FACULTY OF INFORMATION TECHNOLOGY CTU IN PRAGUE

Supervisor's statement of a final thesis

Supervisor: Ing. Adam Valenta
Student: Bc. Maroš Kramár
Thesis title: Implementation of split guiding metric for Isolation Forest
Branch / specialization: Knowledge Engineering
Created on: 5 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 student fulfilled the requirements of the assignment within its designated scope. Additionally, he displayed a proactive approach by exploring the potential of the FairCutForest algorithm for Novelty detection, demonstrating a commendable level of curiosity and initiative.

2. Main written part   97 /100 (A)

Overall, the student has demonstrated excellent work. He has written the thesis in a professional and clear manner, connecting the chapters seamlessly without any unnecessary parts. The student has effectively incorporated information from other studies, unified terminology, and integrated his own ideas into the work while evaluating their validity.
Furthermore, the student's thoroughness in evaluating anomaly detection performance on various real and generated datasets is commendable.

There are a few things that could be better. The statements like:
- "The most typical data",
- "Anomaly scores can also be converted to binary values with thresholding, but are usually preferred because they carry...",
- "Another widely used method that utilizes density,..."
should be provided with information on how widely used or typical something is, etc.

Additionally, it would be interesting to include an analysis of the AUCPR in section 4.2.1, in addition to the evaluation of time efficiency for the gradFindSplit method. This would provide insights into whether the algorithm can achieve similar AUCPR with a smaller

tree size, potentially improving its time efficiency.

Lastly, only one factual error was found: In H2O terminology, MOJO stands for Model ObJect, Optimized.

3. Non-written part, attachments   95 / 100 (A)

The student used two programming languages, Python and Java. He successfully implemented the algorithm in the H2O-3 software, which is a complex platform that takes time to learn. The student has demonstrated that he can create a program based on studies, test its functionality and correctness, and thoroughly test the prototype for its performance. He also checked the validity of previous studies by implementing the algorithm himself. All of his experiments can be repeated and reviewed by other scientists.

He correctly argued for the choice of the FarCutForest algorithm and implemented it in the H2O-3 platform. The implementation in H2O-3 is functional and effectively extends the portfolio of algorithms.

I have one concern about the Java implementation. It looks very similar to the Extended Isolation Forest code. To improve this, we should integrate the algorithm into the final version of the software using inheritance and a more generalized approach. This way, we can reduce the amount of code that needs to be copied from the original implementation.

4. Evaluation of results, publication outputs and awards   95 / 100 (A)

The student validated the university studies used in his work and presented his own ideas, which he evaluated correctly. The implementation in H2O-3 is still in the Proof of Concept phase. For a full production version, the amount of code to be copied will need to be reduced. However, the time efficiency and performance for anomaly detection are already of a quality that can be used in a production environment.

5. Activity of the student

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

No objections.

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

No objections.

The overall evaluation   97 /100 (A)

The student has successfully completed the assignment in its entirety. In addition, he showed interest in exploring the potential of the FairCutForest algorithm for Novelty detection, which went beyond the original scope of the assignment. The student demonstrated the ability to understand the issues in his field, performed a good research and used his programming skills to evaluate the possibilities found. He was able to argue his conclusions convincingly and successfully translate them into a final contribution that represents a contribution to an open-source product.

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.