TDT4173 - Assignment 2

Håkon Ødegård Løvdal

February 23, 2016

1 Theory

Case-Based Reasoning

1)

Case-based reasoning (hereby denoted CBR) methods is characterized by the idea of solving every new problems with the solutions of similar previous/past problems. Stated differently, one could say that CBR uses the knowledge/experience of previous similar problems to solve current problems. The argument for allowing this kind of problem solving, is the theory that similar problems, is likely to have a similar solution.

CBR differs from other machine learning algorithms because it is able to take advantage of specific knowledge from similar previous problems. Other machine learning algorithms rely on a generalized knowledge about a problem domain. In problem domains with rich, specific and complex domains this yields an advantage for CBR.

2)

The terms *experience*, *memory* and *analogy* are terms from the study field of cognitive science. The study on these three terms are major influences to CBR. This is reflected in how CBR organizes its experiences and memories, along with how it uses this information to solve new problems. *Cognitivist learning theory* argues that humans use its previous experiences to make new decisions. CBR tries to solve new problems in this manner, using solutions based of previous solved problems.

3)

In *surface similarity* each of the features is represented by a real number in the interval [0, 1]. The similarity is then calculated with a similarity measure like Euclidian, Manhattan or Cosine similarity. The other alternative, *structural similarity* is much more computationally expensive. The reason for this is that it exploits the domain knowledge to a much more thorough extent. A lot more of the domain knowledge is encoded into the similarity function. The features is often represented in an object-oriented fashion, where the objects can be connected in a graph. This makes *structural similarity* to often provide more precise results than *surface similarity*.

To give an example of this we can look at image recognition. A surface similarity for instance, would only look on colors represented as a 1D vector. A structural similarity on the other hand could look deeper into the image and the meaning of images, the artist and so on. This is a rather naive example, but I did not manage to find any better example...

2 Practical

Case Modeling

See figure 1 for a screenshot of one of the instances.

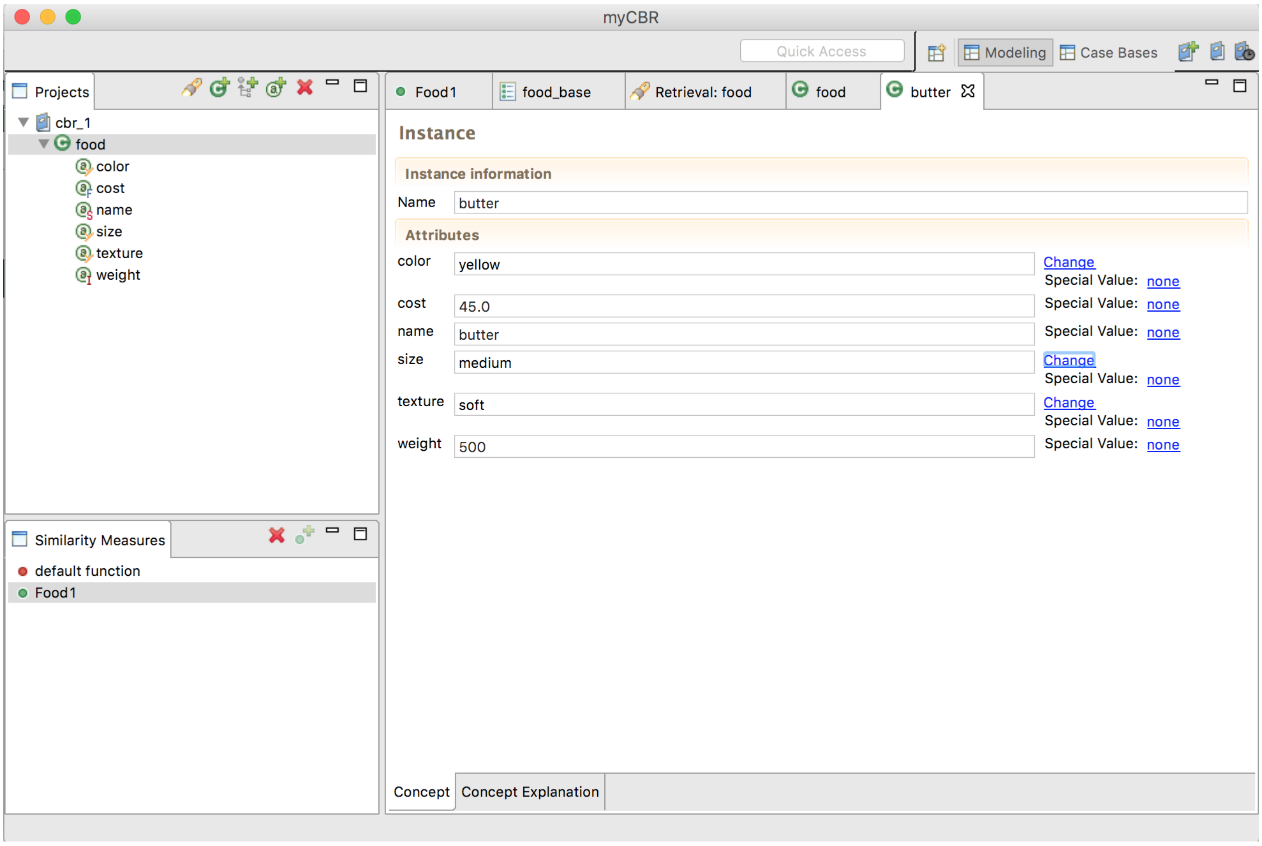


Figure 1: An example of one of the instances.

Case Retrieval

1)

Description of similarity measures (comparison functions) for the different attributes:

color I am interested in an exact match in color, and therefore the result is given a similarity of 1 for a perfect match, and 0 for misses.

cost Like in the exercise description, cost is given the similarity of 1 for an exact match, and otherwise decreasing towards 0 with a 1st degree polynomial function the further away the cost values are apart.

size Since it is hard to determine which size a food item has, I give a 0.5 similarity between all neighboring sizes (*small* ⇔ *medium* and *medium* ⇔ *large*). Exact matches is given a similarity of 1, while complete misses is given 0.

texture It is possible to argue that there is a similarity between for instance *crisp* and *porous*, and therefore I should give them a similarity of 0.5. But for simplicity I have chosen to only reward exact matches with 1 (0 otherwise).

weight Just like cost, weight is awarded with a 1st degree asymmetric polynomial function. Since weight is represented as an integer for gram, the further apart the weight values are it is punished harder. The similarity is 1 for exact matches.

The total similarity function is given by the weighted sum between every attribute, where every attribute except *name* is given the weight of 1.

2)

In this query I try to search for peas. The query is given by: green food, that costs about 10, with a small size, hard texture and about 5 grams in weight. The result is interesting. Since the weight for both *color* and *texture* is 1, both pea and apple gets a similarity of 0.79. For pea; color, size and weight gets a full match. Texture does not match, and gets 0. Cost gets a similarity of 0.9525. This gives a total similarity of 0.79 (see equation 1). See figure 2 for a screenshot of the query.

$$(\frac{1}{5} \times 1) + (\frac{1}{5} \times 0.9525) + (\frac{1}{5} \times 1) + (\frac{1}{5} \times 0) + (\frac{1}{5} \times 1) + = 0.79 \tag{1}$$

For apple; size and texture gets a full match. Cost gets a similarity of 0.98. Weight gets a similarity of 0.971. Color does not match, and gets a similarity of 1. This gives a total similarity of 0.79 (see equation 2)

$$(\frac{1}{5} \times 0) + (\frac{1}{5} \times 0.98) + (\frac{1}{5} \times 1) + (\frac{1}{5} \times 1) + (\frac{1}{5} \times 0.971) + = 0.79 \tag{2}$$

3)

I find it interesting that apple gets the same similarity as peas. The reason for this is the weighted sum, and the fact that both apple and peas match four of the five attributes. Furthermore, in this case, the cost and weight similarity functions are simple 1st degree polynomial functions. One alternative here could be to increase the degree of the polynomial function to punish values further apart from the exact match.

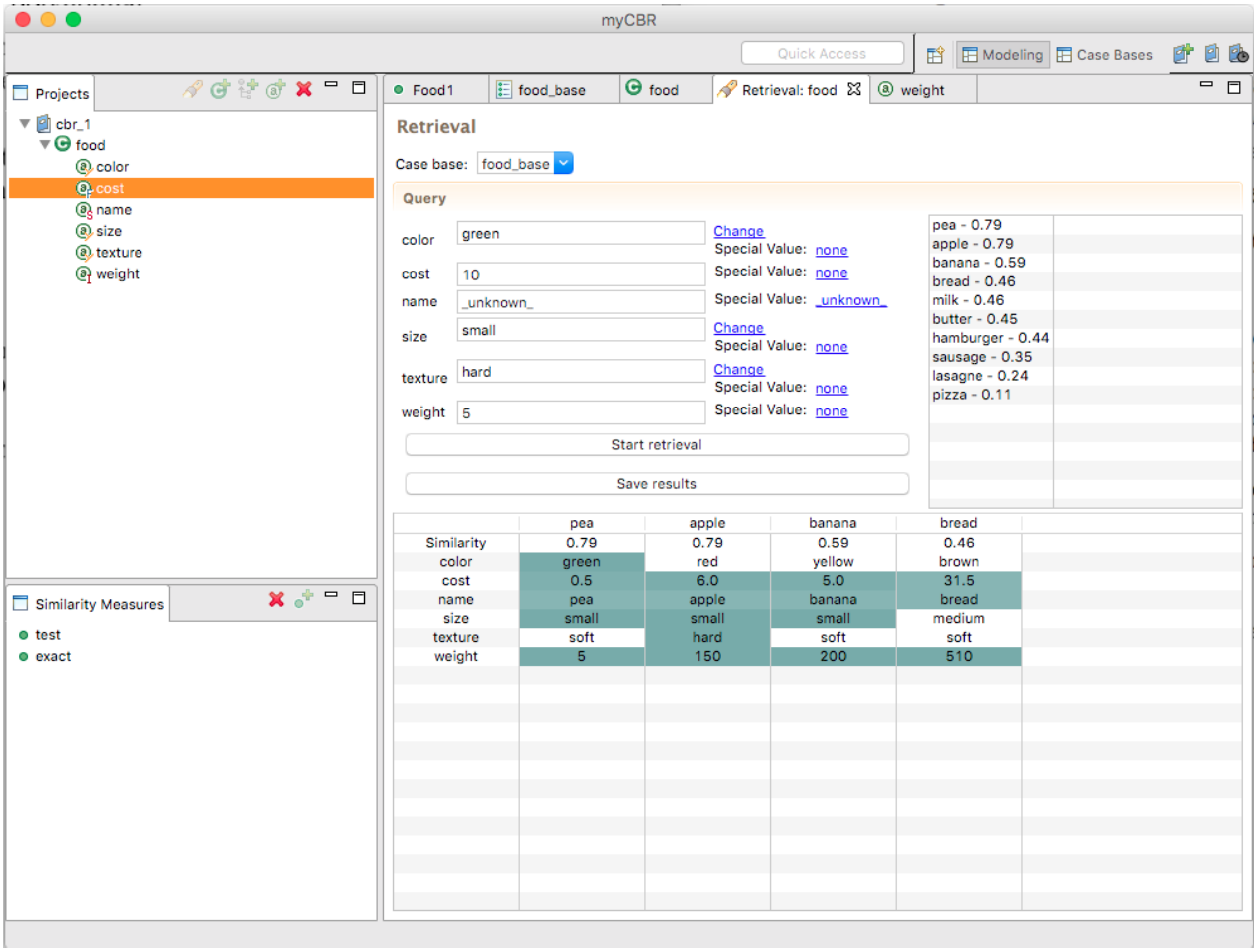
4)

Retrieving In this simple model, it is only one concept (food). The concept contains of six attributes, where only five of them count in the similarity measure. Quite simply, I would get the food item that has the highest similarity measure, given the different similarity functions.

Reusing In the previous query I did not get a good enough match for peas. An example of a reuse here could be to adapt the retrieved solution for peas to fit our target problem. This would for instance include reducing the cost, to ensure that peas would be a better match the next time.

Revising The revise step here, would be to test the new solutions and, if necessary, revise. This would involve to check the correctness, quality and other criteria of the solution. Did I get a better match for peas the next time?

Retaining Given that the revision were successful, I like to retain this new solution/experience to the case base.



	pea	apple	banana	bread
Similarity	0.79	0.79	0.59	0.46
color	green	red	yellow	brown
cost	0.5	6.0	5.0	31.5
name	pea	apple	banana	bread
size	small	small	small	medium
texture	soft	hard	soft	soft
weight	5	150	200	510

Figure 2: Query for subtask 2