



Review report of a final thesis

Student: Bc. Grant Zvolský
Reviewer: MSc. Juan Pablo Maldonado Lopez, Ph.D.
Thesis title: Machine Learning Pipeline for Spectroscopic Data
Branch of the study: Knowledge Engineering

Date: 24. 5. 2018

<i>Evaluation criterion:</i>	<i>The evaluation scale: 1 to 5.</i>
1. Difficulty and other comments on the assignment	1 = extremely challenging assignment, 2 = rather difficult assignment, 3 = assignment of average difficulty, 4 = easier, but still sufficient assignment, 5 = insufficient assignment
<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.)	
<i>Comments:</i> I think the difficulty of the assignment is appropriate.	
<i>Evaluation criterion:</i>	<i>The evaluation scale: 1 to 4.</i>
2. Fulfilment of the assignment	1 = assignment fulfilled, 2 = assignment fulfilled with minor objections, 3 = assignment fulfilled with major objections, 4 = assignment not fulfilled
<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.	
<i>Comments:</i> The assignment was fulfilled.	
<i>Evaluation criterion:</i>	<i>The evaluation scale: 1 to 4.</i>
3. Size of the main written part	1 = meets the criteria, 2 = meets the criteria with minor objections, 3 = meets the criteria with major objections, 4 = does not meet the criteria
<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.	
<i>Comments:</i> The content and size is adequate.	
<i>Evaluation criterion:</i>	<i>The evaluation scale: 0 to 100 points (grade A to F).</i>
4. Factual and logical level of the thesis	90 (A)
<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.	
<i>Comments:</i> The thesis is readable.	
<i>Evaluation criterion:</i>	<i>The evaluation scale: 0 to 100 points (grade A to F).</i>
5. Formal level of the thesis	90 (A)
<i>Criteria description:</i> Assess the correctness of formalisms used in the thesis, the typographical and linguistic aspects, see Dean's Directive No. 26/2017, Article 3.	
<i>Comments:</i> The language used is formal and reflects the academic standard expected at this level.	
<i>Evaluation criterion:</i>	<i>The evaluation scale: 0 to 100 points (grade A to F).</i>
6. Bibliography	80 (B)

Criteria description:

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.

Comments:

Many references to existing software solutions are missing, for instance, <http://hugadams.github.io/scikit-spectra/>. There are also many packages in R that work with spectral data of some type or another, which had not been referenced.

Evaluation criterion:

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

7. Evaluation of results, publication outputs and awards

80 (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:

The thesis surveys existing methods, and compares different pipelines, but it is not clear whether the results improve the state of the art in a significant way.

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:

The results are applicable in practice.

Evaluation criterion:

No evaluation scale.

9. Questions for the defence

Criteria description:

Formulate any question(s) that the student should answer to the committee during the defence (use a bullet list).

Questions:

- The set of features in spectroscopic data is more or less continuous, so if you pick a small subset of features randomly, the probability that the noise in those features correlates with the target variable is very high. This is bad news for genetic algorithms or other heuristic approaches that rely strongly on random selection. How could you prevent this from happening?
- How could deep learning tools and ideas be used for feature extraction? For instance, 1-D convolutional neural networks or autoencoders.

Evaluation criterion:

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

10. The overall evaluation

80 (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:

The thesis presents a comprehensive summary of techniques used in spectroscopic data. However, it would have been nice to see more original contributions, even if they did not perform better. Given the typically small size of the training set in this context, one can not expect any major improvement from the modelling side, but there is ample room for creative feature engineering.

I would have also liked to see implementation of some of the preprocessing methods as extensions of scikit-learn classes, as the author suggests in the end.

Signature of the reviewer: