ABSTRACT

The process of test case generation is a critical component in software testing. However, automatic generation methods based on Large Language Models (LLMs), such as GPT, still face challenges in producing consistent test case outputs when given predefined requirements. These models often generate test cases with unstable formats, structures, and quantities, even for similar input requirements. This study proposes an approach to improve the consistency of test case generation by enhancing the GPT model's understanding of software requirements through a *continual pretraining* process using domain-specific requirement data that has been preprocessed using Natural Language Processing (NLP) techniques.

The dataset used for model training consists of functional requirements from a hospital management system. The GPT model, after undergoing continual domain-adaptive pretraining, is then compared with the base ChatGPT model using several evaluation metrics: requirement coverage, Jaccard similarity, and automated execution via Selenium. The results show that the pretrained model achieved 100% requirement coverage, a Jaccard similarity score of 0.78, and generated consistent test cases that were immediately executable. In contrast, the base model exhibited instability in both the quantity and structure of generated test cases. This study demonstrates that enhancing a GPT model's requirement comprehension through domain-specific continual pretraining can significantly improve the consistency of automatically generated test cases, thus supporting more effective software testing automation.

Keywords: test case generation, GPT, continual pretraining, NLP, requirement coverage, Selenium