Politechnika Wrocławska

Projekt badawczo-rozwojowy w inżynierii oprogramowania

Grupa piątek 9:15 – 13:00

On Improving Line-Level Defect Prediction: An Evaluation and Enhancement of the DeepLineDP Model Using the Defectors Dataset

Patrycja Kałużna (252864)

Jakub Walaszek (252897)

Spis treści:

REPRODUKCJA BADAŃ

Spis treści:

ROZWÓJ BADAŃ

- wybór tematu rozwoju projektu
- wybór zbioru danych
- pytania badawcze
- przygotowanie podzbiorów danych pod pytania badawcze
- infrastruktura badawcza online

- artykuł naukowy

Wybór tematu projektu

REPRODUKCJA BADAŃ

DeepLineDP: Towards a Deep Learning Approach for Line-Level Defect Prediction

Chanathip Pornprasit, *Student Member, IEEE* and Chakkrit (Kla) Tantithamthavorn, *Member, IEEE*

Abstract—Defect prediction is proposed to assist practitioners effectively prioritize limited Software Quality Assurance (SQA) resources on the most risky files that are likely to have post-release software defects. However, there exist two main limitations in prior studies: (1) the granularity levels of defect predictions are still coarse-grained and (2) the surrounding tokens and surrounding lines have not yet been fully utilized. In this paper, we perform a survey study to better understand how practitioners perform code inspection in modern code review process, and their perception on a line-level defect prediction. According to the responses from 36 practitioners, we found that 50% of them spent at least 10 minutes to more than one hour to review a single file, while 64% of them still perceived that code inspection activity is challenging to extremely challenging. In addition, 64% of the respondents perceived that a line-level defect prediction tool would potentially be helpful in identifying defective lines. Motivated by the practitioners' perspective, we present DeepLineDP, a deep learning approach to automatically learn the semantic properties of the surrounding tokens and lines in order to identify defective files and defective lines. Through a case study of 32 releases of 9 software projects, we find that the risk score of code tokens varies greatly depending on their location. Our DeepLineDP is 17%-37% more accurate than other file-level defect prediction approaches; is 47%-250% more cost-effective than other line-level defect prediction approaches; and achieves a reasonable performance when transferred to other software projects. These findings confirm that the surrounding tokens and surrounding lines should be considered to identify the fine-grained locations of defective files (i.e., defective lines).

Infrastruktura badawcza online

REPRODUKCJA BADAŃ

- korzysta ze środowiska Google Colaboratory
- konfiguruje środowisko conda'y
- pozwala na reprodukcję większości badań z wyjątkiem tzw. line-level baselines

https://colab.research.google.com/drive/139uWve5H07uM0SIKZSuevsi-dEjWeK9P?usp=sharing

Infrastruktura badawcza offline

REPRODUKCJA BADAŃ

- ze względu na fakt, że środowisko Google Colaboratory nie pozwala na reprodukcję tzw. line-level baselines została stworzona infrastruktura badawcza offline w postaci skryptu, który reprodukuje te badania lokalnie na komputerze
- sprawdza czy wszystkie wymagane zależności są zainstalowane lokalnie na komputerze i informuje o tym

Infrastruktura badawcza offline

REPRODUKCJA BADAŃ

- reprodukuje tzw. line-level baselines jeśli wszystkie wymagane zależności są zainstalowane
- wyniki reprodukcji tzw. line-level baselines należy ręcznie połączyć razem z pozostałymi wynikami reprodukcji badań

https://github.com/pwr-pbr23/M8/blob/main/Reproduction/DeepLineDP_line_level_baselines_local_reproduction.bat

Przebieg reprodukcji

REPRODUKCJA BADAŃ

- wytrenowano wszystkie modele wykorzystywane w ramach opracowywanego artykułu naukowego oraz dokonano predykcji przy ich pomocy, tj.:

- model DeepLineDP
- modele tzw. file-level baselines
- modele tzw. line-level baselines

na postawie łącznie 32 release'ów pochodzących z 9 otwartoźródłowych projektów

Wyniki reprodukcji

REPRODUKCJA BADAŃ

- wykresy są generowane w środowisku Google Colaboratory o typie runtime'u R
- wszystkie wymagane zależności są instalowane w środowisku
- wszystkie wyniki reprodukcji są pobierane z Dysku Google, a wygenerowane wykresy są zapisywane na nim

Wyniki reprodukcji

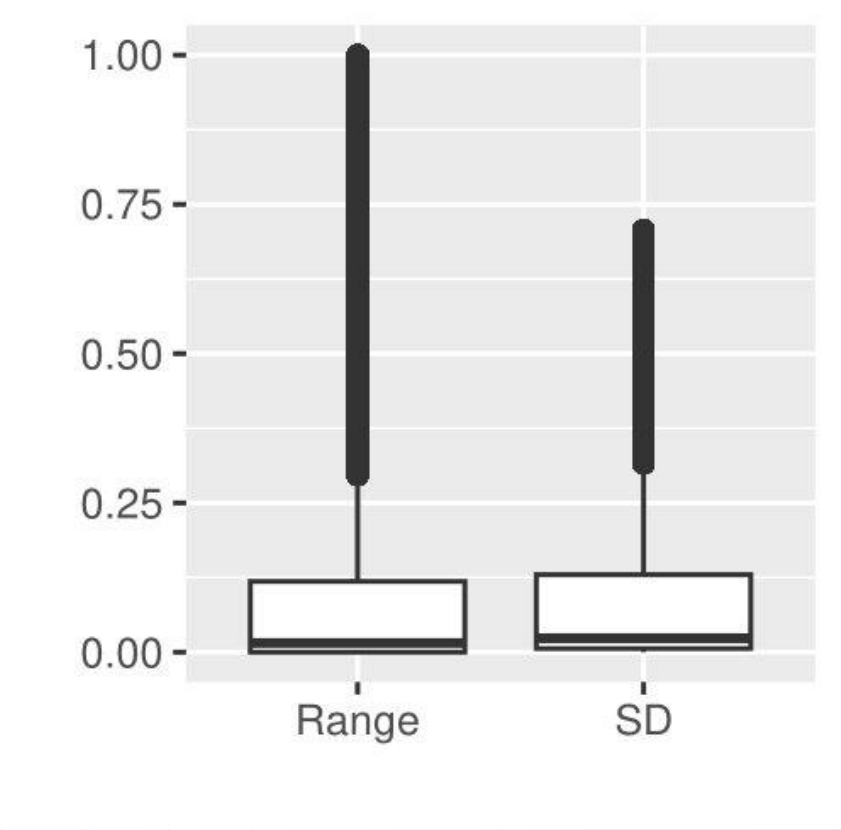
REPRODUKCJA BADAŃ

- R skrypt generujący wykresy wymagał poprawek – udało się poprawić go do tego stopnia, że generuje wykresy dla 2 z 4 pytań badawczych

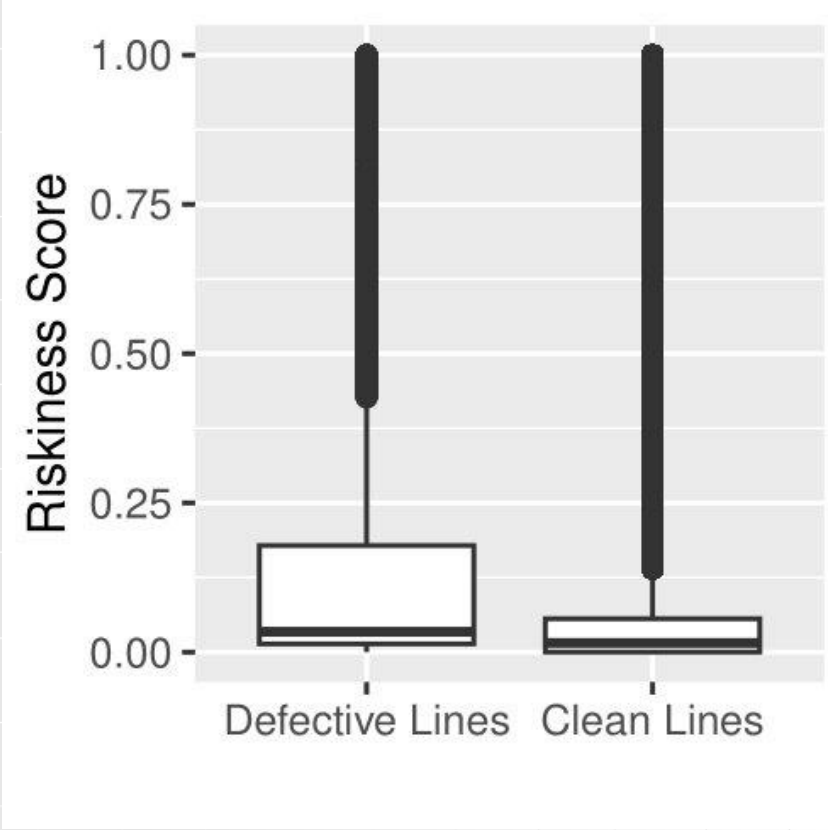
https://colab.research.google.com/drive/1rglM2qt-w5JA-PXk2WnGHK_eAml3Wqae?usp=sharing

Wyniki reprodukcji – wykresy dla RQ1

REPRODUKCJA BADAŃ



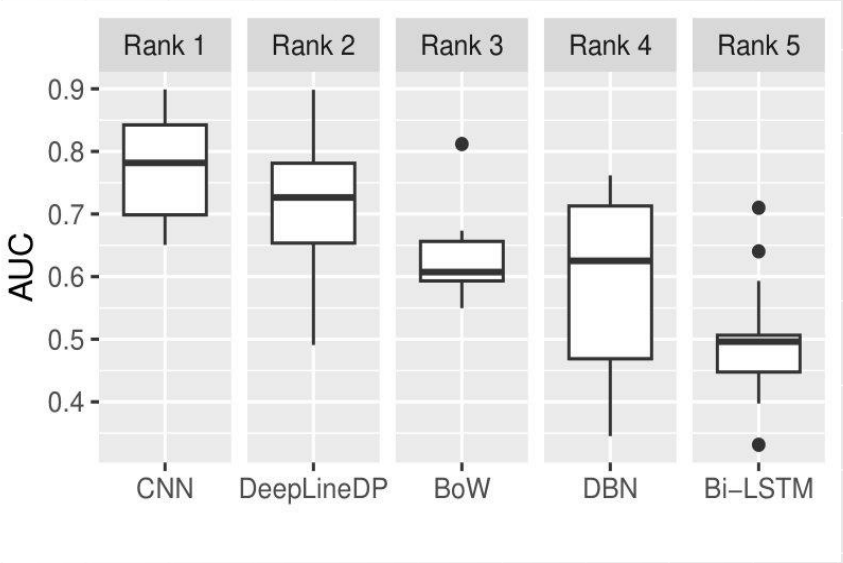
The distribution of the range (i. e., Max-Min) of the risk scores of tokens in a file



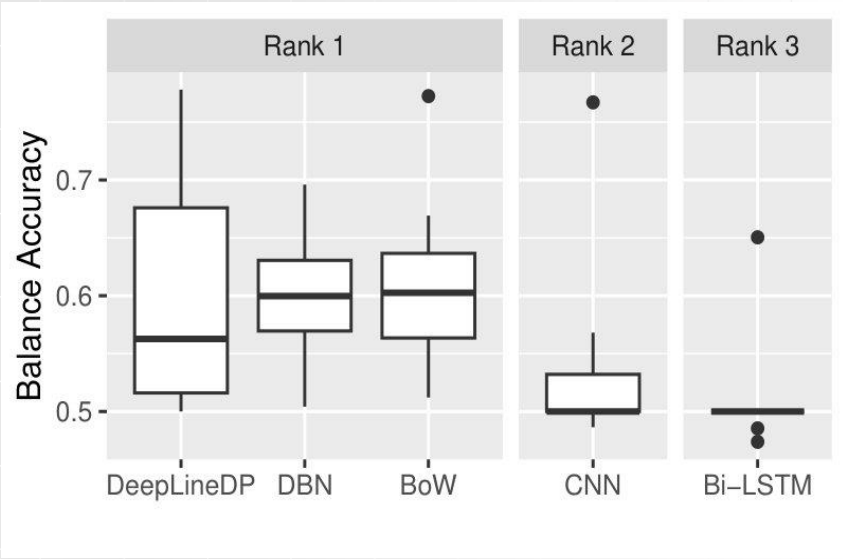
The distribution of the risk scores between actual defective lines and actual clean lines

Wyniki reprodukcji – wykresy dla RQ2

REPRODUKCJA BADAŃ



AUC-ROC

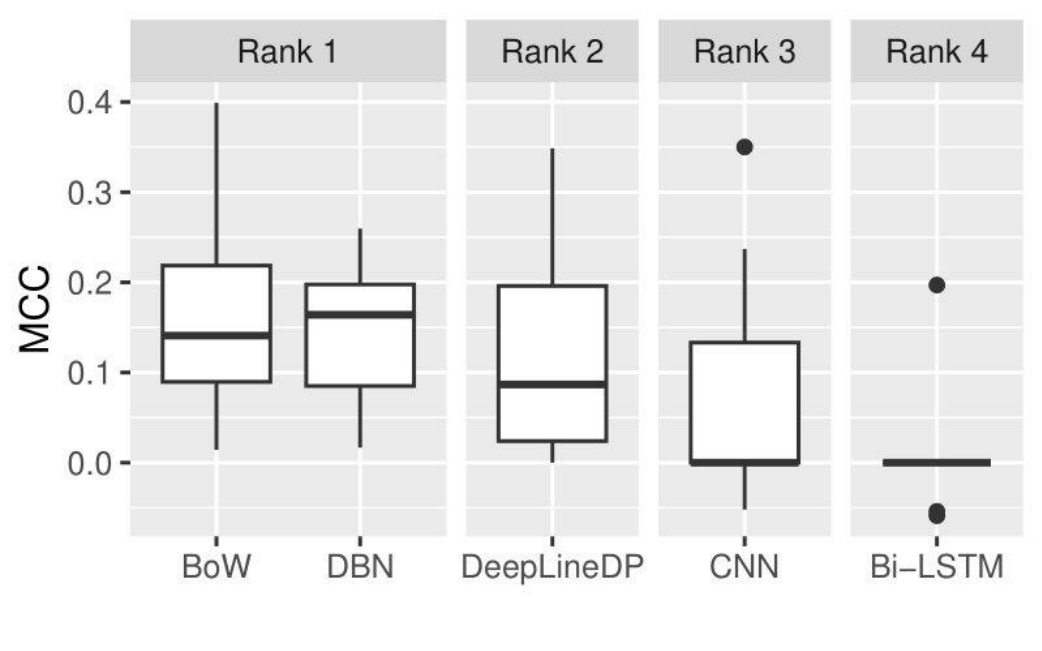


Balanced Accuracy

The ScottKnott ESD ranking and the distributions of the AUC, Balanced Accuracy and MCC of DeepLineDP and the state-of-the-art file-level defect prediction approaches. The higher the values are, the better the approach is.

Wyniki reprodukcji – wykresy dla RQ2

REPRODUKCJA BADAŃ



Mathew Correlation Coefficients (MCC)

The ScottKnott ESD ranking and the distributions of the AUC, Balanced Accuracy and MCC of DeepLineDP and the state-of-the-art file-level defect prediction approaches. The higher the values are, the better the approach is.

Wybór tematu rozwoju projektu

ROZWÓJ BADAŃ

- pomysłem na rozwój badań było poszerzenie zbioru danych wykorzystywanego do wytrenowania modelu DeepLineDP
- wybrano zbiór danych "Defectors" pochodzący z artykułu naukowego "Defectors: A Large, Diverse Python Dataset for Defect Prediction"
- tytuł naszego artykułu naukowego: „On Improving Line-Level Defect Prediction: An Evaluation and Enhancement of the DeepLineDP Model Using the Defectors Dataset”

Wybór zbioru danych

ROZWÓJ BADAŃ

- zbiór danych "Defectors" jest:
- około 2 razy większy niż dotychczas największy zbiór danych wykorzystywany przy predykcji defektów,
- lepiej zbalansowany niż inne zbiory danych (stosunek instancji zawierających defekty do instancji niezawierających ich wynosi około 1:1)
- zawiera dane z 24 projektów pochodzących z 24 organizacji oraz 18 domen takich jak ML, automation i IoT
- oparty o pythonowe projekty kiedy prawie wszystkie zbiory danych są oparte o projekty napisane w Javie
- nowy - artykuł naukowy, z którego pochodzi, został opublikowany 8 marca 2023 roku

Politechnika Wrocławska

Pytania badawcze

ROZWÓJ BADAŃ

(RQ1) Can the DeepLineDP model make better predictions if it is trained on the ”Defectors” dataset which is bigger and its data comes from more number of software projects coming from more diversed domains than in a default dataset?

(RQ2) Can the DeepLineDP model make better predictions if it is trained on the ”Defectors” dataset which is better balanced and its data comes from more number of software projects coming from more diversed domains than in the default dataset?

Politechnika Wrocławska

Pytania badawcze

ROZWÓJ BADAŃ

(RQ3) Can the DeepLineDP model make better predictions if it is trained on the default dataset and its parameters are also tuned?

(RQ4) Can the DeepLineDP model make better predictions if it is trained on the ”Defectors” dataset and its parameters are also tuned?

(RQ5) Can the DeepLineDP model be used to predict defects in software projects written in two different programming languages - Java and Python?

Przygotowanie podzbiorów danych pod pytania badawcze

ROZWÓJ BADAŃ

- ze względu na fakt, że zbiór danych „Defectors” jest w formacie Apache Parquet w celu kompatybilności z posiadanymi skryptami przekonwertowano go na format csv,
- w celu przekonwertowania formatu zbioru danych „Defectors” i jego preprocessingu przygotowano specjalne skrypty

https://github.com/pwr-pbr23/M8/blob/main/Reproduction/Defectors_converting_to_csv_and_preprocessing.ipynb

https://github.com/pwr-pbr23/M8/blob/main/Reproduction/DeepLineDP/script/preprocess_data_python.py

Infrastruktura badawcza online

ROZWÓJ BADAŃ

- korzysta ze środowiska Google Colaboratory
- importuje przekonwertowane i poddane preprocessingowi podzbiory danych z Dysku Google
- konfiguruje środowisko conda'y
- uczy modele i generuje predykcje
- eksportuje wyniki badan na Dysk Google

https://colab.research.google.com/drive/1IPa3uUJq5pp6JZCgie_G34mhz2m8TbqK?usp=sharing

Artykuł naukowy

- przeprowadzono przegląd literatury
- napisano sekcje „Introduction” i „Related work”
- rozpoczęto pisanie sekcji „Methods”

https://www.overleaf.com/project/6401cc2de33881644150cd5f

Politechnika Wrocławska

Dziękujemy za uwagę!

Politechnika Wrocławska

Bibliografia

- Pornprasit C., Tantithamthavorn C., *DeepLineDP: Towards a Deep Learning Approach for Line-Level Defect Prediction*. IEEE Transactions on Software Engineering, tom 49, 2023.
- Mahbub P., Shuvo O., Rahman M. M., Defectors: *A Large, Diverse Python Dataset for Defect Prediction*. arXiv preprint arXiv:2303.04738, 2023.