Summary of Attempt Three – A validation

POSTER: Vulnerability Discovery with Function Representation Learning from Unlabeled Projects

Writer:
Fan Zhang
Tianlei Qi

Date: 30/05/2021

Third prediction attempt

LSTM

Total 2011 samples.

1708 training samples. 303 test samples

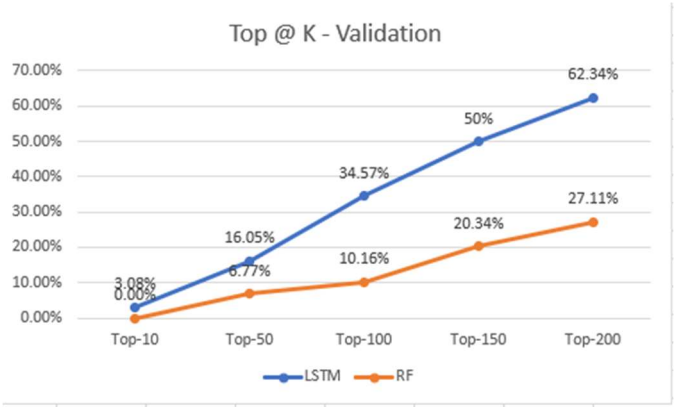
Top-10: 3.08%	Top-1%: 0%
Top-50: 16.05%	Top-5%: 4.93%
Top-100: 34.57%	Top-10%: 9.25%
Top-150: 50%	Top-20%: 21.60%
Top-200: 62.34%	Top-30%: 31.48%

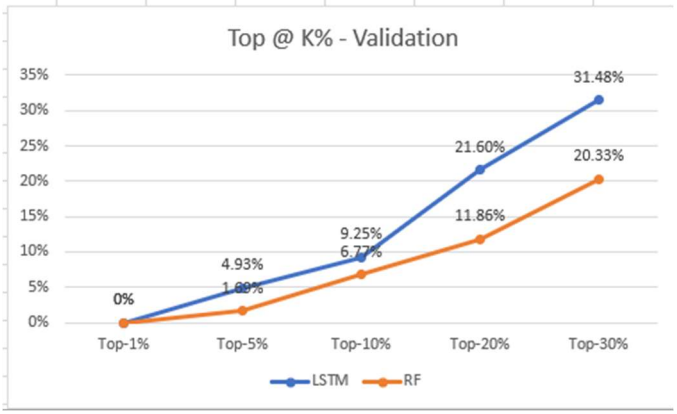
RF

Total 2011 samples.

1498 training samples. 513 test samples (Random)

TOP-10: 0.00%	Top-1%: 0%
TOP-50: 6.77%	Top-5%: 1.69%
TOP-100: 10.16%	Top-10%: 6.77%
TOP-150: 20.34%	Top-20%: 11.86%
TOP-200: 27.11%	Top-30%: 20.33%





What have we done different from previous attempt?

Our previous attempts are using the datasets provided by the paper author. The LSTM and Random Forest results of paper datasets are similar with paper result. In order to validate the methods, we found another dataset from the paper: SySeVR: A Framework for Using Deep Learning to Detect Software Vulnerabilities. SeVC dataset focuses on 1,591 open source C/C++ programs from the NVD and we used them as our dataset. In order to adapt to the data set, we have adjusted our code to extract representations from the functions.

What do you think of the performance of the detection? Is it good or not good? Why?

From the result accuracy point of view, the LSTM method has a better performance compared to the RF method due to the higher Top@K values.

From the efficiency point of view, the LSTM utilized lower CPU resource than the RF method if using same dataset.

What is your conclusion? Could you give the final conclusion about the method based on the current result?

The purpose of this attempt is to validate the credibility of the methods and datasets provided by the paper. So that we can ensure the authenticity of the performance differences between the LSTM and RF shown on the chart.

Through the analysis of the results of the third experiment, we obtained a similar performance trend with that in our second attempt and paper. Based on all the evidence listed above, we could conclude that the paper has provided a valid and convinced vulnerability detection method.