



Supervisor's statement of a final thesis

Supervisor: Dr. Ing. Josef Šivic
Student: Bc. Anton Bushuiev
Thesis title: Machine learning for the design of protein–protein interactions
Branch / specialization: Knowledge Engineering
Created on: 9 June 2023

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 thesis fulfils the assignment very well. Sec. 3.2.2 reviews state-of-the-art machine-learning methods for the design of protein-protein interactions, and Secs. 4.3 and 4.4 identify their limitations and benefits (Objective 1). Chapters 5 and 6 address the identified limitations by preparing two datasets of protein-protein interactions and propose a new algorithm iDist (Sec. 5.2.2) for their analysis and a new self-supervised geometric deep learning model PPIformer (Sec. 6.1) for the design of protein-protein interactions (Objective 2). In Chapter 4 the selected tools are combined to propose promising staphylokinase mutations (Objective 3).

2. Main written part

100/100 (A)

The thesis has an excellent logical structure with a nice thematic flow between the chapters. The thesis starts with an introduction (chapter 1) giving the goals, motivation and challenges. Background chapter (chapter 2) gives the necessary background in both biochemistry and machine learning. The thesis then surveys the related work (chapter 3). Then it describes how the state of the art tools have been applied to staphylokinase design (chapter 4), which motivates the need for a new dataset and new self-supervised machine learning methods. Finally, chapter 5 describes the new dataset and chapter 6 the proposed self-supervised approach. The English is also excellent and understandable. The thesis contains a number of illustrations that help the reader to understand the method and its results.

Overall, the text is going to provide a strong basis for writing the submission to ICLR 2024 (International Conference of Learning Representations, CORE A*), which is one of the top-

tier venues in machine learning and where we would like to submit the extended version of this work in September 2023.

3. Non-written part, attachments

100/100 (A)

Anton's work has contributed to new protein designs of staphylkoinase (chapter 4), which are being validated in the wetlab now. For this, Anton used different state-of-the-art methods and prepared a tool that allowed to combine results from the different methods in a friendly and intuitive way. The tool was used by the protein engineers to come up with new designs. Anton has identified the limitations of existing datasets and proposed a new dataset (chapter 5), which is a strong contribution by itself. He has also designed and implemented a new self-supervised learning method (chapter 6) for protein-protein interactions. The method is novel with respect to the current state of the art.

4. Evaluation of results, publication outputs and awards

95/100 (A)

Overall, Anton contributed both to (i) solving a real problem in protein engineering as well as (ii) proposed, implemented and obtained very promising results with a new original self-supervised method for protein-protein interactions. The results do not quite yet outperform the state-of-the-art but are close (hence 95 not 100).

Anton is contributing to three publications that are in preparation:

- His review of state-of-the-art (chapter 3) is contributing to a review paper "Machine Learning-Guided Protein Engineering" which is in final stages of preparation for the journal ACS Catalysis (IF 13.7).
- His designs from the case study (chapter 4) will be part of a publication in the protein engineering domain, once there are experimental results from the lab.
- The new method (chapter 6) is going to be (after some additional work) submitted to ICLR 2024 in September 2023, which is a top-tier machine learning venue.

5. Activity of the student

- ▶ [1] excellent activity
- [2] very good activity
- [3] average activity
- [4] weaker, but still sufficient activity
- [5] insufficient activity

Anton was very proactive and drove the project forward. He participated in the regular meetings with the protein engineering team at Masaryk University and contributed to the protein designs. He became an invaluable member of our team.

6. Self-reliance of the student

- ▶ [1] excellent self-reliance
- [2] very good self-reliance
- [3] average self-reliance
- [4] weaker, but still sufficient self-reliance
- [5] insufficient self-reliance

Anton was self-driven. With very little help, he has

- surveyed the state-of-the-art in the area,
- identified the best methods,
- used them to propose new protein designs,
- identified the main limitations of the state-of-the-art methods and proposed a new original approach addressing those limitations,
- He also obtained the first very promising results.

Anton has demonstrated an outstanding level of independence and originality.

The overall evaluation

100 /100 (A)

Overall, Anton has shown great motivation, originality and independence.

He has contributed both to (i) solving a real problem in protein engineering (the design of new mutations) as well as (ii) proposed, implemented and obtained very promising results with a new original self-supervised method for protein-protein interactions.

Overall this is an outstanding MSc thesis more akin to the work of a first-year PhD student.

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.

Activity of the student

From your experience with the course of the work on the thesis and its outcome, review the student's activity while working on the thesis, his/her punctuality when meeting the deadlines and whether he/she consulted you as he/she went along and also, whether he/she was well prepared for these consultations.

Self-reliance of the student

From your experience with the course of the work on the thesis and its outcome, assess the student's ability to develop independent creative work.

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.