Preference-Based CBR: A Search-Based Problem Solving Framework

Amira Abdel-Aziz, Weiwei Cheng, Marc Strickert, Eyke Hüllermeier

Computational Intelligence Group
Department of Mathematics and Computer Science
Marburg University, Germany

Philipps Universität Marburg

10/7/2013

Agenda

- **Basic Ideas of Preference-based CBR**
- Case Based Inference (CBI)
- CBR as a Preference-Guided Search
- Case Study

Basic Ideas of Preference-based CBR

- The work done in this paper is a continuation of previous work by (E. Hüllermeier and P. Schlegel, Preference-based CBR: First steps toward a methodological framework-ICCBR 2011).

- Previous work constructed a case-based inference methodology to predict a most plausible candidate solution given a set of preferences on other solutions.

- In this paper we extend previous work by embedding this method in a more general search-based problem solving framework.

Goals and Contributions

A search-based problem solving framework embedding a method for predicting a most plausible candidate solution given a set of preferences on other solutions.

This preference-based framework well accommodates the uncertain and approximate nature of case-based problem solving, by allowing learning from comparing alternative solution pairs instead of providing just a single correct solution at the end.

Conventional CBR

Experience $\longleftrightarrow$ $(x, y) \in \mathbb{X} \times \mathbb{Y}$

x is an element from **problem** space $\mathbb{X}$

y is an element from **solution** space $\mathbb{Y}$

solution y optimally solves problem x

- ❑ Assumes existence of a “correct” solution
- ❑ Potential loss of information
- ❑ Limited guidance in case of failed suggestion

Preference-based Knowledge Representation

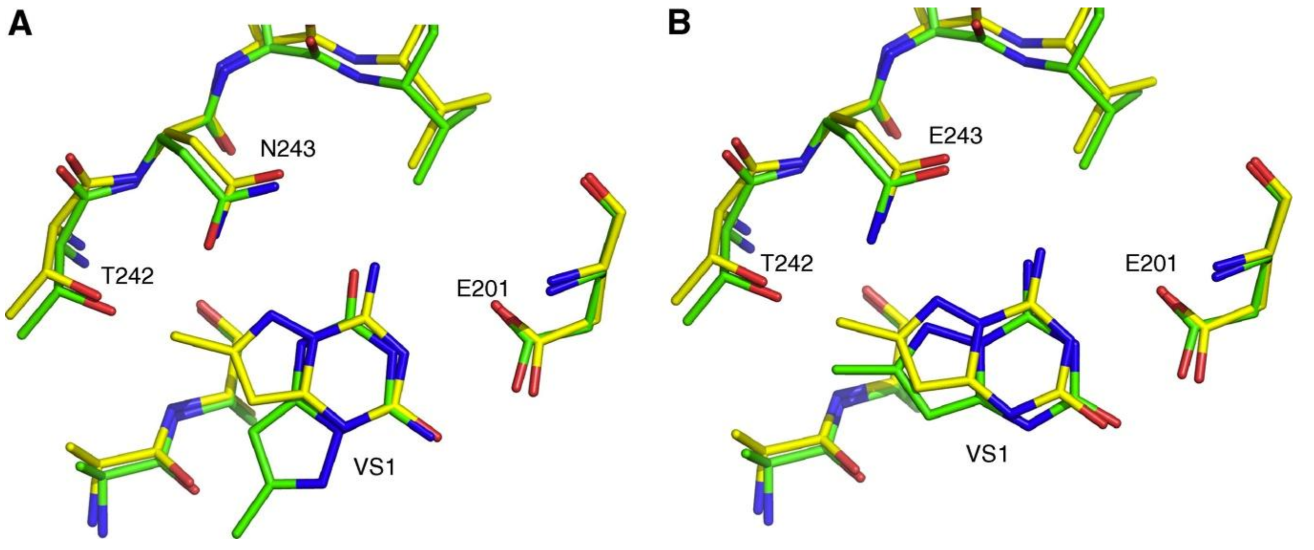
Experience $\longleftrightarrow$ $(x, y \succ z) \in \mathbb{X} \times \mathbb{Y}$

$$y \succ_x z$$

y is more preferred than z as a solution for x

- ❑ It is not required that one of these solutions is optimal
- ❑ No loss of information
- ❑ A ranking of candidate solutions is given for guidance of finding a solution

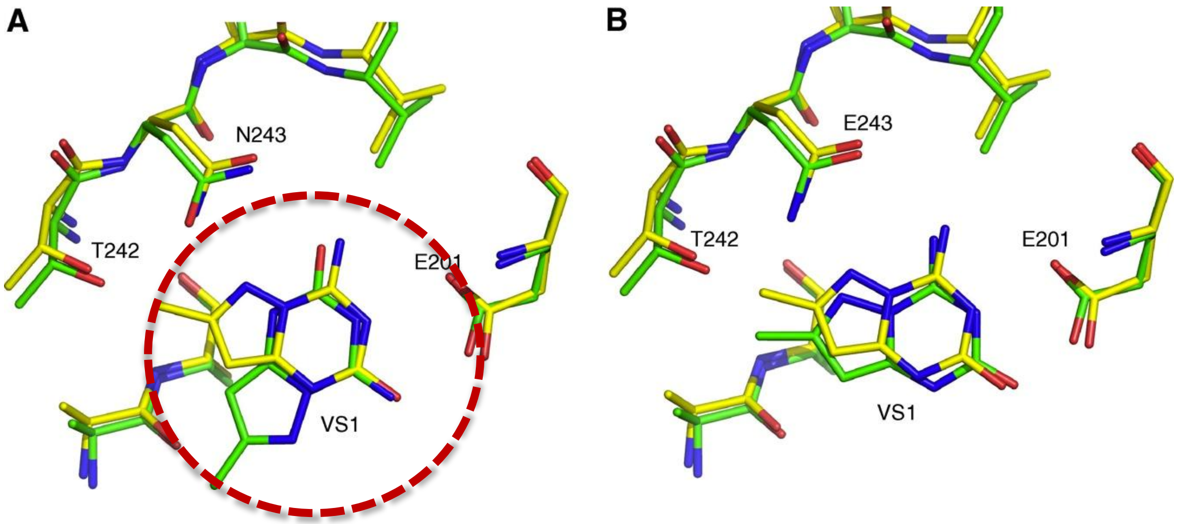
Case-based Representation of Experience



- **Drug discovery:** Finding ligands (small molecules) with high binding affinity to a target protein.

- **CBR perspective**: protein = problem, ligand = solution

Case-based Representation of Experience



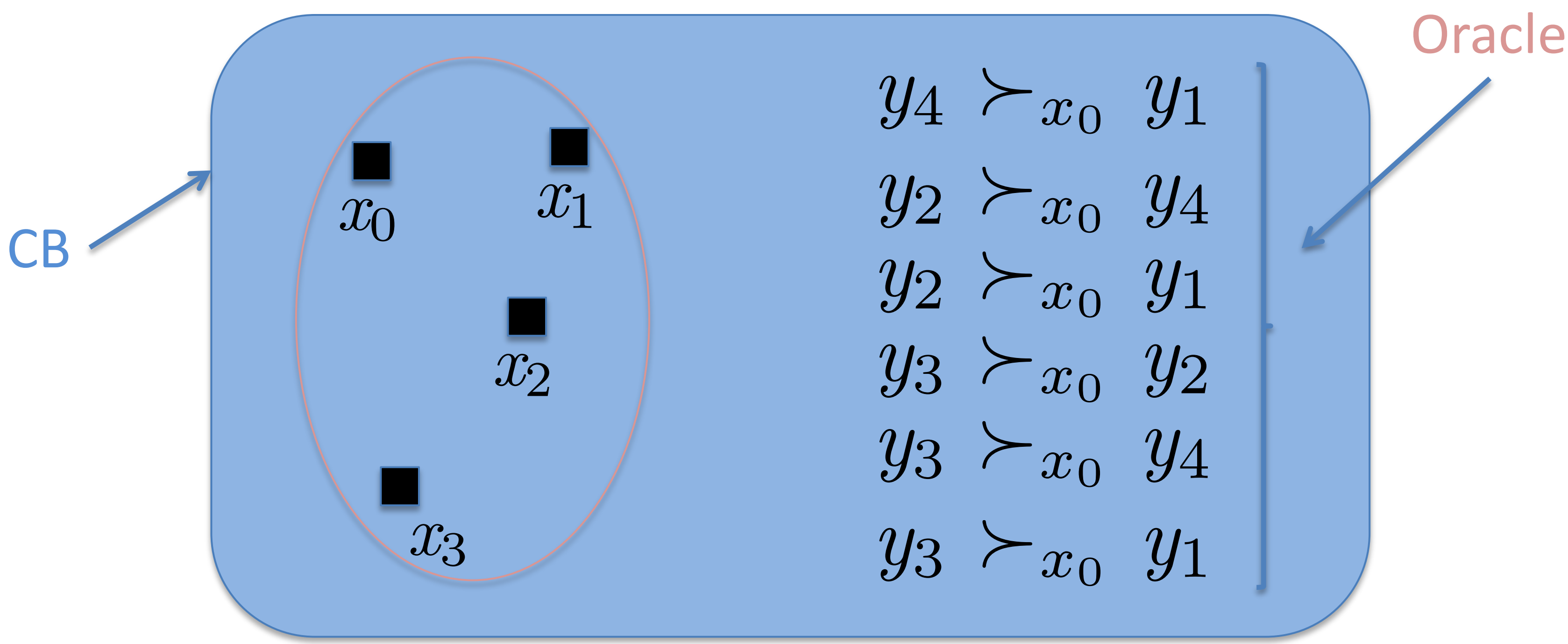
Given a protein as a „problem“, molecule B is preferred as a „solution“ to molecule A.

- Showing two docking poses to a domain expert (chemist, pharmacist), she can easily decide which of the molecules fits better.
- In contrast to this, she will find it difficult to assign a numerical score to an individual molecule.
- Moreover, the notion of „optimality“ is not well defined (the space of molecules is huge and only partly known).

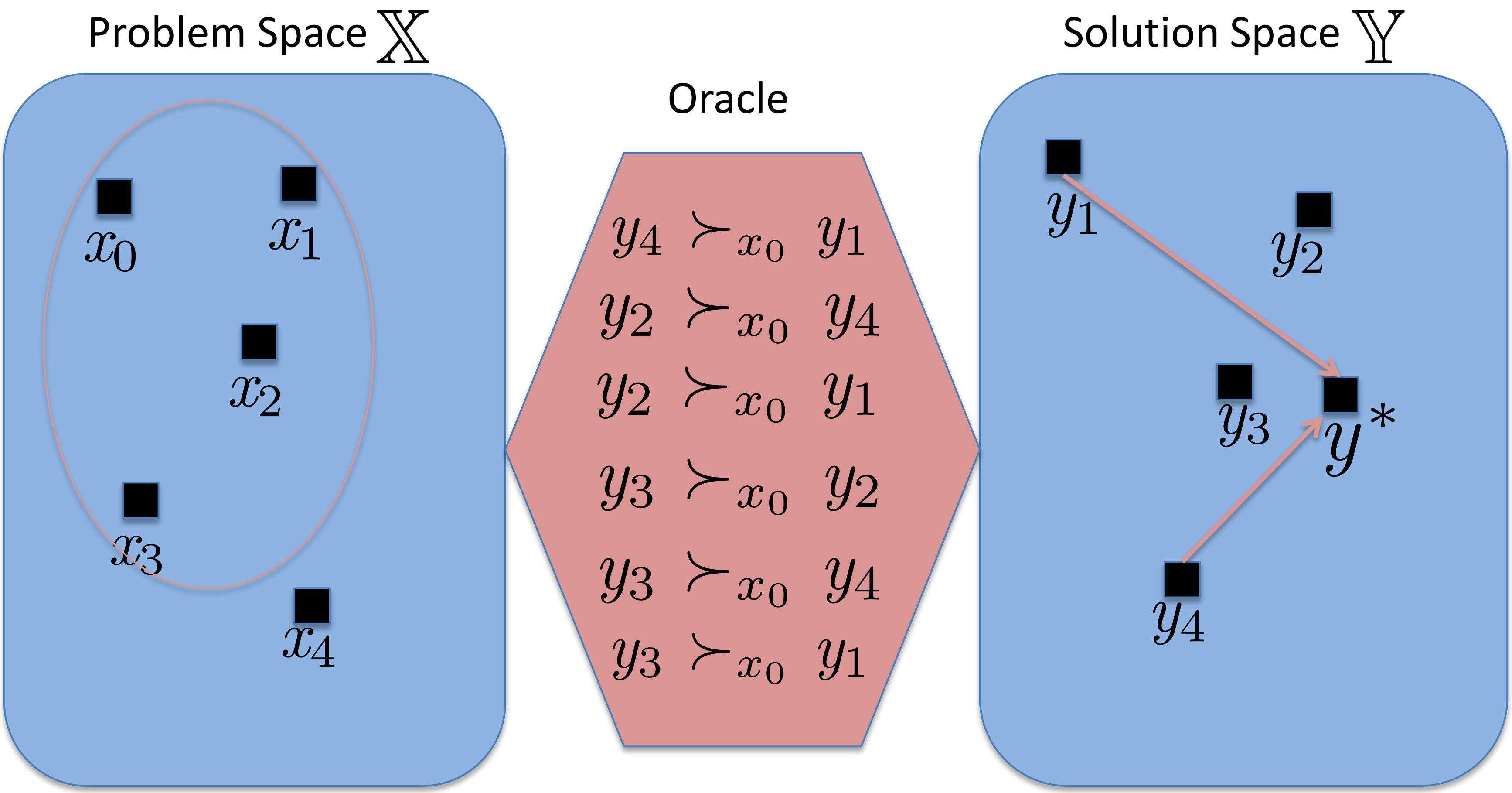
Oracle

Expert providing valid knowledge from which our preferences are created:

- ❑ Expensively computed reference
- ❑ Human expert in a field: pharmacist, doctor, cook, …etc.
- ❑ Expensive computer program



Problem Solving Framework



y^* is the ideal solution for x_0

Assumption: $y_4 \succ_{x_0} y_1$ because $\Delta_Y(y_4, y^*) \leq \Delta_Y(y_1, y^*)$

Agenda

- ❑ Basic Ideas of Preference-based CBR
- ❑ **Case Based Inference (CBI)**
- ❑ CBR as a Preference-Guided Search
- ❑ Case Study

A Discrete Choice Model for Preferences on Solutions

Preferences are created based on the idea that preference of $y \in \mathbb{Y}$

depends on its distance $\Delta_Y(y, y^*) \geq 0$ to an ideal solution y^*

where $\Delta_Y(y, y^*)$ is a "degree of suboptimality" of y

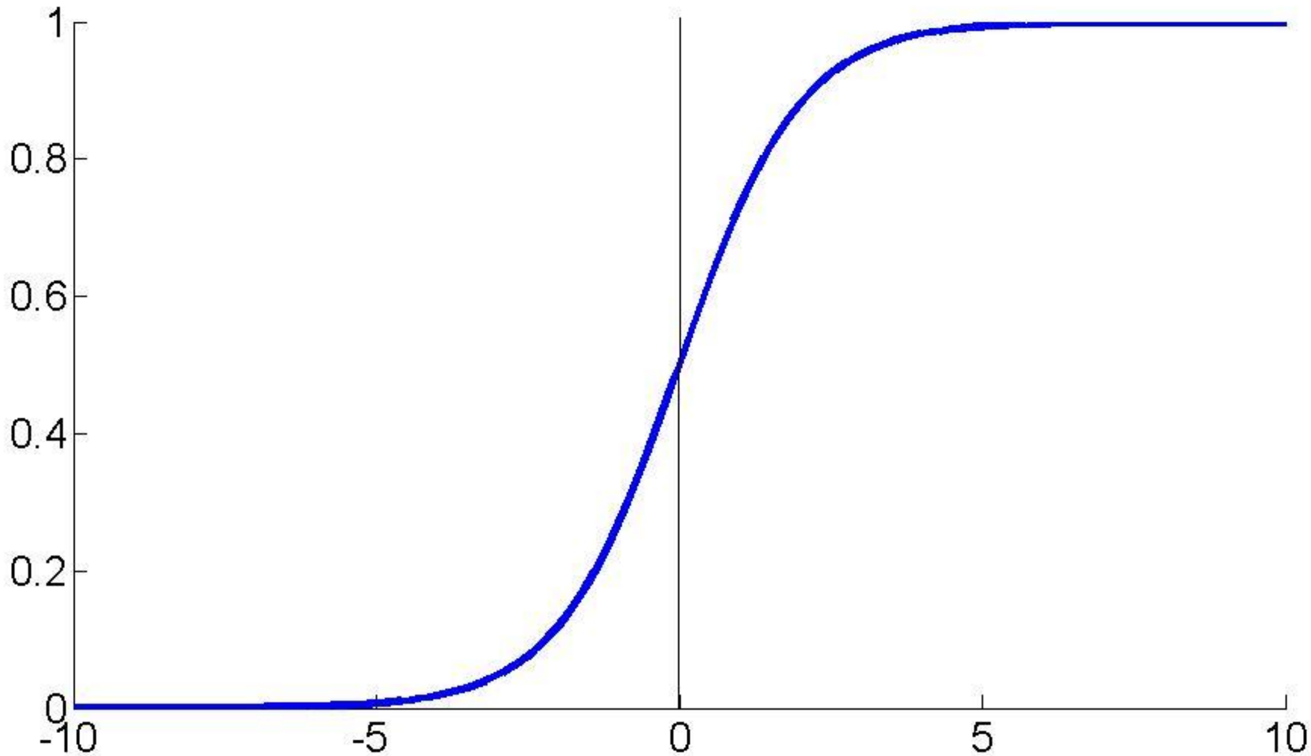
and the probability of observing a preference $y \succ z$

is

$$\mathbf{P}(y \succ z) = \Big(1 + \exp\big(-\beta(\Delta_Y(z, y^*) - \Delta_Y(y, y^*))\big)\Big)^{-1}$$

measure of precision

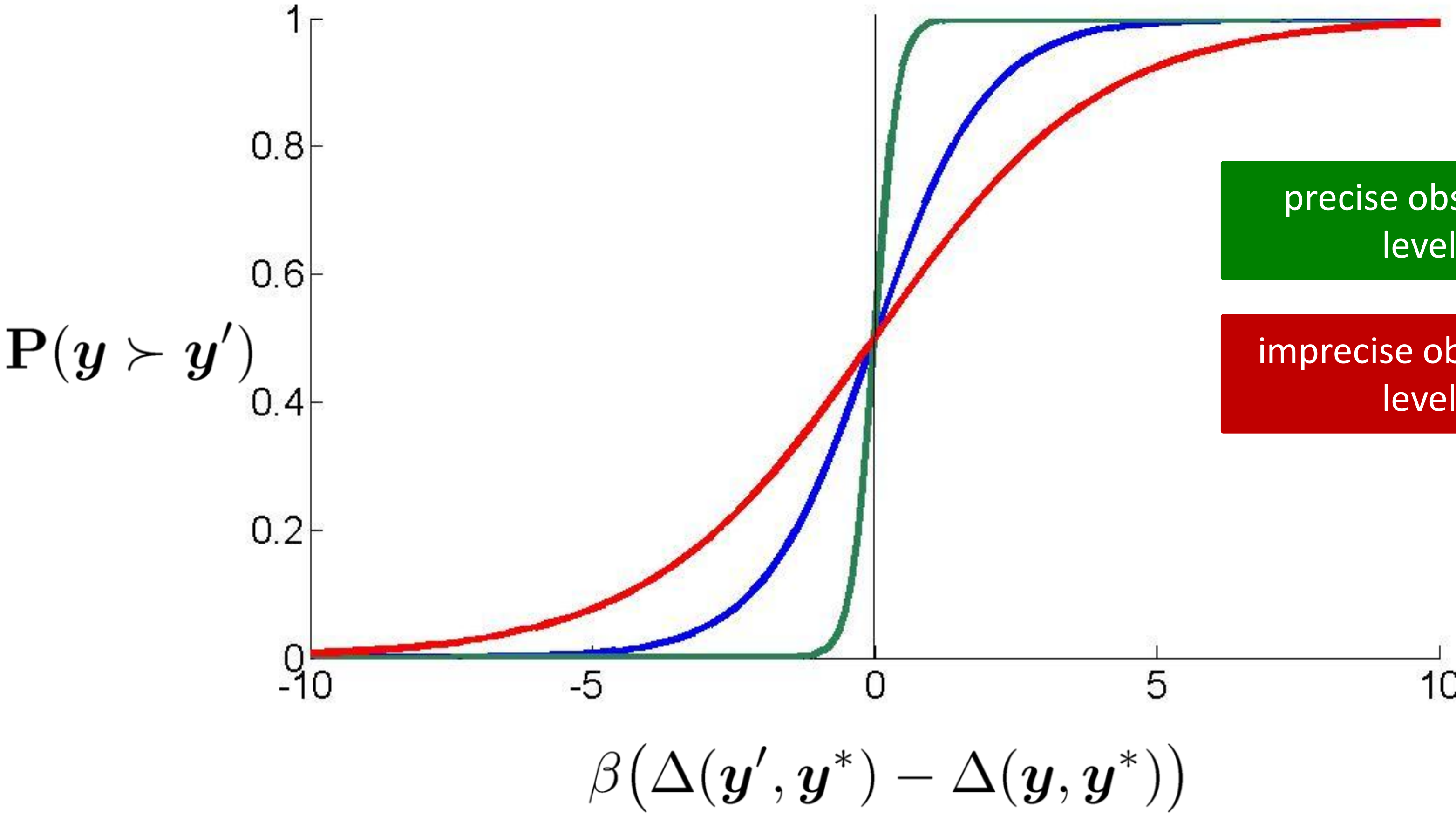
A Discrete Choice Model for Preferences on Solutions



$\mathbf{P}(\boldsymbol{y} \succ \boldsymbol{y}')$

$$\beta\big(\Delta(\boldsymbol{y}', \boldsymbol{y}^*) - \Delta(\boldsymbol{y}, \boldsymbol{y}^*)\big)$$

A Discrete Choice Model for Preferences on Solutions



precise observations, low level of noise

imprecise observations, high level of noise

Case-based Inference (Maximum Likelihood Estimation)

Given set of observed preferences

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

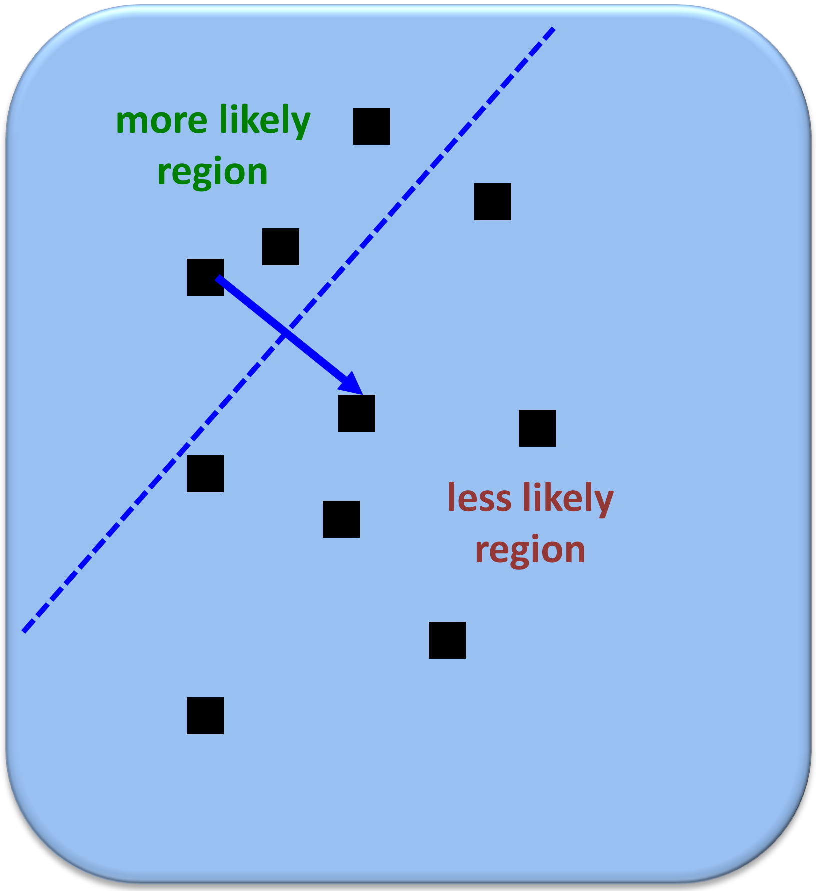
assumed to be representative of current problem x

what is the most plausible "ideal" solution for x

among a given set of candidates $\mathbb{Y}_0 \subset \mathbb{Y}$?

Case-based Inference (Maximum Likelihood Estimation)

Solution Space $\mathbb{Y}$



Each pairwise preference provides a hint at the ideal solution!

Case-based Inference (Maximum Likelihood Estimation)

To estimate parameter vector $\theta^* = (y^*,\ \beta^*) \in \mathbb{Y} \times \mathbb{R}_+$

the log-likelihood of $\theta = (y, \beta)$ is given by

$$\ell(\theta) = \ell(y, \beta) = -\sum_{i=1}^{N} \log\left(1 + \exp\left(-\beta(\Delta(z^{(i)}, y) - \Delta(y^{(i)}, y))\right)\right)$$

CBI (Case Based Inference) equation

The maximum likelihood estimation $\theta_{ML} = (y^{ML}, \beta^{ML})$ of θ^*

is given by

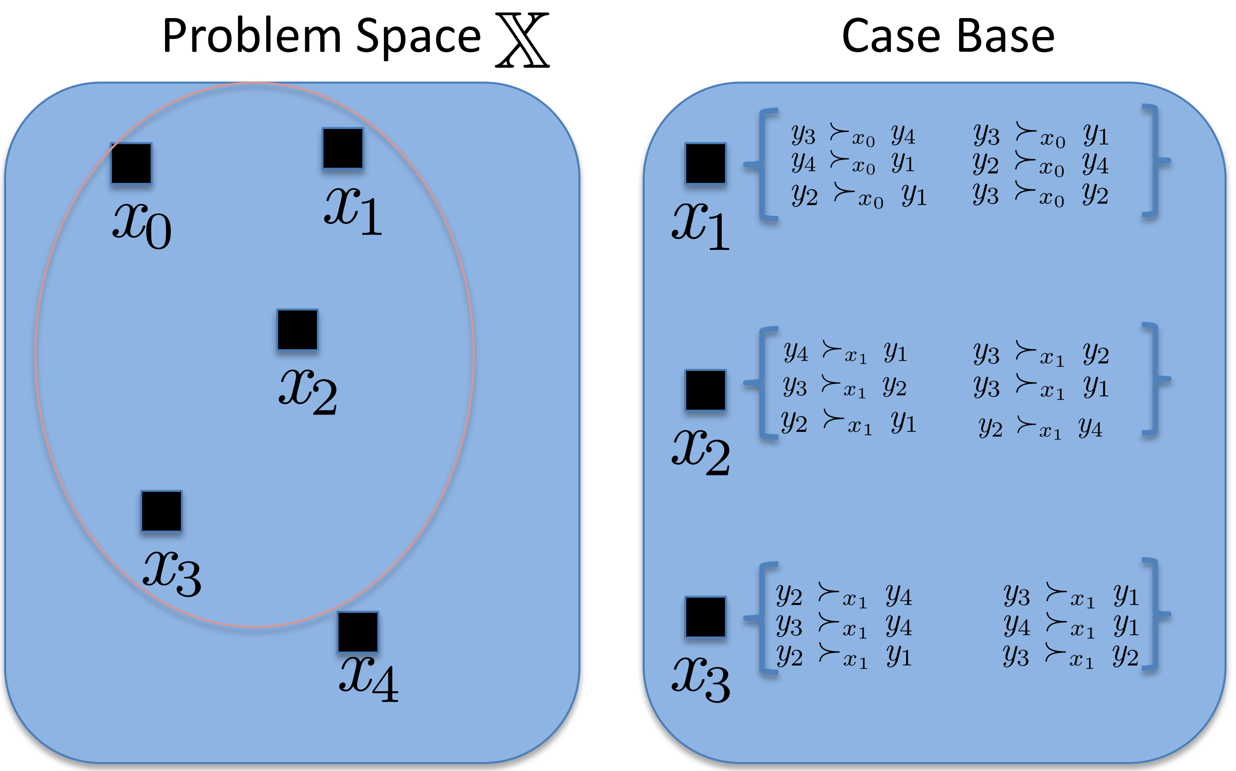
$$\theta_{ML} = (y^{ML}, \beta^{ML}) = \mathrm{argmax}_{y \in \mathbb{Y}, \beta \in \mathbb{R}_+}\ \ell(y, \beta)$$

Agenda

- ❑ Basic Ideas of Preference-based CBR
- ❑ Case Based Inference (CBI)
- ❑ **CBR as a Preference-Guided Search**
- ❑ Case Study

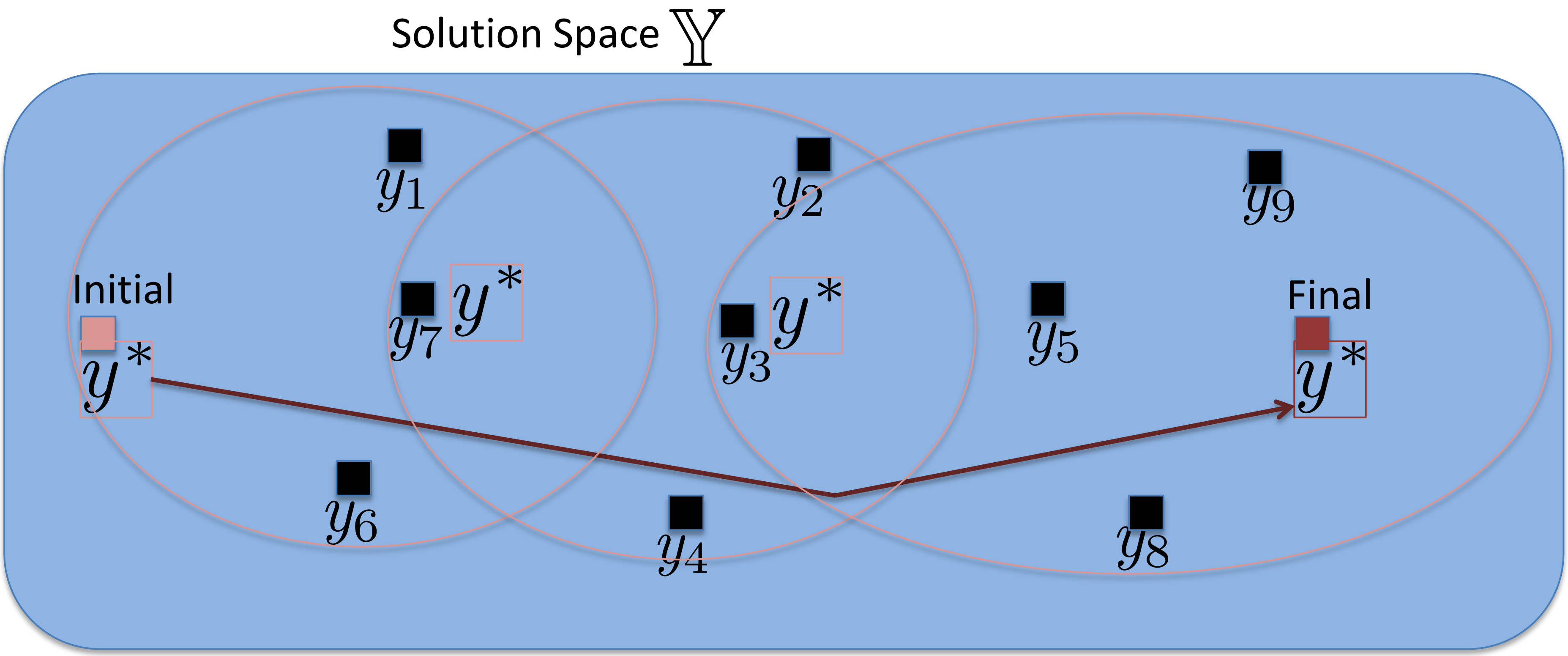
CBR as a Preference-Guided Search

Given a new problem query x_0 retrieve the K nearest neighbors of this problem, those with the smallest Δ_X from x_0



The preferences of the nearest neighbors are collected from the CB and are used to guide the search process.

CBR as a Preference-Guided Search



The search for a solution starts with an initial candidate $y^* \in \mathbb{Y}$ calculated by

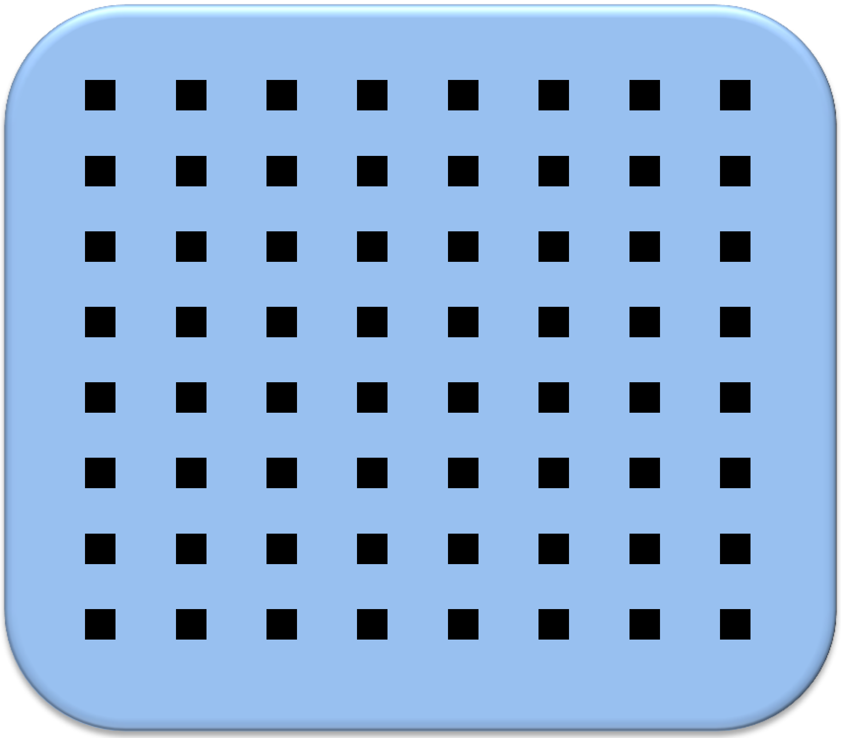
$$y^* \leftarrow \mathrm{CBI}(\mathcal{P},\ \mathbb{Y}_0)$$

Then the solution is iterated L times to reach final $\boxed{y^*}$

L=number of queries to oracle

CBR as a Preference-Guided Search

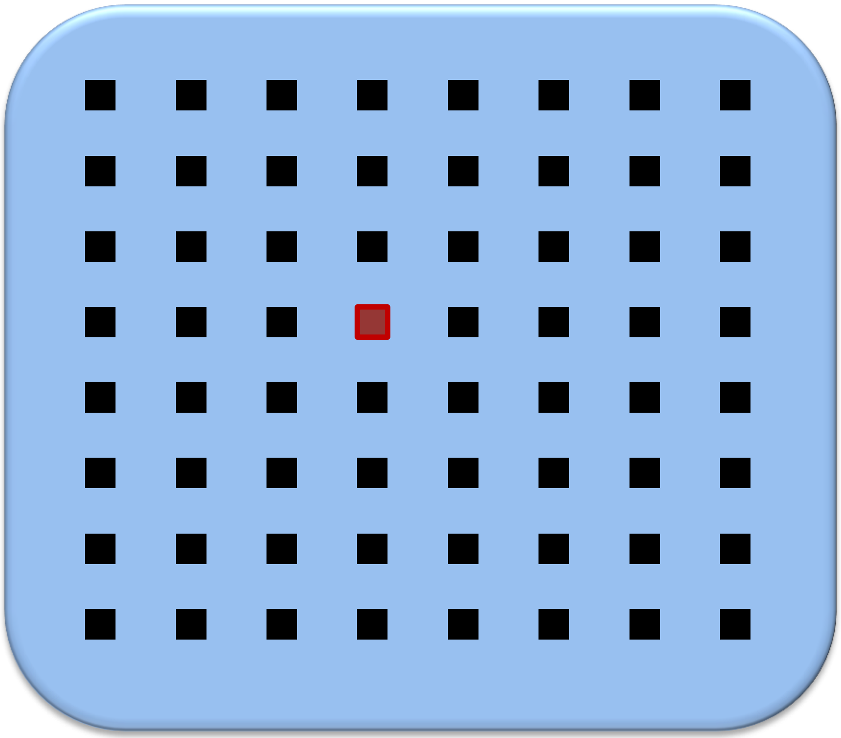
solution space



Case-based problem solving as (heuristic) search in the solution space.

CBR as a Preference-Guided Search

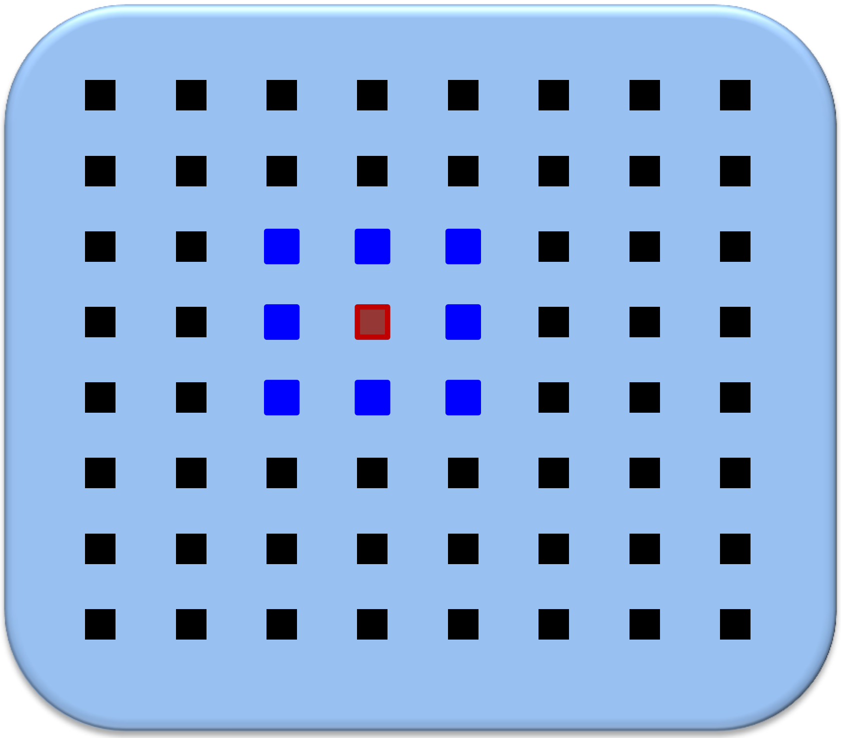
solution space



- Start with an **initial solution**

Case-based problem solving as (heuristic) search in the solution space.

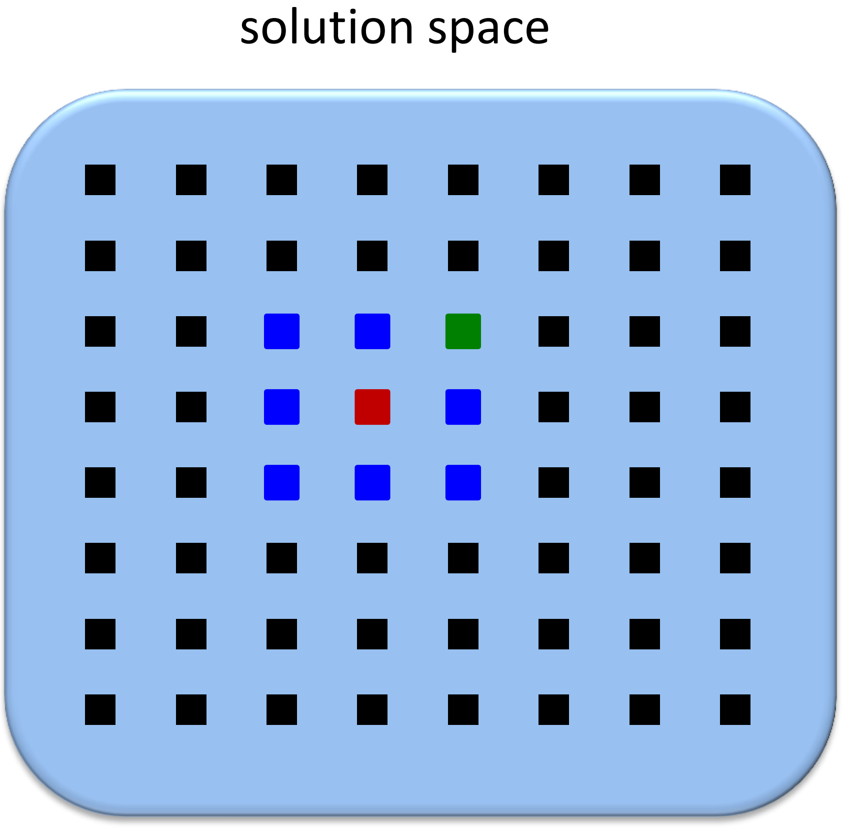
CBR as a Preference-Guided Search



- Start with an initial solution
- Consider the **neighbors** of the current solutions as new candidates.

Case-based problem solving as (heuristic) search in the solution space.

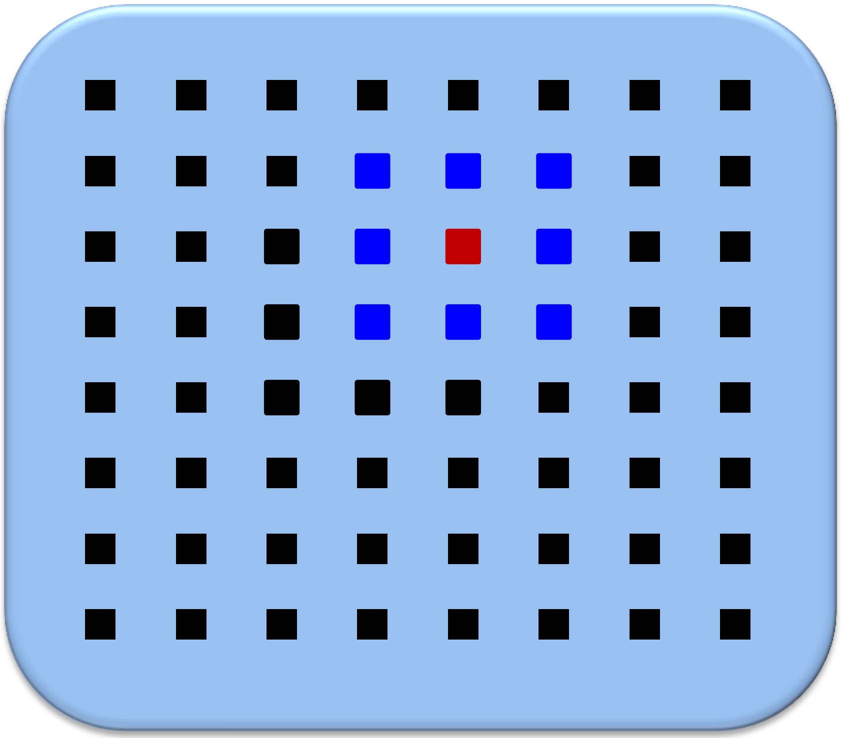
CBR as a Preference-Guided Search



- Start with an initial solution
- Consider the neighbors of the current solutions as new candidates.
- Select a **promising neighbor**, compare with **current solution** and adopt the better one.

Case-based problem solving as (heuristic) search in the solution space.

CBR as a Preference-Guided Search

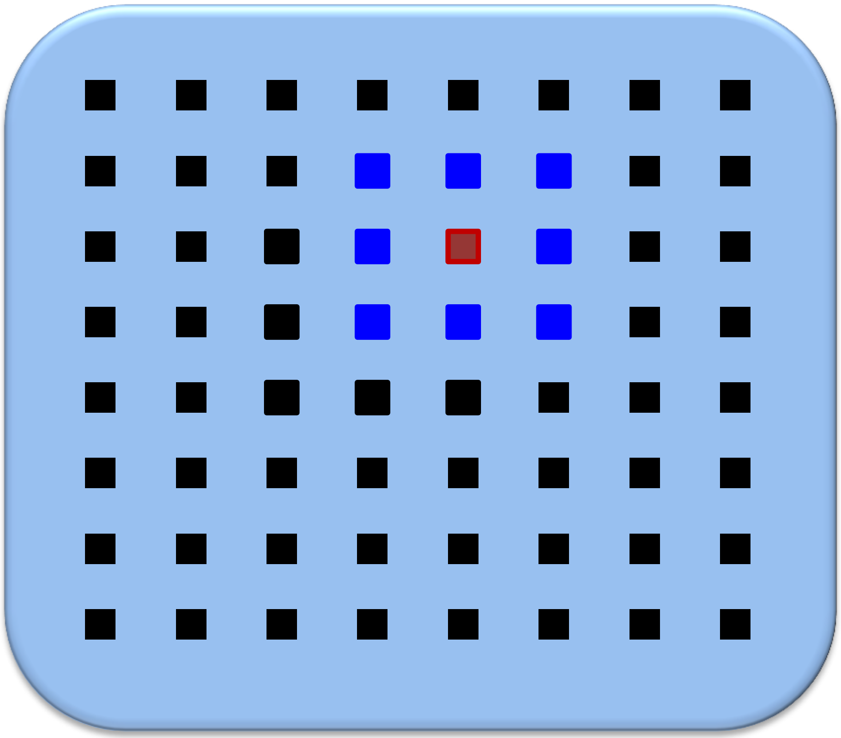


- Start with an initial solution
- Consider the neighbors of the current solutions as new candidates.
- Select a promising neighbor, compare with current solution and adopt the better one.
- Repeat till no further improvement or maximum number of iterations reached.

Case-based problem solving as (heuristic) search in the solution space.

CBR as a Preference-Guided Search

solution space



- **Start with an initial solution**
- Consider the neighbors of the current solutions as new candidates.
- **Select a promising neighbor**, compare with current solution and adopt the better one.
- Repeat till no further improvement or maximum number of iterations reached.

We seek to support these steps through case-based experience!

CBR as a Preference-Guided Search

x1	y12 > y72	y42 > y41	y76 > y21	y42 > y72
x2	y05 > y53	y92 > y43	y32 > y56	y65 > y84
x3	y39 > y37	y33 > y67	y65 > y76	y76 > y37
x4	y72 > y98	y47 > y27	y34 > y34	y76 > y65
x5	y39 > y49	y29 > y81	y32 > y26	y76 > y11
x6	y46 > y11	y46 > y28	y68 > y28	y22 > y42

Problems are stored together with observed pairwise preferences.

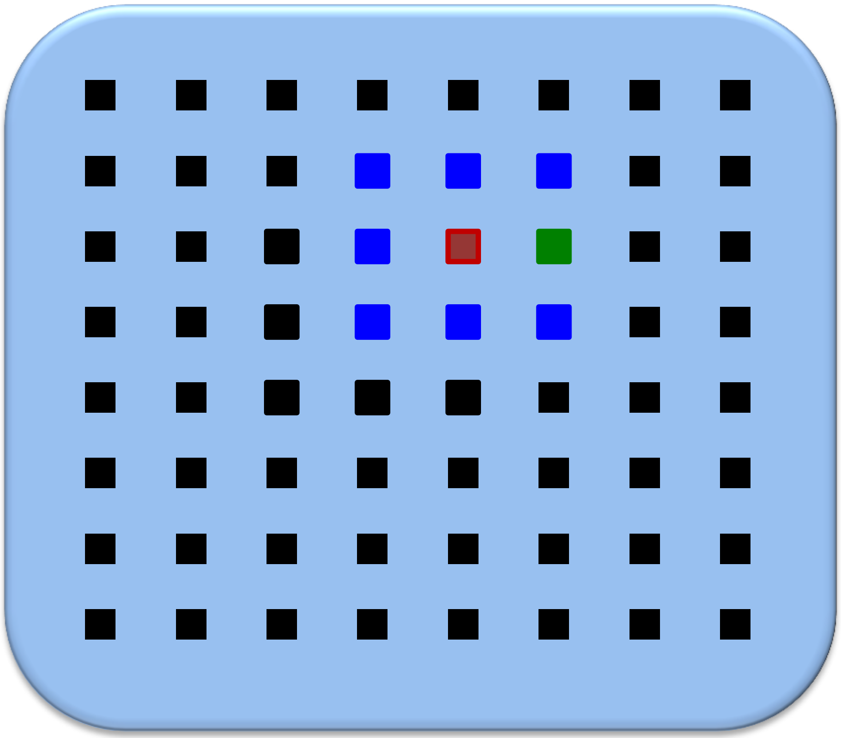
CBR as a Preference-Guided Search

x1	y12 > y72	y42 > y41	y76 > y21	y42 > y72
x2	y05 > y53	y92 > y43	y32 > y56	y65 > y84
x3	y39 > y37	y33 > y67	y65 > y76	y76 > y37
x4	y72 > y98	y47 > y27	y34 > y34	y76 > y65
x5	y39 > y49	y29 > y81	y32 > y26	y76 > y11
x6	y46 > y11	y46 > y28	y68 > y28	y22 > y42

- Given an new problem, find the **nearest neighbors** in the case base and collect the associated preferences into an initial preference set.
- The **initial solution** is then found by applying CBI to this set of preferences (with the complete solution space as candidates).

CBR as a Preference-Guided Search

solution space

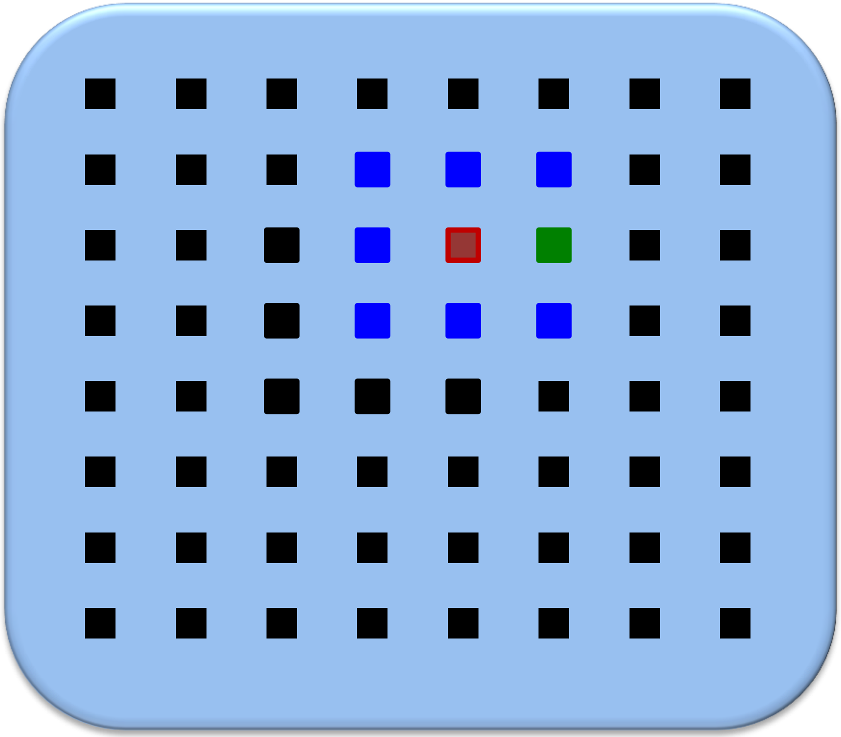


- In each iteration, CBI is applied to the **neighbors of the current solution** to find the **most promising candidate**.

We seek to support these steps through case-based experience!

CBR as a Preference-Guided Search

solution space

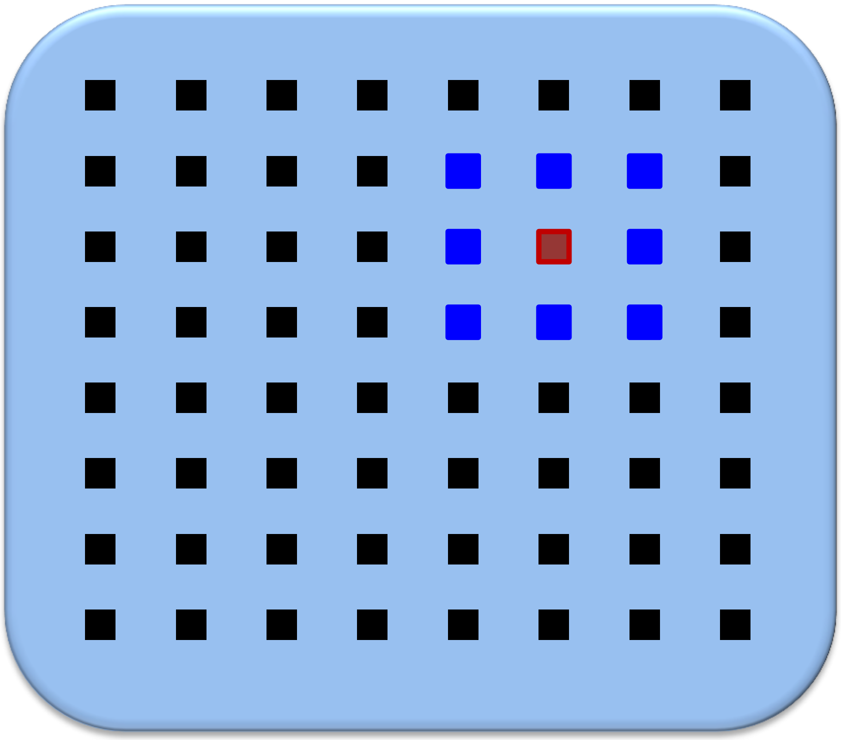


- In each iteration, CBI is applied to the **neighbors of the current solution** to find the **most promising candidate**.
- The two solutions are compared, the better one is adopted, and the **new preference is added to preference set.**

We seek to support these steps through case-based experience!

CBR as a Preference-Guided Search

solution space



- In each iteration, CBI is applied to the **neighbors of the current solution** to find the **most promising candidate**.
- The two solutions are compared, the better one is adopted, and the **new preference is added to preference set.**
- The process stops after a predefined number of iterations, and the current best solution is returned.

We seek to support these steps through case-based experience!

CBR as a Preference-Guided Search

x1	y12 > y72	y42 > y41	y76 > y21	y42 > y72
x2	y05 > y53	y92 > y43	y32 > y56	y65 > y84
x3	y39 > y37	y33 > y67	y65 > y76	y76 > y37
x4	y72 > y98	y47 > y27	y34 > y34	y76 > y65
x5	y39 > y49	y29 > y81	y32 > y26	y76 > y11
x6	y46 > y11	y46 > y28	y68 > y28	y22 > y42
x7	Y62 > y22	Y62 > y81	Y71 > y62	Y77 > y71

The new problem is stored together with the pairwise preferences collected during the problem solving process.

CBR as a Preference-Guided Search

When the search stops, the current best solution

$\boxed{y^*}$ is returned and the case $(x_0, \mathcal{P}_0)$

is added to the case base

problem

set of preferences

Case Base

x_0: $\left\{ \begin{array}{ll} y_3 \succ_{x_0} y_4 & y_3 \succ_{x_0} y_1 \\ y_4 \succ_{x_0} y_1 & y_2 \succ_{x_0} y_4 \\ y_2 \succ_{x_0} y_1 & y_3 \succ_{x_0} y_2 \end{array} \right\}$

x_1: $\left\{ \begin{array}{ll} y_3 \succ_{x_0} y_4 & y_3 \succ_{x_0} y_1 \\ y_4 \succ_{x_0} y_1 & y_2 \succ_{x_0} y_4 \\ y_2 \succ_{x_0} y_1 & y_3 \succ_{x_0} y_2 \end{array} \right\}$

x_2: $\left\{ \begin{array}{ll} y_3 \succ_{x_0} y_4 & y_3 \succ_{x_0} y_1 \\ y_4 \succ_{x_0} y_1 & y_2 \succ_{x_0} y_4 \\ y_2 \succ_{x_0} y_1 & y_3 \succ_{x_0} y_2 \end{array} \right\}$

...

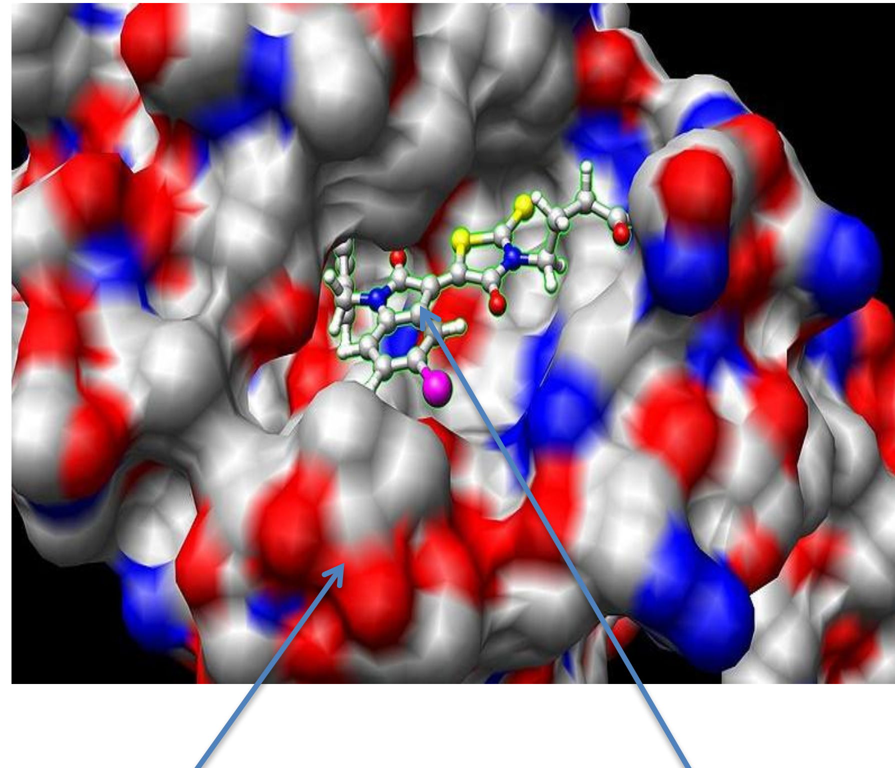
Agenda

- Basic Ideas of Preference-based CBR
- Case Based Inference (CBI)
- CBR as a Preference-Guided Search
- **Case Study**

Applications: Drug Discovery Case Study

❑ Ligand molecules bind to protein surface, thereby blocking or enhancing its biochemical activity

❑ Identification and selection of ligands targeting a specific protein is of high interest for drug development



Experiment:
The search for a ligand that well interacts with a target protein

Drug Discovery Case Study

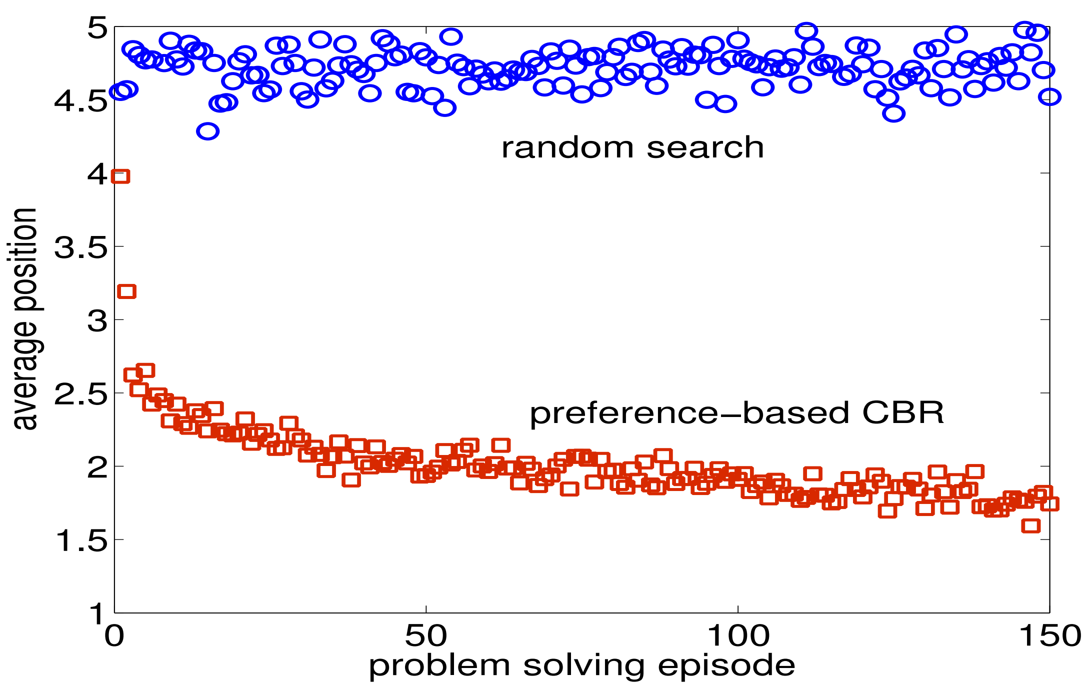
Problem Space $\mathbb{X}$ → *Proteins*

Solution Space $\mathbb{Y}$ → *Ligands*

Δ_X → *Measure which compares proteins in terms of spatial and physicochemical properties of their respective binding sites*

Δ_Y → *Tanimoto similarity of molecules to determine similarity between ligands*

38 ligands



588 proteins

Conclusion

- The preference-based CBR framework presented, is based on representing experience in the form of contextualized **preferences.**

- **Case-based inference** is formalized by means of a probabilistic approach.

- These preferences are used to direct the **problem solving process** that is formalized as a heuristic search process.

- Promising results in two case studies (**set completion** and **drug discovery)**.

Future Work

- Apply the idea of using preferences for guiding the search process to more **sophisticated search methods**.

- Develop effective methods for **case base maintenance**, as number of preferences collected in the course of time may become large.

- **Learning similarity measures** to increase efficiency of the preference-guided search procedure.