Assignment two

Machine Learning

Sigve Skaugvoll, MIT, H2018

Theory

1. **What characterises case-based reasoning (CBR) methods? How are they different from other machine learning approaches?**
    a. Case based reasoning is different because it relies on previous “cases” to find similarity towards the problem at hand. CBR tries to model human problem solving behaviour, because similar problems have similar solutions.

       CBR in the end boils down to structure and methods from CS; Retrieve, Reuse, Revise and Retrain from the general knowledge.

       Another way CBR is different is that it’s a “lazy” learning method. - defer the decision of how to generalize beyond the training data until a new query instance is observed. Instances are typically represented using more rich symbolic descriptions

       Relying on knowledge-based reasoning rather than statistical methods.
2. **Discuss (some of) the ways in which cognitive science has influenced CBR.**
    a. CBR tries to model the way CS has figured out how humans think / reasons to find solutions. Humans tries to find similarity / distance between previous cases and the new problem, the same as CBR. - instance based learning.
    b. CBR uses a form of A script, which is a memory structure that stores general knowledge about typical situation patterns. Such patterns can be MOP and TOP
3. **Methods to evaluate the degree of similarity between two cases are essential in CBR. What is the difference between surface similarity and structural similarity? Give some examples for each approach.**

	Surface Similarity	Structural Similarity
Characteristics	• Similarity is based on surface features • Applicable to different case representations using standard value types (binary, integer, real, string, symbol, etc.)	• Highly dependent on the case representation and heavily uses domain knowledge
Retrieval	• *k*NN • Local-global-principal • SQL Approximation • kd-Tree	• Object-oriented similarity measures • Generalized Cases • MAC/FAC • Graph Editing

a.

i. Surface:

1. Tom chases Jerry and Jerry chases Tom has the same surface-similarity. And the solution to make Tom stop chasing Jerry, would most likely be almost (if not) the same as making Jerry stop chasing Tom.

ii. Structural

1. Someone gets their car stolen - one solution is to call the police to notify / press charges / get help looking, and someone forgets where they parked, and had their child in the car on a hot summer day. - One solution is to call the police as well to get help since it's a critical problem. The 2 problems may look different on the surface, but have the same underlying similarities

4. **Explain how the similarity between cases can be measured when cases are made up of attributes with different data types. Give an example of how this can be done.**

a. We can use the Local-Global principle and apply amalgamation function. This has certain advantages; decomposition of the complex global similarity measure into easy to handle local similarity measures. The drawback is that attribute dependencies have to be considered and modeled with certain "tricks".

Weighted Average:

$$F(s_1,\ldots,s_n)=\sum_{i=1}^{n} w_i \cdot s_i \quad \text{with} \quad \sum_{i=1}^{n} w_i = 1 \ \text{ and } \ s_i = sim_i(q_i,d_i)$$

- Generalization:

$$F(s_1,\ldots,s_n)=\sqrt[\alpha]{\sum_{i=1}^{n} w_i \cdot (s_i)^{\alpha}} \quad \text{with} \quad \alpha \in R^+, \sum_{i=1}^{n} w_i = 1$$

Maximum: $F(s_1,\ldots,s_n)=\max\{s_1,\ldots,s_n\}$

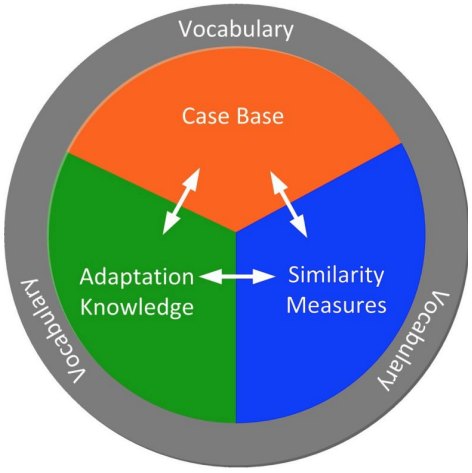
Minimum: $F(s_1,\ldots,s_n)=\min\{s_1,\ldots,s_n\}$

k-Maximum: $F(s_1,\ldots,s_n)=s_{ik}$ with $s_{i1} \geq s_{i2} \geq \ldots \geq s_{in}$

k-Minimum: $F(s_1,\ldots,s_n)=s_{ik}$ with $s_{i1} \leq s_{i2} \leq \ldots \leq s_{in}$

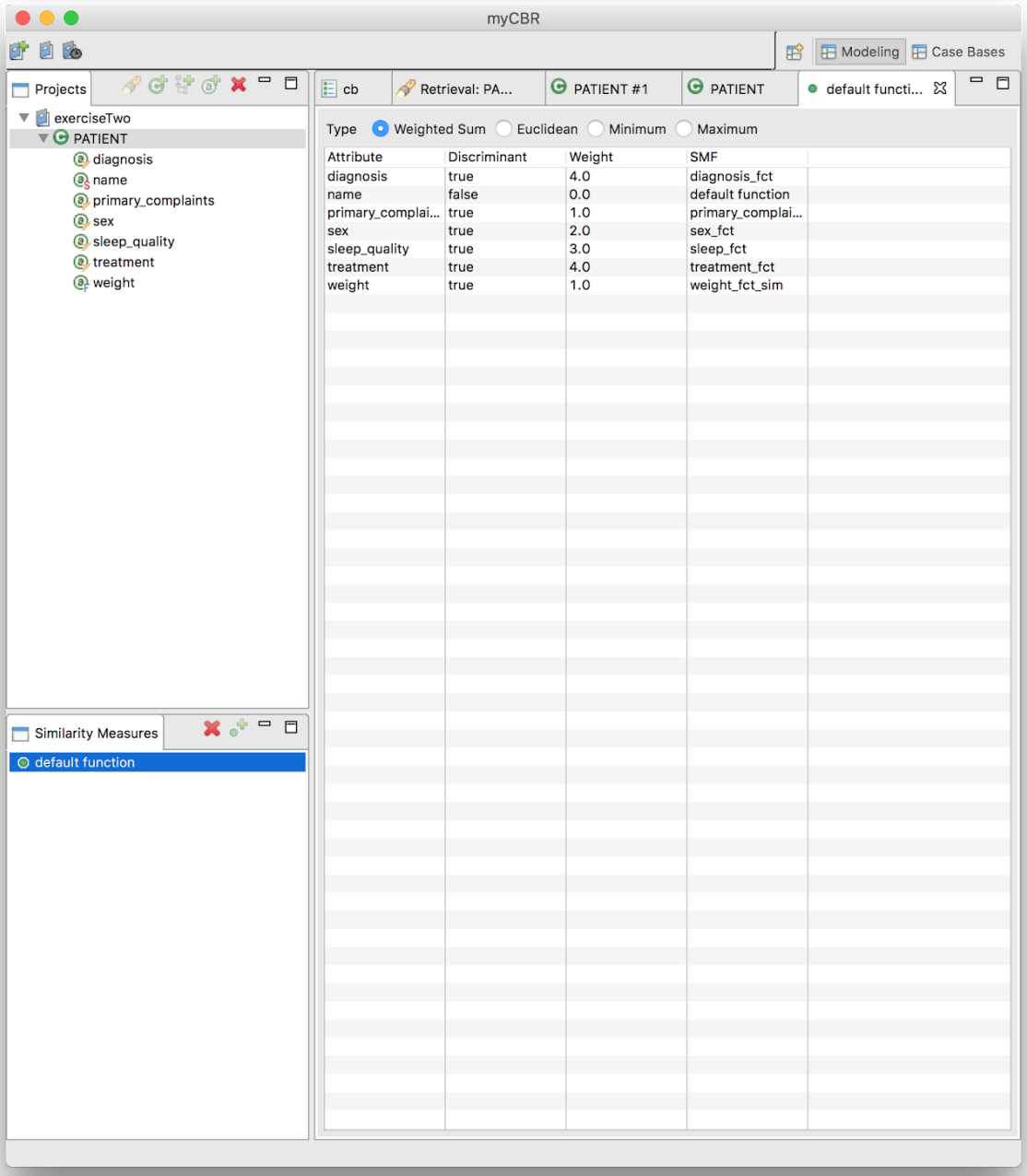
b. HUSK å gi eksempel

5. **What are knowledge containers in the context of CBR? Give a brief explanation of the different containers.**
    a. Define the facts and the rules in such a way that all the intended knowledge is represented. Then these facts and rules are distributed into the knowledge containers - which then becomes the available knowledge distributed into domains. The containers are dependent of each other do be able to solve a problem.
        i. Consists of these four;
            1. Similarity Measures
                a. The retrieval of similar cases is based upon the use of similarity functions (or measures) to compute the distance or similarity of two cases
            2. Case Base
                a. The systems experience is stored as cases within the case base which can be seen as a special form of a database
            3. Vocabulary
                a. The cases themselves, the similarity measures and the adaptation knowledge are composed upon a vocabulary that contains the objects of interests(terms, attributes, concepts).
            4. Adaptation knowledge
                a. Adaptation knowledge is used whenever a retrieved case's solution has to be adapted to be suitable to solve the presented problem.



Practical

Case Modelling



Instance

Instance information

Name	PATIENT #0

Attributes

diagnosis	diarea	Change Special Value: none
name	Andreas	Special Value: none
primary_complaints	tummy_ache	Change Special Value: none
sex	male	Change Special Value: none
sleep_quality	high	Change Special Value: none
treatment	eat_oatmeal	Change Special Value: none
weight	78.5	Special Value: none

Instance

Instance information

Name	PATIENT #2

Attributes

diagnosis	head_ache	Change Special Value: none
name	Martin	Special Value: none
primary_complaints	back_pain	Change Special Value: none
sex	male	Change Special Value: none
sleep_quality	low	Change Special Value: none
treatment	take_pain_killers	Change Special Value: none
weight	80	Special Value: none

Instance

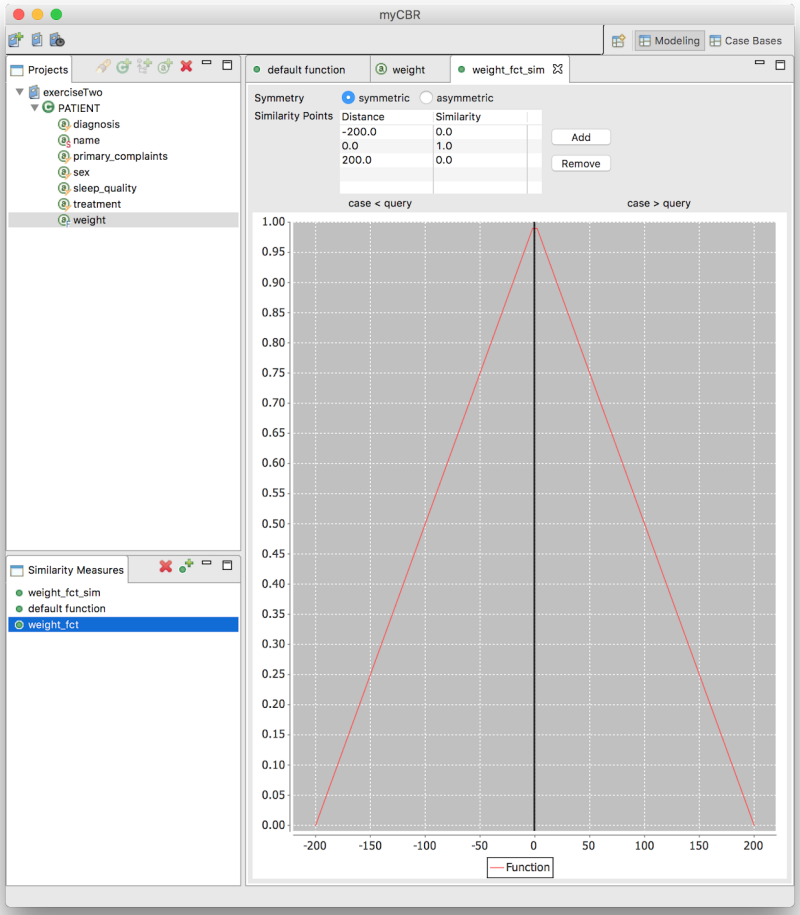
Instance information

Name	PATIENT #3

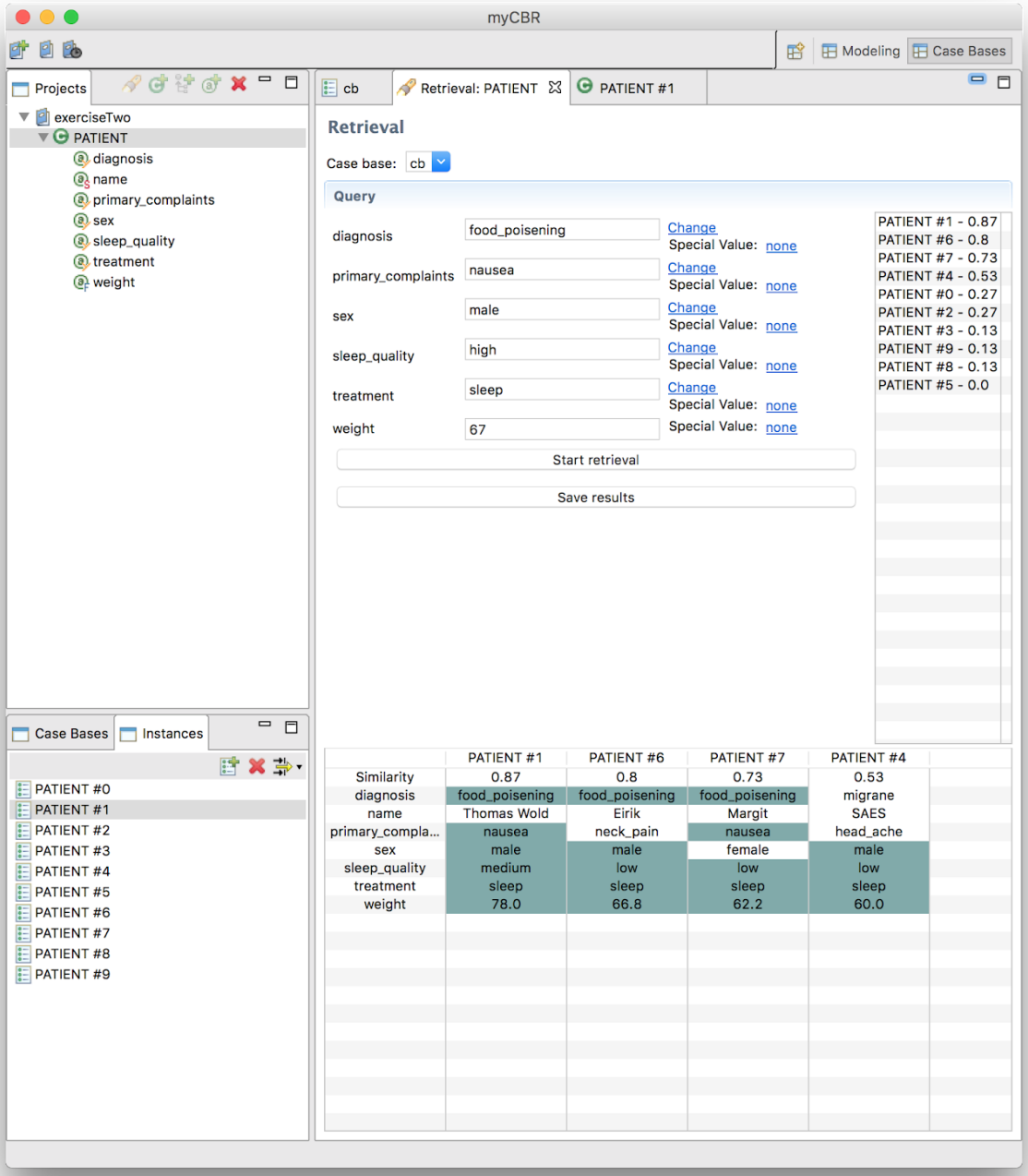
Attributes

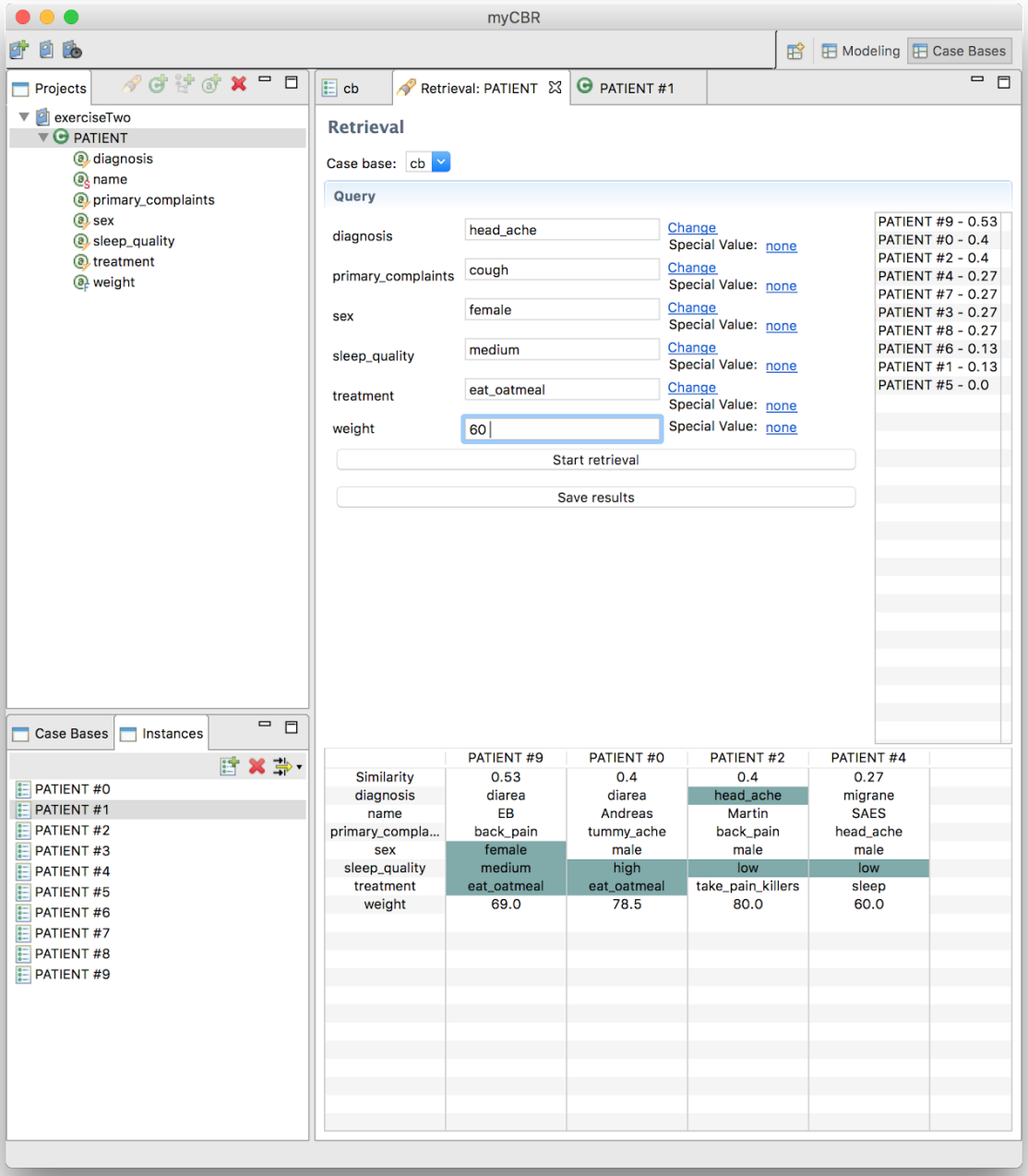
diagnosis	migrane	Change Special Value: none
name	CD	Special Value: none
primary_complaints	head_ache	Change Special Value: none
sex	female	Change Special Value: none
sleep_quality	low	Change Special Value: none
treatment	take_pain_killers	Change Special Value: none
weight	100.0	Special Value: none

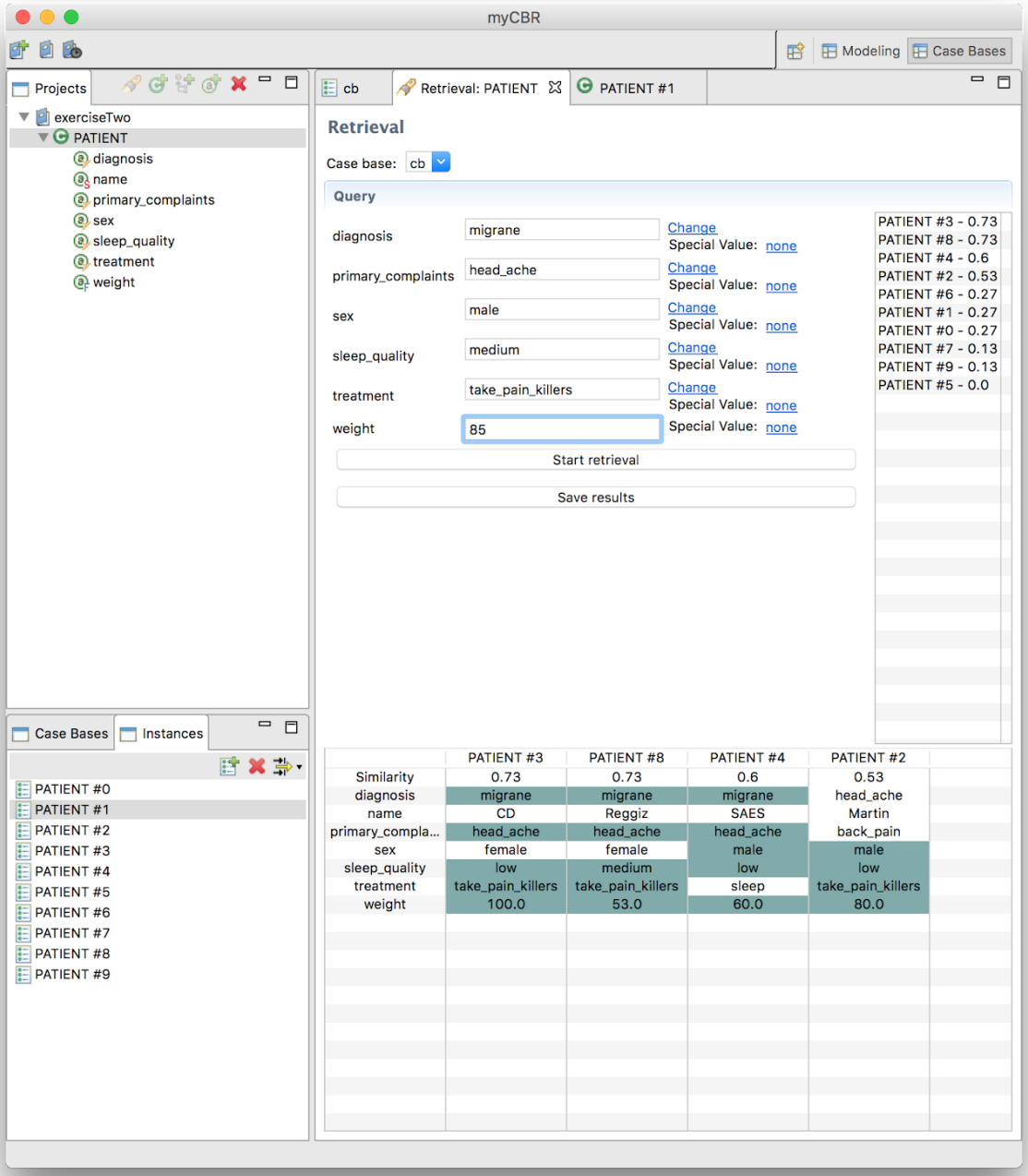
Case Retrieval

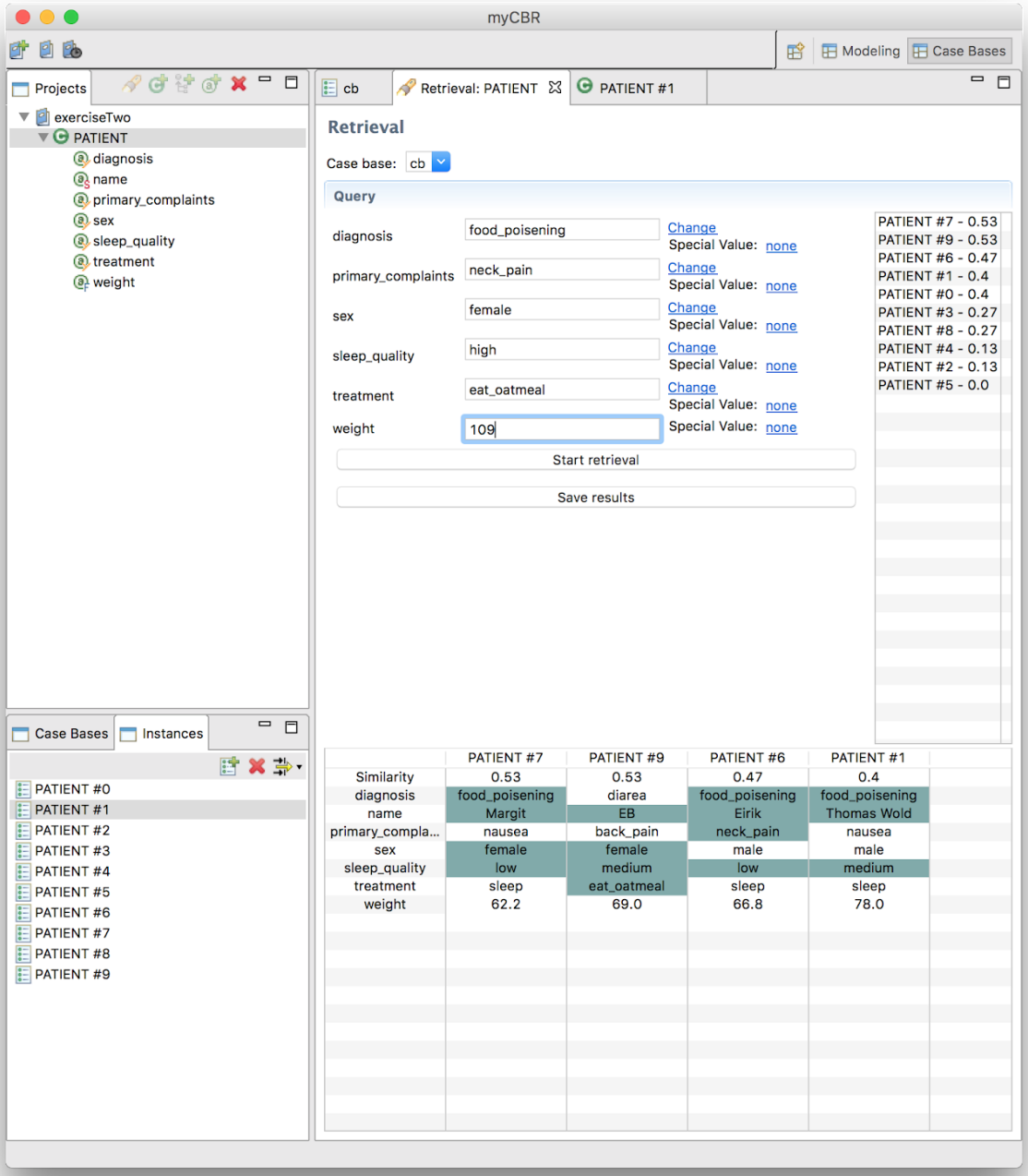


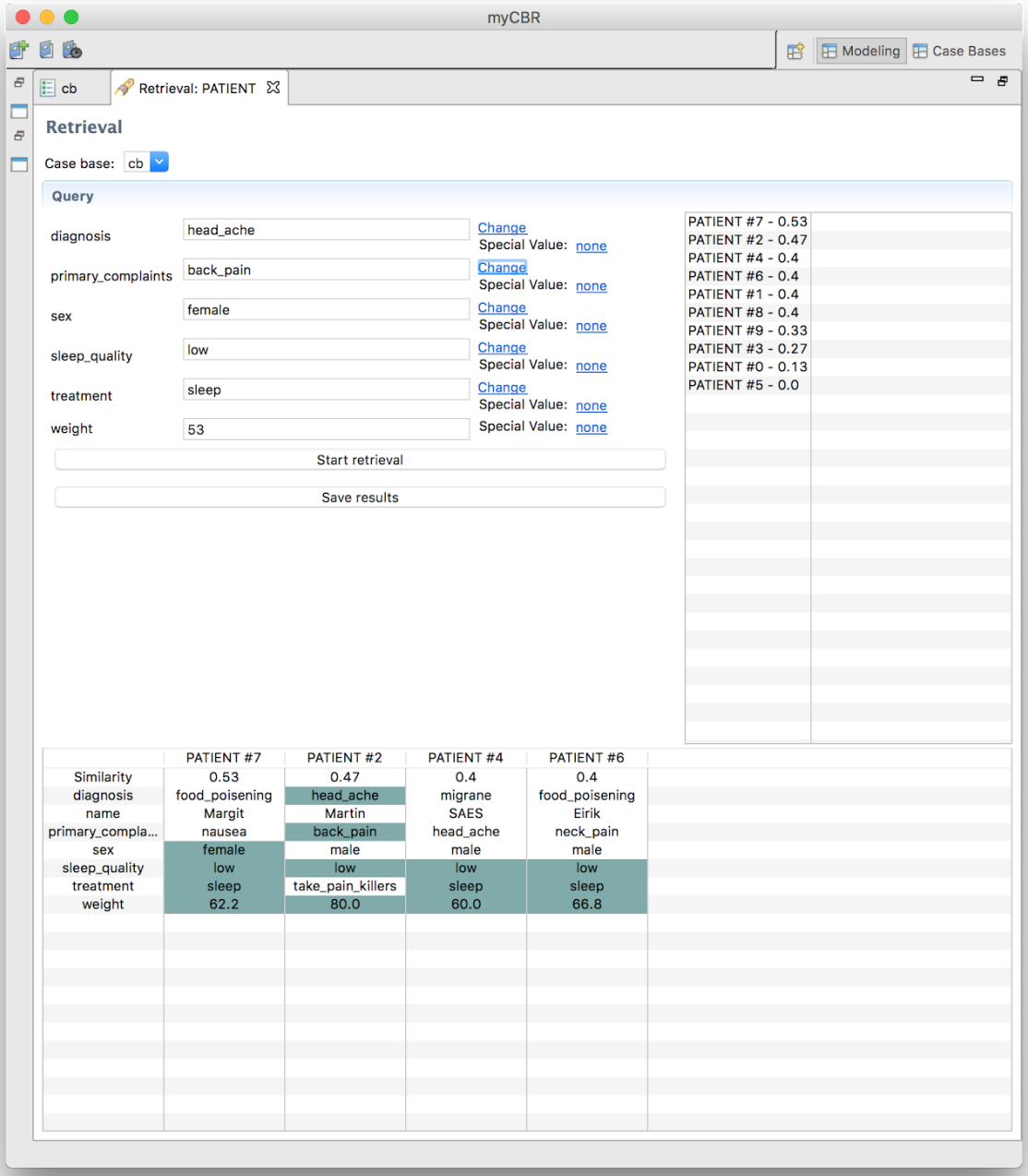
Retrievals:


myCBR
Modeling
Case Bases
Projects
exerciseTwo
PATIENT
diagnosis
name
primary_complaints
sex
sleep_quality
treatment
weight
cb
Retrieval: PATIENT
PATIENT #1
Retrieval
Case base: cb
Query
diagnosis
food_poisening
Change
Special Value: none
primary_complaints
nausea
Change
Special Value: none
sex
male
Change
Special Value: none
sleep_quality
high
Change
Special Value: none
treatment
sleep
Change
Special Value: none
weight
67
Special Value: none
Start retrieval
Save results
PATIENT #1 - 0.87
PATIENT #6 - 0.8
PATIENT #7 - 0.73
PATIENT #4 - 0.53
PATIENT #0 - 0.27
PATIENT #2 - 0.27
PATIENT #3 - 0.13
PATIENT #9 - 0.13
PATIENT #8 - 0.13
PATIENT #5 - 0.0
Case Bases
Instances
PATIENT #0
PATIENT #1
PATIENT #2
PATIENT #3
PATIENT #4
PATIENT #5
PATIENT #6
PATIENT #7
PATIENT #8
PATIENT #9
PATIENT #1 PATIENT #6 PATIENT #7 PATIENT #4
Similarity 0.87 0.8 0.73 0.53
diagnosis food_poisening food_poisening food_poisening migrane
name Thomas Wold Eirik Margit SAES
primary_compla... nausea neck_pain nausea head_ache
sex male male female male
sleep_quality medium low low low
treatment sleep sleep sleep sleep
weight 78.0 66.8 62.2 60.0


myCBR
Modeling
Case Bases
Projects
exerciseTwo
PATIENT
diagnosis
name
primary_complaints
sex
sleep_quality
treatment
weight
cb
Retrieval: PATIENT
PATIENT #1
Retrieval
Case base: cb
Query
diagnosis
head_ache
Change
Special Value: none
primary_complaints
cough
Change
Special Value: none
sex
female
Change
Special Value: none
sleep_quality
medium
Change
Special Value: none
treatment
eat_oatmeal
Change
Special Value: none
weight
60
Change
Special Value: none
Start retrieval
Save results
PATIENT #9 - 0.53
PATIENT #0 - 0.4
PATIENT #2 - 0.4
PATIENT #4 - 0.27
PATIENT #7 - 0.27
PATIENT #3 - 0.27
PATIENT #8 - 0.27
PATIENT #6 - 0.13
PATIENT #1 - 0.13
PATIENT #5 - 0.0
Case Bases
Instances
PATIENT #0
PATIENT #1
PATIENT #2
PATIENT #3
PATIENT #4
PATIENT #5
PATIENT #6
PATIENT #7
PATIENT #8
PATIENT #9
PATIENT #9 PATIENT #0 PATIENT #2 PATIENT #4
Similarity 0.53 0.4 0.4 0.27
diagnosis diarea diarea head_ache migrane
name EB Andreas Martin SAES
primary_compla... back_pain tummy_ache back_pain head_ache
sex female male male male
sleep_quality medium high low low
treatment eat_oatmeal eat_oatmeal take_pain_killers sleep
weight 69.0 78.5 80.0 60.0


myCBR
Modeling
Case Bases
Projects
exerciseTwo
PATIENT
diagnosis
name
primary_complaints
sex
sleep_quality
treatment
weight
cb
Retrieval: PATIENT
PATIENT #1
Retrieval
Case base: cb
Query
diagnosis migrane Change Special Value: none
primary_complaints head_ache Change Special Value: none
sex male Change Special Value: none
sleep_quality medium Change Special Value: none
treatment take_pain_killers Change Special Value: none
weight 85 Special Value: none
Start retrieval
Save results
PATIENT #3 - 0.73
PATIENT #8 - 0.73
PATIENT #4 - 0.6
PATIENT #2 - 0.53
PATIENT #6 - 0.27
PATIENT #1 - 0.27
PATIENT #0 - 0.27
PATIENT #7 - 0.13
PATIENT #9 - 0.13
PATIENT #5 - 0.0
Case Bases
Instances
PATIENT #0
PATIENT #1
PATIENT #2
PATIENT #3
PATIENT #4
PATIENT #5
PATIENT #6
PATIENT #7
PATIENT #8
PATIENT #9
PATIENT #3 PATIENT #8 PATIENT #4 PATIENT #2
Similarity 0.73 0.73 0.6 0.53
diagnosis migrane migrane migrane head_ache
name CD Reggiz SAES Martin
primary_compla... head_ache head_ache head_ache back_pain
sex female female male male
sleep_quality low medium low low
treatment take_pain_killers take_pain_killers sleep take_pain_killers
weight 100.0 53.0 60.0 80.0


myCBR
Modeling
Case Bases
Projects
exerciseTwo
PATIENT
diagnosis
name
primary_complaints
sex
sleep_quality
treatment
weight
cb
Retrieval: PATIENT
PATIENT #1
Retrieval
Case base: cb
Query
diagnosis
food_poisening
Change
Special Value: none
primary_complaints
neck_pain
Change
Special Value: none
sex
female
Change
Special Value: none
sleep_quality
high
Change
Special Value: none
treatment
eat_oatmeal
Change
Special Value: none
weight
109
Change
Special Value: none
Start retrieval
Save results
PATIENT #7 - 0.53
PATIENT #9 - 0.53
PATIENT #6 - 0.47
PATIENT #1 - 0.4
PATIENT #0 - 0.4
PATIENT #3 - 0.27
PATIENT #8 - 0.27
PATIENT #4 - 0.13
PATIENT #2 - 0.13
PATIENT #5 - 0.0
Case Bases
Instances
PATIENT #0
PATIENT #1
PATIENT #2
PATIENT #3
PATIENT #4
PATIENT #5
PATIENT #6
PATIENT #7
PATIENT #8
PATIENT #9
PATIENT #7
PATIENT #9
PATIENT #6
PATIENT #1
Similarity 0.53 0.53 0.47 0.4
diagnosis food_poisening diarea food_poisening food_poisening
name Margit EB Eirik Thomas Wold
primary_compla... nausea back_pain neck_pain nausea
sex female female male male
sleep_quality low medium low medium
treatment sleep eat_oatmeal sleep sleep
weight 62.2 69.0 66.8 78.0



Type Weighted Sum Euclidean Minimum Maximum

Attribute	Discriminant	Weight	SMF
diagnosis	true	4.0	diagnosis_fct
name	false	0.0	default function
primary_complai...	true	1.0	primary_complai...
sex	true	2.0	sex_fct
sleep_quality	true	3.0	sleep_fct
treatment	true	4.0	treatment_fct
weight	true	1.0	weight_fct_sim

The global similarity has a lot to say! About everything to say, when retrieving. Changing the weights and the “discriminant” changes the retrieval results.

On the last query, the result was interesting! I find it quite interesting that **Food_poisening** was more similar than **head_ache**. The reason for this is because the global similarity weights diagnosis and treatment has the same weight. And both patient 7 and 2 has one each of this weight. Sex weights more than primary_complaint, thus patient 7 has more similarity here. The weight 62.2 is closer to 53 then 80. Thus patient 7 has more weight, then patient 2, even though “weight” is the same for both. It would make more sense that primary_complaints has more weight than sex. If we change this set primary_complaint to 3, and sex stays at 2, and take a look at the new result; We can see that this result gives more meaning because patient 2 and 7 swapped places.

	PATIENT #2	PATIENT #7	PATIENT #9	PATIENT #4
Similarity	0.53	0.47	0.41	0.35
diagnosis	head_ache	food_poisening	diarea	migrane
name	Martin	Margit	EB	SAES
primary_compla...	back_pain	nausea	back_pain	head_ache
sex	male	female	female	male
sleep_quality	low	low	medium	low
treatment	take_pain_killers	sleep	eat_oatmeal	sleep
weight	80.0	62.2	69.0	60.0

It’s clear that head_ache should be the most similar case, and weigh primary_complaints more than sex. If we compare patient 2 and 9 (first and third place) we see that both patient has back_pain, patient 9 is closer to the query weight, and is female, but since the diagnosis has more weight, we get the wanted result.

Q: The CBR cycle include retrieve, reuse, revise, and retain. Come up with one or more problems that use the patient concept. Briefly explain how each step in the CBR cycle can be executed with the help of myCBR

myCBR can only help with retrieve, there is no reuse,revise and retain functionality. The reuse and steps after is up to the myCBR user to execute. I guess the user can create new instances with the new retained information and create a case that will help improve the system for further use. myCBR is quite hendy for retrieval. It’s quick and easy to create concepts, add attributes, create similarity functions and query.

The system and concept for PATIENT, can be used by an online Artificial intelligence doctor-agent, help for doctors and or supplement for doctors in third world countries where there is an shortage for doctors. Thus the system can “reason” and then subscribe common (non-lethal) drugs for patients.