

Assist.-Prof. Mag. Dr. Günter Klambauer

Institute for Machine Learning

Johannes Kepler University, Linz

Supervisor statement – Ekaterina Sysoykova Bachelor Thesis

Ekaterina showed her interest in one of our topics in the course of AI in Life Sciences. Her first project was a QSAR-webservice where she gained more and more experience in programming in python and after that she wanted to continue in this area. With her growing interest in protein modeling, we found relatively fast a topic she would devote herself to for her Bachelor-Thesis: Protein solubility prediction from amino-acid sequence.

Ekaterina started with an extensive literature review, which is very well reflected in her final thesis. She writes about the importance of protein solubility prediction as well as related work and first approaches to this issue.

After fixating a dataset, she started implementing an LSTM based approach. She tried to follow the implementation of DeepSol [1], but we agreed on a simpler approach where she would develop the architecture in a more step-by step manner.

She learned a lot using new frameworks like pytorch, as well as understanding the task at hand better. After several iterations and hard work from her side, a rather good predictive outcome could be observed.

She documented her approach and made code available to ensure reproducibility.

Methodological choices are mostly stated. Results with several metrics are reported and discussed as all significance tests have been performed.

In conclusion, I am very pleased with Ekaterina's work and look forward to future collaboration.

Linz, 21-01-2022

Assist.-Prof. Mag.
Dr. Günter Klambauer

[1] Khurana S, Rawi R, Kunji K, Chuang GY, Bensmail H, Mall R. DeepSol: a deep learning framework for sequence-based protein solubility prediction. Bioinformatics. 2018 Aug 1;34(15):2605-2613. doi: 10.1093/bioinformatics/bty166. PMID: 29554211; PMCID: PMC6355112



Přirodovědecká
fakulta
Faculty
of Science

Jihočeská univerzita
v Českých Budějovicích
University of South Bohemia
in České Budějovice

SUPERVISOR'S STATEMENT ON BACHELOR/DIPLOMA* THESIS

Name of the student: Ekaterina Sysoykova
Study program: Bioinformatics Bachelor
Department/Institute: Institute for Machine Learning
Thesis title: Protein solubility prediction using CNN and CNN-LSTM hybrid models

Supervisor: Assist.-Prof. Mag. Dr. Günter Klambauer, Institut für Machine Learning, JKU
Philipp Seidl, MSc, , Institut für Machine Learning, JKU

Supervisor's affiliation: JKU Linz

Point scale¹ Points

(1) FORMAL REQUIREMENTS

Formal and graphical quality of the thesis	0-3	2
Ability to work with literature	0-3	3
Language and stylistics	0-3	3
Formal requirements – points in total		8

(2) PRACTICAL REQUIREMENTS

Fulfillment of the aims	0-3	3
Ability to understand the results, their interpretation, and clarity of the results, discussion, and conclusions	0-3	2
Discussion quality – interpretation of results and their discussion with the literature	0-3	3
Experimental difficulty of the thesis, independence in experimental work	0-3	3
Contribution of the thesis to the knowledge in the field and the possibility to publish the results (after eventual supplementary experiments)	0-3	2
Practical requirements – points in total		13

POINTS IN TOTAL (MAX/AWARDED)

24

21

* Choose one

¹ Mark as: 0-unsatisfactory, 1-satisfactory, 2-average, 3-excellent.

Comments of the supervisor on the student and the thesis:

Conclusion:

In conclusion, I

~~recommend / do not recommend*~~

the thesis for the defense.

In ~~the~~ date 21.1.2022

.....
signature