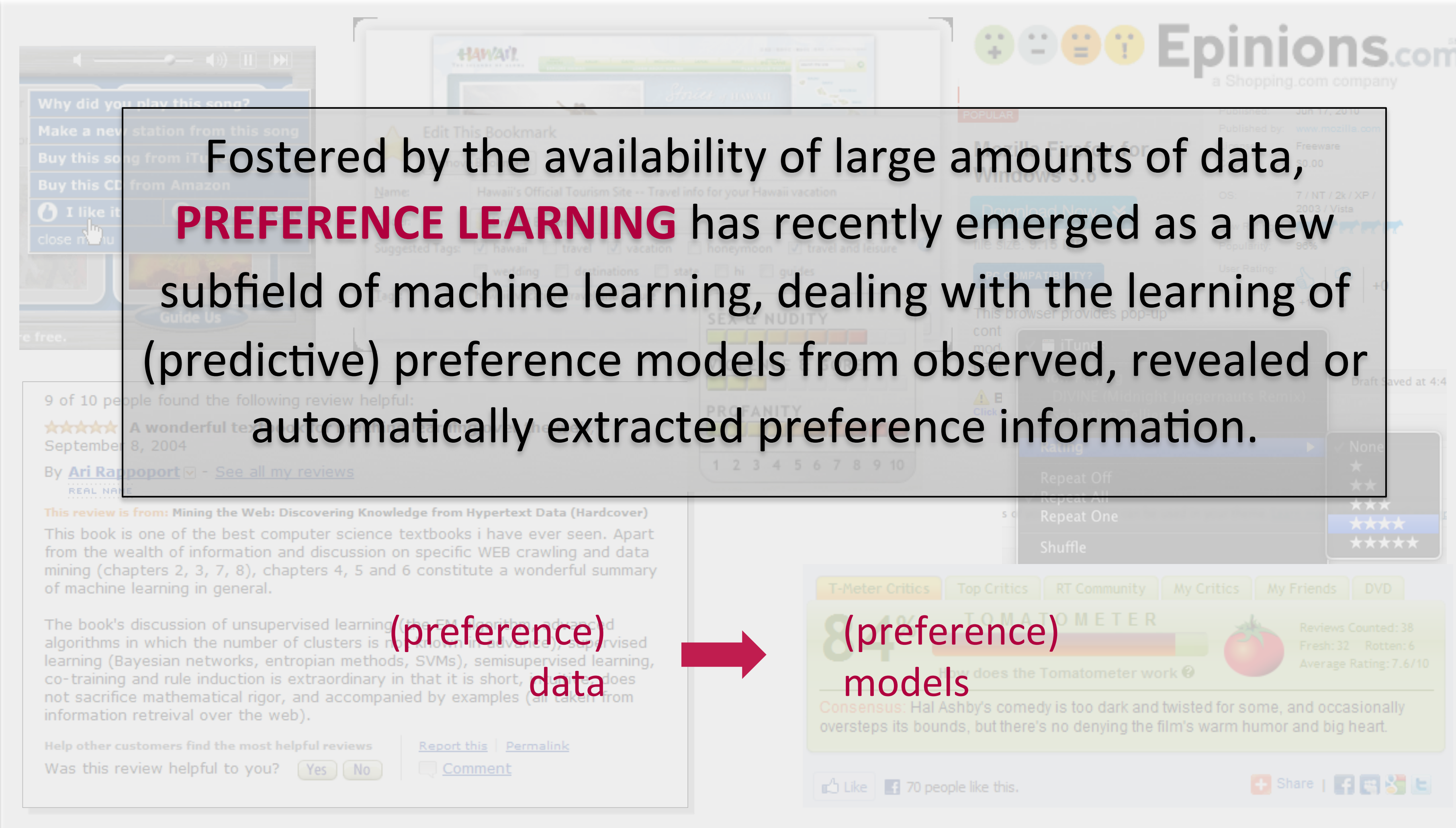
Daniel Baier - aktuelle Themen & Nachrichten - sueddeutsche.de
www.sueddeutsche.de/thema/**Daniel_Baier**
Aktuelle Nachrichten, Informationen und Bilder zum Thema **Daniel Baier** auf sueddeutsche.de.

Daniel Baier | Facebook
de-de.facebook.com/**daniel.baier**.589
Tritt Facebook bei, um dich mit **Daniel Baier** und anderen Nutzern, die du kennst, zu vernetzen. Facebook ermöglicht den Menschen das Teilen von Inhalten mit ...

FC Augsburg: Mein Tag in Bad Gögging: **Daniel Baier**
www.fcaugsburg.de/cms/website.php?id=/index/aktuell/news/...
2. Aug. 2012 – **Daniel Baier** berichtet heute, was für die Profis auf dem Programm stand. Hi FCA- Fans,. heute liegen wieder zwei intensive Trainingseinheiten ...

- *Preferences are not necessarily expressed explicitly, but can be extracted* ***implictly from people's behavior****!*
- *Massive amounts of very* ***noisy data****!*

PREFERENCE LEARNING



Fostered by the availability of large amounts of data, **PREFERENCE LEARNING** has recently emerged as a new subfield of machine learning, dealing with the learning of (predictive) preference models from observed, revealed or automatically extracted preference information.

(preference) data → (preference) models

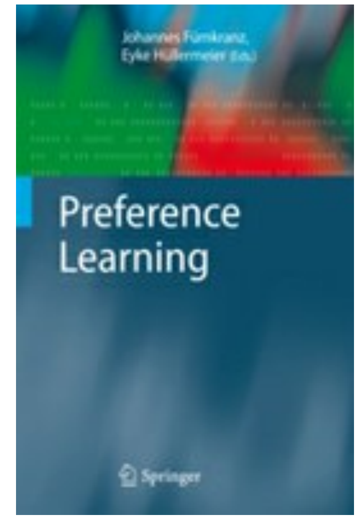
PREFERENCE LEARNING

Tutorials:

- European Conf. on Machine Learning, 2010
- Int. Conf. Discovery Science, 2011
- Int. Conf. Algorithmic Decision Theory, 2011
- European Conf. on Artificial Intelligence, 2012
- Int. Conf. Algorithmic Learning Theory, 2014

Workshops:

- ECML/PDKK 08–10: Workshop on Preference Learning
- ECAI 2012: Workshop on Preference Learning: Problems and Applications in AI
- Dagstuhl Seminar on Preference Learning (2014)



J. Fürnkranz &
E. Hüllermeier (eds.)
Preference Learning
Springer-Verlag 2011

Special Issue on Representing, Processing, and Learning Preferences: Theoretical and Practical Challenges (2011)

Special Issue on Preference Learning (2013)

PL IS AN ACTIVE FIELD

INTELLIGENT SYSTEMS

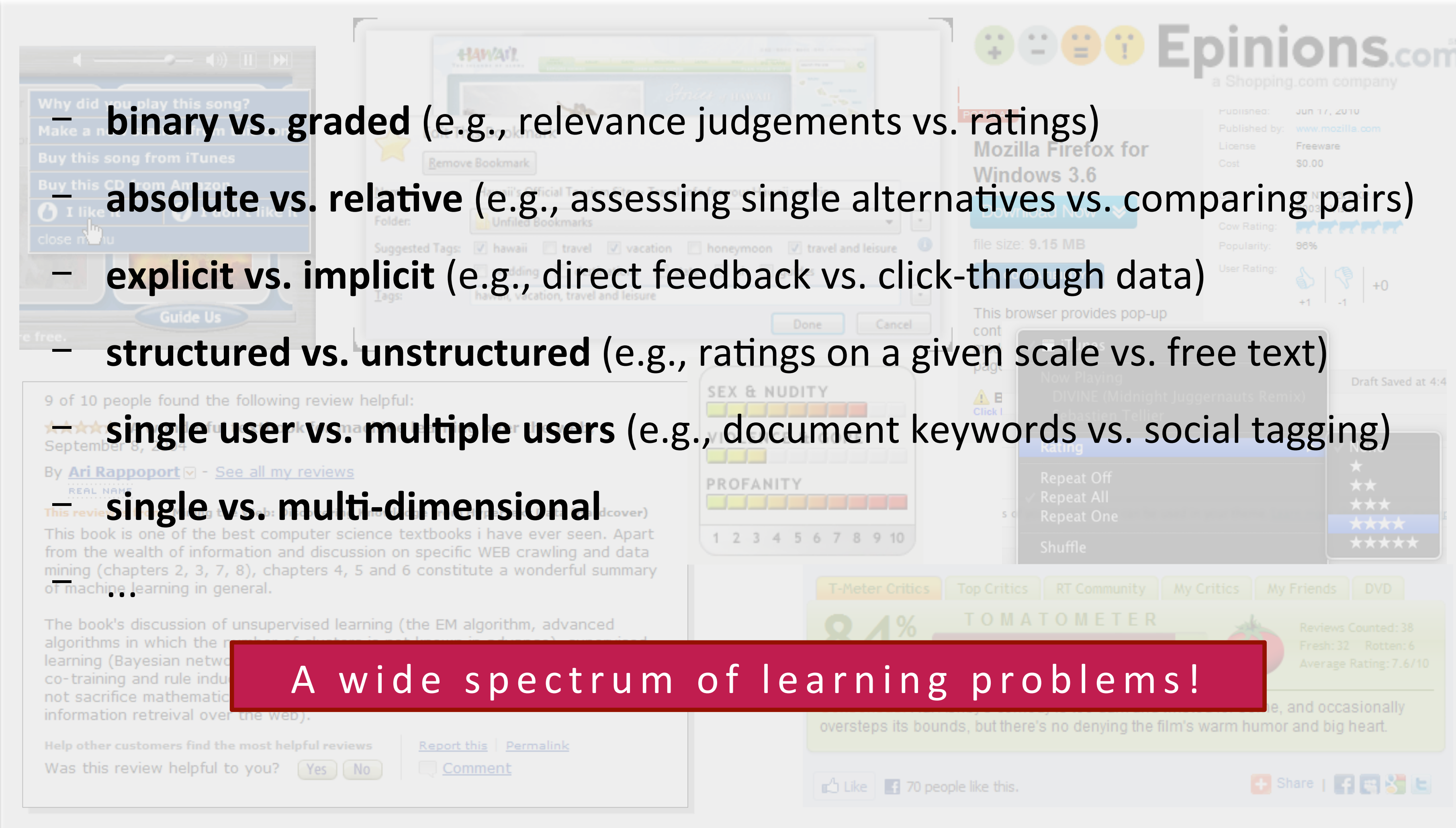
- NIPS 2001: New Methods for Preference Elicitation
- NIPS 2002: Beyond Classification and Regression: Learning Rankings, Preferences, Equality Predicates, and Other Structures
- KI 2003: Preference Learning: Models, Methods, Applications
- NIPS 2004: Learning with Structured Outputs
- NIPS 2005: Workshop on Learning to Rank
- IJCAI 2005: Advances in Preference Handling
- SIGIR 07–10: Workshop on Learning to Rank for Information Retrieval
- ECML/PDKK 08–10: Workshop on Preference Learning
- NIPS 2009: Workshop on Advances in Ranking
- American Institute of Mathematics Workshop in Summer 2010: The Mathematics of Ranking
- NIPS 2011: Workshop on Choice Models and Preference Learning
- EURO 2009-12: Special Track on Preference Learning
- ECAI 2012: Workshop on Preference Learning: Problems and Applications in AI
- DA2PL 2012: From Decision Analysis to Preference Learning
- Dagstuhl Seminar on Preference Learning (2014)
- NIPS 2014: Analysis of Rank Data: Confluence of Social Choice, Operations Research, and Machine Learning

CONNECTIONS TO OTHER FIELDS

INTELLIGENT SYSTEMS

- Structured Output Prediction
- Learning Monotone Models
- Classification (ordinal, multilabel, ...)
- Information Retrieval
- Learning with weak supervision
- **Preference Learning**
- Recommender Systems
- Economics & Decison Science
- Statistics
- Social Choice
- Graph theory
- Optimization
- Operations Research
- Multiple Criteria Decision Making

MANY TYPES OF PREFERENCES



- **binary vs. graded** (e.g., relevance judgements vs. ratings)
- **absolute vs. relative** (e.g., assessing single alternatives vs. comparing pairs)
- **explicit vs. implicit** (e.g., direct feedback vs. click-through data)
- **structured vs. unstructured** (e.g., ratings on a given scale vs. free text)
- **single user vs. multiple users** (e.g., document keywords vs. social tagging)
- **single vs. multi-dimensional**
- ...

A wide spectrum of learning problems!

SUBSET RANKING

INTELLIGENT SYSTEMS

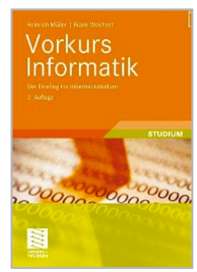
TRAINING

$$\begin{array}{ccc}
(0.74, 1, 25, 165) & \succ & (0.45, 0, 35, 155) \\
(0.47, 1, 46, 183) & \succ & (0.57, 1, 61, 177) \\
(0.25, 0, 26, 199) & \succ & (0.73, 0, 46, 185)
\end{array}$$

Pairwise preferences between objects

$\succ$

$\succ$



→ induction of a RANKING FUNCTION

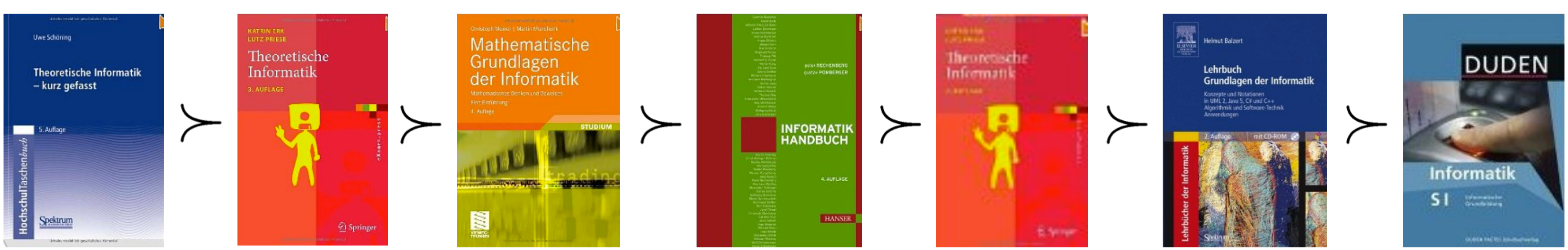
SUBSET RANKING

PREDICTION (ranking a new set of objects)

$$\mathcal{Q} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4, \boldsymbol{x}_5, \boldsymbol{x}_6, \boldsymbol{x}_7, \boldsymbol{x}_8, \boldsymbol{x}_9, \boldsymbol{x}_{10}, \boldsymbol{x}_{11}, \boldsymbol{x}_{12}, \boldsymbol{x}_{13}\}$$

$$\boldsymbol{x}_{10} \succ \boldsymbol{x}_4 \succ \boldsymbol{x}_7 \succ \boldsymbol{x}_1 \succ \boldsymbol{x}_{11} \succ \boldsymbol{x}_2 \succ \boldsymbol{x}_8 \succ \boldsymbol{x}_{13} \succ \boldsymbol{x}_9 \succ \boldsymbol{x}_3 \succ \boldsymbol{x}_{12} \succ \boldsymbol{x}_5 \succ \boldsymbol{x}_6$$

$\mathcal{Q} = \Big\{$  $\Big\}$



COLLABORATIVE FILTERING

PRODUCTS

USERS

	P1	P2	P3	…	P38	…	P88	P89	P90
U1	★		★★★	…		…		★★★	
U2		★★	★	…		…	★		
…				…		…			
U46	?	★★	?	…	?	…	?	?	★★★
…				…		…			
U98	★★★			…		…	★★★		
U99			★	…		…	★★		

PREFERENCE LEARNING TASKS

	OBJECT RANKING	COLLABORATIVE FILTERING
description of alternatives	features	identifier
representation of preference	relative	absolute
predictions	ranking	utility degrees
number of users/models	single	many

PREFERENCE LEARNING TASKS

	representation		type of preference information		
task	context (input)	alternative (output)	training information	prediction	ground truth
collaborative filtering	ID	ID	absolute ordinal	absolute ordinal	absolute ordinal
dyadic prediction	feature	feature	absolute ordinal	absolute ordinal	absolute ordinal
multilabel classification	feature	ID	absolute binary	absolute binary	absolute binary
multilabel ranking	feature	ID	absolute binary	ranking	absolute binary
label ranking	feature	ID	relative binary	ranking	ranking
subset ranking	---	feature	relative binary	ranking	ranking or subset
instance ranking	---	feature	absolute ordinal	ranking	absolute ordinal

OUTLINE

INTELLIGENT SYSTEMS

PART 1	PART 2	PART 3
Preference learning	Label ranking	Preference-based bandit algorithms

SUPERVISED LEARNING

- What kind of **training data** is offered to the learning algorithm?
- What **type of model** (prediction) is the learner supposed to produce?

$$h : \mathcal{X} \to \mathcal{Y}$$

- What is the nature of the **ground truth**, and how is a model assessed?

LOSS FUNCTION

$$\mathcal{L} : \; \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_+$$

$$(y, y^*) \mapsto \text{penalty for predicting } y \text{ if the true outcome is } y^*$$

SUPERVISED LEARNING

- What kind of **training data** is offered to the learning algorithm?
- What **type of model** (prediction) is the learner supposed to produce?

$$h : \mathcal{X} \to \mathcal{Y}$$

- What is the nature of the **ground truth**, and how is a model assessed?

$$\mathcal{R}(h) = \int_{\mathcal{X} \times \mathcal{Y}} \mathcal{L}\left(h(\boldsymbol{x}), y\right) \, d\,\mathbf{P}(X, Y)$$

risk ≈ average penalty caused by the model's predictions

unknown data-generating process

PREFERENCE LEARNING TASKS

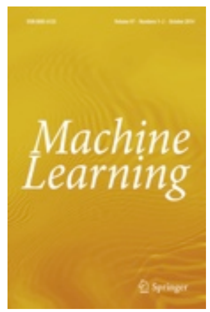
INTELLIGENT SYSTEMS

Preference learning problems are challenging, because

- sought predictions are complex/structured, *top-K ranking*
- supervision is weak (partial, noisy, ...), *clickthrough data*
- performance metrics are hard to optimize, *NDCG@K*
- ...

... mapping instances to **TOTAL ORDERS** over a fixed set of alternatives/labels:

$\texttt{(35,0,...,325)} \longmapsto$



$\succ$

$\succ$

$\succ$

instance $x \in \mathcal{X}$

(e.g., features of a person)

... likes more
... reads more
... publishes more in
...

LABEL RANKING: TRAINING DATA

TRAINING

X_1	X_2	X_3	X_4	preferences
0.34	0	10	174	$A \succ B, C \succ D$
1.45	0	32	277	$B \succ C \succ A$
1.22	1	46	421	$B \succ D, A \succ D, C \succ D, A \succ C$
0.74	1	25	165	$C \succ A \succ D, A \succ B$
0.95	1	72	273	$B \succ D, A \succ D$
1.04	0	33	158	$D \succ A \succ B, C \succ B, A \succ C$

Instances are associated with preferences between labels

... no demand for full rankings!

LABEL RANKING: PREDICTION

PREDICTION				A	B	C	D
0.92	1	81	382	?	?	?	?

new instance

ranking ?

LABEL RANKING: PREDICTION

PREDICTION

				A	B	C	D
0.92	1	81	382	4	1	3	2

A ranking of all labels

new instance

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

LABEL RANKING: PREDICTION

PREDICTION

0.92	1	81	382	4	1	3	2

A ranking of all labels

LOSS

GROUND TRUTH

0.92	1	81	382	2	1	3	4

SPEARMAN

$$\mathcal{L}(\pi, \pi^*) = \sum_{i=1}^{k} \left(\pi(i) - \pi^*(i)\right)^2$$

LOSS

$$\rho = 1 - \frac{6D(\pi, \pi^*)}{k(k^2 - 1)}$$

RANK CORRELATION

LABEL RANKING: PREDICTION

PREDICTION

0.92	1	81	382	4	1	3	2

A ranking of all labels

LOSS

GROUND TRUTH

0.92	1	81	382	2	1	3	4

KENDALL

$$\mathcal{L}(\pi, \pi^*) = \sum_{1 \le i < j \le k} [\![(\pi(i) - \pi(j))(\pi^*(i) - \pi^*(j)) < 0]\!]$$

LOSS

$$\tau = 1 - \frac{4D(\pi, \pi^*)}{k(k-1)}$$

RANK CORRELATION

LEARNING TECHNIQUES

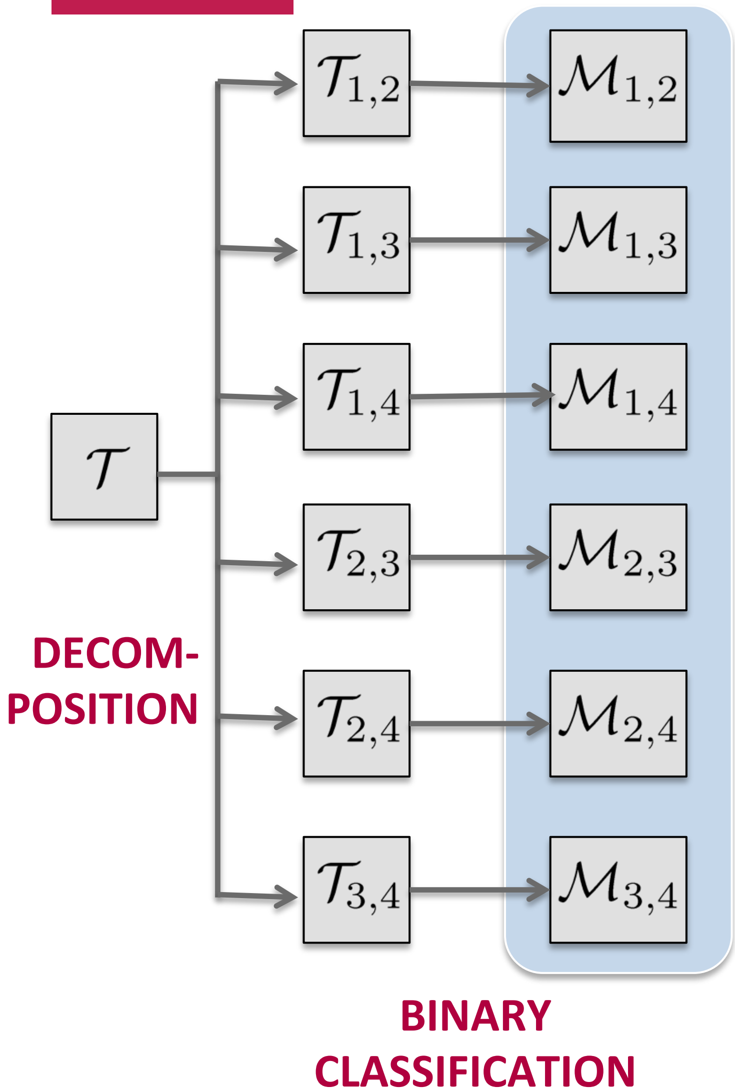
How to learn a label ranker $h: \mathcal{X} \rightarrow \mathcal{S}_k$?

DIFFERENT APPROACHES:

- Reduction to simpler problems (binary classification)
  Transform the problem, so as to make it amenable to standard ML algorithms.
- Extension of (classification) algorithms
  Generalize standard ML algorithms, so as to make them applicable to label ranking data.
- Probabilistic modeling and statistical inference
  Make use of statistical models for rank data and parameter estimation methods.

RANKING BY PAIRWISE COMPARISON

TRAIN



RANKING BY PAIRWISE COMPARISON

INTELLIGENT SYSTEMS

Ranking by Pairwise Comparison (RPC) trains models

$$\mathcal{M}_{i,j} : \mathcal{X} \to [0,1] \qquad\qquad (1 \leq i < j \leq k)$$

Given a query instance $\boldsymbol{x}$, $\mathcal{M}_{i,j}$ is supposed to predict the probability that $y_i \succ y_j$:

$$\begin{aligned}\mathcal{M}_{i,j}(\boldsymbol{x}) &= \mathbf{P}(y_i \succ y_j) \\ &= 1 - \mathbf{P}(y_j \succ y_i)\end{aligned}$$

$\longrightarrow$ decomposition into $k(k-1)/2$ **binary classification problems**

RANKING BY PAIRWISE COMPARISON

Training data (for the label pair A and B):

X_1	X_2	X_3	X_4	preferences	class
0.34	0	10	174	$\mathbf{A} \succ \mathbf{B}$, $B \succ C$, $C \succ D$	1
1.45	0	32	277	$B \succ C$	---
1.22	1	46	421	$B \succ D$, $\mathbf{B} \succ \mathbf{A}$, $C \succ D$, $A \succ C$	0
0.74	1	25	165	$C \succ A$, $C \succ D$, $\mathbf{A} \succ \mathbf{B}$	1
0.95	1	72	273	$B \succ D$, $A \succ D$,	---
1.04	0	33	158	$D \succ A$, $\mathbf{A} \succ \mathbf{B}$, $C \succ B$, $A \succ C$	1

RANKING BY PAIRWISE COMPARISON

Training data (for the label pair A and B):

X_1	X_2	X_3	X_4	class
0.34	0	10	174	1
1.22	1	46	421	0
0.74	1	25	165	1
1.04	0	33	158	1

RANKING BY PAIRWISE COMPARISON

INTELLIGENT SYSTEMS

At prediction time, a query instance is submitted to all models, and the predictions are combined into a binary preference relation:

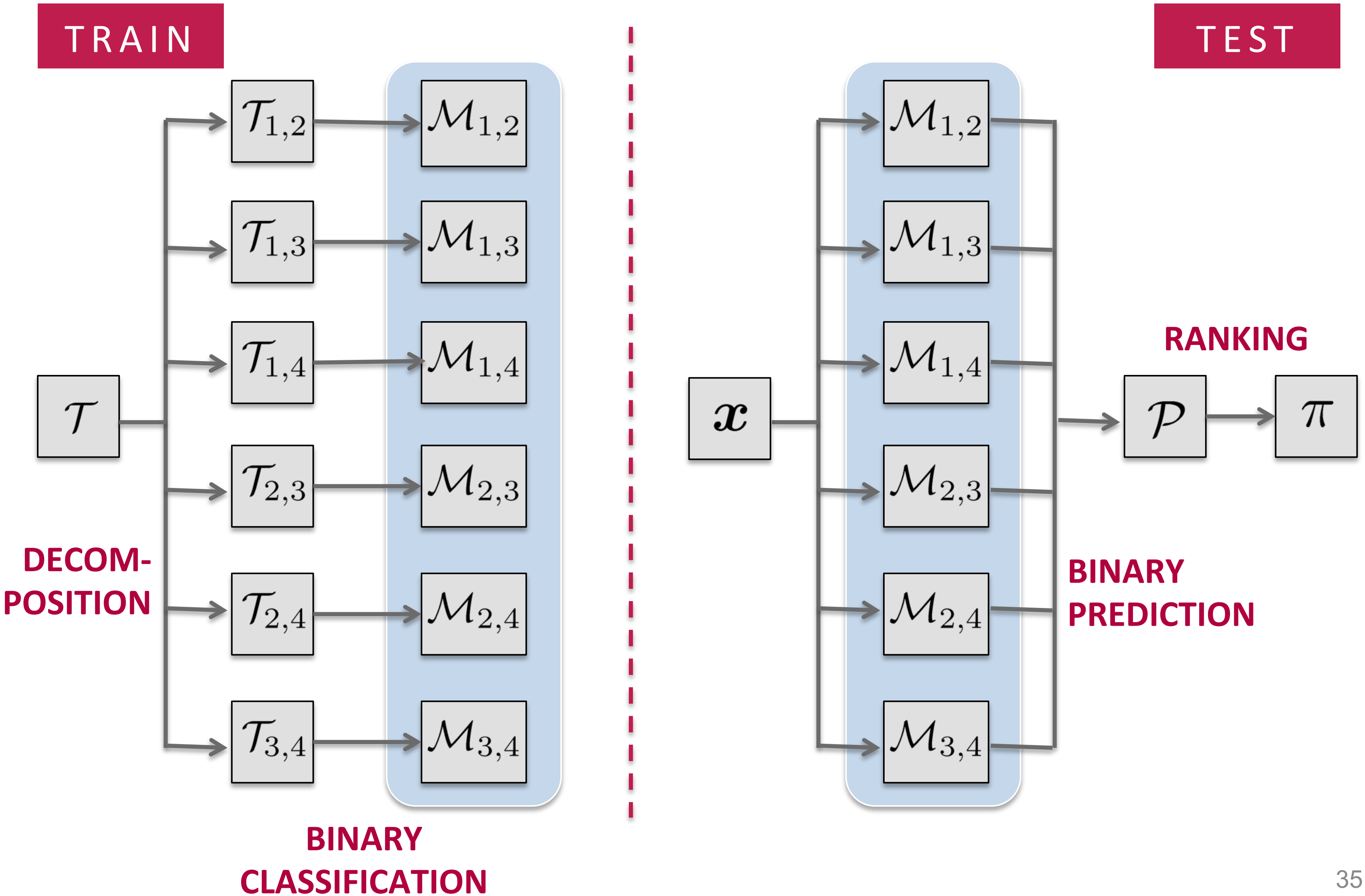
$$\mathcal{P}(i,j) = \begin{cases} \mathcal{M}_{i,j}(\boldsymbol{x}), & i < j \\ 1 - \mathcal{M}_{i,j}(\boldsymbol{x}), & i > j \end{cases}$$

predictions $\mathcal{M}_{i,j}(\boldsymbol{x})$ →

	A	B	C	D
A		0.3	0.8	0.4
B	0.7		0.7	0.9
C	0.2	0.3		0.3
D	0.6	0.1	0.7	

How to produce a ranking on the basis of this preference relation?

RANKING BY PAIRWISE COMPARISON



LOSS DECOMPOSITION

Recall our original goal

$$\mathcal{R}(h) = \int_{\mathcal{X}\times\mathcal{Y}} \mathcal{L}\left(h(x), y\right) \, d\,\mathbf{P}(X,Y) \quad \longrightarrow \quad \min$$

and our representation:

$$h = \mathrm{AGG}\left(\mathcal{M}_{1,2}, \mathcal{M}_{1,3}, \ldots, \mathcal{M}_{k-1,k}\right)$$

Loss decomposition problem: Is it possible to find a suitable loss $\mathcal{L}_p$, to be minimized (in expectation) by the pairwise learners, and an aggregation function AGG, such that $h = \mathrm{AGG}(\mathcal{M}_{1,2}, \ldots, \mathcal{M}_{k-1,k})$ minimizes $\mathcal{L}$ (in expectation)?

MINIMIZING SPEARMAN LOSS

predictions
$\mathcal{M}_{i,j}(\boldsymbol{x})$ →

	A	B	C	D	
A		0.3	0.8	0.4	1.5
B	0.7		0.7	0.9	2.3
C	0.2	0.3		0.3	0.8
D	0.6	0.1	0.7		1.4

B $\succ$ A $\succ$ D $\succ$ C

MINIMIZING SPEARMAN LOSS

Theorem: Suppose the pairwise learners $\mathcal{M}_{i,j}$ yield unbiased probability estimates and let π be a ranking such that

$$\left(\sum_q \mathcal{P}(i,q) > \sum_q \mathcal{P}(j,q) \right) \;\Rightarrow\; (\pi(i) < \pi(j)) \;\; .$$

Then π minimizes risk w.r.t. to the Spearman loss

$$\mathcal{L}(\pi, \pi^*) = \sum_{i=1}^{k} \big(\pi(i) - \pi^*(i)\big)^2 \;\; .$$

LIMITATIONS OF RPC

Proposition: For the following losses, RPC can not guarantee a risk minimizing prediction:

- 0/1 loss

$$\mathcal{L}(\pi, \pi^*) = [\![\pi \neq \pi^*]\!]$$

- Hamming distance

$$\mathcal{L}(\pi, \pi^*) = \sum_{i=1}^{k} [\![\pi(i) \neq \pi^*(i)]\!]$$

- Cayley distance (minimal number of transpositions of any pair of labels needed to turn the first ranking into the second one)
- Ulam distance (minimal number of position changes of labels needed to turn the first ranking into the second one)

OUTLINE

INTELLIGENT SYSTEMS

PART 1	PART 2	PART 3
Preference learning	Label ranking	Preference-based bandit algorithms

MULTI-ARMED BANDITS

„pulling an arm“ ⟷ choosing an option

partial information online learning
sequential decision process

MULTI-ARMED BANDITS

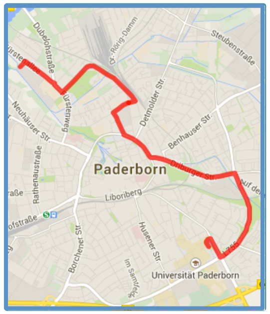
INTELLIGENT SYSTEMS

„pulling an arm" ⟷ putting an advertisement on a website

choice of an option/strategy (arm) yields a **random reward**

partial information online learning
sequential decision process

MULTI-ARMED BANDITS



„pulling an arm“ ⟷ picking a traffic route from source to target

choice of an option/strategy (arm) yields a **random reward**

partial information online learning
sequential decision process

MULTI-ARMED BANDITS

INTELLIGENT SYSTEMS

$X_1 \sim \mathbf{P}_1$	$X_2 \sim \mathbf{P}_2$	$X_3 \sim \mathbf{P}_3$	$X_4 \sim \mathbf{P}_4$	$X_5 \sim \mathbf{P}_5$

„pulling an arm“ $\longleftrightarrow$ choosing an option

choice of an option/strategy (arm) yields a **random reward**

partial information online learning
sequential decision process

MULTI-ARMED BANDITS

$X_1 \sim \mathbf{P}_1$	$X_2 \sim \mathbf{P}_2$	$X_3 \sim \mathbf{P}_3$	$X_4 \sim \mathbf{P}_4$	$X_5 \sim \mathbf{P}_5$

Immediate reward: 2.5
Cumulative reward: 2.5

MULTI-ARMED BANDITS

$X_1 \sim \mathbf{P}_1$ $X_2 \sim \mathbf{P}_2$ $X_3 \sim \mathbf{P}_3$ $X_4 \sim \mathbf{P}_4$ $X_5 \sim \mathbf{P}_5$

Immediate reward: `2.5 3.1`

Cumulative reward: `2.5 5.6`

MULTI-ARMED BANDITS

INTELLIGENT SYSTEMS

$X_1 \sim \mathbf{P}_1$ $\quad X_2 \sim \mathbf{P}_2$ $\quad X_3 \sim \mathbf{P}_3$ $\quad X_4 \sim \mathbf{P}_4$ $\quad X_5 \sim \mathbf{P}_5$

Immediate reward: `2.5 3.1 1.7`

Cumulative reward: `2.5 5.6 7.3`

MULTI-ARMED BANDITS

INTELLIGENT SYSTEMS

$X_1 \sim \mathbf{P}_1$ $X_2 \sim \mathbf{P}_2$ $\boxed{X_3 \sim \mathbf{P}_3}$ $X_4 \sim \mathbf{P}_4$ $X_5 \sim \mathbf{P}_5$

Immediate reward:	2.5	3.1	1.7	3.7	...
Cumulative reward:	2.5	5.6	7.3	11.0	...

maximize cumulative reward → *explore and exploit (tradeoff)*

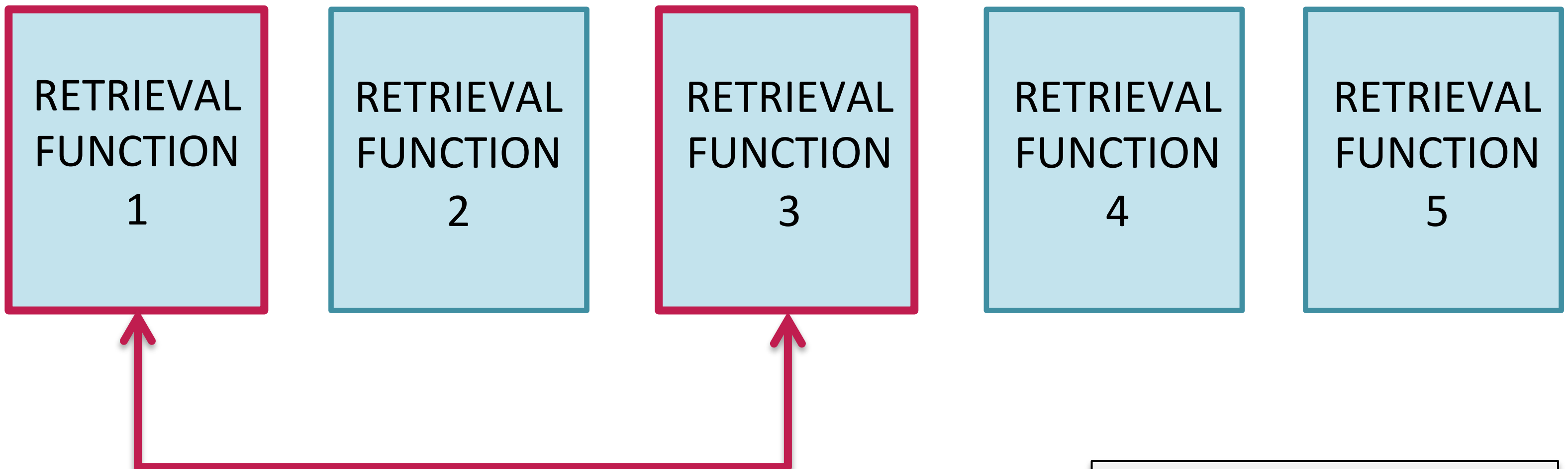
find best option → *pure exploration (effort vs. certainty)*

$X_1 \sim \mathbf{P}_1$	$X_2 \sim \mathbf{P}_2$	$X_3 \sim \mathbf{P}_3$	$X_4 \sim \mathbf{P}_4$	$X_5 \sim \mathbf{P}_5$

In many applications,

- the assignment of (numeric) **rewards to single outcomes** (and hence the assessment of individual options on an absolute scale) is difficult,
- while the **qualitative comparison between pairs of outcomes** (arms/ options) is more feasible.

PREFERENCE-BASED BANDITS

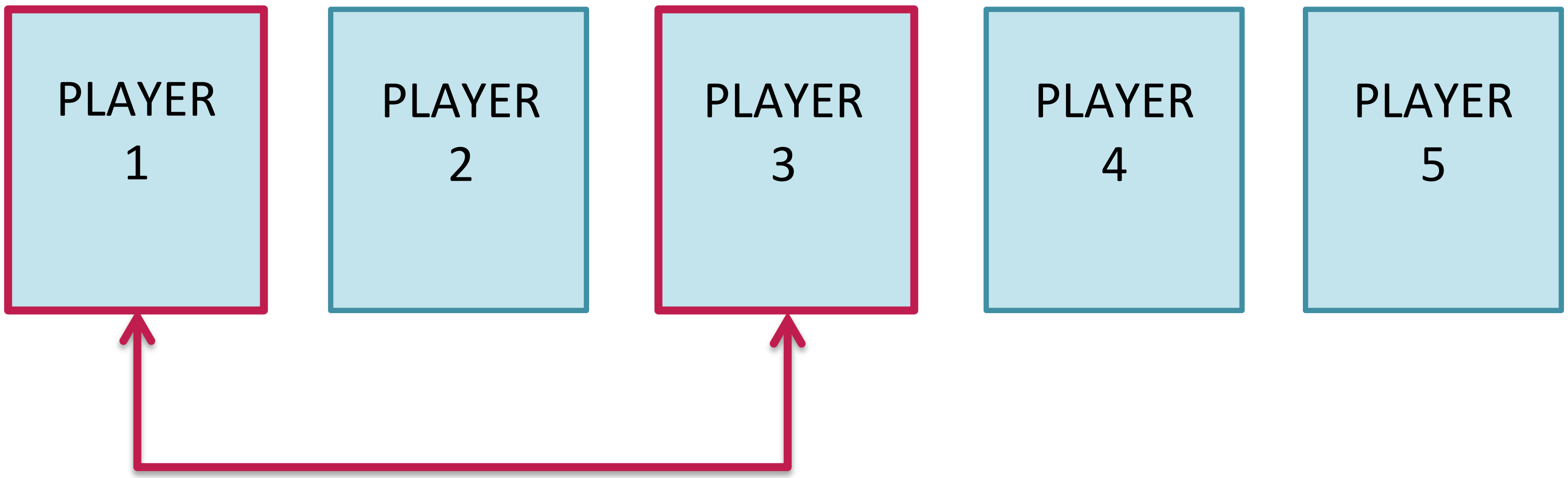


$$X_3 \succ X_1$$

The result returned by the third retrieval function, for a given query, is preferred to the result returned by the first search engine.

Noisy preferences can be inferred from how a user clicks through an **interleaved** list of documents [Radlinski et al., 2008].

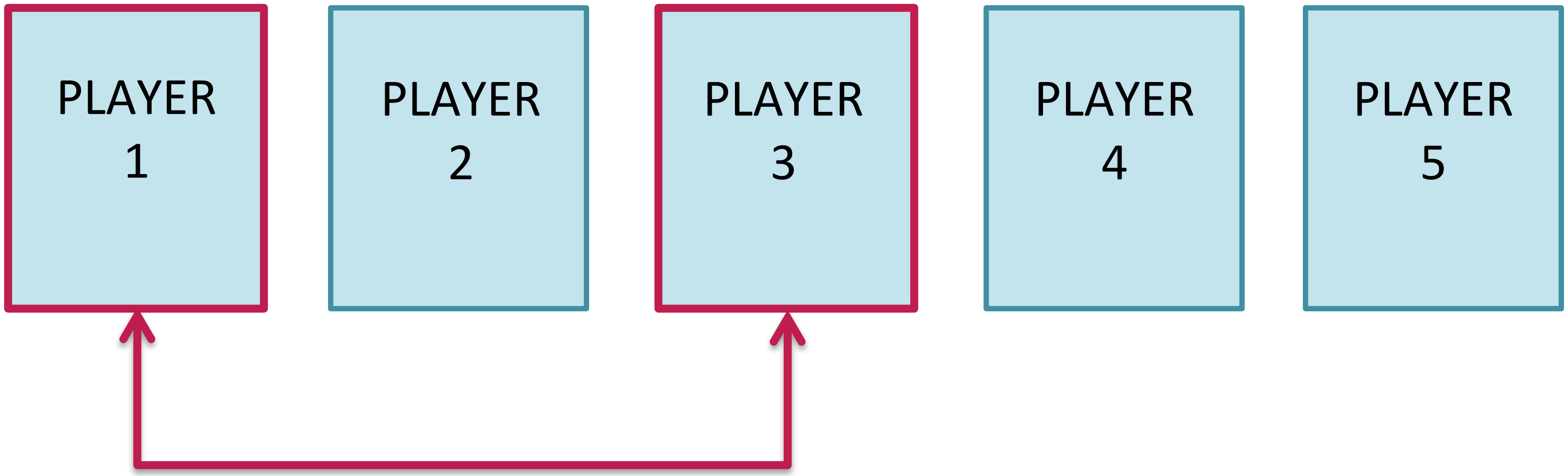
PREFERENCE-BASED BANDITS



$$X_3 \succ X_1$$

Third player has beaten first player in a match.

PREFERENCE-BASED BANDITS



$$X_3 \succ X_1$$

- *This setting has first been introduced as the* ***dueling bandits problem*** *(Yue and Joachims, 2009).*
- *More generally, we speak of* ***preference-based multi-armed bandits*** *(PB-MAB).*

FORMAL SETTING

- fixed set of arms (options) $\mathcal{A} = \{a_1, \ldots, a_K\}$
- **action space** of the learner (agent) $= \{\,(i,j)\,|\,1 \leq i \leq j \leq K\,\}$ (compairing pairs of arms a_i and a_j)
- feedback generated by an (unknown, time-stationary) probabilistic process characterized by a **preference relation**

$$\mathbf{Q} = \begin{bmatrix} q_{1,1} & q_{1,2} & \cdots & q_{1,K} \\ q_{2,1} & q_{2,2} & \cdots & q_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ q_{K,1} & q_{K,2} & \cdots & q_{K,K} \end{bmatrix} ,$$

where

$$q_{i,j} = \mathbf{P}\left(a_i \succ a_j\right)$$

- typically, $\mathbf{Q}$ is reciprocal $(q_{i,j} = 1 - q_{j,i})$

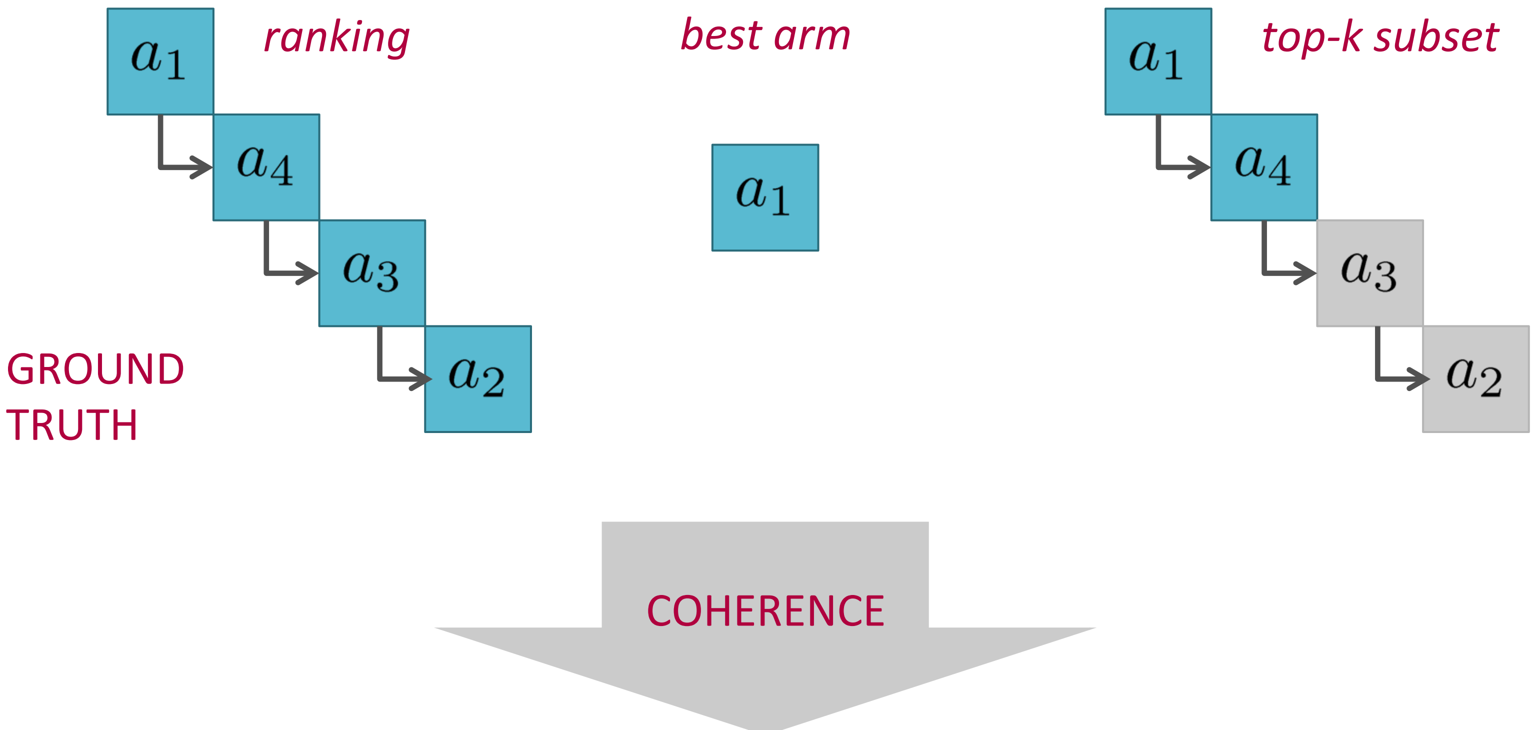
THE PREFERENCE RELATION

- We say arm a_i beats arm a_j if $q_{i,j} > 1/2$.
- The degrees of **distinguishability**

  $$\Delta_{i,j} = q_{i,j} - \frac{1}{2}$$

  quantify the hardness of a PB-MAB task.
- Definition of **regret** is not straightforward.
- Assumptions on properties of $\mathbf{Q}$ are crucial for learning.
- **Coherence:** The pairwise comparisons need to provide hints (even if "noisy" ones) on the target.

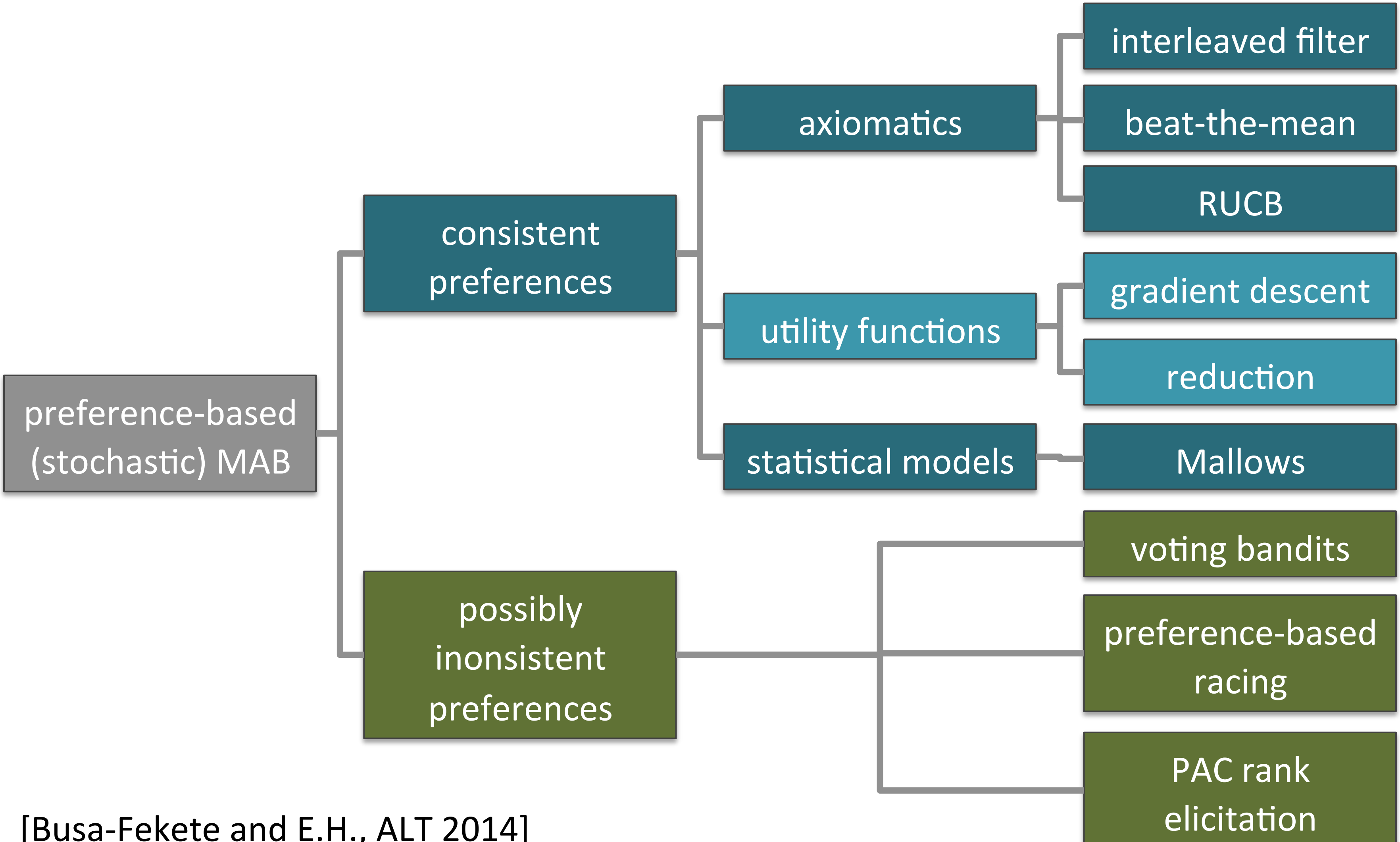
PROPERTIES OF PREFERENCE RELATION



$$\mathbf{Q} = \begin{bmatrix} q_{1,1} & q_{1,2} & \cdots & q_{1,K} \\ q_{2,1} & q_{2,2} & \cdots & q_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ q_{K,1} & q_{K,2} & \cdots & q_{K,K} \end{bmatrix}$$

... the preference relation is derived from, or at least strongly restricted by the target!

OVERVIEW OF METHODS



[Busa-Fekete and E.H., ALT 2014]

Statistical approach

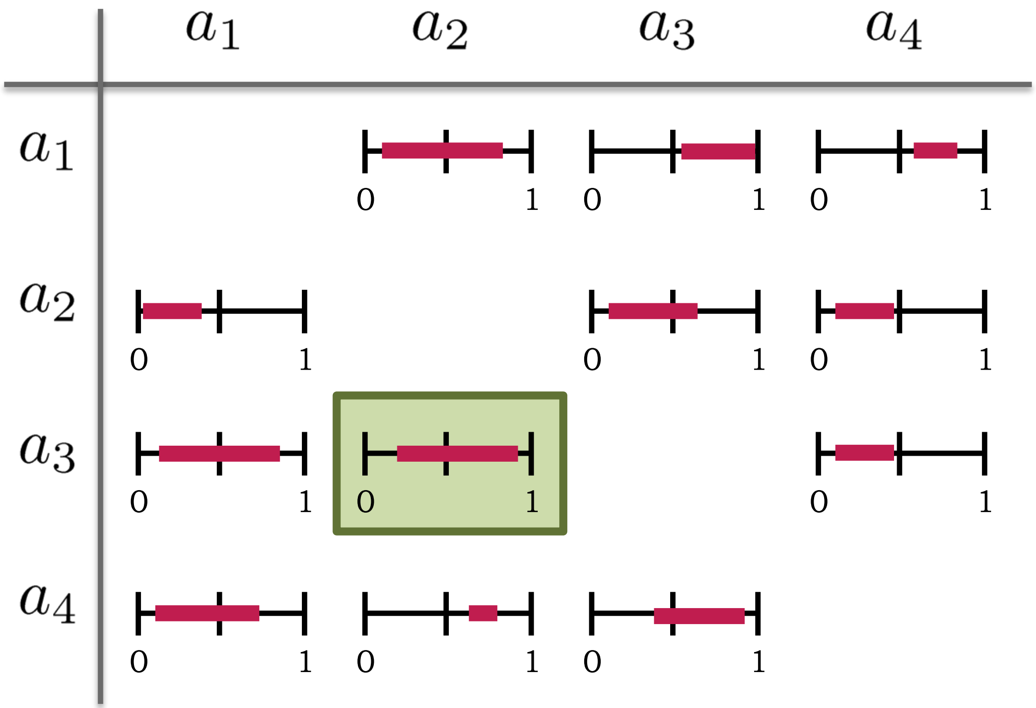
Coherence through statistical assumptions on the data generating process, e.g., pairwise probabilities as marginals of a Mallows model:

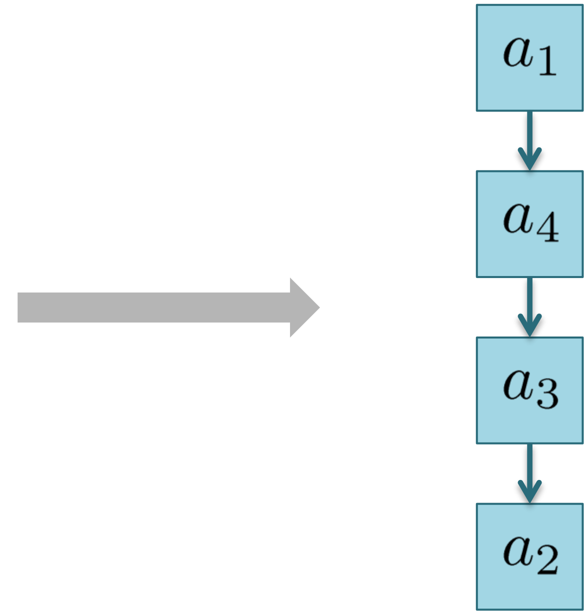
$$q_{i,j} = \mathbf{P}(\, a_i \succ a_j \,) = \sum_{\pi:\, \pi(i) < \pi(j)} \mathbf{P}(\pi \,|\, \pi_0, \theta)$$

$$= \frac{1}{\phi(\pi_0, \theta)} \sum_{\pi:\, \pi(i) < \pi(j)} \exp\big(-\theta \Delta(\pi, \pi_0)\big)$$

$\longrightarrow$ reference ranking π_0 is the natural target!

PAIRWISE SAMPLING





uncertainty about pairwise preferences

translates into

uncertainty about ranking

Busa-Fekete et al. (2014) propose a sampling strategy called **MallowsMPR**, which is based on the **merge sort** algorithm for selecting the arms to be compared.

However, two arms a_i and a_j are not only compared once, but possibly several times until being sure enough:

$$1/2 \notin \left[\widehat{q}_{i,j} - c_{i,j}, \widehat{q}_{i,j} + c_{i,j} \right] \ .$$

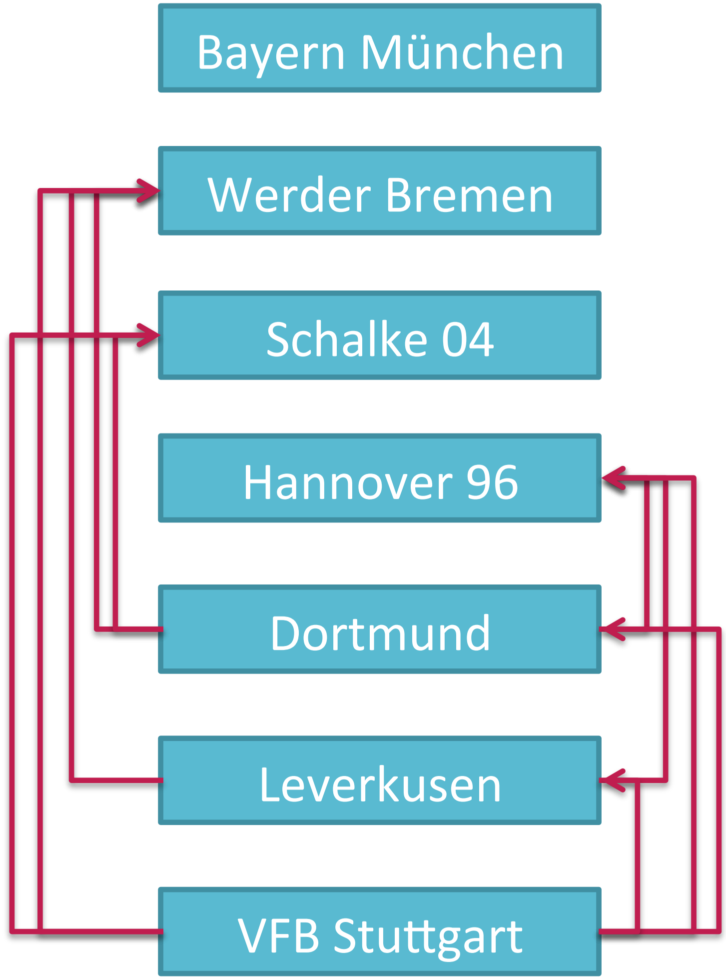
Theorem: For any $0 < \delta < 1$, MallowsMPR outputs the reference ranking π_0 with probability at least $1 - \delta$, and the number of pairwise comparisons taken by the algorithm is

$$\mathcal{O}\left(\frac{K \log_2 K}{\rho^2} \log \frac{K \log_2 K}{\delta \rho}\right) \quad ,$$

where $\rho = \frac{1-\phi}{1+\phi}$, $\phi = \exp(-\theta)$.

EMPIRICAL VALIDATION

- In general, the approach performs quite well compared to baselines.
- However, it may fail if the underlying data is not enough „Mallowsian“ ...



SUMMARY

INTELLIGENT SYSTEMS

- Growing interest in **preferences in AI** and **preference learning**
- Focus so far on **rank-based preference models** („learning-to-rank")
- **Online preference learning** not yet strongly developed
- Preference-based online learning with multi-armed bandits (PB-MAB):
  - **emerging** research topic,
  - no complete and **coherent framework** so far,
  - many **open questions and problems** (e.g., necessary conditions for bounds on regret or sample complexity, lower bounds, verifying model assumptions, generalizations to large (structured) set of arms, contextual bandits, adversarial setting, etc., ...)

SELECTED REFERENCES

- J. Fürnkranz and E. H. (eds.). **Preference Learning**, Springer, 2011.
  General introduction to preference learning

- E. Hüllermeier, J. Fürnkranz, W. Cheng, K. Brinker**. Label ranking by learning pairwise preferences**. Artif. Intell., 172, 2008.
  The LPC approach presented in this talk.

- R. Busa-Fekete, B. Szorenyi, E. H. **Preference-Based Rank Elicitation using Statistical Models: The Case of Mallows.** Proc. ICML-2014, Int. Conf. Machine Learning, 2014.
  PB-MAB method briefly sketched in this talk.

- R. Busa-Fekete and E. H. **A Survey of Preference-based Online Learning with Bandit Algorithms**. Proc. ALT-2014, Int. Conf. Algorithmic Learning Theory, Bled, 2014.
  Survey paper on preference-based bandits.

SELECTED LITERATURE (PB-MAB)

- N. Ailon, K. Hatano, and E. Takimoto. Bandit online optimization over the permutahedron. CoRR, abs/1312.1530, 2014.
- N. Ailon, Z. Karnin, and T. Joachims. Reducing dueling bandits to cardinal bandits. ICML 2014.
- P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. Machine Learning, 47:235-256, 2002.
- **R. Busa-Fekete and E. Hüllermeier. A Survey of Preference-based Online Learning with Bandit Algorithms. Proc. ALT-2014, Int. Conf. Algorithmic Learning Theory, Bled, 2014.**
- **R. Busa-Fekete, E. Hüllermeier, and B. Szorenyi. Preference-based rank elicitation using statistical models: The case of Mallows. ICML 2014.**
- R. Busa-Fekete, B. Szorenyi, and E. Hüllermeier. PAC rank elicitation through adaptive sampling of stochastic pairwise preferences. AAAI 2014.
- R. Busa-Fekete, B. Szorenyi, P. Weng, W. Cheng, and E. Hüllermeier. Top-k selection based on adaptive sampling of noisy preferences. ICML 2013.
- W.W. Cohen, R.E. Schapire and Y. Singer. Learning to order things. J. of Artif. Intelligence Res., 10:243–270, 1999.
- J. Duchi, L. Mackey, and M. Jordan. On the consistency of ranking algorithms. ICML 2010.
- J. Fürnkranz and E. Hüllermeier, editors. Preference Learning. Springer-Verlag, 2011.
- E. Hüllermeier, J. Fürnkranz, W. Cheng, K. Brinker. Label ranking by learning pairwise preferences. Artif. Intell., 172, 2008.
- F. Radlinski, M. Kurup, and T. Joachims. How does clickthrough data reflect retrieval quality? CIKM 2008.
- T. Urvoy, F. Clerot, R. Feraud, and S. Naamane. Generic exploration and k-armed voting bandits. ICML 2013.
- Y. Yue, J. Broder, R. Kleinberg, and T. Joachims. The K-armed dueling bandits problem. Journal of Computer and System Sciences, 78(5):1538-1556, 2012.
- Y. Yue and T. Joachims. Interactively optimizing information retrieval systems as a dueling bandits problem. ICML 2009.
- Y. Yue and T. Joachims. Beat the mean bandit. ICML 2011.
- M. Zoghi, S. Whiteson, R. Munos, and M. de Rijke. Relative upper confidence bound for the k-armed dueling bandit problem. ICML 2014.

SELECTED LITERATURE (PL)

- E. Hüllermeier, J. Fürnkranz, W. Cheng and K. Brinker. Label ranking by learning pairwise preferences. Artificial Intelligence, 172, 2008.
- W. Cheng, J. Hühn and E. Hüllermeier. Decision tree and instance-based learning for label ranking, ICML-09, Montreal, 2009.
- W. Cheng, K. Dembczynski and E. Hüllermeier. Label ranking using the Plackett-Luce model. ICML-10, Haifa, Israel, 2010.
- W. Cheng, W. Waegeman, V. Welker and E. Hüllermeier. Label ranking with partial abstention based on thresholded probabilistic models. NIPS 2012.
- J. Fürnkranz, E. Hüllermeier, W. Cheng, S.H. Park. Preference-Based Reinforcement Learning: A Formal Framework and a Policy Iteration Algorithm. Machine Learning, 89, 2012.
- E. Hüllermeier and J. Fürnkranz. On predictive accuracy and risk minimization in pairwise label ranking. J. Computer and System Sciences , 76, 2010.
- E. Hüllermeier and P. Schlegel. Preference-based CBR: First steps toward a methodological framework. ICCBR-11, London, 2011.
- R. Akrour, M. Schoenauer, M. Sebag. Preference-Based Policy Learning, ECML 2011.
- W.W. Cohen, R.E. Schapire and Y. Singer. Learning to order things. Journal of Artificial Intelligence Research, 10:243–270, 1999.
- O. Dekel, C.D. Manning, Y. Singer. Log-Linear Models for Label Ranking. NIPS-2003.
- D. Goldberg, D. Nichols, B.M. Oki and D. Terry. Using collaborative filtering to weave and information tapestry. Communications of the ACM, 35(12):61–70, 1992.
- S. Har-Peled, D. Roth and D. Zimak. *Constraint classification: A new approach to multiclass classification*. Proc. ALT-2002.
- D.R. Hunter. MM algorithms for generalized Bradley-Terry models. The Annals of Statistics , 32(1):384–406, 2004.
- S. Vembu and T. Gärtner. Label ranking: a survey. In: Preference Learning. J. Fürnkranz and E. Hüllermeier (eds.), Springer-Verlag, 2011.