RESEARCH SUMMARY

Understanding the Limits of Artificial Intelligence for Warfighters

Volume 2, Distributional Shift in Cybersecurity Datasets

JOSHUA STEIER, ERIK VAN HEGEWALD, ANTHONY JACQUES, GAVIN S. HARTNETT, LANCE MENTHE

To access the full report, visit www.rand.org/t/RRA1722-2

ISSUE

The occurrence of distributional shift can reduce the performance of machine-learning (ML) systems. This issue is of particular relevance to cybersecurity datasets because the signatures of cyberattacks can shift rapidly and unpredictably in many different ways—the data environment is both high-dimensional and highly nonstationary. In seeking a solution to the detection of such a shift along with mitigation, we can create and enhance ML models so that they are more robust and effective. Therefore, detecting and mitigating the adverse effects of distributional shift is paramount to effectively defending against cyberattacks.

APPROACH

We chose several publicly available cybersecurity benchmark datasets, specifically those for network intrusion detection and malware classification, to investigate distributional shift. Each dataset is a state-of-the-art benchmark dataset used by researchers to study ML algorithm performance and to further research in the field. We used statistics-based methods and database segmentation to detect the shift, and we proposed mitigation methods.

Furthermore, we created a deep neural network for network intrusion detection and decision tree-based methods for malware classification. We used the open-source framework TensorFlow to create and use these tools and evaluated ML algorithm performance on new and recent data.

KEY FINDINGS

- Cybersecurity datasets suffer from distributional shift, especially in standard network intrusion detection and malware classification datasets.
- Distributional shift can be characterized in multiple ways, and the ease of detection depends on the dataset.
- Although data quality is important in training ML algorithms, the recency of the data is also significant.
- Cases in which the data must be recent to be useful limit the data available for training, which in turn bound artificial intelligence (AI) performance.

continued

RECOMMENDATIONS

- Dataset segmentation tests, such as those demonstrated in this report, should be performed to determine the significance of distributional shift for any AI system for cybersecurity. From these tests, a data decay rate can be estimated, which yields a rough estimate of the AI system's effective shelf life, after which it must be retrained with more recent data.
- Well-known statistical tests should be performed on the dataset as an additional measure to detect distributional shift.

RAND PROJECT AIR FORCE

RAND Project AIR FORCE (PAF), a division of the RAND Corporation, is the Department of the Air Force's (DAF's) federally funded research and development center for studies and analyses, supporting both the United States Air Force and the United States Space Force. PAF provides DAF with independent analyses of policy alternatives affecting the development, employment, combat readiness, and support of current and future air, space, and cyber forces. For more information, visit PAF's website at www.rand.org/paf.