Informatics Institute of Technology

In Collaboration With

University of Westminster, UK

University of Westminster, Coat of Arms

BugScope: A Deep Learning Approach to Bug Localization using Neural Networks

A dissertation by

Ms. Bodiyabaduge Fernando
w1761086 / 2019322

Supervised by:
Mrs. S. R. B. N. Helmini

Submitted in partial fulfilment of the requirements for the BEng in Software Engineering degree at the University of Westminster

May 2023

ABSTRACT

In this thesis, we present a novel deep-learning approach to bug localization using neural networks, by employing the pre-trained model, GraphCodeBERT and training it with the XLCoST dataset. The primary goal of this research is to improve the efficiency and accuracy of bug localization in software development, which is a critical aspect of the software maintenance process. The suggested method aims to address the drawbacks of conventional Information Retrieval (IR) and Convolutional Neural Network (CNN) based approaches, which often struggle to accurately identify the location of bugs within a large codebase.

To accurately capture the complex links between code snippets and their related descriptions, our model combines the GraphCodeBERT transformer, a cutting-edge pre-trained model, with cross-attention mechanisms. In addition, we incorporate a customized weighting scheme that further refines the model's ability to identify relevant code components associated with specific bugs. The model's ability to concentrate on the most important features of the code is improved by this weighting method which enhances the model's capacity to focus on the most pertinent aspects of the code.

Another contribution of this research is the integration of data from attachments in bug reports, such as log files, screenshots, and configuration files, which provides additional contextual information for the model to accurately localize bugs. By considering this supplementary information, the model is better equipped to identify relevant code components and isolate the root cause of the bug.

The results of our experiments show that the proposed deep learning strategy surpasses existing IR and CNN-based methods in terms of efficiency and accuracy. The success of this approach lays the groundwork for future research in the field of software engineering, with potential applications in other areas, such as code comprehension, code generation, and automated debugging.

Keywords: Bug Localization, Deep Learning, Source Code Analysis, GraphCodeBERT