



Review report of a final thesis

Reviewer: Ing. Tomáš Kalvoda, Ph.D.
Student: Bc. Roman Bushuiev
Thesis title: Self-supervised machine learning for the interpretation of molecular mass spectrometry data
Branch / specialization: Knowledge Engineering
Created on: 5 June 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

The assignment is a challenging one. It requires the student to master an extensive domain-specific knowledge from chemistry to which he was not directly exposed during his studies at FIT. The assignment can be considered fulfilled.

2. Main written part

85 / 100 (B)

The main written part consists of almost 90 pages of dense text. The logical structure of the exposition is fine. The author first introduces necessary notions, reviews related work, and then describes his methods and results. The text is pleasant to read. However, at few places, it is not clear what the author had in mind (e.g., on page 17, the sentence "...express both x_i and y_i in terms of y_i " is vague at best; it is not clear what the author has in mind when speaking about "continuous vector").

The typography and quality of figures and diagrams are excellent. There are only some minor mistakes like "-" vs. "--" and malformed index on page 17 in the definition of the function class.

The bibliography is impressive (144 entries). All items have the required properties and are related to the topic.

My main problem with the written part is that it does not describe the software (python package, scripts, notebooks) the author created and used to obtain the presented results. The `_only_` place where the author references this code is in the appendix "Contents of enclosed CD". I find it quite insufficient as it goes against the possible reproducibility of

the results (how does one obtain the dataset, which script should he run and in what order, what is the purpose of each module, etc.). These aspects could have been described at least in an additional Appendix.

3. Non-written part, attachments

75 /100 (C)

There is a zip archive attached to the thesis. It contains the source code of the LaTeX thesis itself and various Python and bash scripts and Jupyter notebooks.

The LaTeX source code of the thesis is fine, but the student somehow managed to include the pdf of his bachelor thesis (main.pdf). I guess this is just a mistake.

The software part of the thesis (the content of the msml directory) consists of circa 3000 lines of Python code, 450 lines of bash scripts, and 37 Jupyter notebooks (!). Unfortunately, it is not clear how one should use these modules/scripts/notebooks. No "README" or any other pipeline description (like Makefile) exists. The code itself looks genuine and functional. It contains sparse comments and only a few docstrings. Sometimes these are not very helpful, like "TODO: Ask Anton why" in train.sh.

4. Evaluation of results, publication outputs and awards

95 /100 (A)

To the best of my understanding, the results are sound, and the methods employed are adequate for the task. In my opinion, the student has exceeded the usual level of the Master's thesis and proved he can work on cutting-edge research.

The overall evaluation

85 /100 (B)

The written part is excellent but is completely separated from the computational part of the thesis. Then the software part itself is challenging to understand. This situation prevents me from giving the best grade possible.

Questions for the defense

It follows from the text that you had to work with a large dataset (order of TB). Where have you run your computations and experiments? Are there some computational bottlenecks?

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.