



Review report of a final thesis

Reviewer: Ing. Josef Koumar
Student: Bc. Lukáš Jančíčka
Thesis title: Concept drift and model degradation in network traffic classification
Branch / specialization: Knowledge Engineering
Created on: 18 January 2024

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 student studied the network traffic monitoring and detection of data and concept drifts from peer-reviewed journal or conference articles. In the master thesis, the student analyzed the three datasets (MAWIIlab, CESNET-QUIC22, and CESNET-TLSyear23). The student performs analysis of the distribution and detection of data/concept drifts on all of these long-term and big-data datasets. Furthermore, the student made significant observations on these datasets, which resulted in building his own drift detector. The implemented framework for drift detection was well tested, and the results make students' work significant in this domain. This work deviates significantly from the standards of the faculty in terms of difficulty and execution.

2. Main written part 100/100 (A)

The master thesis has the usual structure, is well written, and has an accurate scope. I do not find any mistakes in the typography and language. Citations contain mainly journal and conference articles, citations of software, and some RFC. The citation style is correct, and the sources are highly relevant to the domain. Overall, the text of the master thesis is of high quality, and the text is comprehensible to the reader.

3. Non-written part, attachments 100/100 (A)

The code of the framework is well-written in Python and documented using docstrings. Furthermore, the framework is easy to use for the user and performs what may be expected. Moreover, the non-written part also includes analysis in Jupyter Notebooks. I

would recommend using more markdown cells in Jupyter Notebooks for future work. However, this trifle is insignificant and negligible compared to the other results.

4. Evaluation of results, publication outputs and awards 100/100 (A)

The results are ready to use in practice. The student demonstrates it on pages 46 to 49. The framework is suitable to include in Active Learning Framework to turn on retraining wherever drift is detected. The student demonstrates this suitability in Figure 4.1, and the results are significant to the overall improvement of the Active Learning Framework. Furthermore, the student performed an analysis of existing concept drifts in the network traffic classification domain, which is a hot open challenge in this domain. This work is suitable to be published in peer-reviewed journals or conference proceedings.

The overall evaluation 100/100 (A)

Overall, this master's thesis significantly outperforms the average master's thesis. The thesis is well written and has significant contributions to network traffic monitoring domain, where the concept of drift detection is now a hot open challenge. Therefore, I recommend awarding the dean's prize for an excellent master's thesis.

Questions for the defense

What are the most significant conclusions from the drift analysis in this domain?

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.