

TDT4173: Machine Learning and Case-Based Reasoning

Assignment 2

February 14, 2018

Kerstin Bach and Amar Jaiswal
Department of Computer Science
Norwegian University of Science and Technology (NTNU)

- **Delivery deadline: February 28, 2018** by 24:00.
- **This assignment counts towards 3 % of your final grade.**
- Solutions must be submitted individually.
- Deliver your solution on *Blackboard* before the deadline.
- Please upload your report as a PDF file, and package your project file and screenshots into an archive (e.g. zip, rar, tar).

Objective: Gain insight into (a) core concepts in case-based reasoning, and (b) how case-based reasoning can be put into practice.

1 Theory [1.5 points]

1. **[0.2 points]** What characterises case-based reasoning (CBR) methods? How are they different from other machine learning approaches?
2. **[0.4 points]** Discuss (some of) the ways in which cognitive science has influenced CBR.
3. **[0.3 points]** 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.
4. **[0.3 points]** 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.
5. **[0.3 points]** What are *knowledge containers* in the context of CBR? Give a brief explanation of the different containers.

2 Practical [1.5 points]

Purpose

This exercise is designed to give an overview of how a concept could be modeled in a CBR system, the student will also learn how an attribute could be created with different types of similarity measures. It will provide an intuitive understanding for local and global similarity measures in the context of CBR case modeling.

Introduction

The utility of case-based reasoning (CBR) system depends on cases within its case-base and its ability to retrieve the relevant cases in response to an indexed case. The similarity serves as its core concept starting from case modeling, retrieval and adaptation in a CBR systems.

The idea behind this task is to create a case-base with multiple cases for individual patient, which will act as an instances of patient concept with their user-defined attributes. Every patient instance should have at least 4-6 relevant attributes such as name, age, sex, primary_complaint, diagnosis, treatment etc. These instances of the patient will be our prior knowledge/cases for our CBR system. When a new patient is searched in the case-base, the CBR system retrieves patients based upon their local and global similarity with respect to the searched case. The retrieved similar patient's diagnosis/treatment would be adapted or reused in the diagnosis/treatment of the new patient.

The entire concept is based on the hypothesis of similar patients could be treated with similar treatment/interventions.

Information

For this assignment, you will experiment with using an existing CBR program, myCBR, to get hands-on experience with how a CBR system may work. The myCBR¹ tool is a stand-alone application that you will need to download and install.

1. **There is a demo video² on myCBR tool which could be used as reference for this exercise.**
2. **Make sure you save your work after every task, since you may encounter some remaining bugs in the program.**

¹The program can be downloaded from <http://www.idi.ntnu.no/~kerstinb/mycbr/builds/>, where you can also find a number of example projects to play around with. Additionally, make sure that you have the latest JDK and JRE installed before running myCBR.

² <http://www.idi.ntnu.no/~kerstinb/mycbr/demo-video/myCBR-demo.mov>

Include screenshots from myCBR for the tasks as well as the resulting project file when submitting your answer.

Case Modelling [0.5 points]

- a) Create a new concept called `PATIENT`, which will be used as the basis for the rest of this assignment.
- b) Create 4-6 relevant attributes for the `PATIENT` concept, including `name`, `weight`, and `sleep_quality`.
 - `name` – a purely descriptive string.
 - `weight` – a float value representing the weight of the patient.
 - `sleep_quality` – a symbol type with at least *low*, *medium*, and *high* as allowed values.
- c) Create ten or more instances of the `PATIENT` concept, for each instance you must fill out all of the attributes.
- d) Include a screenshot of 2-3 of the instances in your report.

Case Retrieval [1 point]

- e) **[0.3 points]** Create a global similarity measure for the `PATIENT` concept. It should ignore the `name` attribute. Create a few different similarity measures for the `weight` attribute, including one where the similarity is 1 for an exact match, and otherwise decreasing linearly towards 0 the further away the weight values are apart. Select the other attribute similarity modes and comparison functions as you see fit.
- f) **[0.3 points]** Perform at least five different retrieval queries. Explain the returned similarity scores for the top 3 results for one of your most interesting queries. Include a screenshot of each retrieval result in your report. Were there any strange or unexpected results? Explain your answer.
- g) **[0.4 points]** 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.

Acknowledgment

This assignment is based on the original work of Kerstin Bach (Associate Professor) and Aleksander Rognhaugen (TA for TDT4173 in year 2017)