Improving Fault Localization with Pre-training

Zhuo ZHANG, Ya LI, Jianxin XUE, Xiaoguang MAO

Frontiers of Computer Science, DOI: 10.1007/s11704-023-2597-8

Problems & Ideas

- Problems of fault localization:
  - Existing fault localization techniques mainly focus on coverage information of statements and rarely consider program tokens and their relationships responsible for failures

- Ideas:
  - A new perspective to customize a pre-trained model for fault localization to learn representations from the training samples. It could learn the features of tokens in an execution. PetraFL uses slicing technology and test cases to extract code snippets as the training samples

Experimental results and Conclusions

- Experimental results:
  - An expempirical research on real large-sized programs with 14 stateof-the-art fault localization approaches
  - Improve fault localization effectiveness
- Conclusions:PetraFL is verified as able to improve fault localization effectiveness

MLP-FL	14	CNN-FL	12	BiLSTM-FL	14
DeepFL	11	FLUCCS	11	ER5	13
GP02	10	GP03	11	Dstar	9
ER1’	10	GP19	13	Ochiai	10

Wilcoxon-Signed-Rank test of PetraFL versus baselines