Machine Learning based Code Smell Detection through WekaNose

Umberto Azadi  Francesca Arcelli Fontana  Marco Zanoni
u.azadi@campus.unimib.it {arcelli, marco.zanoni}@disco.unimib.it
Università degli Studi di Milano Bicocca - Milan, Italy
Evolution of Software SystEms and Reverse Engineering Lab

• **Why this tool?**
  Code smells can be subjectively interpreted, the results provided by detectors are usually different, the agreement in the results is scarce and a benchmark for the comparison of these results is not yet available.

• **What is the problem?**
  The main approaches used to detect code smells are based on the computation of a set of metrics. However code smell detectors often use different metrics and/or different thresholds, according to their detection rules. As a result of this inconsistency the number of detected smells can increase or decrease accordingly, and this makes hard to understand when a certain characteristic identifies a code smell or not.

• **What is it for?**
  WekaNose is a tool that allows to perform an experiment to study code smell detection through machine learning techniques. The experiment’s purpose is to select rules and/or obtain trained algorithms, that can classify an instance (method or class) as affected or not by a code smell. These rules have the main advantage of being extracted through an example-based approach, rather than a heuristic-based one.

**Flow graph of the experiment**

**DATASET CREATION**
- Define the code smell
- Access a collection of heterogeneous systems
- Extract code metrics from all systems
- Labeling of the selected candidates
- Use code smell advisors to sample candidates

**MACHINE LEARNING TECHNIQUES**
- Choose the machine learning algorithms
- Perform machine learning parameter optimizations
- Compare machine learning algorithms with each other

**Main features:**
- supports a semi-automated work-flow which aims to study the code smells through a machine learning approach;
- exploits supervised machine learning techniques, to support a learn-by-example process;
- exploits the full interpretability of code smells.

**Future Development:**
- take in consideration more metrics;
- make WekaNose compatible with already existing code smell detectors;
- develop a benchmark platform for comparing code smells detection results and the performances of different tools;
- extend WekaNose by considering other code smells or other issues as antipatterns and architectural smells.

For any further information:
Website: [http://essere.disco.unimib.it/wiki/wekanose](http://essere.disco.unimib.it/wiki/wekanose)
Video-Demo Tutorial: [https://www.youtube.com/watch?v=cUKwipHZDuY](https://www.youtube.com/watch?v=cUKwipHZDuY)

ICSE 2018 Posters – 40th International Conference on Software Engineering
Gothenburg, Sweden