Hybrid Adaptive Ensemble Model with Dynamic Feature Augmentation for Software Defect Prediction in Highly Imbalanced and Evolving Datasets

Salahuddin Shaikh
Tianfu Jiangxi Laboratory, Chengdu Sichuan, China.

ABSTRACT

This study introduces an innovative hybrid adaptive ensemble model with dynamic feature augmentation for predicting software defects in datasets that are highly imbalanced and continuously changing. Software defect prediction is an essential component of software quality assurance; however, it encounters substantial challenges stemming from the intrinsic class imbalance in defect datasets and the evolving nature of software projects, resulting in concept drift. Our proposed model tackles these problems by using an adaptive ensemble framework that combines several machine learning classifiers (SVM, Random Forest, and Gradient Boosting). The model may adjust to changing data distributions by assigning each classifier a dynamic weight according to its current performance. To further integrate features into the model in real time, we use Variational Autoencoders (VAEs). The temporal adaptation technique, including drift detection, ensures the model's functionality persists despite changes in the software environment. Our hybrid model surpasses conventional techniques on benchmark NASA datasets, attaining substantial improvements in accuracy, recall, and F1-score. This enhances accuracy and reliability in forecasting the occurrence of software defects. These advancements not only contribute to the robustness of the forecasting system but also pave the way for more proactive maintenance strategies. By using real-time data and learning that adjusts over time, software teams can fix problems before they get worse, which helps the project run better overall.

REFERENCES

1. Zhao, Y., & Chen, H. (2023). Deep Incremental Learning of Imbalanced Data for Just-In-Time Software Defect Prediction. arXiv preprint arXiv:2310.12289.
2. Khakhar, P., & Dubey, R. K. (2022). The integrity of machine learning algorithms against software defect prediction. In Artificial Intelligence and Machine Learning for EDGE Computing (pp. 65- 74). Academic Press.
3. Thomas, R. N., & Roopam, G. (2020). A survey on machine learning approaches and its techniques. In 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), 1–6.
4. Tufano, M., Watson, C., Bavota, G., Penta, M. D., White, M., & Poshyvanyk, D. (2019). An empirical study on learning bug-fixing patches in the wild via neural machine translation. ACM Transactions on Software Engineering and Methodology (TOSEM), 28(4), 1-29.
5. Zimmermann, T., & Nagappan, N. (2008, May). Predicting defects using network analysis on dependency graphs. In Proceedings of the 30th international conference on Software engineering (pp. 531-540).