MANDIANT
NOW PART OF Google Cloud

Efficient Malware Analysis Using Metric Embeddings

Authors: Ethan Rudd, David Krisiloff, Daniel Olszewski, Edward Raff, James Holt

Presenter: Ethan Rudd

Staff Data Scientist

About This Project

Authors + Affiliations

- A collaboration between Mandiant Inc. and UMD Laboratory for Physical Sciences (LPS).

- Began during a Mandiant internship (2020); later sponsored by LPS.

- Current Affiliations:
  - Mandiant: Ethan Rudd, David Krisiloff
  - Booz Allen Hamilton: Edward Raff
  - LPS: James Holt
  - University of Florida: Daniel Olszewski

Motivation

ML Malware Analysis: Existing and Idealized Solutions

Commercial ML Malware Analysis Solutions → Idealized Generic Embeddings → Our Approach: Metric Learning

Motivation

Issues with Existing Solutions

Commercial ML Malware Analysis Solutions → Idealized Generic Embeddings → Our Approach: Metric Learning

- Lots of industry focus on detection.
- Numerous other ML malware analysis use-cases
  - Classification, information retrieval, and analysis contextualization.
- Issues with training task-specific representations:
  - Time + resource intensive
  - Limited number of labeled samples
  - Model update + storage complexity

a) Malware Family

Shlayer	ZeuS	Agent Tesla

b) Attribute tags

Downloader	Ransomware	Spyware

c) ATT&CK TTP Summarization

Initial Access	Execution	Defense Evasion

d) Exploited Vulnerability Analysis

CVE-2022-0010	CVE-2021-0002	CVE-2022-0067

e) Authorship Attribution

APT-12	APT-33	APT-42

Examples of additional malware analysis use cases and labeling.

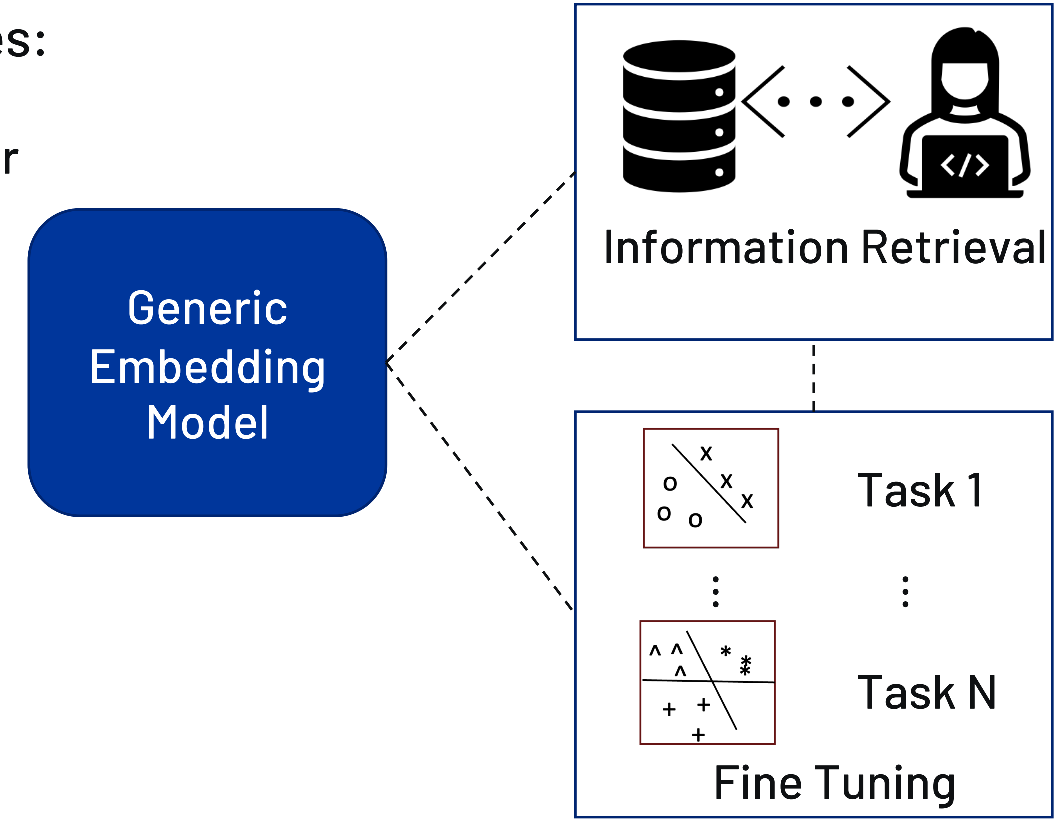
Motivation

ML Malware Analysis: Existing and Idealized Solutions

Commercial ML Malware Analysis Solutions → **Idealized Generic Embeddings** → Our Approach: Metric Learning

Learn a base model with the following attributes:

- **Incorporate contextual/semantic data** useful for multiple problem scenarios
- **Transferable** to different analysis tasks with minimal additional telemetry/labeling
- **Low-dimensional** output representation
  - portability
  - transfer training efficiency
  - Indexing and information retrieval support



Motivation

ML Malware Analysis: Existing and Idealized Solutions

Commercial ML Malware Analysis Solutions → Idealized Generic Embeddings → **Our Approach: Metric Learning**

- Use **<u>metric learning</u>** to arrive at a generic representation where neighboring samples are contextually and semantically similar

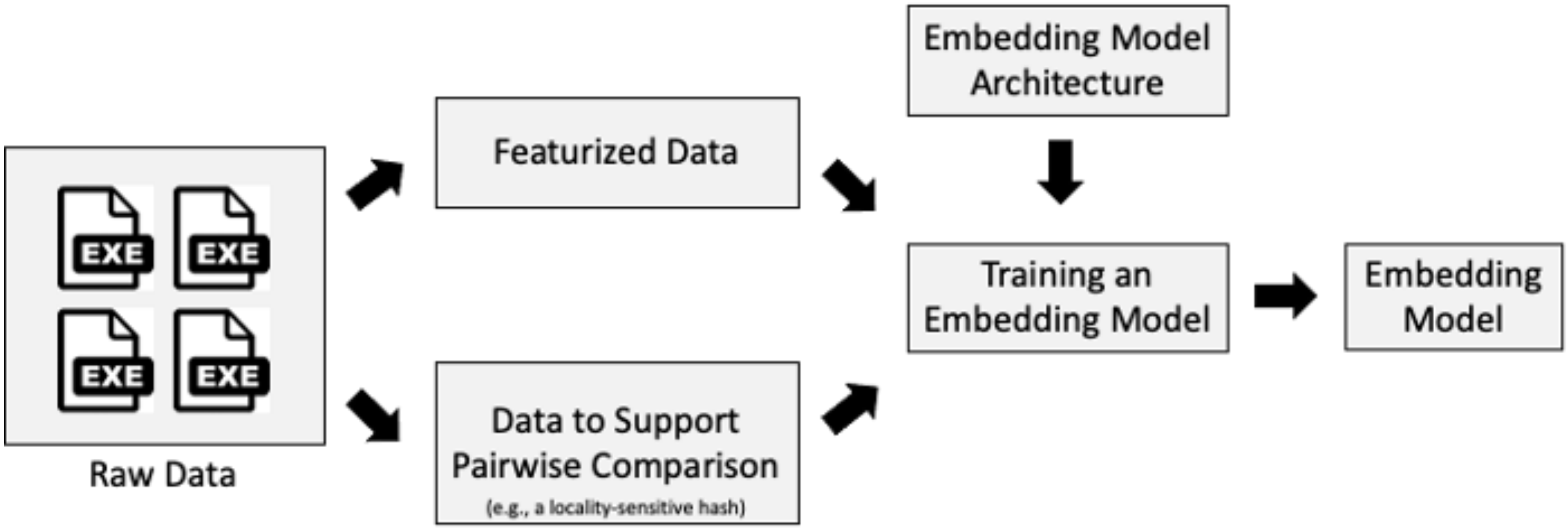
- Incorporate enrichment from multiple data sources to arrive at a generic embedding space
  - No assumption about downstream task labels a priori

- Use the learnt embedding for multiple downstream tasks, e.g.:
  - Fine-tuning novel classifiers
  - Retrieval via some distance measure

- Disclaimer: we make no strict metric assumption

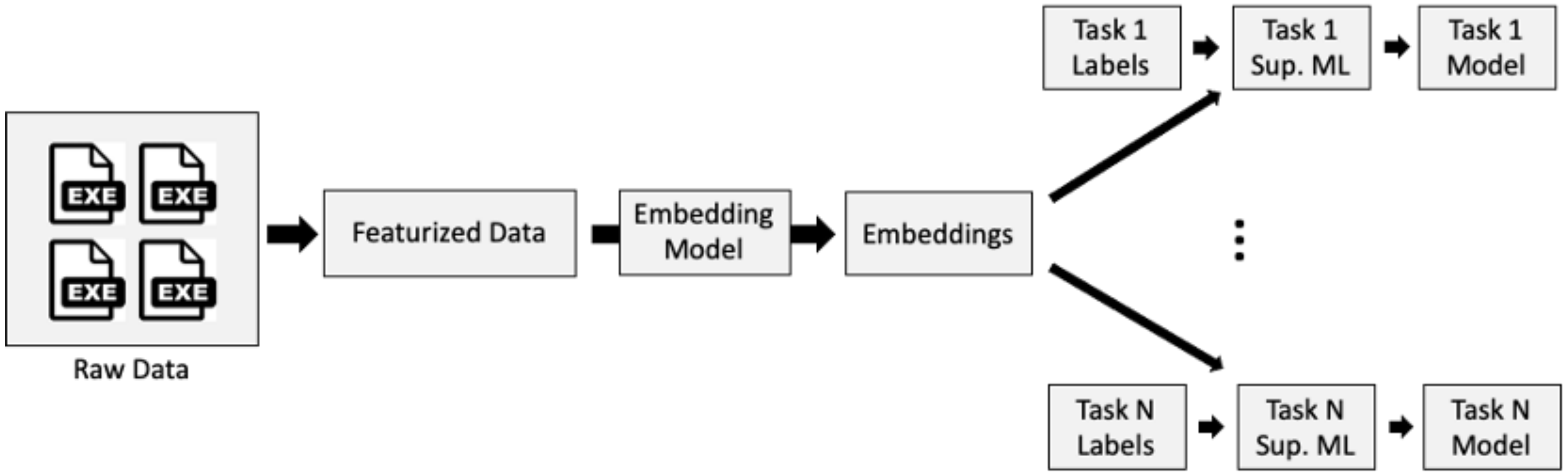
System Design

Upstream Training



System Design

Downstream Use



Defining Similarity

CAPA: Malware Capabilities Analysis

Mandiant's CAPA Project:

- Open source tool released by Mandiant's FLARE team

- Utilizes a variety of disassembly rules/heuristics to output capabilities and MITRE ATT&CK tactics utilized by different executable formats
  - Current support for PE, ELF, .NET, and shellcode files

Repository: https://github.com/mandiant/capa
Blog Post: https://www.mandiant.com/resources/capa-automatically-identify-malware-capabilities

CAPA

Defining Similarity

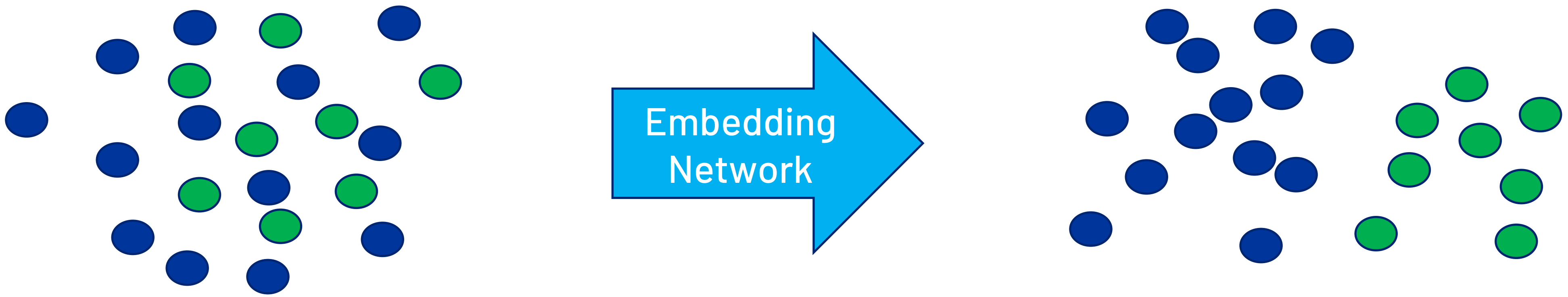
Example CAPA Output

```
$ capa Lab01-01.dll_
+------------------------+------------------------------------------------------------------------------------+
| md5                    | 290934c61de9176ad682ffdd65f0a669                                                   |
| path                   | Lab01-01.dll_                                                                      |
+------------------------+------------------------------------------------------------------------------------+

+------------------------------------------------------+------------------------------------------------------+
| CAPABILITY                                           | NAMESPACE                                            |
|------------------------------------------------------+------------------------------------------------------|
| receive data                                         | communication                                        |
| send data                                            | communication                                        |
| initialize Winsock library                           | communication/socket                                 |
| receive data on socket                               | communication/socket/receive                         |
| send data on socket                                  | communication/socket/send                            |
| connect TCP socket                                   | communication/socket/tcp                             |
| create TCP socket                                    | communication/socket/tcp                             |
| act as TCP client                                    | communication/tcp/client                             |
| check mutex                                          | host-interaction/mutex                               |
| create mutex                                         | host-interaction/mutex                               |
| resolve DNS                                          | host-interaction/network/dns/resolve                 |
| create process                                       | host-interaction/process/create                      |
+------------------------------------------------------+------------------------------------------------------+
```

Metric Embeddings

Intuition



Input
Space

Embedding
Space

Enriching Metric Embeddings w/ CAPA Outputs

Enrichment Approach 1: Contrastive Loss

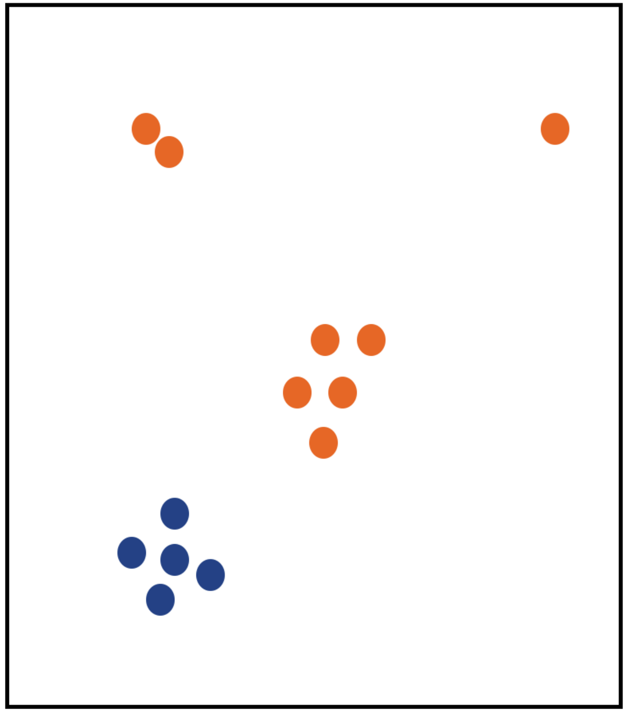
- Contrastive loss idea:
  - Push together / pull apart pairs of positively / negatively associated samples

$$L_{contrastive} = [d_p - m_{pos}]_+ + [m_{neg} - d_n]_+$$

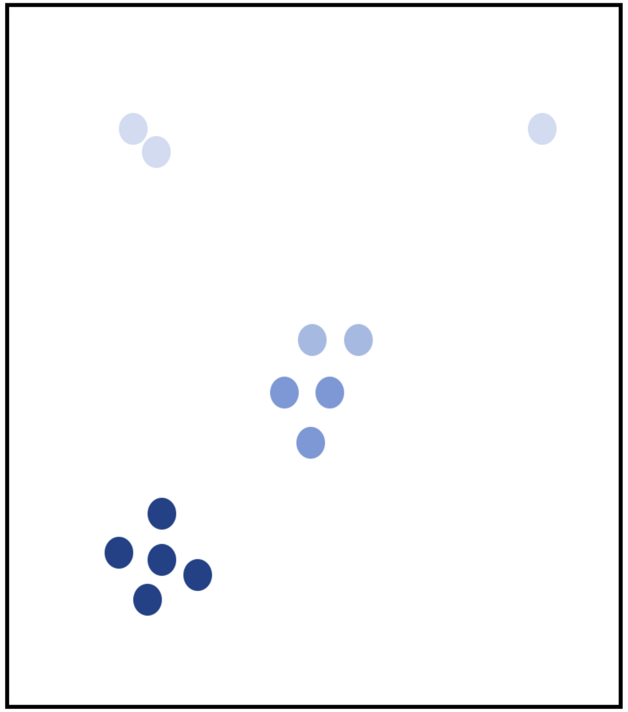
- CAPA Enrichment
  - Form distinct CAPA clusters
  - apply contrastive loss on an in vs. out of cluster basis
- Issue: does not incorporate notion of inter-cluster "similarity"
  - i.e., some clusters are more similar than others

Enriching Metric Embeddings w/ CAPA Outputs

Coarse Enrichment vs. Fine-Grained Enrichment



Coarse Approach: Contrastive Learning



Fine-Grained Approach: ?

Enriching Metric Embeddings w/ CAPA Outputs

Enrichment Approach 2: Spearman Rank Loss

- Utilizes recent research on teaching neural nets differentiable sorting/ranking

Blondel, Mathieu, et al. "Fast differentiable sorting and ranking." *International Conference on Machine Learning*. PMLR, 2020.

URL: http://proceedings.mlr.press/v119/blondel20a/blondel20a.pdf

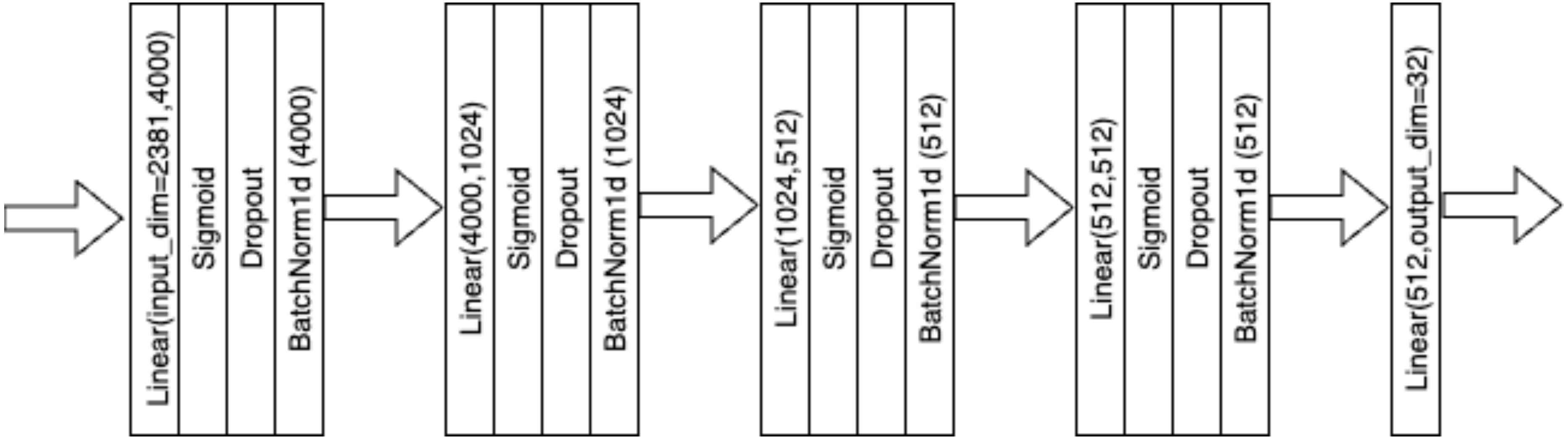
- Loss Function: Spearman's Rank Correlation Coefficient

$$r = 1 - \frac{6\sum\left(R(X_i) - R(Y_i)\right)^2}{n(n^2-1)}$$

- Aims to measure the degree to which similarity ranks predicted by the network differ from the ground truth
- CAPA Enrichment
  - Ground truth established via Jaccard similarity of CAPA capabilities

Base Architecture

Embedding Network



Experimental Evaluation

Experimental Protocols

- Metric embeddings derived using CAPA v1 telemetry for the training partition of the EMBER dataset.
- Fine tuning was performed over extracted embeddings under five different experimental regimes.

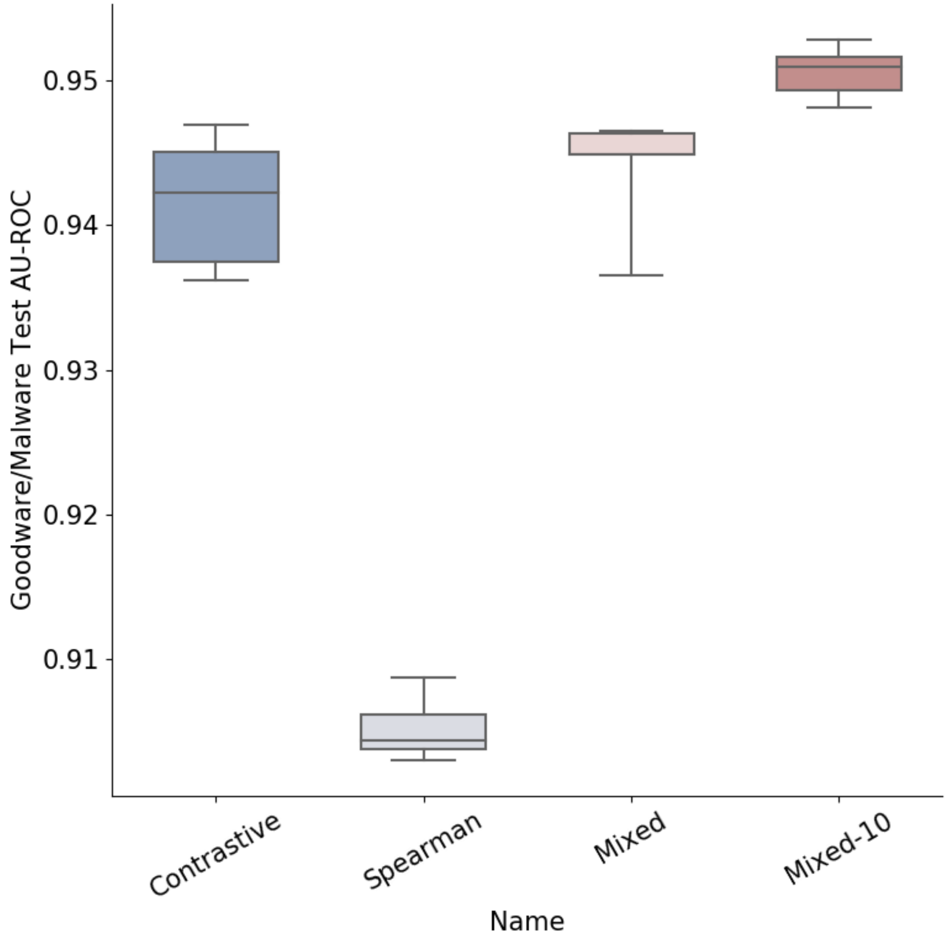
- Fine Tuning on EMBER 2018
1) Fine Tune on EMBER 2018 Train; Test EMBER 2018 (Malicious/Benign)
2) Fine Tune on EMBER 2018 Train; Test EMBER 2018 (Malware Family)
3) Fine Tune on EMBER 2018 Train; Test SOREL-20M (Malicious/Benign)

- Fine Tuning on SOREL-20M
4) Fine Tune on SOREL-20M Train; Test SOREL-20M (Malicious/Benign)
5) Fine Tune on SOREL-20M Train; Test SOREL-20M (Semantic Attribute Tags)

Experimental Evaluation

Experiment 1: EMBER Fine Tune; EMBER Eval (Malicious/Benign)

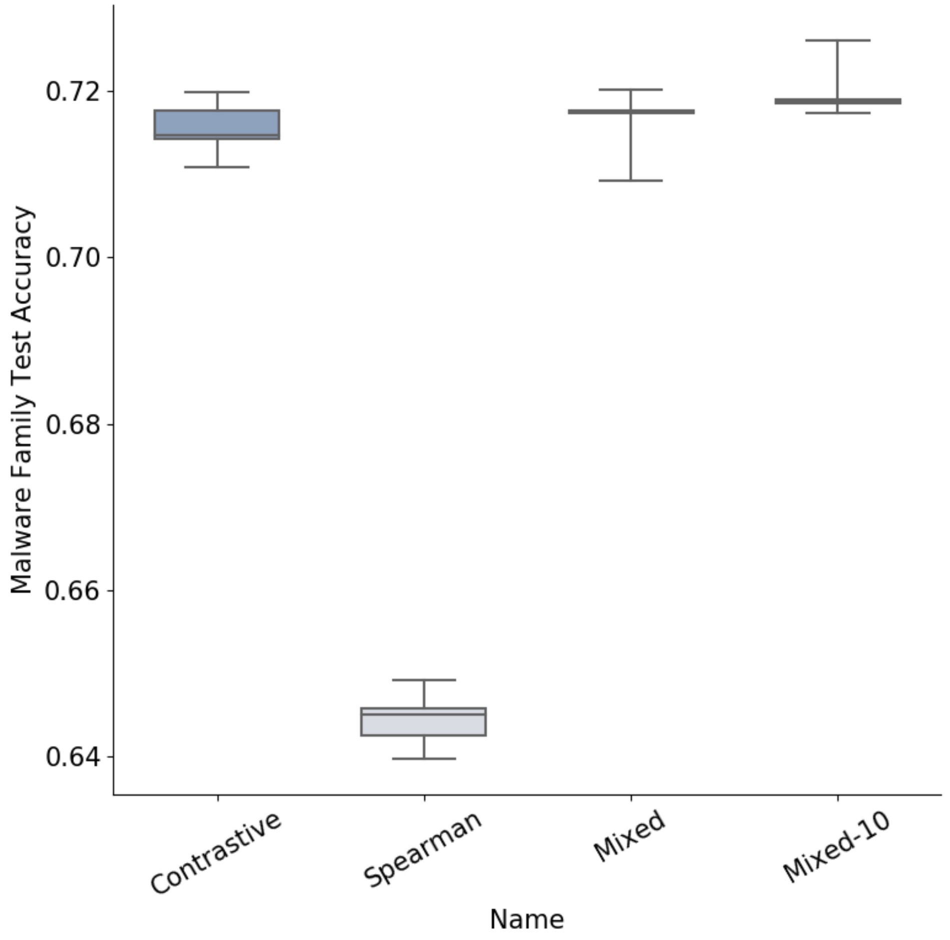
- Mixed Spearman + Contrastive Loss outperforms other metric learning loss functions
- Underperforms "baseline" (~0.995 AUC)



Experimental Evaluation

Experiment 2: EMBER Fine Tune; EMBER Eval (Malware Family)

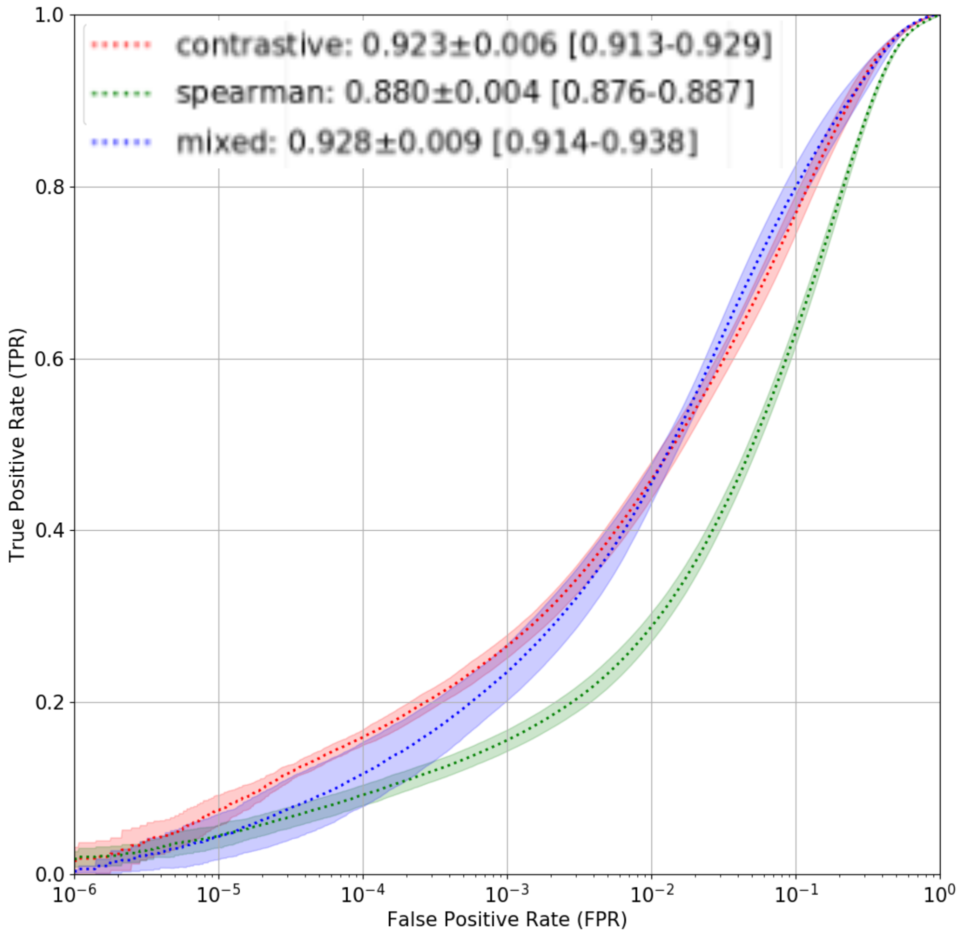
- Consistent order w/ results from malicious/benign tasks
- Within striking distance of "baseline" (73.3% Accuracy)



Experimental Evaluation

Experiment 3: EMBER Fine Tune SOREL-20M Eval (Malicious/Benign)

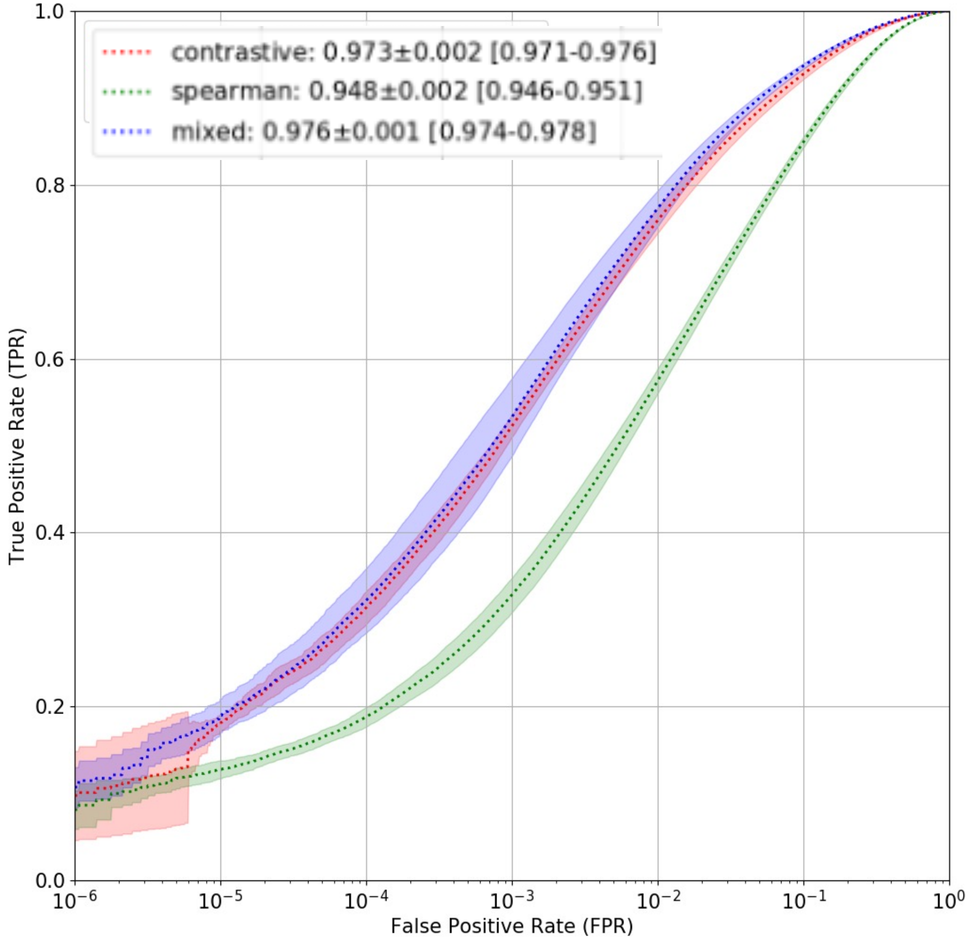
- No direct training on SOREL
- Again, consistent ordering w.r.t. loss combinations
- Performance degradation



Experimental Evaluation

Experiment 4: Transfer SOREL-20M; Eval SOREL-20M (Malicious/Benign)

- Results are again consistent in ordering w/ prior experimentation
- SOREL-20M lightGBM benchmark: 0.981 ± 0.002



Experimental Evaluation

Experiment 5: Transfer SOREL-20M; Eval SOREL-20M (Semantic Tags)

- Utilized lightGBM classifier trained on embeddings
- Similar trend on loss magnitudes

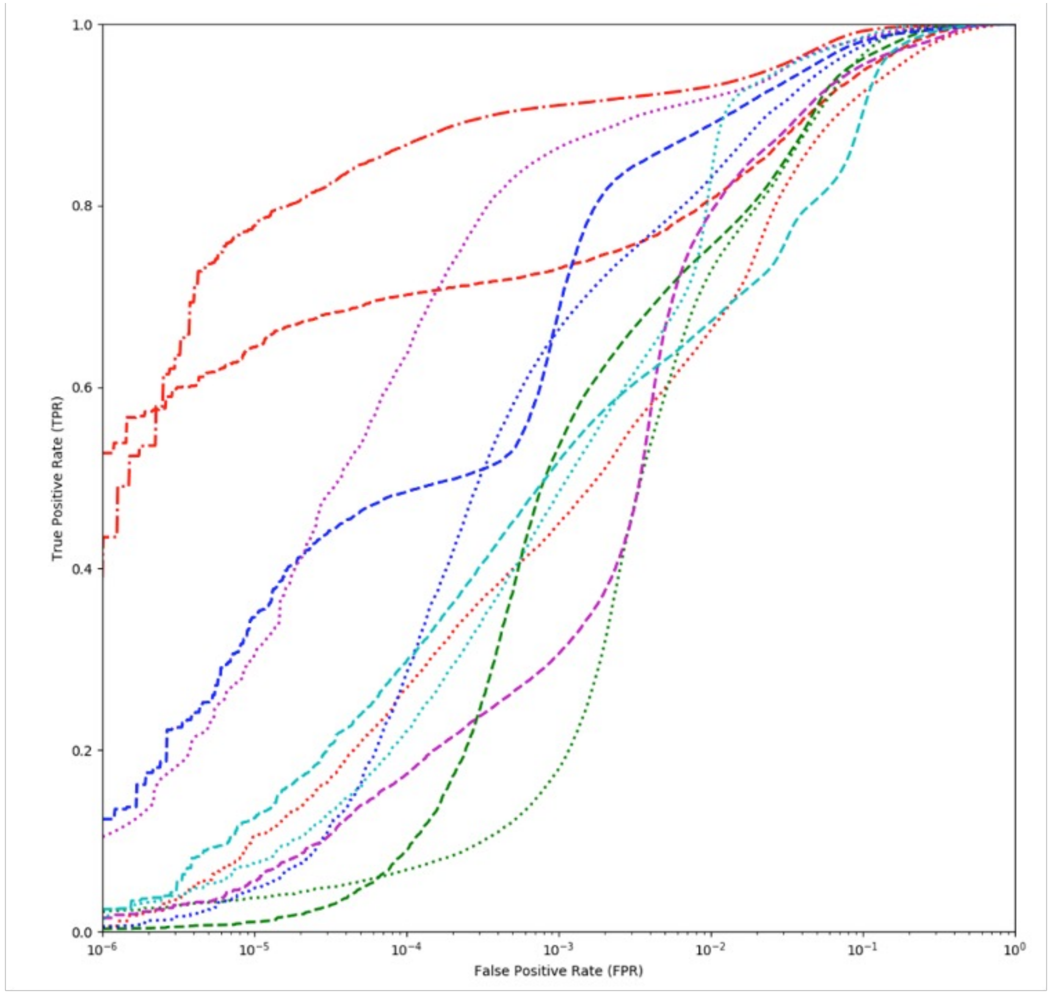
	Contrastive	Spearman	Mixed-10
Adware	**0.917 ± 0.005**	0.883 ± 0.005	**0.917 ± 0.002**
Crypto Miner	**0.976 ± 0.004**	0.962 ± 0.001	**0.976 ± 0.003**
Downloader	0.832 ± 0.007	0.798 ± 0.005	**0.835 ± 0.004**
Dropper	0.819 ± 0.009	0.773 ± 0.005	**0.824 ± 0.011**
File Infector	0.878 ± 0.003	0.834 ± 0.005	**0.885 ± 0.007**
Flooder	**0.982 ± 0.006**	0.981 ± 0.003	0.979 ± 0.003
Installer	0.957 ± 0.003	0.929 ± 0.002	**0.962 ± 0.002**
Packed	**0.783 ± 0.003**	0.742 ± 0.004	0.779 ± 0.013
Ransomware	0.977 ± 0.003	0.959 ± 0.002	**0.978 ± 0.003**
Spyware	**0.848 ± 0.010**	0.776 ± 0.003	0.846 ± 0.014
Worm	**0.877 ± 0.014**	0.804 ± 0.014	**0.877 ± 0.014**

SOREL Semantic Tagging AU-ROCs

Experimental Evaluation

Comparison to SOREL-20M FFNN (multi-objective)(Semantic Tags)

- lightGBM transfer under-perform multi-objective network



Conclusions

Takeaways and Directions for Future Work

- Introduced two approaches to enriching metric embeddings with CAPA data: fine-grained (Spearman) and coarse (contrastive)
- Consistent with multi-objective literature, combining approaches and balancing loss magnitude improved performance
- Storage savings comparison
  - SOREL-20M Features ~172 GB; SOREL-20M embeddings ~2.1 GB
  - Allows for rapid iteration/testing in resource-constrained scenarios
- Could further improve performance by incorporating other label info
  - E.g., Malicious/Benign, Attribute Tags, ATT&CK Tactics, etc.
- Future work
  - resiliency of metric embedding approaches to concept drift
  - embeddings on other data -- beyond malware

MANDIANT
NOW PART OF Google Cloud

Thank You