World Academy of Science, Engineering and Technology
International Journal of Computer and Systems Engineering
Vol:17, No:10, 2023

A Genetic Algorithm Based Ensemble Method with Pairwise Consensus Score on Malware Cacophonous Labels

Authors : Shih-Yu Wang, Shun-Wen Hsiao

Abstract : In the field of cybersecurity, there exists many vendors giving malware samples classified results, namely naming after the label that contains some important information which is also called AV label. Lots of researchers relay on AV labels for research. Unfortunately, AV labels are too cluttered. They do not have a fixed format and fixed naming rules because the naming results were based on each classifiers' viewpoints. A way to fix the problem is taking a majority vote. However, voting can sometimes create problems of bias. Thus, we create a novel ensemble approach which does not rely on the cacophonous naming result but depend on group identification to aggregate everyone's opinion. To achieve this purpose, we develop an scoring system called Pairwise Consensus Score (PCS) to calculate result similarity. The entire method architecture combine Genetic Algorithm and PCS to find maximum consensus in the group. Experimental results revealed that our method outperformed the majority voting by 10% in term of the score.

Keywords : genetic algorithm, ensemble learning, malware family, malware labeling, AV labels
Conference Title : ICCST 2023 : International Conference on Computical Science and Technology
Conference Location : Rome, Italy
Conference Dates : October 09-10, 2023