

Deep Learning for Cyber Security in Scientific Computing

Steven Young, Robert Patton, Thomas Karnowski, Derek Rose, Thomas Potok

Oak Ridge National Laboratory

Research performed using scientific computing has a profound effect on society, on national budgets, on the economy, and upon lives of individuals. Thus, scientific computing resources are prime targets for malicious attacks that poison results or prevent research from being completed. Without trust in the results of scientific research, it cannot be acted upon. Thus, developing systems to protect scientific resources is important not only maintaining the availability of these systems, but in maintaining the public's trust in the results obtained by utilizing them.

Deep Learning (DL) provides a powerful tool by which the massive amounts of data produced by scientific computing systems can be leveraged for cyber security. DL methods have been applied to large computer vision data sets for classification and general object recognition with great success [1,3,6] along with other domains such as speech recognition [2], hardware prognostics [5], and traffic prediction [4]. DL methods have been successful in these domains because DL relies on local proximity (typically spatial and/or temporal) among patterns to find and construct higher order patterns. Similar proximity properties exist in scientific computing systems, but new developments are needed to enable the application of DL to this domain.

Although much of the DL literature focuses on image classification problems, it has been successfully applied to many other domains. One successful application is traffic prediction where mean residual error (MRE) was improved by more than 30% over more traditional machine learning techniques such as shallow neural networks, support vector machines, and radial basis function neural networks [4]. This success in prediction using sensor data is exciting when you consider the volume of data that can be collected from scientific computing systems, and that prediction tasks do not need a human to create a labeled data set. Successes like this indicate that DL methods could be used to identify patterns of malicious behavior on scientific systems. Additionally, DL codes exist that can target either GPUs or CPUs. This would allow a DL code being used for cyber security to be used on a variety of systems and would also allow the code to better utilize unused resources on heterogeneous systems.

If labeled training data were available, DL could be used to explicitly classify malicious behavior without the need for an expert to create rules for what defines malicious behavior. Any sensor data or log data available from the system (e.g. temperature, utilization, memory usage, etc.) could be used as input to the DL algorithm. The DL would then learn the spatial and temporal dependencies in the data, such as the relationships in temperatures between nodes across time, in order to classify behavior as normal or malicious. Many benefits would result from such a system. It would reduce the amount of manual effort required to identify patterns of malicious behavior on scientific systems and could utilize unused resources on HPC systems whether they are CPUs or GPUs. If DL could bring the improvements in performance to cyber security that it has brought to computer vision and speech recognition, it could provide improved trust in scientific systems.

References

- [1] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Delving Deep into Rectifiers: Surpassing Human-Level Performance on Imagenet Classification." arXiv Preprint arXiv:1502.01852, 2015. <http://arxiv.org/abs/1502.01852>.
- [2] Hinton, Geoffrey, Li Deng, Dong Yu, George Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, et al. "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups." IEEE Signal Processing Magazine 29, no. 6 (November 2012): 82–97. doi:10.1109/MSP.2012.2205597.
- [3] Ioffe, Sergey, and Christian Szegedy. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift." arXiv Preprint arXiv:1502.03167, 2015. <http://arxiv.org/abs/1502.03167>.
- [4] Lv, Yisheng, Yanjie Duan, Wenwen Kang, Zhengxi Li, and Fei-Yue Wang. "Traffic Flow Prediction With Big Data: A Deep Learning Approach." IEEE Transactions on Intelligent Transportation Systems, 2014, 1–9. doi:10.1109/TITS.2014.2345663.
- [5] Ng, Andrew. "Deep Learning: What's Next." GPU Technology Conference, March 19, 2015.
- [6] Wu, Ren, Shengen Yan, Yi Shan, Qingqing Dang, and Gang Sun. "Deep Image: Scaling up Image Recognition." arXiv Preprint arXiv:1501.02876, 2015. <http://arxiv.org/abs/1501.02876>.