An Efficient One-Class SVM for Novelty Detection in IoT

Kun Yang*, Samory Kpotufe*, Nick Feamster✦
Columbia University*, University of Chicago✦

Overview

- One-Class Support Vector Machines (OCSVM) have been applied in Internet of Things (IoT) for novelty detection, due to their flexibility in fitting complex nonlinear boundaries between normal and novel data [1].
- Conventional OCSVMs introduce significant memory requirements and are computationally expensive at prediction time as the size of the train set grows.

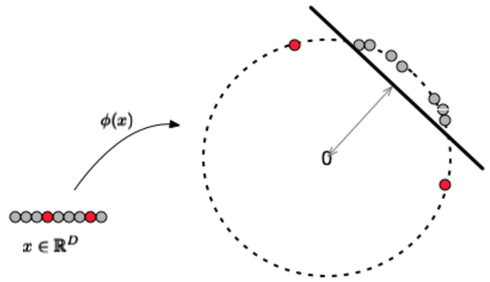

Figure 1. OCSVM maps datapoints $x \in \mathbb{R}^D$ as $\phi(x)$ in infinite-dimensional space, where $\mathbb{R}^D$ is linear separable.

- This work extends so-called Nystrom and (Gaussian) Sketching approaches to OCSVM, by combining these methods with clustering and Gaussian mixture models to achieve significant speedups in prediction time and space in IoT settings, without sacrificing detection accuracy.

Projection back to $\mathbb{R}^d$

- Nystrom and Kernel Johnson-Lindenstrauss (KJL) (a.k.a. Gaussian sketching) use matrix P and subsample S_m (of size m) for projection ϕ' back to $\mathbb{R}^d$.

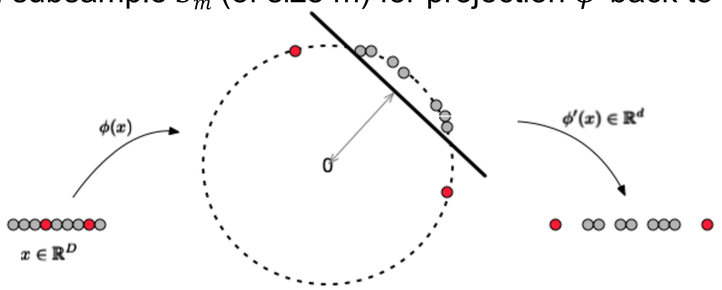

Figure 2. Nystrom and KJL projection back to $\mathbb{R}^d$ while maintaining cluster structures.

Efficient Detection Procedure

The extended one-class Nystrom (OC-Nystrom) and (OC-KJL) approaches are summarized as follows. Given a Gaussian kernel K with bandwidth h, embedding choices m, $d \ll$ training size n:

Training: Given normal data $\{Xi\}_{i=1}^n \in \mathbb{R}^D$ do:
- Embed data as $\{\phi'(X_i)\}_{i=1}^n$ into $\mathbb{R}^d$ via Nystrom or KJL
- Estimate a GMM density f with k components on $\{\phi'(Xi)\}_{i=1}^n$ (k is passed in , or chosen via Quickshift++ on $\{\phi'(X_i)\}_{i=1}^n \in \mathbb{R}^d$)
- **Return** GMM f, along with projection ϕ' (i.e., matrix P and subsample S_m) □

Testing: Given query $x \in \mathbb{R}^D$, and model (ϕ', f), do:
- Embed x as $\phi'(x)$ into $\mathbb{R}^d$
- Flag x as novelty iff $f(\phi'(x)) \leq$ threshold t □

Experiment

- Evaluate on two publicly available traffic traces (UNB and MAWI) and one trace collected on a private consumer IoT device (Smart Fridge (SFRIG)).
- Parse bidirectional flows from the datasets and extract interarrival times (IAT) and packet sizes (SIZE) as features.

Table 1. OCSVM performance. Time is in seconds over 100 datapoints, and ñ is the number of support vectors.

Dataset	UNB	MAWI	SFRIG
AUC	0.62±0.03	0.99±0.00	0.93±0.00
Train time	0.04±0.00	0.06±0.00	0.03±0.00
Test time	0.03±0.00	0.04±0.01	0.02±0.00
Space: ñ · (D + 1)	121785.60 ±188.12	305463.60 ±687.54	60105.60 ±39.00

Table 2. Preserved AUC (method over OCSVM), train time and test time speedups (OCSVM over method), and space reduction (OCSVM over method).

Dataset		OC-KJL	OC-Nystrom	OC-KJL +QS	OC-Nystrom +QS
UNB	AUC	1.43±0.04	1.56±0.01	1.40±0.05	1.51±0.06
	Train time	0.59±0.06	1.10±0.11	0.43±0.04	0.59±0.06
	Test time	13.41±4.15	19.40±6.00	31.11±9.63	26.52±8.21
	Space	21.49±0.03	22.44±0.03	22.00±0.03	21.86±0.03
MAWI	AUC	0.99±0.02	0.98±0.02	0.98±0.01	0.94±0.03
	Train time	1.46±0.04	1.20±0.03	0.65±0.02	0.94±0.03
	Test time	29.71±1.60	27.29±1.47	39.24±2.12	50.70±2.74
	Space	23.44±0.05	23.55±0.05	23.81±0.05	23.95±0.05
SFRIG	AUC	1.00±0.01	0.98±0.02	0.99±0.01	0.98±0.01
	Train time	0.81±0.06	0.92±0.07	0.46±0.03	0.44±0.03
	Test time	16.93±1.76	20.97±2.18	25.99±2.70	24.89±2.59
	Space	18.60±0.01	20.01±0.01	19.31±0.01	19.06±0.01

Summary

Experimental results show that the proposed approaches have

- Significant reduction in detection time and space. Typical detection time speedups (w.r.t. the baseline OCSVM) in factors between 14 to 30 times. Typical space complexities decrease by factors of 20 or more w.r.t. OCSVM.
- Similar or improved detection performance.

Reference

1. **Amaal Al Shorman, Hossam Faris, and Ibrahim Aljarah. 2020. Unsupervised intelligent system based on one class support vector machine and Grey Wolf optimization for IoT botnet detection. Journal of Ambient Intelligence and Humanized Computing 11, 7 (2020), 2809–2825.**