Assessing Modality Selection Heuristics to Improve Multimodal Deep Learning for Malware Detection

Farzana Ahamed Bhuiyan, Katherine Brown, Md Bulbul Sharif, Douglas A. Talbert, PhD
Tennessee Tech University, Cookeville, TN

Introduction

With the growing use of Android devices, security threats are also increasing. While there are some existing malware detection methods, cybercriminals continue to develop ways to evade these security mechanisms. Thus, malware detection systems also need to evolve to meet this challenge. This work is a step towards achieving that goal. Malware detection methods need as much information as possible about the potential malware, and a multimodal approach can help in this regard by combining different aspects of an Android application. Multiple modalities can improve classification by providing complementary information, however, the use of all available modalities does not necessarily maximize algorithm performance. Thus, multimodal machine learning could benefit from a mechanism to guide the selection of modalities to include in a multimodal model. This work uses a malware detection problem to compare multiple heuristics for this selection process and the assumptions behind them.

Multimodal Learning

- Modality refers to how something occurs or is experienced.
- When it involves various modalities, a research problem is described as multimodal.
- Helps us to get different aspects of an event of interest.

Data Collection

- We built a labeled multimodal dataset for malware analysis for android devices.
- The dataset consists of several hundred known malicious applications as well as benign applications.
- The malicious applications were verified and confirmed to be of a malicious content by using well-known studied dataset.
- These datasets were examined via VirusTotal to confirms its classification.

Experimental Methodology

Framework

- Raw data Extraction Process
- Modality Extraction Process
- Feature Vector Generation Process
- Detection Process

Available Modalities

- String Modality
- Method Opcode Modality
- Method API Modality
- Shared Library Function Opcode Modality
- Manifest Modality

Modality Selection

- All combination of modalities?
- Heuristic Selection?
  - Forward Stepwise Selection using heuristics
    - ❖ The maxDifference heuristic
    - ❖ The maxSimilarity heuristic
    - ❖ The maxAccuracy heuristic

Greedy Forward Stepwise Selection

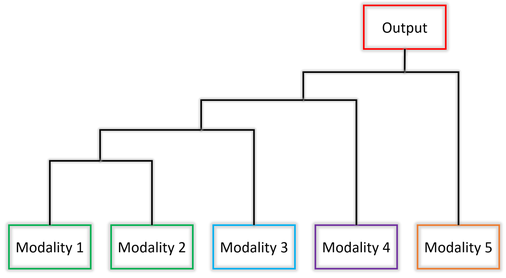


Table 1: Configuration of our Multimodal Neural Network with just two modalities

Layers	Input Shape	Number of Units	Activation Function
Input	(None, 39512)	39512	ReLU
Hidden	(None, 39512)	2000	ReLU
Hidden	(None, 2000)	1000	ReLU
Merge	[(None, 1000), (None, 1000)]	2000	ReLU
Hidden	(None, 2000)	100	ReLU
Hidden	(None, 100)	10	ReLU
Output	(None, 10)	1	Sigmoid

Results

Table 2: Performance Measure for Forward Selection using maxDifference Heuristic

Modality	Accuracy	Precision	Recall	F-score
String + Function opcode	0.9731	0.98	0.97	0.97
String + Function opcode + Manifest	0.9819	0.98	0.98	0.98
String + Function opcode + Manifest + Method opcode	0.9802	0.97	0.97	0.97
String + Function opcode + Manifest + Method opcode + Method API	0.9594	0.96	0.96	0.96

Table 3: Performance Measure for Forward Selection using maxSimilar Heuristic

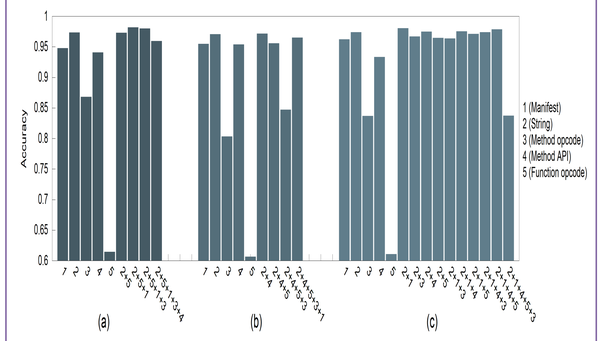
Modality	Accuracy	Precision	Recall	F-score
String + Method API	0.9716	0.97	0.97	0.97
String + Method API + Function opcode	0.9559	0.96	0.96	0.96
String + Method API + Function opcode + Method opcode	0.8473	0.75	0.81	0.77
String + Method API + Function opcode + Method opcode + Manifest	0.9652	0.97	0.96	0.96

Table 4: Performance Measure for Forward Selection using maxAccurate Heuristic

Modality	Accuracy	Precision	Recall	F-score
String + Manifest	0.9804	0.98	0.98	0.98
String + Method opcode	0.9667	0.97	0.96	0.97
String + Method API	0.9750	0.98	0.97	0.97
String + Function opcode	0.9648	0.97	0.96	0.97
String + Manifest + Method opcode	0.9638	0.97	0.96	0.97
String + Manifest + Method API	0.9755	0.97	0.97	0.97
String + Manifest + Function opcode	0.9711	0.97	0.97	0.97
String + Manifest + Method API + Method opcode	0.9741	0.98	0.97	0.97
String + Manifest + Method API + Function opcode	0.9785	0.98	0.96	0.97
String + Manifest + Method API + Function opcode + Method opcode	0.8375	0.96	0.96	0.96

- Our experiments show that selecting modalities with low predictive correlation works better than the other examined heuristics.
- This heuristic method can improve the stability and accuracy of our malware detection algorithms while reducing the overall cost.

Accuracy Comparison using Three Different Heuristics



(a) maxDifferent Heuristic (b) maxSimilar Heuristic (c) maxAccurate Heuristic

Discussion

- The combination found may not be the absolute best one since the first selected model, even though it is the best from a single unimodal model point of view, may not be the ideal one when multiple models are combined in a more sophisticated way.
- We got an improved overall performance and have a simpler model with fewer connections than just using every modality together.

Conclusion

- Using the greedy step-wise forward selection, we get the optimal accuracy for the maxDifference heuristic
- Using this heuristic, we reduced our modalities from five to three.
- We do not need highly accurate unimodal models, we need models that make different kinds of errors
- High accuracy can be accomplished if different models misclassify different training examples, even if the unimodal classifier accuracy is low

References

1. Kim, T.; Kang, B.; Rho, M.; Sezer, S.; and Im, E. G. 2018. A mul- timodal deep learning method for android malware detection using various features. *IEEE Transactions on Information Forensics and Security* 14(3):773–788.
2. 2. Lahat, D.; Adali, T.; and Jutten, C. 2015. Multimodal data fusion: an overview of methods, challenges, and prospects. *Proceedings of the IEEE* 103(9):1449–1477.
3. Liu, K.; Li, Y.; Xu, N.; and Natarajan, P. 2018. Learn to combine modalities in multimodal deep learning. *arXiv preprint arXiv:1805.11730*.