TDT4173: Machine Learning and Case-Based Reasoning

Assignment 2

February 4, 2016

- **Delivery deadline: February 25, 2016** by 22:00.

- Solutions must be submitted individually.

- Deliver your solution on *itslearning* 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).

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

1 Theory

Case-Based Reasoning

1. What characterises case-based reasoning (CBR) methods? How are they different from other machine learning approaches?

2. In what ways has cognitive science influenced CBR?

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 from each approach.

2 Practical

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 myCBR[1] tool is a stand-alone application that you will need to download and install. **Make sure you save your work after every task, since you may encounter some remaining bugs in the program**.

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

---

[1] The program can be downloaded from `http://mycbr-project.net/download.html`, where you can also find a number of example projects to play around with.

Case Modelling

1. Create a new concept called `food`, which will be used as the basis for the rest of this assignment.

2. Create 4-6 relevant attributes for the `food` concept, including `name`, `cost`, and `size`.

   - `name` – a purely descriptive string.
   - `cost` – a float representing the approximate value of the item.
   - `size` – a symbol type with at least *small*, *medium*, and *large* as allowed values.

3. Create at least 10 instances of the `food` concept, including *banana*, *pea*, and *hamburger*.

4. Include a screenshot of one of the instances in your report.

Case Retrieval

1. Create a global similarity measure for the `food` concept. It should ignore the `name` attribute. Create different similarity measures for the `cost` attribute, including one where the similarity is 1 for an exact match, and otherwise decreasing towards 0 the further away the cost values are apart. Select the other attribute similarity modes and comparison functions as you see fit.

2. Perform at least 5 different retrieval queries. Explain the returned similarity scores for the top 3 results for one of your most interesting queries. Include a screenshot.

3. Are there any strange or unexpected results? What caused / can cause such difficulties?

4. Describe how a full CBR cycle can be performed manually for one of your queries.