Prediction Based Deep Autoencoding Model for Anomaly Detection

Zhanzhong Pang[1], Xiaoyi Yu[1], Jun Sun[1] and Inakoshi Hiroya[2]
1 Fujitsu R&D Center Co., Ltd, Beijing, China;
2 Fujitsu Laboratories Ltd, Kawasaki, Japan
{pangzhanzhong, yuxiaoyi, sunjun}@cn.fujitsu.com, inakoshi.hiroya@jp.fujitsu.com

Abstract:

Latent variables and reconstruction error are the two important features generated from an auto encoder. We propose a method combining these two features together for anomaly detection. The proposed architecture comprises of two networks. To compress and rebuild an input, a deep auto encoder is utilized where low dimensional latent variables and reconstruction error can be obtained, and compactness loss is introduced to maintain a low intra-variance in latent variables for normal class. Meanwhile multi-layer perceptron (MLP) network which takes the generated latent variables as input is established aiming at predicting its corresponding reconstruction error. By introducing MLP network, anomalies sharing similar reconstruction error yet different distribution of latent variables to normal data or vice versa can be further separated. The prediction error form MLP network is used as final score for anomaly detection. Experiments on several benchmarks including image and multivariable datasets demonstrate the effectiveness and practicability of this new approach when comparing with several up-to-data algorithms.

Background

☐ Problem

Anomaly detection without anomalies for training

☐ Traditional solutions

Self-reconstruction model and statistical analysis

☐ The proposed method

Auto encoder with prediction mechanism by jointly utilizing latent variables and reconstruction error.

☐ Example application

Performance evaluation and event recognition

Flowchart

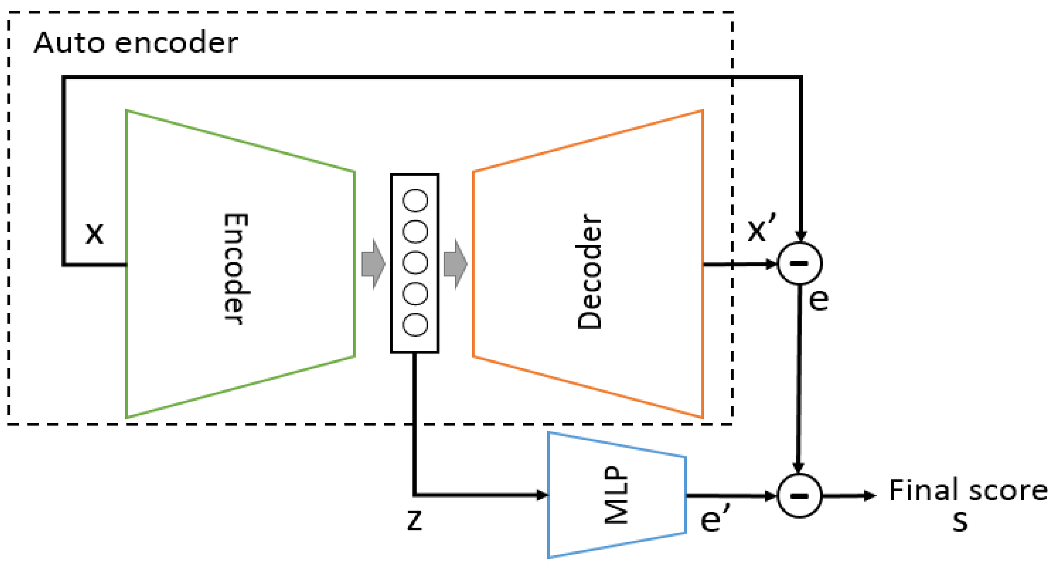

Fig.1. Overview of the proposed structure

Step 1 Low dimensional latent variables by encoder with constraint from compactness loss.

Step 2 Reestablishment of the input by decoder with reconstruction error.

Step 3 Prediction mechanism by MLP for connecting latent variables and reconstruction error.

- ◆ Latent variables as input.
- ◆ Reconstruction error as ground truth for guidance.

Step 4 Prediction error for anomaly detection

The Proposed Method

Deep auto encoder(DeAE)

- ◆ Denoising auto encoder / convolutional denoising auto encoder
- ◆ Compactness loss

$$L_C = \frac{1}{nk}\sum_{i=1}^{n} d_i^T d_i \,, \boldsymbol{where}\ d_i = \|z_i - m_i\|_2 \,, \theta_m = \frac{1}{n}\sum_{i=1}^{n} z_i$$

In a batch with size n, the compactness loss L_C is defined as the average distance of the latent variables $z_i \in R^k$.

Prediction mechanism(MLP)

◆ Connections between latent variables and reconstruction error only in normal data will be learned.

$p_n \rightarrow e_{n1}\ and\ e_{n2}$

$p_a \not\rightarrow e_{a1}\ and\ e_{a2}$

▲- separable only using reconstruction error.
★- separable using proposed method

Latent variable	Reconstruction error			
	e_{n1}	e_{n2}	e_{a1}	e_{a2}
Normal p_n	▲★	★	--	--
Abnormal p_a	--	--	▲★	★

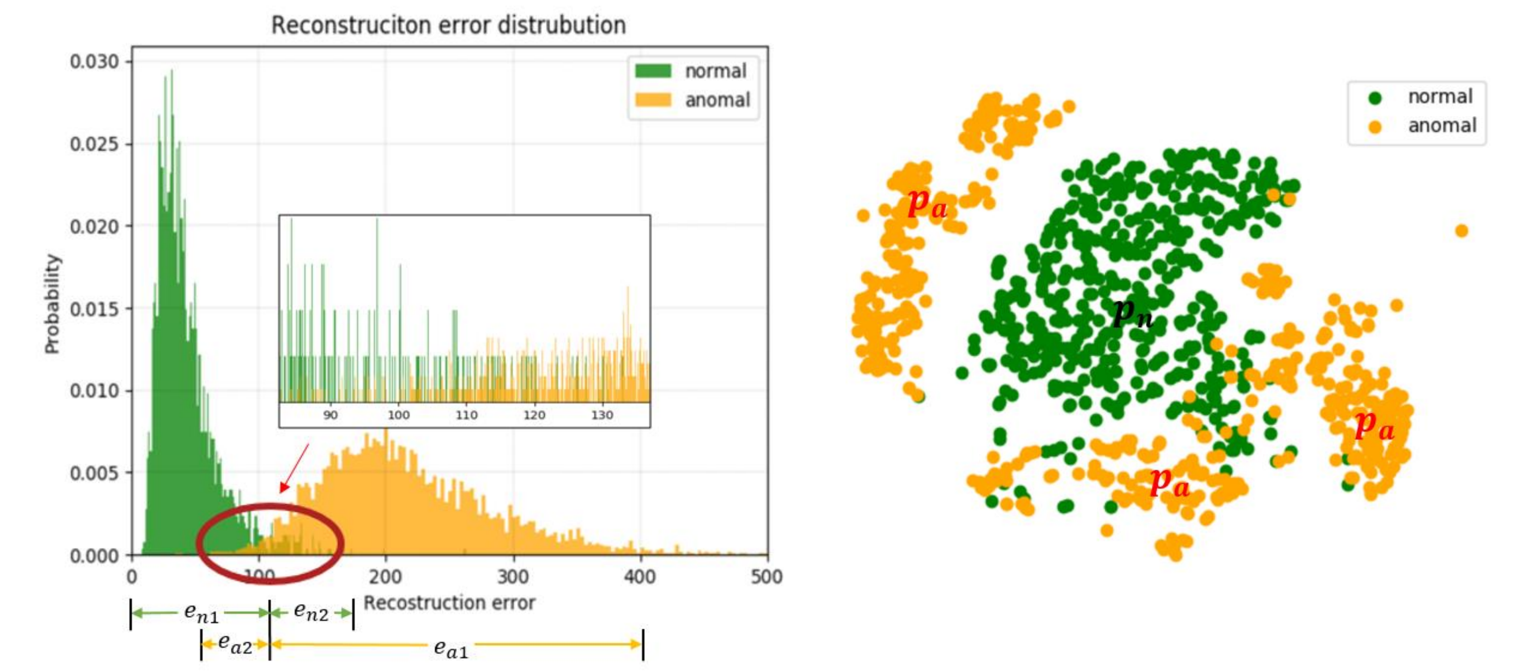

Fig.2. Distributions of reconstruction error and latent variables on MNIST dataset: (a) construction error distribution. e_n for reconstruction error from normal data, e_a for anomalies. The reconstruction error can be further divided into two groups for each class according to its corresponding value. (b) latent variables distribution of samples in the bounding box of (a) (denoted as p)

Evaluation

◆ Minimize the cost function:

$$J(\theta_e, \theta_d, \theta_m) = J_{DeAE}(\theta_e, \theta_d) + \lambda_3 J_{MLP}(\theta_m)$$

$$J_{DeAE}(\theta_e, \theta_d) = \frac{1}{n}\sum_{i=1}^{n}\|x_i - x_i'\|_2 + \lambda_1 \sum_{i=1}^{l}\theta_i^2 + \lambda_2 L_C, \quad J_{MLP}(\theta_m) = \frac{1}{n}\sum_{i=1}^{n}\|e_i - e_i'\|_2$$

$\theta_e, \theta_d, \theta_i$ and θ_m are network parameters in encoder, decoder, fully connected layers of DeAE, and MLP. e_i and e_i' are reconstruction error and the predicted one from MLP.

◆ Prediction error s from MLP for anomaly detection.

$$s_i = \|e_i - e_i'\|_2$$

$$\text{prediction} = \begin{cases} Target\ class & if\ \ s < \tau \\ Anomaly & otherwise \end{cases}$$

τ is the predefined threshold.

◆ Average precision, recall and F1 score.

Experimental results

Test on two kinds of datasets

◆ On image datasets: MNIST and CIFAR-10

Randomly choose one class as normal data. Treat other 9 classes as anomalies.

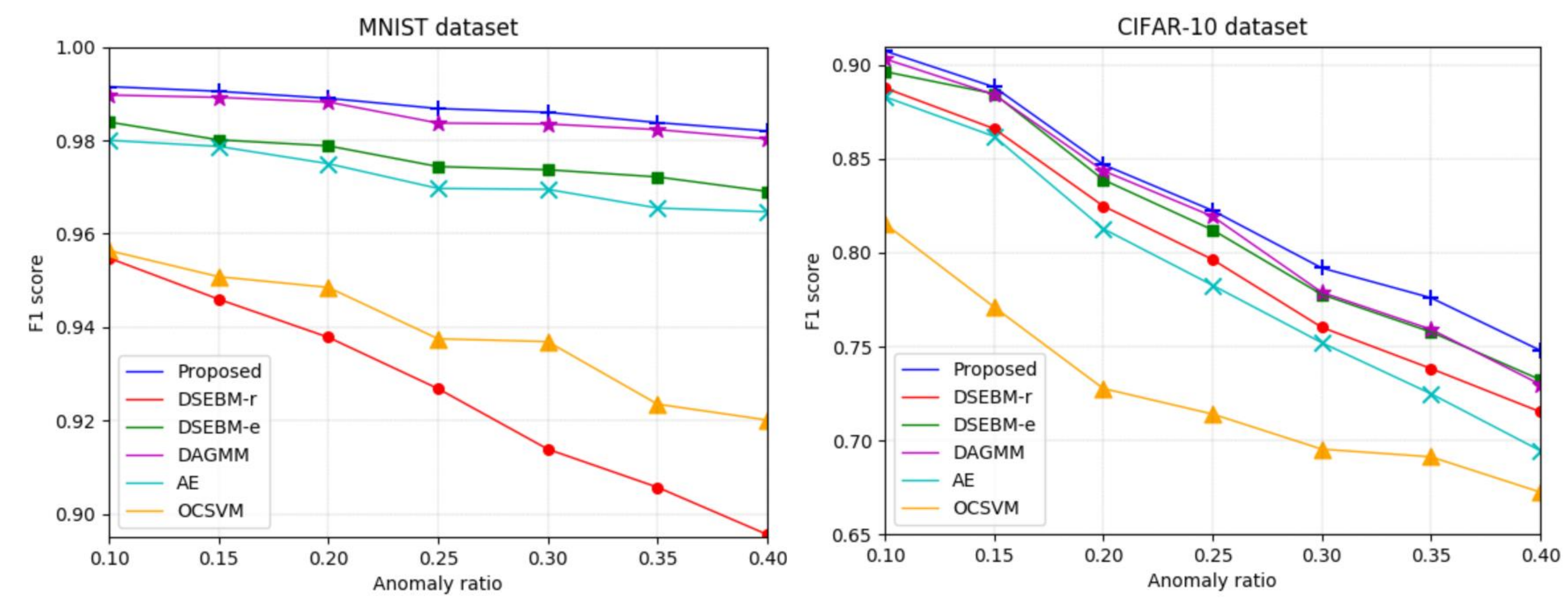

Fig.3. F1 scores on two images datasets with different anomaly ratio

◆ On multi-variable datasets: KDD, Thyroid and Arrhythmia

Treat one class or combination of several classes as anomalies as same as [2].

Table.1. Average scores on three datasets with different models

Method	KDD			Thyroid			Arrhythmia		
	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec	F1
OCSVM	0.7457	0.8523	0.7954	0.3639	0.4239	0.3887	0.6251	0.4545	0.5263
DSEBM-e[1]	0.8619	0.6446	0.7399	**0.6811**	0.5054	0.5802	0.6054	0.5294	0.5650
DAGMM[2]	0.9711	0.9414	0.9559	0.6573	0.5053	0.5714	0.6569	0.4697	0.5487
AE	0.9495	0.8897	0.9185	0.6197	0.4731	0.5366	0.6111	0.5012	0.5493
Proposed	**0.9779**	**0.9582**	**0.9679**	0.6760	**0.5161**	**0.5854**	**0.6727**	**0.5606**	**0.6115**

[1] Zhai, S.: "Deep Structured Energy Based Model for Anomaly Detection," International Conference on Machine Learning, pp. 1100-1109, 2016.
[2] Zong, B.: "Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly Detection," 6-th International Conference on Learning Representations, 2018.