Adaptive Online Learning for Software Defect Prediction with Defect Overlooking Consideration

Vanama Nikhith

23VV1F0027

February, 2025

Abstract

- This study explores online learning techniques for adaptive software defect prediction.
- Introduces a novel contrastive learning-based approach to address defect overlooking.
- Evaluates the effectiveness of incremental learning models across real-world datasets.

Introduction

- Software defect prediction helps in identifying defect-prone code modules early.
- Traditional static models fail to adapt to evolving software environments.
- Online learning enables models to continuously update with new data.

Why Online Learning?

- Incremental learning adapts to evolving defect patterns in software projects.
- Reduces memory consumption by learning from new data without retraining.
- Enhances real-time defect detection capabilities.

Addressing Defect Overlooking

- Defect overlooking occurs due to class imbalance and feature drift.
- Uses SMOTE to handle imbalanced datasets.
- Applies cost-sensitive learning to reduce misclassification of defects.

Contrastive Learning Approach

- Leverages pre-trained models (e.g., CodeBERT, GraphCodeBERT) for feature representation.
- Uses contrastive loss to improve defect classification by learning meaningful embeddings.
- Enhances defect detection by distinguishing between defective and non-defective code snippets.

Evaluation Metrics

- Uses Precision, Recall, F1-score, and AUC for evaluation.
- Compares batch learning vs. online learning models.
- Evaluates impact of contrastive learning on defect classification.

Experimental Results

- Online learning outperforms batch learning in adapting to evolving defect patterns.
- SMOTE improves F1-score by mitigating class imbalance.
- Contrastive learning enhances defect classification accuracy.

The End