Preference-based CBR: First Steps Toward a Methodological Framework

Eyke Hüllermeier and Patrice Schlegel
Knowledge Engineering & Bioinformatics Lab
Department of Mathematics and Computer Science
Marburg University, Germany

Philipps Universität Marburg

Goals and Contributions

Our paper makes a first step toward a

preference-based

methodological framework of CBR.

- in-between high-level models (like CBR cycle) and concrete implementations
- sufficiently general and abstract, so as to allow for the development of generic algorithms, for analyzing formal properties, etc.
- sufficiently concrete, so as to support the development of specific applications

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]

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]

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]

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]

→ applies to AI in general and to CBR in particular!

→ modeling case-based experience in terms of preferences!

Case-based Representation of Experience

The standard representation of experience in terms of **problem/solution** pairs

$$(\boldsymbol{x}, \boldsymbol{y}) \in \mathbf{X} \times \mathbf{Y}$$

may cause disadvantages (consider, e.g., the cooking domain):

- assumes **existence of „correct“** (and perhaps even unique) **solution**
- assumes that a certain level of **optimality can be proved**
- a single solution does not necessarily reflect the whole experience gathered during a problem solving episode (**loss of information**)
- provides **limited guidance** if a retrieved solution fails

Case-based Representation of Experience

Our basic idea is to replace experiences of the form $(\boldsymbol{x}, \boldsymbol{y})$, meaning

"solution y (optimally) solves problem x",

by "contextualized preferences" of the form $\boldsymbol{y} \succeq_{\boldsymbol{x}} \boldsymbol{y}'$, meaning

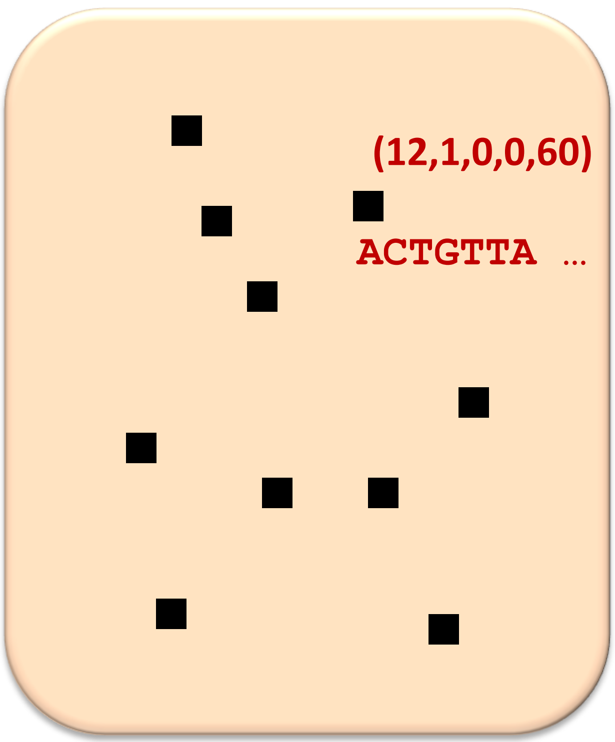
"y is better (more preferred) than y' as a solution for x".

- This is relatively **weak, qualitative knowledge**, which is easy to acquire.
- Thus, the above problems (existence of correct solutions, proof of optimality, loss of information, limited guidance) can be alleviated.
- Suggests recommendation for a new problem in the form of a **ranking**:

$$\boldsymbol{y}_1 \succeq_{\boldsymbol{x}} \boldsymbol{y}_2 \succeq_{\boldsymbol{x}} \boldsymbol{y}_3 \succeq_{\boldsymbol{x}} \cdots \succeq_{\boldsymbol{x}} \boldsymbol{y}_n$$

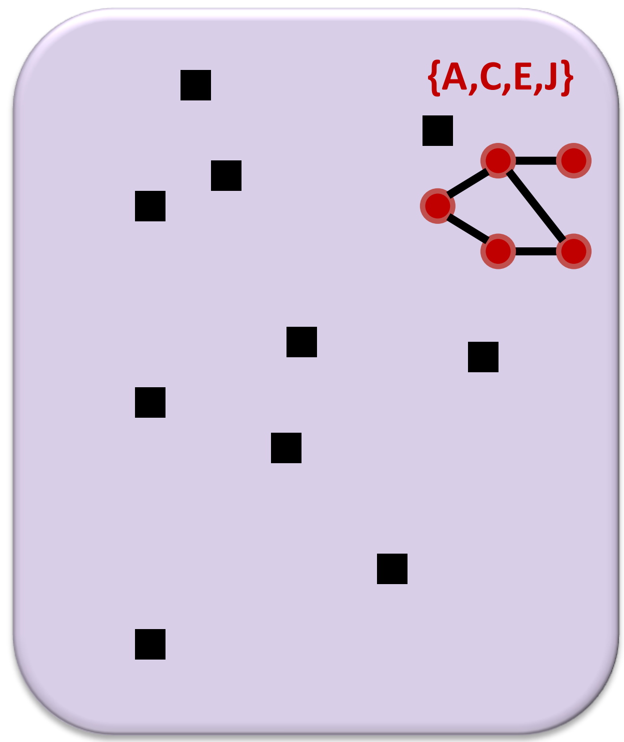
A Formal Framework

problem space $\mathbf{X}$



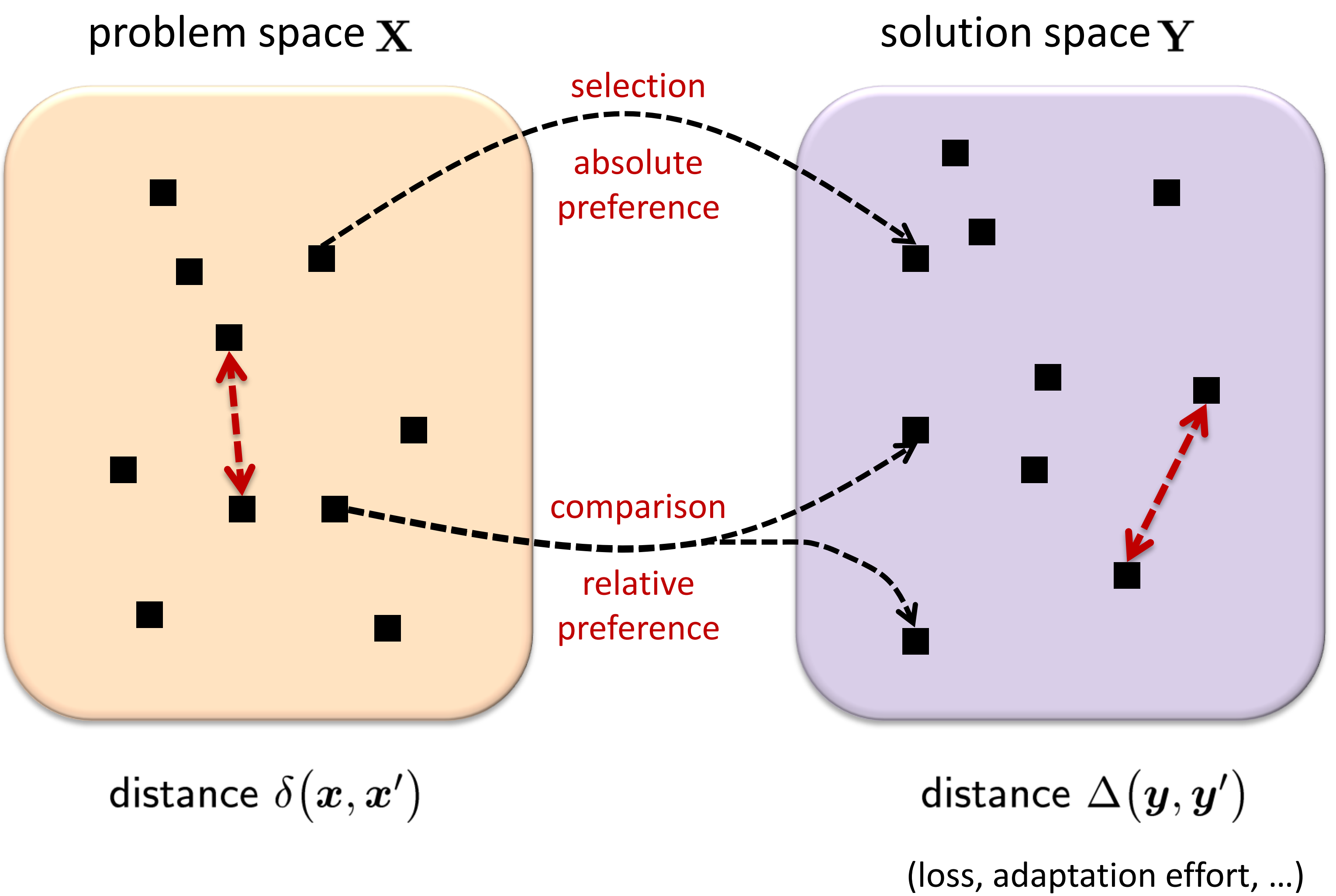
retrieval

solution space $\mathbf{Y}$



inference

A Formal Framework



Preferences on Solutions: A Discrete Choice Model

Assuming an "ideal" solution $\boldsymbol{y}^*$ (given query $\boldsymbol{x}$), Δ induces a ranking on $\mathbf{Y}$:

$$\boldsymbol{y} \succeq_{\boldsymbol{x}} \boldsymbol{y}' \quad \Leftrightarrow \quad \Delta(\boldsymbol{y}, \boldsymbol{y}^*) \leq \Delta(\boldsymbol{y}', \boldsymbol{y}^*)$$

Comparisons are made according to the **latent utility** of $\boldsymbol{y}$, namely

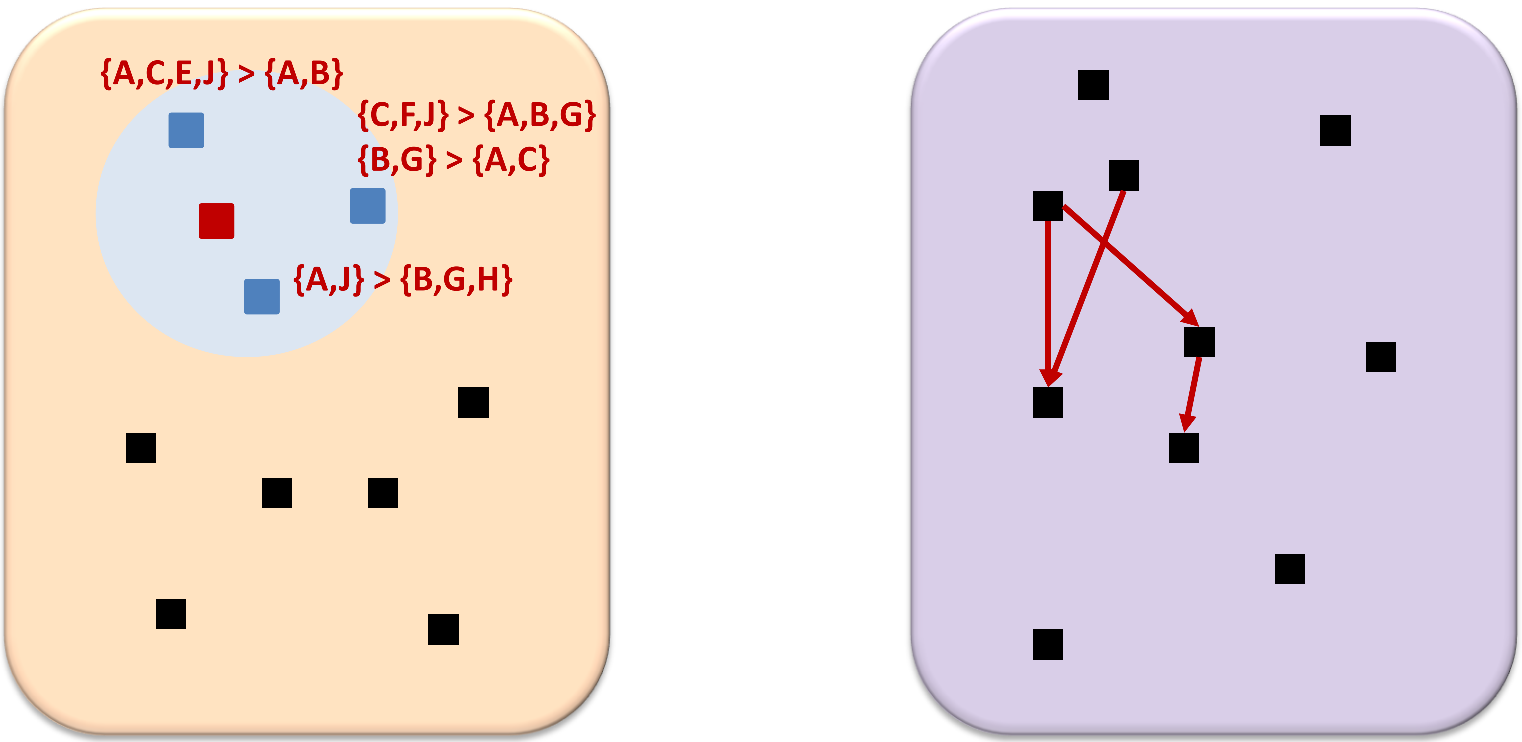
$$U(\boldsymbol{y}) = -\beta\, \Delta\big(\boldsymbol{y}, \boldsymbol{y}^*\big) + \epsilon \ ,$$

which leads to the **logit model of discrete choice**:

$$\mathbf{P}(\boldsymbol{y} \succ \boldsymbol{y}') = \frac{1}{1 + \exp\Big(\underset{\substack{\uparrow \\ \text{precision parameter}}}{-\, \beta}\big(\Delta(\boldsymbol{y}', \boldsymbol{y}^*) - \Delta(\boldsymbol{y}, \underset{\substack{\uparrow \\ \text{ideal solution}}}{\boldsymbol{y}^*})\big)\Big)}$$

Case-based Inference: A Probabilistic Approach

problem space $\mathbf{X}$ solution space $\mathbf{Y}$



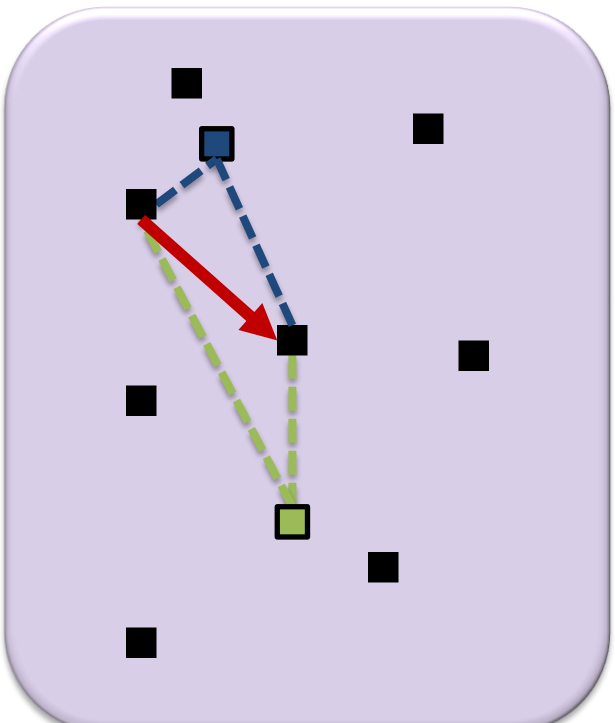
retrieval inference

observed preferences $\mathcal{D} = \{\boldsymbol{y}^{(i)} \succ \boldsymbol{z}^{(i)}\}_{i=1}^{N}$

Case-based Inference: A Probabilistic Approach

Maximum-likelihood estimation:

$$\ell(\boldsymbol{y}, \beta \,|\, \mathcal{D}) = -\sum_{i=1}^{N} \log\left(1 + \exp\left(-\beta\underbrace{(\Delta(\boldsymbol{z}^{(i)}, \boldsymbol{y}) - \Delta(\boldsymbol{y}^{(i)}, \boldsymbol{y})}_{\text{penalty if negative}})\right)\right) \quad \to \max$$



solution space $\mathbf{Y}$

Case-based Inference: A Probabilistic Approach

Maximum-likelihood estimation:

$$\ell(\boldsymbol{y}, \beta \,|\, \mathcal{D}) = -\sum_{i=1}^{N} \log\left(1 + \exp\left(-\beta(\Delta(\boldsymbol{z}^{(i)}, \boldsymbol{y}) - \Delta(\boldsymbol{y}^{(i)}, \boldsymbol{y})\right)\right) \quad \rightarrow \text{max}$$

Case-based inference = finding the most likely „ideal" solution on the basis of the „neighbor preferences"

- fixing a solution, the log-likelihood becomes a one-dimensional (convex) function of the precision parameter → *simple numerical optimization*
- this optimization is embedded in a heuristic search in the (discrete) solution space (hill climbing, genetic algorithms, …)

Case-based Inference: A Probabilistic Approach

To start with a good initial solution, we solve an extension of the generalized median problem:

$$\boldsymbol{y}_{med} = \arg\min_{\boldsymbol{y} \in \mathbf{Y}} \left(\sum_{i=1}^{N} \Delta(\boldsymbol{y}, \boldsymbol{y}^{(i)}) - \sum_{i=1}^{N} \Delta(\boldsymbol{y}, \boldsymbol{z}^{(i)}) \right)$$

preferred solutions

non-preferred solution

The difficulty of this problem strongly depends on the structure of the solution space and the definition of the distance measure!

Case-based Inference: A Probabilistic Approach

Example: $C = \{c_1, c_2, \ldots, c_n\}$ is a finite set of items, and $\mathbf{Y}$ is the "subset solution space" $\mathbf{Y} = 2^C$ (solutions are subsets of C).

The median problem is then easy to solve for the **Hamming distance**

$$\Delta_H(A,B) = \big|(A \setminus B) \cup (B \setminus A)\big| \ ,$$

more difficult for the **F-measure** (polynomial)

$$\Delta_F(A,B) = 1 - \frac{2\big|A \cap B\big|}{\big|A\big| + \big|B\big|} \ ,$$

and very difficult for the **Jaccard distance** (NP-hard)

$$\Delta_J(A,B) = \frac{\big|(A \setminus B) \cup (B \setminus A)\big|}{\big|A \cup B\big|} \ .$$

Case-based Inference: A Probabilistic Approach

Comparison (relative preference):

$$\mathbf{P}(\boldsymbol{y} \succ \boldsymbol{y}') = \frac{\exp\big(-\beta\Delta(\boldsymbol{y},\boldsymbol{y}^*)\big)}{\exp\big(-\beta\Delta(\boldsymbol{y},\boldsymbol{y}^*)\big) + \exp\big(-\beta\Delta(\boldsymbol{y}',\boldsymbol{y}^*)\big)}$$

$$= \frac{1}{1+\exp\Big(-\beta\big(\Delta(\boldsymbol{y}',\boldsymbol{y}^*) - \Delta(\boldsymbol{y},\boldsymbol{y}^*)\big)\Big)}$$

Selection (absolute preference):

$$\mathbf{P}(\boldsymbol{y}) = \frac{\exp\big(-\beta\Delta(\boldsymbol{y},\boldsymbol{y}^*)\big)}{\sum_{\boldsymbol{y}'\in\mathbf{Y}} \exp\big(-\beta\Delta(\boldsymbol{y}',\boldsymbol{y}^*)\big)}$$

$$= \frac{1}{1+\sum_{\boldsymbol{y}'\in\mathbf{Y}\setminus\{\boldsymbol{y}\}} \exp\Big(-\beta\big(\Delta(\boldsymbol{y}',\boldsymbol{y}^*) - \Delta(\boldsymbol{y},\boldsymbol{y}^*)\big)\Big)}$$

→ „classical“ CBR as a special case

Experiments

Experiments with benchmark data from **multi-label classification**.

X1	X2	X3	X4	Y1	Y2	Y3	Y4
0.34	0	10	174	0	0	**1**	0
1.45	0	32	277	0	**1**	0	0
1.22	1	46	421	0	0	0	**1**
0.74	1	25	165	0	**1**	0	0
0.95	1	72	273	**1**	0	0	0
1.04	0	33	158	0	0	**1**	0
0.92	1	81	382	**1**	0	0	0

Experiments

Experiments with benchmark data from **multi-label classification**.

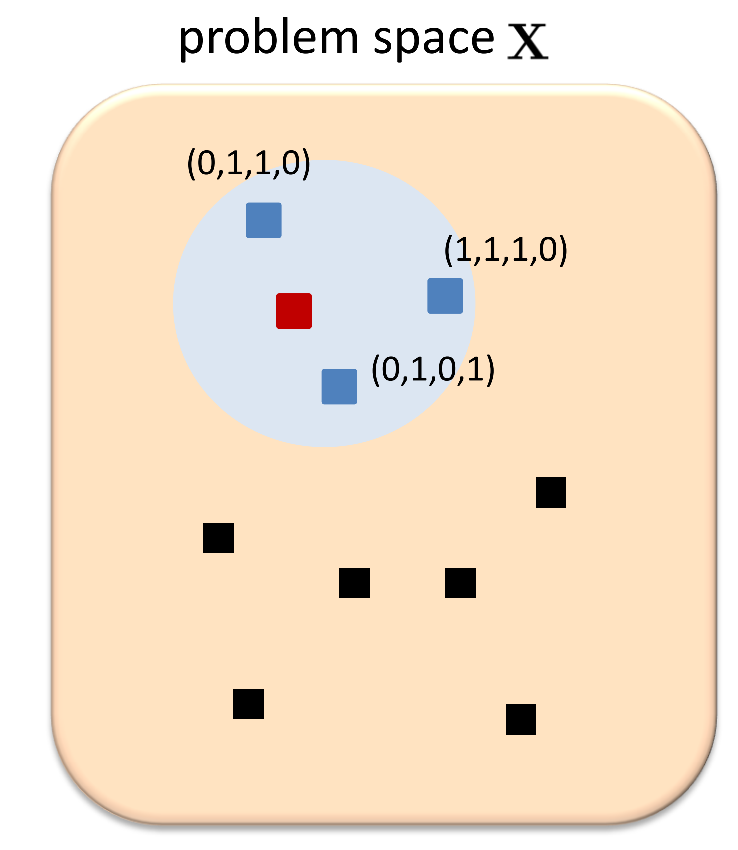
X1	X2	X3	X4	Y1	Y2	Y3	Y4
0.34	0	10	174	0	**1**	**1**	0
1.45	0	32	277	0	**1**	0	**1**
1.22	1	46	421	0	0	0	1
0.74	1	25	165	0	**1**	**1**	**1**
0.95	1	72	273	**1**	0	**1**	0
1.04	0	33	158	**1**	**1**	**1**	0
0.92	1	81	382	**1**	**1**	0	0

Experiments

Experiments with benchmark data from **multi-label classification**.

X1	X2	X3	X4	Y1	Y2	Y3	Y4	subset
0.34	0	10	174	0	**1**	**1**	0	{Y2, Y3}
1.45	0	32	277	0	**1**	0	**1**	{Y2, Y4}
1.22	1	46	421	0	0	0	1	{Y4}
0.74	1	25	165	0	**1**	**1**	**1**	{Y2, Y3, Y4}
0.95	1	72	273	**1**	0	**1**	0	{Y1, Y3}
1.04	0	33	158	**1**	**1**	**1**	0	{Y1, Y2, Y3}
0.92	1	81	382	**1**	**1**	0	0	{Y1, Y2}

Experimental Setup

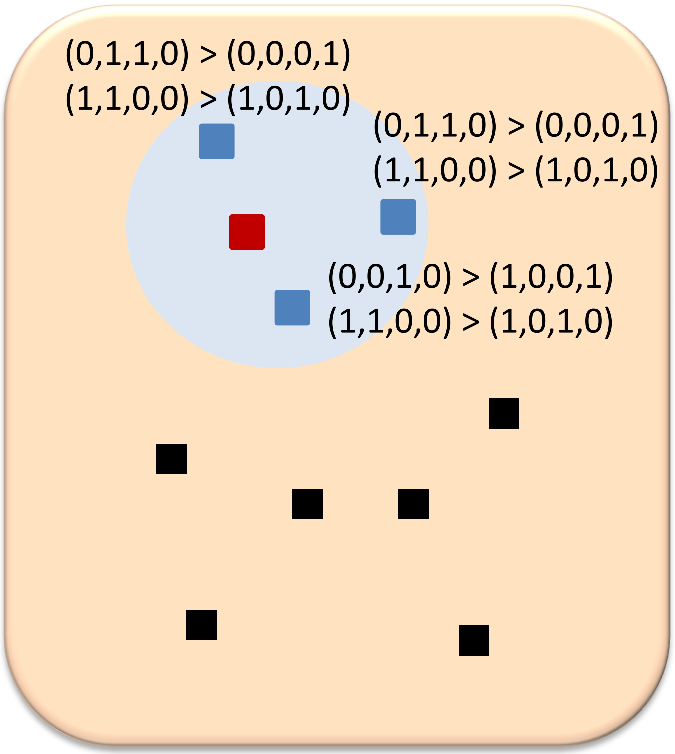


standard NN estimation

– estimation by the median of the neighbor labelings
– corresponds to the „absolute preference“ variant of our method

Experimental Setup

problem space $\mathbf{X}$



preference-based CBR

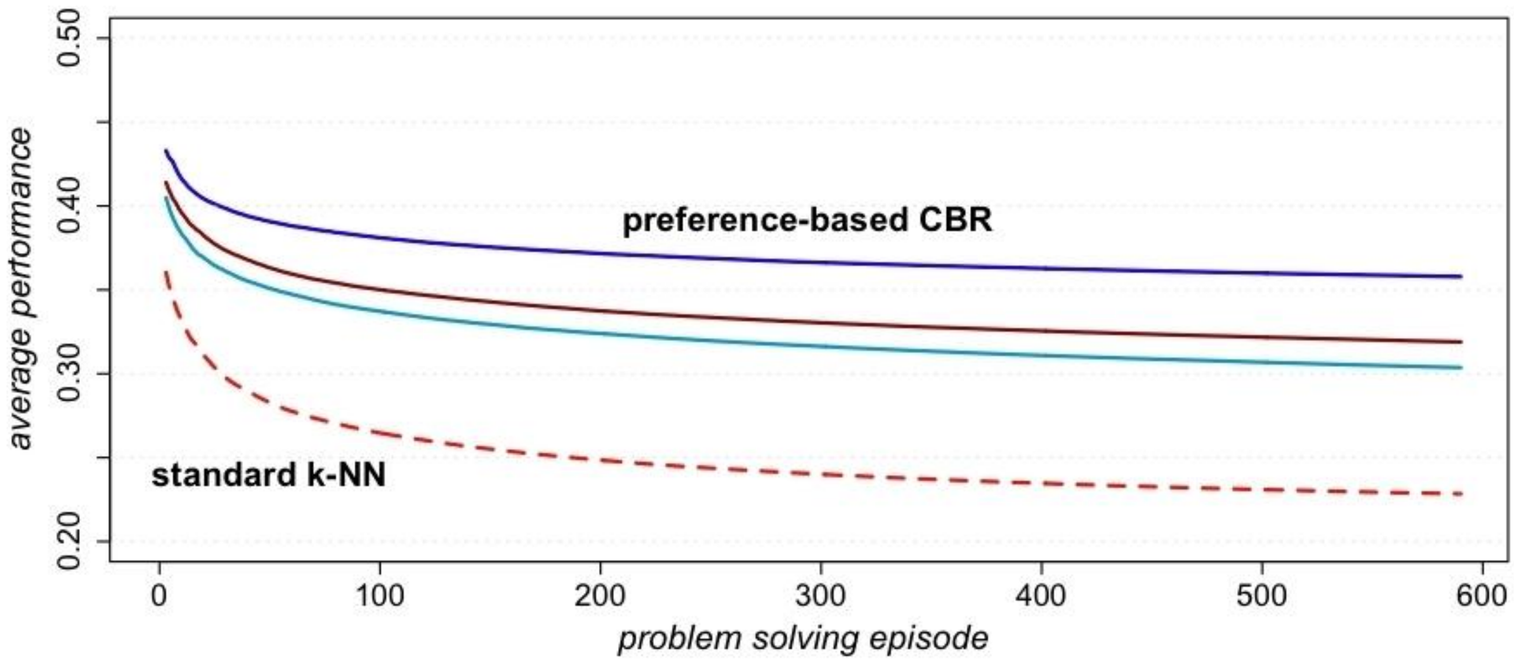
true labelings are replaced by preferences generated according to our probabilistic model

$$\mathbf{P}(\boldsymbol{y} \succ \boldsymbol{y}') = \frac{e^{-\beta\Delta(\boldsymbol{y},\boldsymbol{y}^*)}}{e^{-\beta\Delta(\boldsymbol{y},\boldsymbol{y}^*)} + e^{-\beta\Delta(\boldsymbol{y}',\boldsymbol{y}^*)}}$$

- **indirect supervision**: each preference (indirectly) hints at the true solution
- β controls the **reliability** of this information (level of noise)

Experimental Results (Emotions Data)

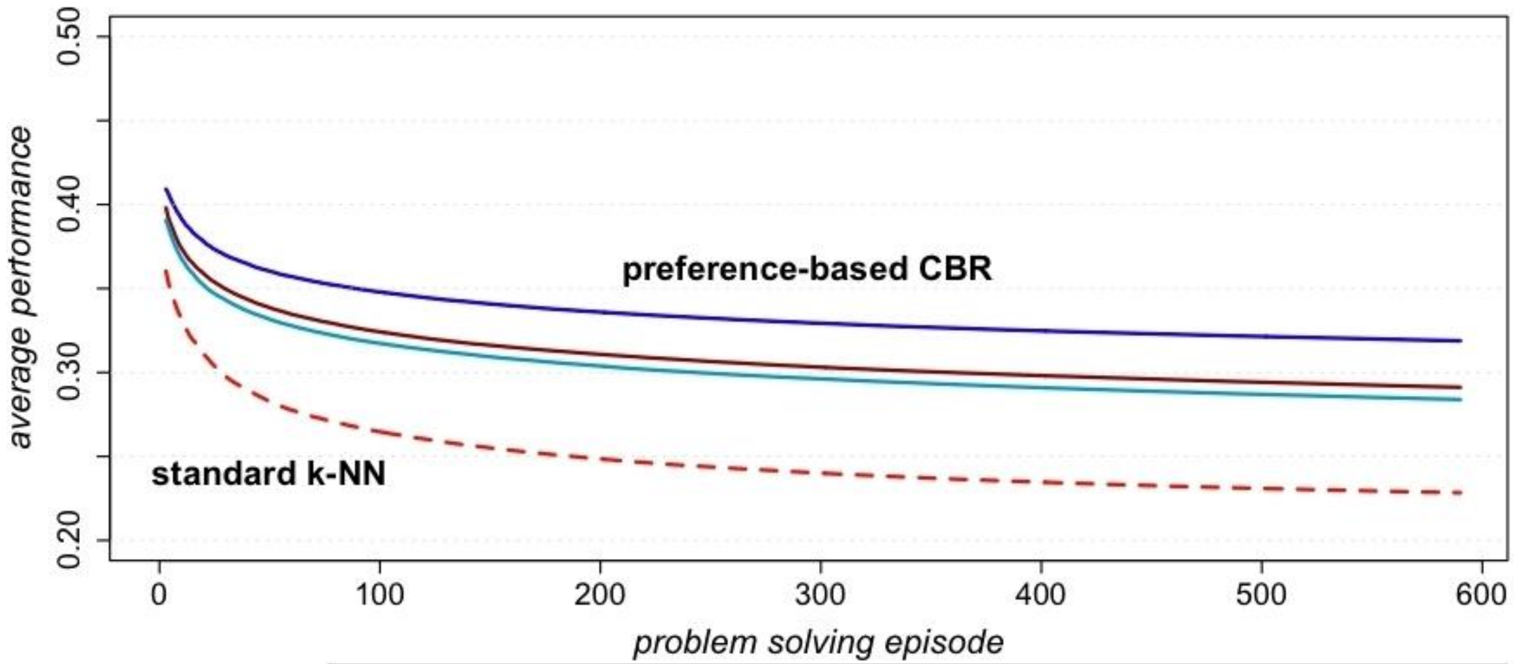
Songs described by 72 features. The task is to predict the emotions that apply: amazed-surprised, happy-pleased, relaxing-calm, quiet-still, sad-lonely and angry-aggressive.



Hamming loss on emotions data: Standard 3-NN (dashed) and preference-based CBR with 7 preferences per case and β = 5, 10, 20.

Experimental Results (Emotions Data)

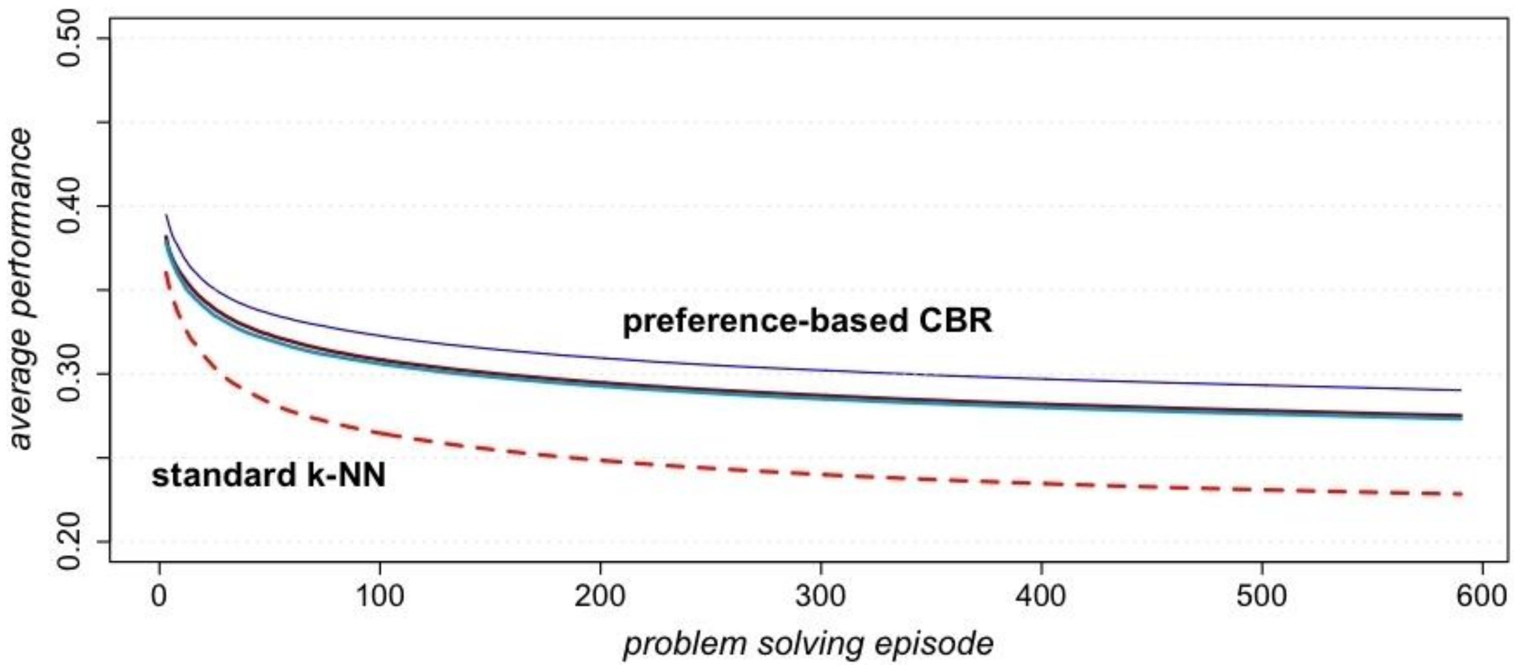
Songs described by 72 features. The task is to predict the emotions that apply: amazed-surprised, happy-pleased, relaxing-calm, quiet-still, sad-lonely and angry-aggressive.



Hamming loss on emotions data: Standard 3-NN (dashed) and preference-based CBR with 15 preferences per case and β = 5, 10, 20.

Experimental Results (Emotions Data)

Songs described by 72 features. The task is to predict the emotions that apply: amazed-surprised, happy-pleased, relaxing-calm, quiet-still, sad-lonely and angry-aggressive.



Hamming loss on emotions data: Standard 3-NN (dashed) and preference-based CBR with 30 preferences per case and β = 5, 10, 20.

Summary & Conclusions

- Our goal is a **methodological framework of preference-based CBR** disposing of a sound theoretical basis while accommodating a wide spectrum of potential applications.
- In this work, our focus was on **case-based inference**, for which we developed a probabilistic method.
- Ideally, a user can easily “parameterize” the framework by choosing the **type of solution space** and the **distance measure** on this space, while the methods themselves are completely generic.
- Our approach still needs to be instantiated for **different types of solution spaces**.
- Besides, **other CBR issues** need to be addressed (case based maintenance, efficient retrieval, etc.)