FACULTY OF INFORMATION TECHNOLOGY CTU IN PRAGUE

Review report of a final thesis

Student:	Bc. Jan Koza
Reviewer:	Ing. Tom Kocmi
Thesis title:	Semi-supervised learning of deep neural networks
Branch of the study:	System Programming

Date: 21. 1. 2020

Evaluation criterion:	*The evaluation scale: 1 to 4.*
1. Fulfilment of the assignment	***1 = assignment fulfilled,*** *2 = assignment fulfilled with minor objections,* *3 = assignment fulfilled with major objections,* *4 = assignment not fulfilled*

Criteria description:
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.

Comments:

The thesis assignment contains six steps, which the author followed properly. Especially, I'd like to appreciate the detailed statistical assessment of method comparison.

As for the assignment instruction number 4 which states "to propose at least one modification, hybridization or other extension". I would like the author to clarify the fulfilment of this assignment (see questions for defence).

Evaluation criterion:	*The evaluation scale: 0 to 100 points (grade A to F).*
2. Main written part	*60 (D)*

Criteria description:
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. 26/2017, 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.

Comments:

The thesis is written in excellent English with only a few errors, which I appreciate. The text of the thesis is well structured and easy to follow. The list of bibliography is adequate and I find all formal aspects fulfilled such as correct use of acronyms or correct citation style. Although, as for the formal aspects, the author has not met the recommended 50 – 150 pages as specified by "Dean's Directive No. 26/2017, Art. 3.".

The thesis can be divided into two major parts. In the first part (Chapters 1 and 2), the author explains machine learning principles and basics of artificial neural networks, followed by an explanation of semi-supervised learning, its application by ?-model and Temporal Ensembling, that are used in the experimental part of the thesis and also related works on malware detection. The second part is focused on model design and experiments on presented datasets. The author first explains Moon-Shaped Data as a baseline experiment, to confirm that the implementation is correct and is able to learn on baseline dataset. Then he introduces the malware detection dataset provided by Avast, which he analyses and statistically evaluates the performance of his implemented architecture.

However, there are some shortcoming to this thesis. Some concepts are not explained or cited as they are first mentioned, for example, "stochastic augmentation" first mentioned on page 14 is described on page 21. Some other concepts, that are important for the thesis, are not described at all nor they are cited: dropout, stratified random sampling, chained denoising autoencoders or ramp up/down period.
I would appreciate if the hyperparameter setting would be described on one place instead of various pages (for example optimizer setting is on page 19 and 20, but the rest of the model settings is on page 23).
However, I see as the most critical problem of the thesis a lack of explanation of the two main methods used in the paper ?-model and Temporal Ensembling, which are described only briefly by one paragraph and a picture without proper explanation. This makes it hard to understand and since it is a core of the work I believe a proper explanation would be needed.
Overall, a thorough description of the mentioned parts could extend this rather short master thesis and improve its coherence.

Evaluation criterion:	*The evaluation scale: 0 to 100 points (grade A to F).*
3. Non-written part, attachments	*90 (A)*

Criteria description:
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.

Comments:

The thesis contains source codes of implementation of used architecture and its experiments. The source code is clearly structured and well documented.

Unfortunately, I was unable to find the "Avast malware dataset" and thus test the code. There is no proper citation of the dataset in the thesis and no mention, where to obtain the data (there is only a mention in README file, that the data could not be appended to the thesis). I understand, that the data could be proprietary, however, I lack further explanation or the citation of the dataset. The code also does not contain any trained models to be used for evaluation.

Thus, I have not been able to run the scripts and replicate the results.

Evaluation criterion:	*The evaluation scale: 0 to 100 points (grade A to F).*
4. Evaluation of results, publication outputs and awards	*90 (A)*

Criteria description:
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.

Comments:

This criterion is not applicable to this thesis. The thesis is evaluated on the proprietary dataset without citing other works evaluated on it. Thus it cannot be compared with other works and replicated by other researchers.
However, I do not think, that publication of results is a necessary criterion for a successful defence.

Evaluation criterion:	*No evaluation scale.*
5. Questions for the defence	

Criteria description:
Formulate questions that the student should answer during the Presentation and defence of the FT in front of the SFE Committee (use a bullet list).

Questions:

1) Can you clarify which extensions have you proposed (Assignment instruction no. 4)? If those are suggestions on page 40 to use similarity measures: have you considered their interaction with your normalization process (before applying PCA)? Furthermore, how would you compare distances of various features, as they are anonymized and each of them can have a different distribution in their own space?
2) On page 19, you mention that MLP is "a universal model" as a reason for not using CNN in your work. What do you mean by term "universal model"? I agree that the CNN would probably not work on this dataset, but can you clarify why?
3) On page 30: you have done hyperparameter search on individual values separately and then used the best value for each, have you checked if when used together, they also obtain the best results?

Evaluation criterion:	*The evaluation scale: 0 to 100 points (grade A to F).*
6. The overall evaluation	*80 (B)*

Criteria description:
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.

Comments:

The author showed, that he understands the problematics and is able to implement neural architectures. The thesis is written in excellent English and is mostly well structured. The experiments are well-described and their design is solid. Especially, I appreciate that the author first evaluated his implementation over trivial Moon shape recognition task to confirm, that his implementation is able to learn properly. Furthermore, I value greatly the thorough statistical analysis of the results.

My main concerns are that the thesis feels quite short with many explanations of concepts missing. The reader is then left to study cited articles or even search for them due to a lack of citation. Another issue is the lack of replicability due to the use of proprietary dataset without a proper explanation of the source of data or other publications evaluated on the dataset.

Signature of the reviewer: