Semi-Supervised Anomaly Detection

Yongchan, Choi

2018.08.09

Introduction

- Unsupervised learning (SVDD, one-class SVM)
- Supervised learning (classifier model)
- These models often fail to match the required detection rates.

Introduction

1. Why the unsupervised learning paradigm is needed to solve anomaly detection
2. Proposed model (SSAD)
3. Active learning

Data distiribution

- Let's consider non-stationary outlier distribution

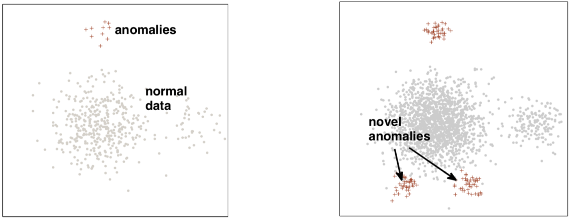


Figure 3: Left: training data stems from two clusters of normal data (gray) and one small anomaly cluster (red). Right: two additional anomaly clusters (red) appear in the test data set.

Data distiribution

- Performance of 4 models.
- Left : Identical training and test dsitribution
- Right : In test data, there is a different anomaly cluster.

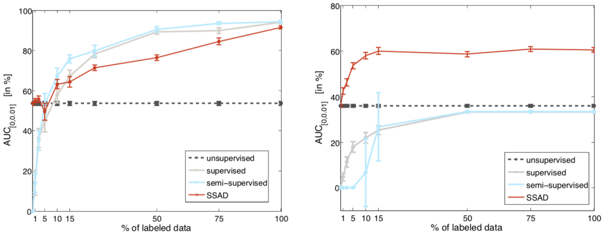


Figure 2: Left: The standard supervised classification scenario with identical training and test distributions. Right: The anomaly detection setting with two novel anomaly clusters in the test distribution.

SVDD

- SSAD(proposed model) is based on SVDD.
- Before explaining SSAD(proposed model), Let's remind SVDD.

SVDD

- We are given n observations $x_1, \ldots, x_n \in \mathcal{X}$.
- The underlying assumption is that the bulk of the data stems from the same (unknown) distribution and we call this part of the data normal.
- Compute a hypersphere with radius R and center $\mathbf{c}$.
- $f(x) = \|\phi(x) - \mathbf{c}\|^2 - R^2$
- x is treated as normal data if $f(x) < 0$
- Point lying outised of the ball(i.e. $f(x) > 0$) are considered anomalous

SVDD

- Use slack variable ξ
- $\min_{R,\mathbf{c},\xi} R^2 + \eta_u \sum_{i=1}^{n} \xi_i$
  subject to (i) $\|\phi(x_i) - \mathbf{c}\|^2 \le R^2 + \xi_i$
  (ii) $\xi_i \ge 0$
- Above OP can be solved equivalently in dual space using the representation $\mathbf{c} = \sum_{i=1}^{n} \alpha_i \phi(x_i)$

Proposed model

- In addition to the n unlabeled data $x_1, \ldots, x_n \in \mathcal{X}$, we are now given m labeled observations $(x_1^*, y_1^*, \ldots, x_m^*, y_m^*) \in \mathcal{X} \times \mathcal{Y}$
- Nomial data are encoded $y^* = 1$ and anomalies are encoded $y^* = -1$
- We want to place anomalies outside of the ball

Proposed model

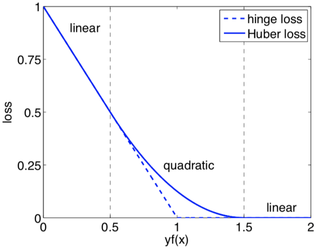
- $\min_{R,\gamma,\mathbf{c},\xi} R^2 - \kappa\gamma + \eta_u \sum_{i=1}^{n} \xi_i + \eta_l \sum_{j=n+1}^{n+m} \xi_j$
  subject to (i) $\|\phi(x_i) - \mathbf{c}\|^2 \leq R^2 + \xi_i$
  (ii) $y_j^*(\|\phi(x_j^*) - \mathbf{c}\|^2 - R^2) \leq -\gamma + \xi_j^*$
  (iii) $\xi_i \geq 0$
  (iv) $\xi_j^* \geq 0$
- The inclusion of negatively labeled data renders the above optimization problem non-convex.

Remedy for OP

- Idea : Translate above equation into an unconstrained problem
- Resolve the slack term as follows (Chapelle and Zien, 2005)
- $\xi_i = l(R^2 - \|\phi(x_i) - \mathbf{c}\|^2)$
- $\xi_j^* = l(y_j^*(R^2 - \|\phi(x_i) - \mathbf{c}\|^2 - \gamma)), \quad l(t) = max\{-t, 0\}$
- $\mathbf{c} = \sum_{i=1}^{n} \alpha_i \phi(x_i) + \sum_{j=n+1}^{n+m} \alpha_j y_j^* \phi(x_j^*)$

Re-fomulate optimization problem

- $min_{R,\gamma,\alpha}\{R^2 - \kappa\gamma + \eta_u \sum_{i=1}^{n} l_\epsilon(R^2 - k(x_i,x_i) + (2e_i-\alpha)' K\alpha + \eta_l \sum_{j=n+1}^{n+m} l_\epsilon(y_j^*(R^2 - k(x_j^*,x_j^*) + (2e_j^*-\alpha)' K\alpha) - \gamma)\}$
- $K = (k_{ij})_{1\le i,j\le n}$ denotes the kernel matrix given by $k_{ij} = k(x_i,x_j) =< \phi(x_i), \phi(x_j) >$
- $e_1, \ldots, e_{n+m}$ is the standard base of $\mathbb{R}^{n+m}$
- Use gradient-based optimization tool.



Active learning for SSAD

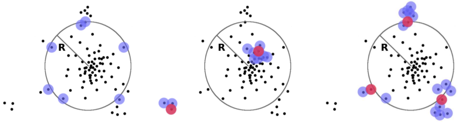
1. Borderline points
2. Novel anomaly classes
3. Combind

1 $x^{'} = \arg\min_{x \in \{x_1, \ldots, x_n\}} \frac{\|f(x)\|}{\max_k \|f(x_k)\|} = \arg\min_{x \in \{x_1, \ldots, x_n\}} \|R^2 - \|\phi(x) - \mathbf{c}\|^2\|$

2 Let $A = (a_{ij})_{i,j=1,\ldots,n+m}$ be adjacent matrix of training data.

- Introduce an extended labeling $\bar{y}_1, \ldots, \bar{y}_{n+m}$ defining $\bar{y}_i = 0$ if unlabeled data, $\bar{y}_j = y_j$ for labeled instance
- $x^{'} = \arg\min_{x \in \{x_1, \ldots, x_n\}} \frac{1}{2k} \sum_{j=1}^{n+m} (\bar{y}_j + 1) a_{ij}$

3 $x^{'} = \arg\min_{x \in \{x_1, \ldots, x_n\}} \delta \frac{\|f(x)\|}{c} + \frac{1-\delta}{2k} \sum_{j=1}^{n+m} (\bar{y}_j + 1) a_{ij}, \quad \delta \in [0, 1]$



(a) margin strategy (b) cluster strategy (c) combined strategy

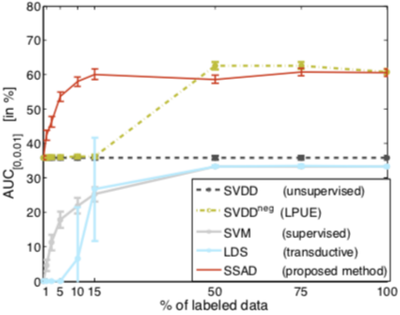
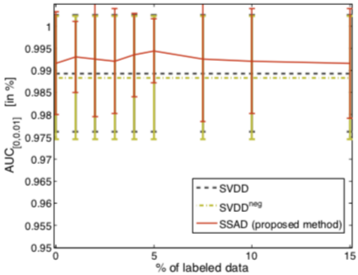
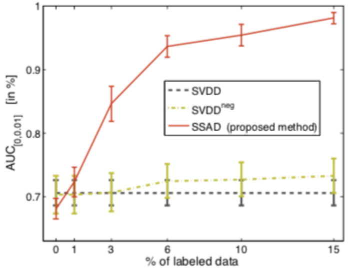


Figure 6: Performance of various unsupervised, supervised and semi-supervised methods in the anomaly detection setting.



(a) Detection accuracies of regular attacks.



(b) Detection accuracies of cloaked attacks.