

Supervisor's statement of a final thesis

Czech Technical University in Prague

Faculty of Information Technology

Student: Bc. Jan Veselý
Supervisor: Ing. Jan Černý
Thesis title: Analýza chování modelů strojového učení
Branch of the study: Knowledge Engineering

Date: 2. 6. 2017

<p><i>Evaluation criterion:</i></p> <p>1. Difficulty and other comments on the assignment</p> <p><i>Criteria description:</i> Characterize this final thesis in detail and its relationships to previous or current projects. Comment what is difficult about this thesis (in case of a more difficult thesis, you may overlook some shortcomings that you would not in case of an easy assignment, and on the contrary, with an easy assignment those shortcomings should be evaluated more strictly.)</p> <p><i>Comments:</i> The difficulty of this assignment lies in the ability to be able to design and implement things in a general way so that it can work for any type of models and also for multiple use cases from automated model testing between different code versions to performance evaluation. Difficulty from the algorithmic point of view is average.</p>	<p><i>The evaluation scale: 1 to 5.</i></p> <p>1 = extremely challenging assignment, 2 = rather difficult assignment, 3 = assignment of average difficulty, 4 = easier, but still sufficient assignment, 5 = insufficient assignment</p>
<p><i>Evaluation criterion:</i></p> <p>2. Fulfilment of the assignment</p> <p><i>Criteria description:</i> Assess whether the thesis meets the assignment statement. In Comments indicate parts of the assignment that have not been fulfilled, completely or partially, or extensions of the thesis beyond the original assignment. If the assignment was not completely fulfilled, try to assess the importance, impact, and possibly also the reason of the insufficiencies.</p> <p><i>Comments:</i> All goals of the assignment were addressed and completed in the thesis.</p>	<p><i>The evaluation scale: 1 to 4.</i></p> <p>1 = assignment fulfilled, 2 = assignment fulfilled with minor objections, 3 = assignment fulfilled with major objections, 4 = assignment not fulfilled</p>
<p><i>Evaluation criterion:</i></p> <p>3. Size of the main written part</p> <p><i>Criteria description:</i> Evaluate the adequacy of the extent of the final thesis, considering its content and the size of the written part, i.e. that all parts of the thesis are rich on information and the text does not contain unnecessary parts.</p> <p><i>Comments:</i> I would like to see more research around metrics for model performance measuring and more detailed analysis of this area, all other parts of the thesis are described adequately.</p>	<p><i>The evaluation scale: 1 to 4.</i></p> <p>1 = meets the criteria, 2 = meets the criteria with minor objections, 3 = meets the criteria with major objections, 4 = does not meet the criteria</p>
<p><i>Evaluation criterion:</i></p> <p>4. Factual and logical level of the thesis</p> <p><i>Criteria description:</i> Assess whether the thesis is correct as to the facts or if there are factual errors and inaccuracies. Evaluate further the logical structure of the thesis, links among the chapters, and the comprehensibility of the text for a reader.</p> <p><i>Comments:</i> Thesis is comprehensible and easily understandable.</p>	<p><i>The evaluation scale: 0 to 100 points (grade A to F).</i></p> <p>85 (B)</p>
<p><i>Evaluation criterion:</i></p> <p>5. Formal level of the thesis</p> <p><i>Criteria description:</i> Assess the correctness of formalisms used in the thesis, the typographical and linguistic aspects, see Dean's Directive No. 14/2015, Article 3.</p> <p><i>Comments:</i> I don't have any major objections on the form of the thesis.</p>	<p><i>The evaluation scale: 0 to 100 points (grade A to F).</i></p> <p>75 (C)</p>
<p><i>Evaluation criterion:</i></p> <p>6. Bibliography</p> <p><i>Criteria description:</i> Evaluate the student's activity in acquisition and use of studying materials in his thesis. Characterize the choice of the sources. Discuss whether the student used all relevant sources, or whether he tried to solve problems that were already solved. Verify that all elements taken from other sources are properly differentiated from his own results and contributions. Comment if there was a possible violation of the citation ethics and if the bibliographical references are complete and in compliance with citation standards.</p>	<p><i>The evaluation scale: 0 to 100 points (grade A to F).</i></p> <p>65 (D)</p>

Comments:

There are only few sources mentioned in the thesis and research part of the metrics could be more detailed.

Evaluation criterion:

The evaluation scale: 0 to 100 points (grade A to F).

7. Evaluation of results, publication outputs and awards

85 (B)

Criteria description:

Comment on the achieved level of major results of the thesis and indicate whether the main results of the thesis extend published state-of-the-art results and/or bring completely new findings. Assess the quality and functionality of hardware or software solutions. Alternatively, evaluate whether the software or source code that was not created by the student himself was used in accordance with the license terms and copyright. Comment on possible publication output or awards related to the thesis.

Comments:

Result of the thesis is a well designed and well programmed system for analyzing behaviour of machine learning models. The system is designed as a general framework which allows to use it for different type of models such as prediction models which works with model matrix or behavioral models working with transactional data. System is also easily scalable and adding new metrics is very easy due to well defined interfaces and project structure.

Second goal was also met by implementing baseline metrics that are most commonly used and also implementing metrics to monitor changes in models prediction with new model deployment. This is used as advanced testing layer to discover bugs before releasing new model source code to the production.

Only downside is that those baseline metrics does not perform well enough on some use cases and there is certainly a space for future work regarding improving those or implementing new ones.

Student used proper technologies suitable for the tasks.

Evaluation criterion:

No evaluation scale.

8. Applicability of the results

Criteria description:

Indicate the potential of using the results of the thesis in practice.

Comments:

Whole system currently runs in production and is processing thousands of jobs per day and sending tens of millions of requests per day to the running models.

Evaluation criterion:

The evaluation scale: 1 to 5.

9. Activity and self-reliance of the student

9a:

1 = excellent activity,

2 = very good activity,

3 = average activity,

4 = weaker, but still sufficient activity,

5 = insufficient activity

9b:

1 = excellent self-reliance,

2 = very good self-reliance,

3 = average self-reliance,

4 = weaker, but still sufficient self-reliance,

5 = insufficient self-reliance.

Criteria description:

Review student's activity while working on this final thesis, student's punctuality when meeting the deadlines and consulting continuously and also, student's preparedness for these consultations. Furthermore, review student's independency.

Comments:

Honza was approaching the task with great responsibility and was meeting the deadlines and also checking the deadlines for himself. It was never necessary to remind him that he should do something and he was always prepared for the meetings regarding the thesis.

Evaluation criterion:

The evaluation scale: 0 to 100 points (grade A to F).

10. The overall evaluation

85 (B)

Criteria description:

Summarize the parts of the thesis that had major impact on your evaluation. The overall evaluation **does not** have to be the arithmetic mean or any other formula with the values from the previous evaluation criteria 1 to 9.

Comments:

Thesis itself represents a very good work in regard of solving challenging task from the research to the production deployment. Thesis contains analysis of the task, service and API design, application of proper technologies and components, implementation of the service, analysis of different metrics and their implementation, experimental evaluation of those metrics and also application to another use cases like prediction models or advanced testing of new model versions to prevent bugs to reach production during source code changes.

My only minor objection is that more time could be spend to test more metrics or try to implement new ones if generally used metrics are not good enough.

Signature of the supervisor: