



Review report of a final thesis

Reviewer: Ing. Karel Klouda, Ph.D.
Student: Bc. Yuliia Syzon
Thesis title: Standard Decision Trees in Machine Learning
Branch / specialization: Knowledge Engineering
Created on: 7 February 2023

Evaluation criteria

1. Fulfillment of the assignment

- ▶ [1] assignment fulfilled
- [2] assignment fulfilled with minor objections
- [3] assignment fulfilled with major objections
- [4] assignment not fulfilled

All points of the assignment have been met. I have concerns about some of them in the text below, but the concerns are not so serious to call into question the fulfillment of any of the assignment points.

2. Main written part

75 /100 (C)

The text is written in understandable English, although certainly not flawless (missing articles, skeletal expressions, etc.). I like the gradual building of the theoretical foundations in Chapter 1: the introduction of the necessary terminology, the research of approaches to building decision trees and the subsequent design of the implementation (Chapter 2) and its testing (Chapter 3).

There are a lot of places in the text where I don't like or understand the established notation, the formulation of the technical details of the algorithms, and so on. For example, in the first chapter, the lower and upper case letters m/M and n/N are used indistinguishably. Moreover, this notation is not used anywhere later in the text. In Definition 1.1.7, "Impurity(children)" is used but not defined anywhere. In the definition of Chi-square, n_i is used, but in the explanation, it is denoted as f_i etc. The notion of optimality from Definition 1.1.8 is very vague. The description of MapReduce is so brief to be hard to understand. On page 29, there is a symbol $V[\text{Ö}2]$ which I have no idea what it means; Figure 1.17 does not help.

3. Non-written part, attachments

83 /100 (B)

The resulting implementation is in Java. The code looks clear and (hopefully) meets the implementation standards for the H2O framework (I cannot verify this). It's unclear where the requirements R1 to R9 (page 39) came from and which of them were implemented. Based on the list in section 3.5, probably not all of them.

4. Evaluation of results, publication outputs and awards

80 /100 (B)

The resulting implementation seems to be working and ready to be incorporated into the H2O platform. This is a good result which certainly required a lot of work. The implementation evaluation in Chapter 3 is unconvincing. The testing datasets are really small and simple (and their the description is too brief).

The overall evaluation

78 /100 (C)

In view of the above, I propose evaluating the thesis as good, i.e. with a grade of C.

Questions for the defense

Why did you include Definition 1.1.10 of Min-Depth Optimal Tree? Is there always such a tree?

Can you explain how were the R1 to R9 requirements selected?

Instructions

Fulfillment of the assignment

Assess whether the submitted FT defines the objectives sufficiently and in line with the assignment; whether the objectives are formulated correctly and fulfilled sufficiently. In the comment, specify the points of the assignment that have not been met, assess the severity, impact, and, if appropriate, also the cause of the deficiencies. If the assignment differs substantially from the standards for the FT or if the student has developed the FT beyond the assignment, describe the way it got reflected on the quality of the assignment's fulfilment and the way it affected your final evaluation.

Main written part

Evaluate whether the extent of the FT is adequate to its content and scope: are all the parts of the FT contentful and necessary? Next, consider whether the submitted FT is actually correct – are there factual errors or inaccuracies?

Evaluate the logical structure of the FT, the thematic flow between chapters and whether the text is comprehensible to the reader. Assess whether the formal notations in the FT are used correctly. Assess the typographic and language aspects of the FT, follow the Dean's Directive No. 52/2021, Art. 3.

Evaluate whether the relevant sources are properly used, quoted and cited. Verify that all quotes are properly distinguished from the results achieved in the FT, thus, that the citation ethics has not been violated and that the citations are complete and in accordance with citation practices and standards. Finally, evaluate whether the software and other copyrighted works have been used in accordance with their license terms.

Non-written part, attachments

Depending on the nature of the FT, comment on the non-written part of the thesis. For example: SW work – the overall quality of the program. Is the technology used (from the development to deployment) suitable and adequate? HW – functional sample. Evaluate the technology and tools used. Research and experimental work – repeatability of the experiment.

Evaluation of results, publication outputs and awards

Depending on the nature of the thesis, estimate whether the thesis results could be deployed in practice; alternatively, evaluate whether the results of the FT extend the already published/known results or whether they bring in completely new findings.

The overall evaluation

Summarize which of the aspects of the FT affected your grading process the most. The overall grade does not need to be an arithmetic mean (or other value) calculated from the evaluation in the previous criteria. Generally, a well-fulfilled assignment is assessed by grade A.