Towards Robustness of Deep Program Processing Models – Detection, Estimation and Enhancement

HUANGZHAO ZHANG, Peking University, China
ZHIYI FU, Peking University, China
GE LI*, Peking University, China
LEI MA, University of Alberta, Canada
ZHEHAO ZHAO, Peking University, China
HUA'AN YANG, Peking University, China
YIZHE SUN, Peking University, China
YANG LIU, Nanyang Technological University, Singapore
ZHI JIN*, Peking University, China

Deep learning (DL) has recently been widely applied to diverse source code processing tasks in the software engineering (SE) community, which achieves competitive performance (*e.g.*, accuracy). However, the robustness, which requires the model to produce consistent decisions given minorly perturbed code inputs, still lacks systematic investigation as an important quality indicator. This paper initiates an early step and proposes a framework CARROT for robustness detection, measurement, and enhancement of DL models for source code processing. We first propose an optimization-based attack technique CARROT_A to generate valid adversarial source code examples effectively and efficiently. Based on this, we define the robustness metrics and propose robustness measurement toolkit CARROT_M, which employs the worst-case performance approximation under the allowable perturbations. We further propose to improve the robustness of the DL models by adversarial training (CARROT_T) with our proposed attack techniques. Our in-depth evaluations on three source code processing tasks (*i.e.*, functionality classification, code clone detection, defect prediction) containing more than 3 million lines of code and the classic or SOTA DL models, including GRU, LSTM, ASTNN, LSCNN, TBCNN, CodeBERT and CDLH, demonstrate the usefulness of our techniques for ❶ effective and efficient adversarial example detection, ❷ tight robustness estimation, and ❸ effective robustness enhancement.

The full paper of our work is published and available at https://dl.acm.org/doi/10.1145/3511887. Huangzhao Zhang would attend the conference and present this paper.

ACM Reference Format:
Huangzhao Zhang, Zhiyi Fu, Ge Li, Lei Ma, Zhehao Zhao, Hua'an Yang, Yizhe Sun, Yang Liu, and Zhi Jin. 2022. Towards Robustness of Deep Program Processing Models – Detection, Estimation and Enhancement. *ACM Trans. Softw. Eng. Methodol.* 1, 1, Article 1 (January 2022), 1 page. https://doi.org/10.1145/3511887

*Corresponding authors.

Authors' addresses: Huangzhao Zhang, zhang_hz@pku.edu.cn, Peking University, China; Zhiyi Fu, ypfzy@pku.edu.cn, Peking University, China; Ge Li, lige@pku.edu.cn, Peking University, China; Lei Ma, ma.lei@acm.org, University of Alberta, Canada; Zhehao Zhao, zhaozhehao@pku.edu.cn, Peking University, China; Hua'an Yang, chrisyoung@pku.edu.cn, Peking University, China; Yizhe Sun, yizhe@pku.edu.cn, Peking University, China; Yang Liu, yangliu@ntu.edu.sg, Nanyang Technological University, Singapore; Zhi Jin, zhijin@pku.edu.cn, Peking University, China.

1049-331X/2022/1-ART1
https://doi.org/10.1145/3511887