MLinter: Learning Coding Practices from Examples - Dream or Reality?

Corentin Latappy*[1]

[1]LaBRI – Université de Bordeaux (Bordeaux, France) – France

Résumé

Coding practices are increasingly used by software companies. Their use promotes consistency, readability, and maintainability, which contribute to software quality. Coding practices were initially enforced by general-purpose linters, but companies now tend to design and adopt their own company-specific practices. However, these company-specific practices are often not automated, making it challenging to ensure they are shared and used by developers. Converting these practices into linter rules is a complex task that requires extensive static analysis and language engineering expertise. In this paper, we seek to answer the following question: can coding practices be learned automatically from examples manually tagged by developers? We conduct a feasibility study using CodeBERT, a state-of-the-art machine learning approach, to learn linter rules. Our results show that, although the resulting classifiers reach high precision and recall scores when evaluated on balanced synthetic datasets, their application on real-world, unbalanced codebases, while maintaining excellent recall, suffers from a severe drop in precision that hinders their usability.

*Intervenant