
Master Thesis Topic

Extending and Evaluating Alhazen using different Machine Learning approaches

Motivation and Background

Alhazen [1] is a tool that generates a diagnosis that explains why a bug occurs. The tool generates a model, currently a decision tree, based on observed inputs. Afterward, it extracts constraints from the model and generates new inputs which violate those constraints. These new inputs can be executed and the behavior of the software under test can be observed. This information then serves to refine the existing model. More information about Alhazen can be found in the original authors presentation at ESEC/FSE 2020 under [2].

Goals

In this thesis, we want to replace the decision trees in Alhazen with a more advanced, and more powerful machine learning technique. While such a machine learning model may provide better predictions, it is also much harder to interpret or explain. This means, that in contrast to a decision tree, we cannot simply extract constraints needed for the input generation. Therefore, we want to use SHAP [3] (or a similar tool) to extract such constraints.

Description of the Task

- Replace the decision tree underlying Alhazen's approach with another Machine Learning (ML) approach.
- Use SHAP (or similar tool) to extract constraints from the chosen ML algorithm.
- Translate those extracted constraints into the format that the existing generator in Alhazen operates on.
- Evaluate the new results, and compare them to the original Alhazen.

Research Type

Theoretical Aspects: *****

Industrial Relevance: *****

Implementation *****

Prerequisite

The student should be enrolled in the master of computer science program, and has completed the required course modules to begin a masters thesis.

Skills required

Programming skills in Java, Python, and/or Kotlin; understanding of, or willingness to learn, the software engineering and software analysis foundations needed for the project.

References

- [1] [KHS+20] Kampmann, A., Havrikov, N., Soremekun, E. O., & Zeller, A. (2020, November). When does my program do this? learning circumstances of software behavior. In *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (pp. 1228-1239).
- [2] <https://youtu.be/oUaJupWIWqQ>
- [3] <https://github.com/slundberg/shap>

Contacts

Sebastian Müller, Humboldt-Universität zu Berlin, Institut für Informatik, Lehrstuhl Software Engineering, Unter den Linden 6, 10099 Berlin, Germany

Application

Please contact me during my office hours or send me an email with the title: “[ThesisProject]-ExtendingAlhazen” to se-career@informatik.hu-berlin.de