Measuring the Stylistic Inconsistency in Software Projects using Hierarchical Agglomerative Clustering

**ABSTRACT**

Background: Although many software engineering methodologies and guidelines are provided, it is common that developers apply their very own programming styles to the source code being produced. These individually preferred programming styles are more comprehensive for themselves, but may well conflict with each other. Thus, the problem of stylistic inconsistency is inevitable during the software development process involving multiple developers, the result is undesirable and that will significantly degrade program readability and maintainability. Aims: Given limited understanding in this regard, we perform an empirical analysis for the purpose of quantitatively measuring the inconsistency degree of programming style within a software project team. Method: We first propose stylistic fingerprints, which are represented as a set of attribute-counting-metrics, in an attempt to characterize different programming styles. Then we adopt the hierarchical agglomerative clustering (HAC) technique to quantitatively measuring the proximity of programming style based on six C/C++ open source projects chosen from different application domains. Results: The empirical results demonstrate the feasibility and validity of our fingerprinting methodology. Moreover, the proposed clustering procedure utilizing HAC algorithm with dendrograms is capable of effectively illustrating the inconsistency degree of programming style among source files, which is significant for future research. Conclusions: This study proposed an effective and efficient approach for analyzing programming style inconsistency, supported by a sound theoretical basis for dealing with such a problem. Ultimately improving program readability and therefore reduce the maintenance overhead for software projects.

This paper will be presented at The 12th International Conference on Predictive Models and Data Analytics in Software Engineering, September 7, 2016, Ciudad Real, Spain

Supervisor: Dr Jacky Keung

Research Interests: Software Engineering; Programming Style; Empirical Experiments; Data Mining and Analytics; Metrics and Measurements; Machine Learning

All are welcome!