ROSE at ESEC/FSE 2018

Abstract Submission

TITLE A replication of "Easy over Hard: A Case Study on Deep Learning"

WHO Authors of the reproduced work: Wei Fu and Tim Menzies [1]
Title of the paper: Keep It Simple: Is Deep Learning Good for Linguistic Smell Detection?
Authors: Sarah Fakhoury, Venera Arnaoudova, Cedric Noiseux, Foutse Khomh, and Giuliano Antoniol

WHAT A replication of a comparative study investigating the performance of deep learning techniques versus traditional machine learning techniques for software engineering tasks.

WHY Although deep learning is reported to produce results comparable and even superior to human experts for certain complex tasks, they are costly approaches that impose high time and space constraints. Additionally, deep learning techniques have been shown to produce similar results to traditional machine learning techniques for some tasks. Thus, researchers and practitioners should be careful when selecting machine learning models for their problems and be sure to explore the full potential of traditional machine learning models via hyper parameter tuning before turning to more taxing approaches which may not always be a good fit for software engineering tasks.

HOW Our comparative study follows the same methodology as the original authors, by comparing deep learning and sufficiently tuned traditional machine learning algorithms. We use the same deep learning algorithm, CNN, and explore similar tuning algorithms for the traditional machine learning techniques.

WHERE The original work by Fu and Menzies explores the performance of deep learning for linking questions on Stack Overflow whereas our work considers the task of linguistic smell detection. Additionally, Fu and Menzies use SVM as the traditional machine learning algorithm whereas we use Random Forest, SVM, J48, and Naive Bayes. Fu and Menzies use Differential Evolution to tune the parameters of SVM whereas we use basic parameter tuning for all traditional machine learning algorithms and we use more sophisticated tuning, based on Bayesian optimization, to find the best model and features. The original work uses the results for the CNN task from previous work by Xu et al. [2] and does not implement the CNN architecture directly. Therefore, to implement our own CNN we replicate the proposed CNN architecture and tuning described by Zhang and Wallace [3].

DISCUSSION The proper tuning of traditional machine learning algorithms is widely explored and documented in the literature. Replicating the methodology of Fu and Menzies was straightforward and easily adaptable to the task of linguistic smell detection using pre-existing tools for hyperparameter tuning. However, the tool and details on CNN architecture implementation were not provided in the work by Xu et al. and subsequently, the work by Fu and Menzies. Therefore, we used the architecture described by Zhang and Wallace [3], which may or may not correspond to the best hyperparameter setting and architecture and thus could potentially introduce implementation bias when results are compared between studies. Given that we confirm the results by Fu and Menzies, our findings could be interpreted as being more robust because they hold despite being tested on a different software engineering task and network configuration. Still, providing an online appendix and sharing implementation code is essential to facilitate the reproducibility and comparison of results.

[1] W. Fu and T. Menzies, “Easy over hard: A case study on deep learning,” in Proc. of the Joint Meeting on Foundations of Software Engineering (FSE), 2017, pp. 49–60

[2] Bowen Xu, Deheng Ye, Zhenchang Xing, Xin Xia, Guibin Chen, and Shanping Li. 2016. Predicting semantically linkable knowledge in developer online forums via convolutional neural network. In Proceedings of the 31st IEEE/ACM International Conference on Automated Software Engineering. ACM, 51–62.

[3] Y. Zhang and B. Wallace, “A sensitivity analysis of (and practitioners’ guide to) convolutional neural networks for sentence classification ,” CoRR, 2015. [Online]. Available: http://arxiv.org/abs/1510.03820