Lecture 14: Case-Based Reasoning

Dr. Roman V Belavkin

BIS3226

Contents

1 Case-Based Reasoning

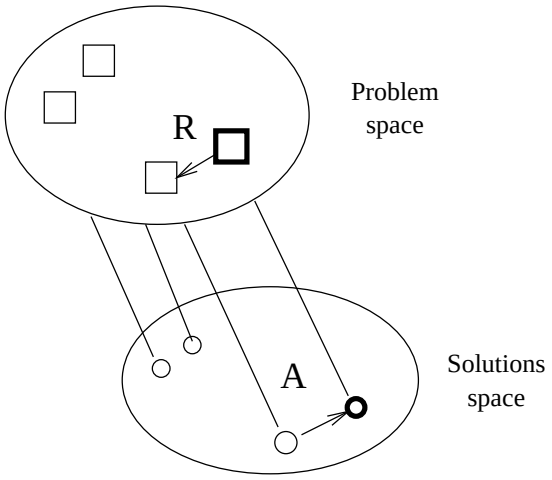
Case-Based Reasoning Systems

- Instead of facts and rules, CBR systems use *cases* and their entire *solutions*.

Case:	Age	Gender	M. Income (£ K)	M. Expenses (£ K)	Home owner	Credit score
1	21	0	2	1	0	3
2	18	1	1	2	0	1
3	50	1	6	2	1	5
4	23	0	3	1	1	4
5	40	1	3	2	0	2

- Cases can have quite complex descriptions using symbolic and numerical values
- CBR is based on the concepts of *similarity* and *analogy*. Similarity is used to find similar cases, and analogy is used to find solutions for similar cases.

Problem (Case) and Solution Spaces



R : retrieval of a similar case from problem space

A : adaptation of a solution from solution space

Metric as a Measure of Similarity

- Operation of CBR systems is based on similarity between the cases and similarity between the solutions.
- The similarities can be represented by metrics d_c and d_s on the spaces of cases and solutions:

$$\begin{aligned} \mathbf{v} &= (21, 0, 2, 1, 0) \\ \mathbf{w} &= (18, 1, 1, 2, 0) \\ \mathbf{x} &= (50, 1, 6, 2, 1) \\ \mathbf{y} &= (23, 0, 3, 1, 1) \\ \mathbf{z} &= (40, 1, 3, 2, 0) \end{aligned}$$

$$d_c(\mathbf{z}, \mathbf{x})$$

$$\begin{aligned} \mathbf{v} &= (3) \\ \mathbf{w} &= (1) \\ \mathbf{x} &= (5) \\ \mathbf{y} &= (4) \\ \mathbf{z} &= (2) \end{aligned}$$

$$d_s(\mathbf{z}, \mathbf{x})$$

Remark 1. *The choice of metric is important, because generally different metrics will produce different results.*

2 Operation of CBR Systems

Operation Cycle of CBR

In general, can be described using four **RE**s:

- **RE**trieve the most similar case or cases.
- **RE**use the case(s) to attempt to solve the problem.
- **RE**vise the proposed solution if necessary.
- **RE**tain the new solution as a part of a new case.

Retrieval and Reuse

- To solve a new case, a CBR systems **retrieves** an old similar case. There are two main methods to retrieve similar cases

  Nearest-neighbour : is based on comparing the cases using some distance (e.g. Euclidean distance)

  $$d(a,b) = \sqrt{(a_1 - b_1)^2 + \cdots + (a_m - b_m)^2}$$

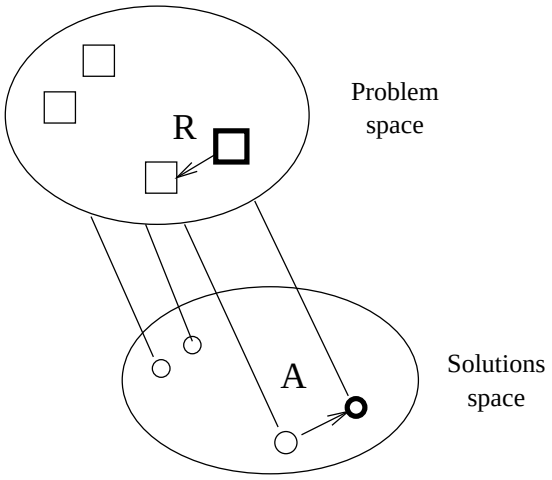
  where $a = (a_1, \ldots, a_m)$ and $b = (b_1, \ldots, b_m)$ are two cases.

  Inductive retrieval : is based on learning which feature of the case carries the most useful information to predict the solution (highest information gain).

- The solution of a retrieved case is **reused** to for the new case.

Revise and Retain

- Because the new case is most likely different from all the old cases, the solution of a retrieved case can be **revised** or **adapted**.
- If the solution was a success, then the new case is **retained** in the database together with the new solution.



3 Discussion, Examples and Applications

Rule-Based vs Case-Based

- In rule-based systems a solution is achieved through an application of many rules, inference of facts, etc.
- In case-based systems the whole problem definition (case) is compared with similar problems, and the entire solution is applied at once.
- Rule-based systems usually work better for well-defined problems that do not change with time.
- Case-based systems can be used where problems are less understood and are dynamic.

Examples of CBR Systems

- CASPIAN publicly available CBR shell built at Aberystwyth:

  `http://www.aber.ac.uk/ dcswww/Research/mbsg/cbrprojects/`

- myCBR:

  `http://mycbr-project.net/`

- Other CBR shells and tools can be found here:

  `http://cbrwiki.fdi.ucm.es/wiki/index.php/Tools`