Machine Learning for Network Analysis

IA 473 – Assignment Four

Jesse Russell

10/27/2021

Until recently, network admins have had to perform a tedious and time-consuming job when analyzing their organization's network. From data flow analysis, to intrusion detection, to anomaly detection, to malware analysis, these administrators worked tirelessly to ensure that the system or network runs smoothly. This process requires time, knowledge, and manpower that can ultimately slowdown a company's productivity when performed by a human. That's where machine learning (ML) comes into play. ML is a method of handling network analysis that allows for computers to do the heavy lifting of performing these tedious tasks, freeing up the network administrator's time to put out the other fires that may spring up across the company. ML works by first reading the training data, creating a model based on that training data, making a prediction based upon the data model, and then checking whether the prediction is accurate. Once the accuracy is verified, it's ready to be implemented. If the accuracy is not up to required standards, however, new training data will be fed to the machine until the desired level of accuracy is achieved. The more data the machine has, the more accurately it will be able to predict the behavior of unknown data sets. (Alqudah and Yassen) Alqudah and Yassen describe two types of machine learning: supervised and unsupervised. Supervised learning is fed labelled data, which allows the platform to accurately predict input and output and adjust its learning accordingly. Meanwhile, unsupervised learning feeds the platform raw and unlabeled input data only, without much need for human intervention. This can be used for gathering large demographic data that doesn't necessarily have to be organized and associated with other data sets. While this breakthrough technology can change how we handle network security, there are plenty of advantages and drawbacks to discuss while adopting this model.

First, why should you adopt machine learning into your company's network? As mentioned in the introduction, ML automates tedious tasks while possibly saving the company time and money. ML can be used for many tasks in network analysis, but it mainly boils down to two main concepts: Network Based Defense and Intrusion Detection. When implemented for network defense, the trained AI may be able to analyze packets coming in and out of the network as well as filter any unapproved packets that could cause harm. This is helpful for network administrators because computers can see certain patterns in packets that humans would miss. For instance, if 1,000 packets came into the network and 20 of them came from the same external source, a machine would be able to organize all the packets from that source and try to determine what the purpose of these packets are. If this were to be done manually, the 20 similar packets may be lost in the larger sea of packets and the patterns would be extremely difficult to figure out. Using ML for this purpose may also allow the system to filter malicious packets that are sent sporadically over a long period of time, such as weeks or months. Attackers may be sending small packets to avoid detection and using ML would be able to see the patterns that will mostly likely be missed without it. The machine can then use these patterns to improve its database of likely malicious actors and filter them out easier when they come up again. A machine learning environment can also determine how much of a threat a certain packet or group of packets presents to the network and filter them accordingly. The machine may see an unrecognized packet, but it can look at the metadata, including source, protocol, port, destination, etc. to determine if it needs to be filtered or not. If a packet says it's an http packet but uses a different port, this may raise a red flag for further review by the algorithm.

On top of filtering packets, ML can take a more passive approach and handle intrusion detection. During this process, the algorithm will constantly be looking at the regular usage of the network and adjust itself if any authorized or unauthorized changes occur. Using this method of intrusion detection, it can also determine if there is any suspicious activity in the network baseline, then raise a red flag for either human review or the network defense system to deal with. Besides finding suspicious activity, the system can detect normal changes to the network, such as new computers or new departments that were added to the network. This allows the company to expand as much as they need to without redefining new network baseline parameters for the IDS. This again can save time and money for the network admins to use for other important tasks. The longer the algorithm is in place, the better idea it can get about the normal operations of the business. This way, if a certain subnet or department is using an unusually high amount of bandwidth, this may be an indicator of intrusion and may warrant further investigation.

ML is extremely malleable and can be customized to whatever specifications the company needs, all in a fraction of the amount of time it would take for a human to do the same job. Using ML as opposed to a preprogrammed system also allows the algorithm to adjust itself, without much human intervention. Using a static system would require regular audits and updates, which would require hiring more manpower and time wasted. ML diminishes this requirement by essentially maintaining itself through constant reeducation and updates. This does not mean, however, that these algorithms will completely circumvent the need for a network administrator. These methods require human knowledge and action to set up and audit. It’s unlikely that this tech will make the network admin job obsolete.

As previously mentioned, implementing AI/machine learning can be immensely helpful for keeping the company's network afloat and consistent. Though, with heavy reliance on technology to perform tasks, there are plenty of downsides to consider when determining if ML is right for your needs. First, machine learning can be expensive to implement. Many companies have an existing network infrastructure that was built long before ML was widely available. Completely rebuilding this from the ground up in order to adopt ML can be costly, and may not even provide the necessary savings return to make the rebuild worth it. If a company is small and doesn't have much traffic to analyze, they may not even need to spend the money and time in implementing it. It may actually end up costing them more in the long run, than if they were to just hire an Administrator to maintain the IDS and IPS systems. The most expensive part of implementing the system is the large amount of data needed to produce accurate results. Most AI systems fully implemented in the real world need roughly 100,000 – 1 Million datasets for strong prediction confidence. (Dimesional Research) Producing this data can be highly costly and ineffective for smaller companies. Each company should run a cost benefit analysis before looking into implementing an ML system. According to the same study by Dimensional Research, 81% of survey participants said that training AI was more difficult than anticipated. Unforeseen circumstances like this can halt business operations until the AI can be properly trained by knowledgeable ML professionals, which can cost even more money. Finding these experts to hire can also be a herculean task that may not provide a return for certain industries.

Bibliography

Alqudah, Nour and Qussai Yassen. "Machine Learning for Traffic Analysis: A Review." *Science Direct* (2020): 911-916.

Dimesional Research. "Artificial Intelligence and Machine Learning Projects Are Obstructed by Data Issues." Global Survey. 2019.