Label Ranking with Partial Abstention using Ensemble Learning

Weiwei Cheng & Eyke Hüllermeier

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

Label Ranking (an example)

Learning geeks' preferences on hotels

	label ranking
geek 1	Golf ≻ Park ≻ Krim
geek 2	Krim ≻ Golf ≻ Park
geek 3	Krim ≻ Park ≻ Golf
geek 4	Park ≻ Golf ≻ Krim
new geek	???

where the geek could be described by feature vectors,
e.g., (*gender*, *age*, *place of birth*, *is a professor*, ...)

Label Ranking (an example)

Learning geeks' preferences on hotels

	Golf	Park	Krim
geek 1	1	2	3
geek 2	2	3	1
geek 3	3	2	1
geek 4	2	1	3
new geek	?	?	?

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

1: Golf 2: Park 3: Krim

Label Ranking (more formally)

Given:

- a set of training instances $\{\mathbf{x}_k \mid k = 1 \ldots m\} \subseteq \mathbf{X}$
- a set of labels $\mathcal{L} = \{l_i \mid i = 1 \ldots n\}$
- for each training instance $\mathbf{x}_k$: a set of *pairwise preferences* of the form $l_i \succ_{\mathbf{x}_k} l_j$ (for <u>some</u> of the labels)

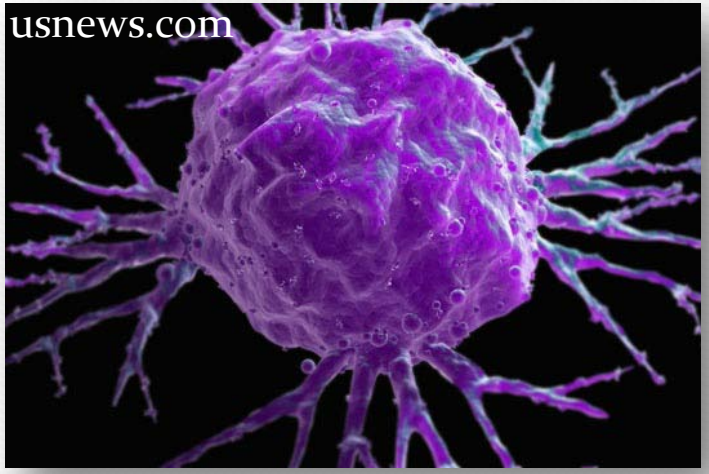
Find:

- A ranking function ($\mathcal{X} \rightarrow \Omega$ mapping) that maps each $\mathbf{x} \in \mathbf{X}$ to a ranking $\succ_{\mathbf{x}}$ of $\mathcal{L}$ (permutation $\pi_{\mathbf{x}}$) and generalizes well in terms of a loss function on rankings (e.g., *Kendall's tau*)

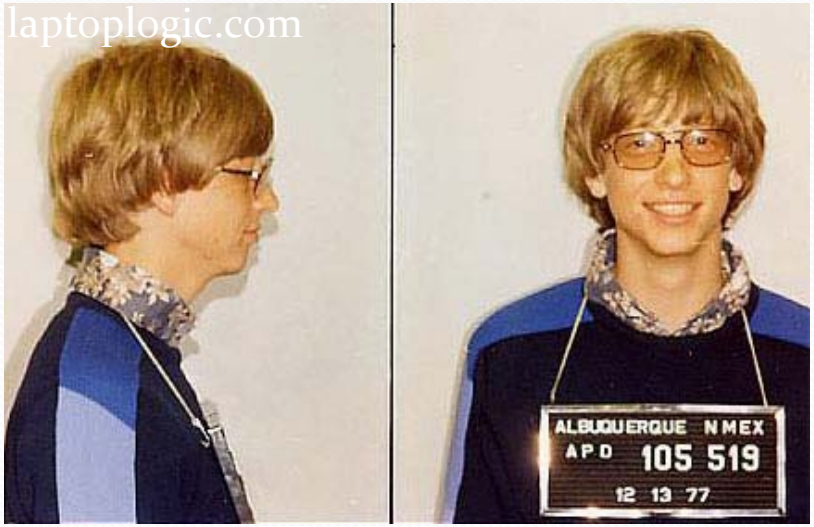
Existing Approaches

- Constraint classification
  Har-Peled , Roth, and Zimak, *NIPS-03*
- Log linear models for label ranking
  Dekel, Manning, and Singer, *NIPS-03*
- Label ranking by learning pairwise preferences
  Hüllermeier, Fürnkranz, Cheng, and Brinker, *Artificial Intelligence*
- Decision tree and instance-based learning for label ranking
  Cheng, Hühn, and Hüllermeier, *ICML-09*

Learning with Reject Option



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

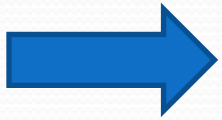


Label Ranking with Reject Option

For **each pair of labels** a and b, the learner can

- predict $a \succ b$ or $b \succ a$, or
- abstain from prediction (reject option).

The learner should be consistent (transitivity).



partial orders

Label Ranking Ensemble

Create a “committee of experts”

Label Ranking Ensemble

- For a query, setup a label ranking ensemble of size *k*

$$\succ_1 , \succ_2 , \ldots\ldots , \succ_k .$$

- Define a partial order with

$$(\lambda_i \succeq \lambda_j) \Longleftrightarrow \forall l \in \{1, 2, \ldots, k\} : \lambda_i \succ_l \lambda_j$$

✓ $$(\lambda_i \succeq \lambda_j) \Longleftrightarrow \frac{\#\{l \in \{1, 2, \ldots, k\} \mid \lambda_i \succ_l \lambda_j\}}{k} \geq t.$$

$$0.5 < t \leq 1$$

Two Problems

$$(\lambda_i \succeq \lambda_j) \Longleftrightarrow \frac{\#\{l \in \{1, 2, \ldots, k\} \mid \lambda_i \succ_l \lambda_j\}}{k} \geq t.$$

problem	Transitivity If $a \succ b$ and $b \succ c$, then $a \succ c$.	No cycle If $a \succ b$ and $b \succ c$, then not $c \succ a$.
solution	Get transitive closure with Marshall's algorithm.	*to be solved*

Proposition

Given a set of total orders on a finite set $\mathcal{L}$, denote by P_{ab} the proportion of orders in which a precedes b. Then, for any triple of elements $a, b, c \in \mathcal{L}$, we have

$$P_{ca} \leq 2 - P_{ab} - P_{bc}.$$

S.t. $(P_{ab} \geq 2/3) \wedge (P_{bc} \geq 2/3) \Longrightarrow (P_{ca} \leq 2/3)$

Choosing t > 2/3, we can guarantee $\succeq$ acyclic.

Experimental Setting

dataset	#instance	#attribute	#labels
iris	150	4	3
wine	178	13	3
glass	214	9	6
vowel	528	10	11
vehicle	846	18	4

Evaluation metrics

$$C(\succeq, \succ) = \frac{\#\text{concordant label pairs} - \#\text{discordant label pairs}}{\#\text{concordant label pairs} + \#\text{discordant label pairs}}$$

prediction *true ranking*

Experimental Results

threshold	iris	wine	glass	vowel	vehicle
original	0.868±0.093	0.884±0.078	0.793±0.070	0.324±0.028	0.809±0.034
0.7	0.919±0.066	0.918±0.079	0.847±0.055	0.436±0.034	0.851±0.032
0.8	0.921±0.064	0.956±0.057	0.869±0.055	0.478±0.039	0.872±0.031
0.9	0.940±0.050	0.971±0.049	0.892±0.054	0.515±0.045	0.896±0.031
1.0	0.950±0.045	0.995±0.019	0.928±0.046	0.563±0.056	0.926±0.027

ensemble size of 10

Our Contributions

- A first attempt on label ranking with reject option;
- Output a reliably partial ranking with ensemble learning.

Follow up!

The End

Thanks!

Google “kebi germany” for more info.