CS.760

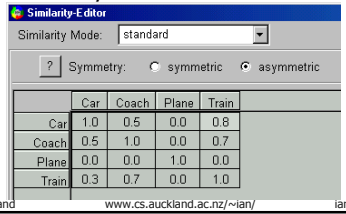
Case-Based Reasoning 3
Dr. Ian Watson

Contents

- Similarity
  - We'll use CBR-Works as an exemplar www.tecinno.com
    - Symbols
    - Numbers
    - Intervals
    - Sets
    - Ordered sets
    - Strings
    - Taxonomy

Similarity

- Symbols (unordered)
  - Similarity defined by developer
  - Similarity values stored in a decision table



	Car	Coach	Plane	Train
Car	1.0	0.5	0.0	0.8
Coach	0.5	1.0	0.0	0.7
Plane	0.0	0.0	1.0	0.0
Train	0.3	0.7	0.0	1.0

Similarity

- Symmetric vs asymmetric similarity
  - In symmetric similarity the result is independent of the role of the values being compared
    - Sim(Car, Train) = 0.8
    - Sim(Train,Car) = 0.8
  - In asymmetric similarity the role is important
    - Sim(Car,Train) = 0.8
    - Sim(Train,Car) = 0.3

Similarity

- Numbers
  - Step function
    - Computes a *sim* of 1 if the difference is < step point otherwise a *sim* of 0
  - Polynomial function
    - Computes a *sim* between 0 & 1 using a polynomial curve (ie the closer to the query the greater the sim)
  - Function with a smooth step point
    - *sim* will be in (0.0; 1.0) in respect to the distance between query- and case-value.

Similarity

- Perfect options
  - *Greater is perfect*
    - More horsepower
    - More days holiday
  - *Less is perfect*
    - Less money
    - Less time

Similarity

- Intervals
  - if the intervals in query and case do not intersect the similarity is higher the closer the gap
  - if the intervals intersect the similarity is higher the closer the bounds
  - if the case completely covers the query the
  - similarity is 1
  - if the query completely covers the case the similarity is higher the closer the bounds

Similarity

- Sets – three similarity methods

Similarity	Description
intersection	The similarity computes to the proportion of the intersection of two SymbolSets to its union.
case-inclusion	The similarity equals to **1**, if the case is included in the query. Entries included in case but not in query lower the similarity.
query-inclusion	The similarity equals to **1**, if the query is included in the case. Entries included in query but not in case lower the similarity.

Similarity

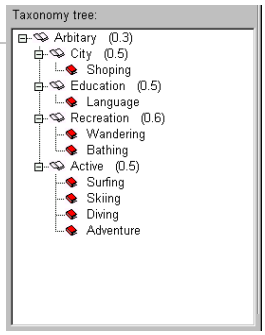
- Ordered Symbols
  - Similar to numbers
  - The symbols are mapped to a numeric range

Similarity

- Strings
  - exact match - two strings are similar if they are spelled the same way
  - spelling check - compares the number of letters in two strings which are the same (Useful for strings consisting of one word only)
  - word-count - counts the number of matching words of two cases. (Useful for strings consisting of several words).

Similarity

- Taxonomy
  - A classification hierarchy defines similarity for concepts
  - Inner nodes of the tree are assigned similarity values
  - Leaves under a node will share the nodes similarity



Retrieve, but efficiently ...

- Efficient case retrieval is essential for large case bases
- Different approaches depending on the:
  - case representation
  - size of the case base
- Organization of the case base:
  - Linear lists, only for small case bases
  - Index structures for large case bases
    - Kd-trees: index structure for large case bases (Wess)
    - Retrieval nets: index structure for textual (conversational) CBR (Lenz)