TDT4174: Assignment 2

Stian Ismar

1 Theory

1.1

CBR has its roots in the cogitive psycholohist Roger Schank's work in the 1980s. His model of dynamic memory was the basis for the earliest CBR systems. Cognitive science is the study of the mind and how it functions. CBR uses the same type of reasoning of the human mind, in which memories are stored as cases.

1.2

Case-based reasoning (CBR) differs from other machine learning approaches in the way the problems are solved. In CBR, new problems are solved by using the assumption that similar problems have similar solutions. Older problems and solutions are stored and used to solve similar problems. The previous experiences are called "cases". Compared to other ML techniques, CBR uses a cognitive approach for modelling how humans solve problems by using previous situations.

1.3

Surface similarity is when a new case is compared to an existing one, and where one matches on the actual attributes, not the values. Objects of the same structure can naturally be compared by the attributes values, since the attributes are the same. An example would be a car-problem-object with the following attributes: Problem, Car, Year, Battery voltage, mileage, age of engine.
Structural similarity is when one compares a new case to an old one with regards to the structures. This can be the number of attributes, the name of the attributes etc. In other words, in structural comparison, the values are not compared. An example of structural similarity would be to compare two different objects to see if the solution for the stored one can be used for the new one as well.

1.4

The similarity between cases can be measured when cases are made up of attributes with different data types like so:

Case 1	Case 2
Integer	Integer
Float	Float
Symbol	Symbol
Float	Float

The corresponding attribut types are matched against each other.

1.5

Knowledge containers in CBR are represented by four containers:

- **Vocabulary**: The cases, the similarity measures and adaption knowledge make up the vocabulary container.
- **Similarity Measure**: The comparison of two cases is based upon the distance or similarity of these cases.
- **Case Base**: The system experience- a special form of database
- **Adaptation Knowledge**: This knowledge is used when a retrieved case's solution has to be adapted to be suitable to solve the presented problem.

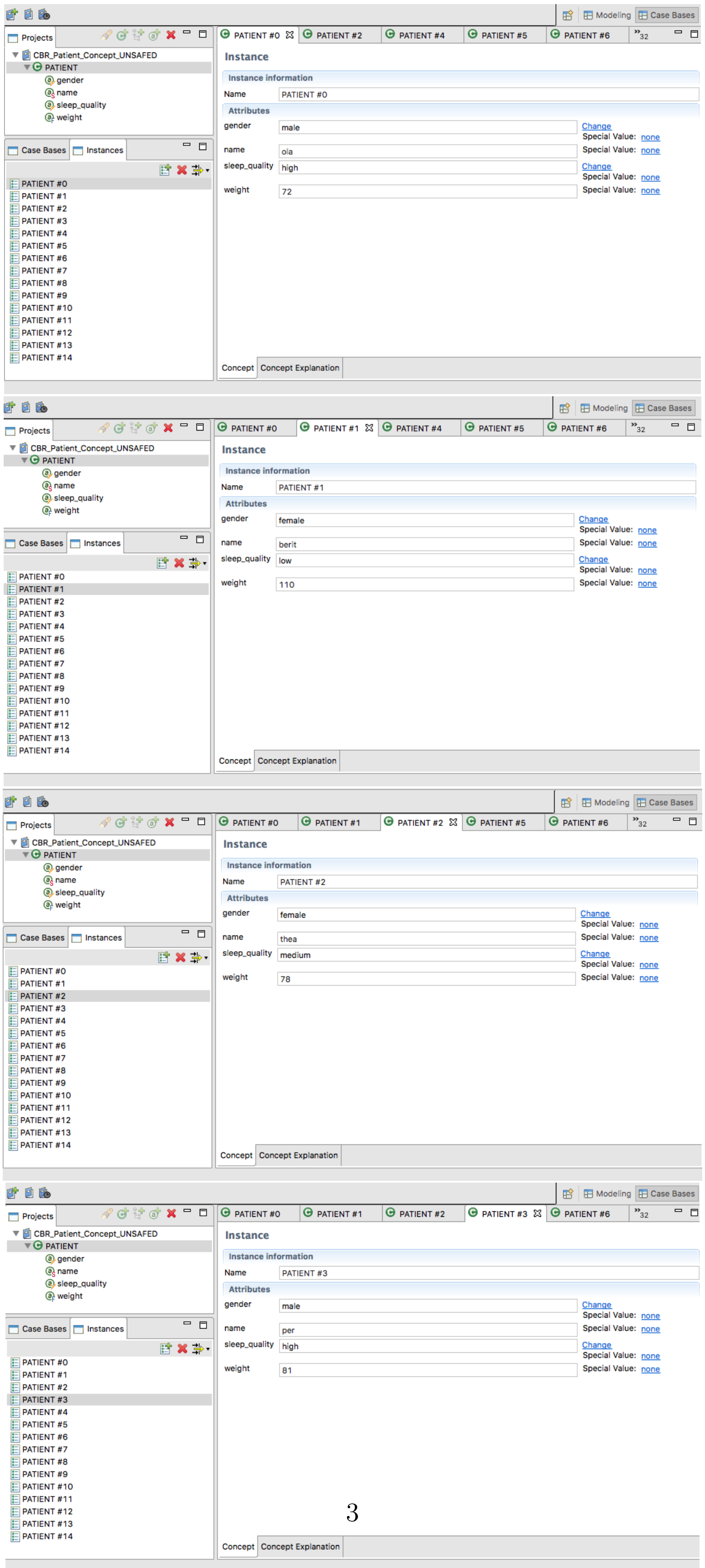
2 Practical

2.1 Case Modelling

I made a new concept called patient, and added the attributes:

- gender (symbol)- female/male
- name (String)
- sleepquality (String)
- weight (Float)- between 1 and 150.

a,b,c) 14 instances were made. All of the 4 attributes were filled out. The weights were rounded.



Figur 1: Screenshots of 4 of the 14 instances.

2.2 Case Retrieval

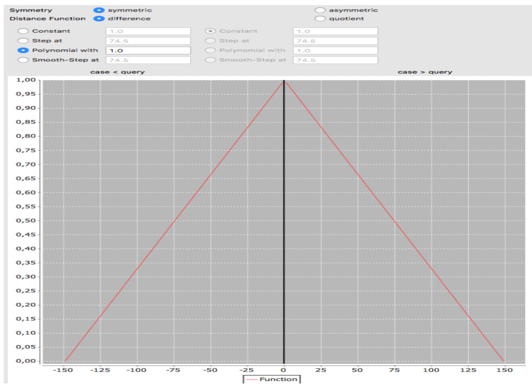
e)
A global similarity measure was made for the patient concept. It ignores the name attribute.

Type: (•) Weighted Sum ( ) Euclidean ( ) Minimum ( ) Maximum

Attribute	Discriminant	Weight	SMF
gender	true	1.0	gender_fnct
name	false	1.0	default function
sleep_quality	true	4.0	sleep_quality_fnct
weight	true	6.0	weight_func

Figur 2: The global similarity function.

For the **weight** attribute, a new polynomial measure function was made. If it is a perfect match, it will return 1. If not, it will be <1. See figure 2. It said in the task to create several different similarity measures, but I chose only to create one for weight, since only one can be selected in the global similarity measure for each attribute.



Figur 3: The similarity function for the weights attribute.

The **gender** similarity function was set up like in figure 3, using a symmetric symbol function.

	female	male
female	1.0	0.0
male	0.0	1.0

Figur 4: The similarity function for the gender attribute.

For the **sleepQuality** similarity function, it was set up using a symmetric symbolic function. I put some overlap for the different symbolic attributes. This is because a patient with medium sleep quality might overlap with someone with low or high. This might be relevant when giving a diagnose.

	high	low	medium
high	1.0	0.0	0.3
low	0.0	1.0	0.3
medium	0.3	0.3	1.0

Figur 5: The similarity function for the sleep attribute.

f)

Five different retrieval queries were done.

Query no.	gender	sleepQuality	weight
1	male	low	108
2	female	high	110
3	female	high	74
4	male	high	55
5	female	high	55

The results of the five queries are on the next page.

These results were expected because I put in the instances myself. I deliberately connected a high weight with lower sleep quality. What I should have done in retrospective, was to have another attribute measuring hours of physical activity per week. That would have made the results more interesting. I can imagine a database with a lot of instances would have been interesting to find correlations with using myCbr.

For query number two (See table above), the first result had a similarity of 0.85. This is because it matched on the gender attribute, as well as the sleep quality. I can see that the weight of 110 for the retrieval was closer to 69, which is the maximum case weight for a female instance. That is why the next results scored lower on similarity. The third result was a male, and therefore only had 0.8 similarity to the query.

	PATIENT #14	PATIENT #4	PATIENT #8	PATIENT #1
Similarity	0.99	0.96	0.96	0.9
gender	male	male	male	female
name	paul	erlend	kristoffer	berit
sleep_quality	low	low	low	low
weight	110.0	98.0	97.0	110.0

	PATIENT #10	PATIENT #7	PATIENT #3	PATIENT #6
Similarity	0.85	0.85	0.8	0.78
gender	female	female	male	male
name	nina	elisabeth	per	alex
sleep_quality	high	high	high	high
weight	69.0	68.0	81.0	76.0

	PATIENT #10	PATIENT #7	PATIENT #6	PATIENT #0
Similarity	0.98	0.98	0.9	0.9
gender	female	female	male	male
name	nina	elisabeth	alex	ola
sleep_quality	high	high	high	high
weight	69.0	68.0	76.0	72.0

	PATIENT #11	PATIENT #0	PATIENT #6	PATIENT #3
Similarity	0.97	0.94	0.92	0.9
gender	male	male	male	male
name	leon	ola	alex	per
sleep_quality	high	high	high	high
weight	62.0	72.0	76.0	81.0

	PATIENT #7	PATIENT #10	PATIENT #11	PATIENT #0
Similarity	0.95	0.95	0.88	0.85
gender	female	female	male	male
name	elisabeth	nina	leon	ola
sleep_quality	high	high	high	high
weight	68.0	69.0	62.0	72.0

Figur 6: Screen shots of 5 queries.

g)

Another problem with similar problem description are problems with a corresponding set of adaption rules. One problem description that uses the patient concept could be a problem with similar attributes. Another solution could be formed by using the attributes to draw a conclusion by interpolation the known attributes. The patient concept could be used for determining if a person has insomnia (After adding some more attributes like blood pressure, time it takes to fall asleep, stress levels etc.). If a patient checks out on a set of variables in a case where the Solution is Person has insomnia", the patient could get this diagnosis.

This could be done by adding a case in the CB with a problem description similar to the patient-concept, and having the solution be Person with insomnia".

The retrieve-step would find the most relevant cases from the CB and reuse the them to propose a number of solutions. The cases are retreived using the defined similarity functions. In the revise step, the solution is tested. After successfully adaption the solution to the target problem, it is stored as a new solution. I did not do the last to steps of the CBR cycle in myCBR.