World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:16, No:11, 2022

Code Embedding for Software Vulnerability Discovery Based on Semantic Information

Authors : Joseph Gear, Yue Xu, Ernest Foo, Praveen Gauravaran, Zahra Jadidi, Leonie Simpson

Abstract : Deep learning methods have been seeing an increasing application to the long-standing security research goal of automatic vulnerability detection for source code. Attention, however, must still be paid to the task of producing vector representations for source code (code embeddings) as input for these deep learning models. Graphical representations of code, most predominantly Abstract Syntax Trees and Code Property Graphs, have received some use in this task of late; however, for very large graphs representing very large code snip- pets, learning becomes prohibitively computationally expensive. This expense may be reduced by intelligently pruning this input to only vulnerability-relevant information; however, little research in this area has been performed. Additionally, most existing work comprehends code based solely on the structure of the graph at the expense of the information contained by the node in the graph. This paper proposes Semantic-enhanced Code Embedding for Vulnerability Discovery (SCEVD), a deep learning model which uses semantic-based feature selection for its vulnerability classification model. It uses information from the nodes as well as the structure of the code graph in order to select features which are most indicative of the presence or absence of vulnerabilities. This model is implemented and experimentally tested using the SARD Juliet vulnerability test suite to determine its efficacy. It is able to improve on existing code graph feature selection methods, as demonstrated by its improved ability to discover vulnerabilities.

Keywords : code representation, deep learning, source code semantics, vulnerability discovery
Conference Title : ICPC 2022 : International Conference on Program Comprehension
Conference Location : Amsterdam, Netherlands
Conference Dates : November 03-04, 2022