Anomaly Detection of SVM under Poisoning Attack

W.D. Samanwickrama[a*], I.J. Amadoru[b], and L.D.R.D Perera[a]

Faculty of Applied Sciences, Wayamba University of Sri Lanka, Sri Lanka[a]
Faculty of Agriculture and Plantation Management, Wayamba University of Sri Lanka, Sri Lanka[b]

dumindu@wyb.ac.lk*

Abstract

Machine Learning (ML) applications have been adopted heavily due to the use of artificial intelligence systems, cloud computing, social media and smart computing. Venders integrate ML into products across various industries. ML systems train models periodically to enhance the ad hoc functionality. However, data poisoning has been identified as a challenge in ML, and this can be occurred by an injection attack or a label flipping. A hacker needs to know the existing system, and they have to craft and transplant compromised data points avoiding outlier regions for an injection attack. In label flipping, the hackers should access training data systems and make alterations or periodically insert data into the systems through a proper channel trending towards wrong decision making. Support Vector Machine (SVM) is an algorithm which is commonly used in ML, but the algorithm has become a target for data poisoning. The objective of this study was to develop early identification parameters in order to check whether a SVM model is attacked by data poisoning or not.

Danmini Doorbell (DDb) data in University of California, Irvine (UCI) machine learning repository was used in this experiment. Each record contained N=115 features which were generated by the publishers of the dataset using the raw attributes of network traffic. The top 20 (n) was picked from the total N features using the Gini index obtained by the Random Forest algorithm in order to test them according to a reduced feature set architecture. Accuracy and kappa were calculated using One Class SVM model to identify a poisoning attack on the training data set.

Table 1: The change of accuracy and kappa values in data poisoning

	Sample	**A**	**B**	**C**	**D**	**E**	**F**
All features 115	kappa	0.8876	0.8894	0.8697	0.2349	0.2112	-0.0423
	Accuracy	0.9951	0.9951	0.9941	0.8909	0.8800	0.8657
Top 20 features	kappa	0.8851	0.4254	0.2261	0.1135	0.0997	-0.0446
	Accuracy	0.9950	0.9502	0.8847	0.7900	0.7783	0.8243
	% poisoning	0	0.0099	0.0199	0.2	0.2999	1

According to the results, in the all-features set architecture, the accuracy and the kappa values were decreased with the increase in data poisoning (Table 1). Similar results were also obtained in the reduced-features set architecture as well.

The results revealed that the SVM anomalous behavior due to data poisoning can be identified by checking accuracy and kappa on training data sets. However, in order to make firm applications it should be further investigated with different data sets and algorithms.

Keywords: Accuracy; Data poisoning; Kappa; Machine learning