INDIAN INSTITUTE OF INFORMATION TECHNOLOGY ALLAHABAD

Subject – Data Mining and Warehousing

Topic – Deep Support Vector Data Description for Unsupervised and Semi-Supervised Anomaly Detection

Report By -

IIT2018108 – Ravi Kumar Sharma

Introduction :

Anomaly detection (AD) (Chandola et al., 2009; Pimentel et al., 2014) is the task of identifying unusual samples in data. This task lacks a supervised learning objective and AD methods typically formulate an unsupervised problem to

find a "compact" description of the "normal" class, e.g. finding a set of small measure that contains most of the data

as in one-class classification (Moya et al., 1993). Samples that deviate from this description are deemed anomalous.

2. Deep Support Vector Data Description

Here, we introduce a generalization of *Deep Support Vector Data Description (Deep SVDD)* to the more general semi-supervised AD setting that contains the unsupervised Deep SVDD method (Ruff et al., 2018) as a special case.

2.1. Unsupervised Deep SVDD

For input space $\mathcal{X} \subseteq \mathbb{R}^d$ and output space $\mathcal{F} \subseteq \mathbb{R}^p$, let $\phi(\cdot\,; \mathcal{W}) : \mathcal{X} \to \mathcal{F}$ be a neural network with $L \in \mathbb{N}$ hidden layers and weights $\mathcal{W} = \{\boldsymbol{W}^1, \dots, \boldsymbol{W}^L\}$. The objective of Deep SVDD is to learn a neural network transformation ϕ that minimizes the volume of a data-enclosing hypersphere with radius $R > 0$ and fixed center $\boldsymbol{c} \in \mathcal{F}$ in output space $\mathcal{F}$. Given $n \in \mathbb{N}$ (unlabeled) training samples $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \in \mathcal{X}$, the *Soft-Boundary Deep SVDD* objective is defined by

$$\min_{R,\mathcal{W}} \; R^2 + \frac{1}{\nu n} \sum_{i=1}^{n} \max\{0, \|\phi(\boldsymbol{x}_i; \mathcal{W}) - \boldsymbol{c}\|^2 - R^2\}. \quad (1)$$

Points mapped outside the sphere ($\|\phi(\boldsymbol{x}_i; \mathcal{W}) - \boldsymbol{c}\|^2 > R^2$) get penalized and the network weights $\mathcal{W}$ are optimized such that most of the data falls within the hypersphere centered at $\boldsymbol{c}$. Minimizing the volume of the sphere via R^2 enforces this learning process. In consequence, normal points get closely mapped to the hypersphere center, whereas anomalies are mapped further away or outside the sphere. Hyperparameter $\nu \in (0, 1]$ controls this trade-off between volume and boundary violations (Ruff et al., 2018).

If the unlabeled training data $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ is not polluted, i.e. if most of the training examples are normal, the simplified *One-Class Deep SVDD* objective, which penalizes the mean squared distance of *all* the mapped data points (not just the outliers), is preferable:

$$\min_{\mathcal{W}} \; \frac{1}{n} \sum_{i=1}^{n} \|\phi(\boldsymbol{x}_i; \mathcal{W}) - \boldsymbol{c}\|^2. \quad (2)$$

2.2. Semi-Supervised Deep SVDD

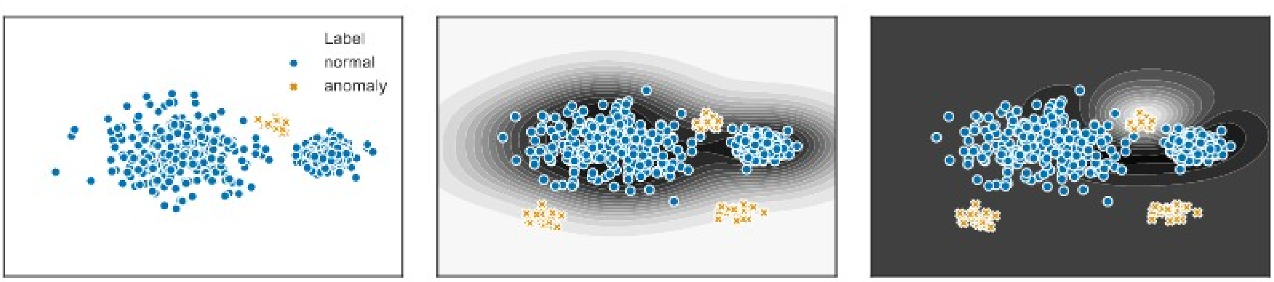
Now we assume we also have access to $m \in \mathbb{N}$ labeled samples $(\tilde{\boldsymbol{x}}_1, \tilde{y}_1), \ldots, (\tilde{\boldsymbol{x}}_m, \tilde{y}_m) \in \mathcal{X} \times \mathcal{Y}$ in addition to the $n \in \mathbb{N}$ unlabeled samples $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathcal{X}$ with $\mathcal{X} \subseteq \mathbb{R}^d$ and $\mathcal{Y} = \{-1, +1\}$. We denote $\tilde{y} = +1$ for known normal examples and $\tilde{y} = -1$ for known anomalies.

We establish a *Semi-Supervised Deep SVDD (SS-DSVDD)* generalization by extending the objectives (1) and (2) with terms that enables learning from labeled data. We formulate the *Soft-Boundary SS-DSVDD* problem as

$$\begin{aligned} \min_{R, \mathcal{W}} \; & R^2 + \frac{1}{\nu(n+m)} \sum_{i=1}^{n} l\left(R^2 - \|\phi(\boldsymbol{x}_i; \mathcal{W}) - \boldsymbol{c}\|^2\right) \\ & + \frac{\eta}{\nu(n+m)} \sum_{j=1}^{m} l\left(\tilde{y}_j \left(R^2 - \|\phi(\tilde{\boldsymbol{x}}_j; \mathcal{W}) - \boldsymbol{c}\|^2\right)\right), \end{aligned} \tag{3}$$

where $l(z) = \max\{0, -z\}$ is the hinge loss. That is, we require normal examples ($\tilde{y} = +1$) to lie inside the hypersphere and labeled anomalies ($\tilde{y} = -1$) to lie outside. We achieve this by penalizing accordingly: if a labeled anomaly lies *inside* the sphere, the penalty is given by $R^2 - \|\phi(\tilde{\boldsymbol{x}}_j; \mathcal{W}) - \boldsymbol{c}\|^2$ and $\|\phi(\tilde{\boldsymbol{x}}_j; \mathcal{W}) - \boldsymbol{c}\|^2 - R^2$ elsewise. If a labeled data point is already mapped onto the correct side, there is no penalty. To generalize (2), we propose the following *One-Class SS-DSVDD* objective:

$$\begin{aligned} \min_{\mathcal{W}} \; & \frac{1}{n+m} \sum_{i=1}^{n} \|\phi(\boldsymbol{x}_i; \mathcal{W}) - \boldsymbol{c}\|^2 \\ & + \frac{\eta}{n+m} \sum_{j=1}^{m} \left(\|\phi(\tilde{\boldsymbol{x}}_j; \mathcal{W}) - \boldsymbol{c}\|^2\right)^{\tilde{y}_j}. \end{aligned} \tag{4}$$



(a) Training data (b) Unsupervised One-Class Model (c) Supervised Model

Conclusion :

We have generalized Deep SVDD to the more generalsemi-supervised setting in this work. The resulting Semi-Supervised Deep SVDD is an end-to-end deep method forsemi-supervised anomaly detection on high-dimensionaldata. We demonstrated experimentally, that SS-DSVDDsignificantly improves detection performance already withonly small amounts of labeled data. Our results suggest thatsemi-supervised approaches to AD should be preferred inapplications where some labeled information is available.