World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:16, No:06, 2022

Multi-Label Approach to Facilitate Test Automation Based on Historical Data

Authors : Warda Khan, Remo Lachmann, Adarsh S. Garakahally

Abstract : The increasing complexity of software and its applicability in a wide range of industries, e.g., automotive, call for enhanced quality assurance techniques. Test automation is one option to tackle the prevailing challenges by supporting test engineers with fast, parallel, and repetitive test executions. A high degree of test automation allows for a shift from mundane (manual) testing tasks to a more analytical assessment of the software under test. However, a high initial investment of test resources is required to establish test automation, which is, in most cases, a limitation to the time constraints provided for quality assurance of complex software systems. Hence, a computer-aided creation of automated test cases is crucial to increase the benefit of test automation. This paper proposes the application of machine learning for the generation of automated test cases. It is based on supervised learning to analyze test specifications and existing test implementations. The analysis facilitates the identification of patterns between test steps and their implementation with test automation components. For the test case generation, this approach exploits historical data of test automation projects. The identified patterns are the foundation to predict the implementation of unknown test case specifications. Based on this support, a test engineer solely has to review and parameterize the test automation components instead of writing them manually, resulting in a significant time reduction for establishing test automation. Compared to other generation approaches, this ML-based solution can handle different writing styles, authors, application domains, and even languages. Furthermore, test automation tools require expert knowledge by means of programming skills, whereas this approach only requires historical data to generate test cases. The proposed solution is evaluated using various multi-label evaluation criteria (EC) and two small-sized real-world systems. The most prominent EC is ‘Subset Accuracy’. The promising results show an accuracy of at least 86% for test cases, where a 1:1 relationship (Multi-Class) between test step specification and test automation component exists. For complex multi-label problems, i.e., one test step can be implemented by several components, the prediction accuracy is still at 60%. It is better than the current state-of-the-art results. It is expected the prediction quality to increase for larger systems with respective historical data. Consequently, this technique facilitates the time reduction for establishing test automation and is thereby independent of the application domain and project. As a work in progress, the next steps are to investigate incremental and active learning as additions to increase the usability of this approach, e.g., in case labelled historical data is scarce.

Keywords : machine learning, multi-class, multi-label, supervised learning, test automation
Conference Title : ICNLPU 2022 : International Conference on Natural Language Processing and Understanding
Conference Location : Barcelona, Spain
Conference Dates : June 09-10, 2022