Anomaly Detection with Adaptive-AutoEncoder Ensemble

Song Minsoo

Yonsei University

Table of Contents

Intro

- Task
  - Anomaly Detection in tabular dataset
- Goal of paper
  - Improvement in meric (roc auc, pr auc) with proposed approachs
  - New approach to model training, not model architecture

Task

- Unsupervised Learning (no label)
    - Experiments with generated dataset
- Semi-supervised Learning (know only normal)
    - Experiments with benchmark dataset

AutoEncoder

- An autoencoder has two main parts
  - an encoder(ϕ) that maps the input-data(X) into the latent-value(Z)
  - a decoder(ψ) that maps the latent-value to a **reconstruction of the input-data**

$$\phi : \; X \to Z$$

$$\psi : \; Z \to X$$

$$\phi, \psi = \arg\min_{\phi,\psi} \|X - (\psi \circ \phi) X\|^2$$

AutoEncoder

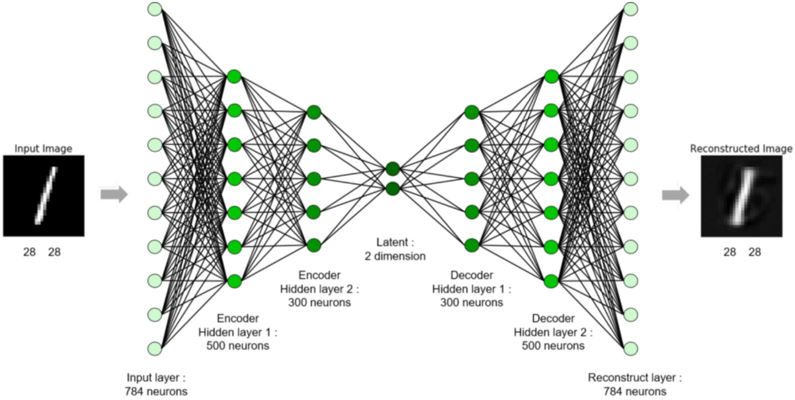

Figure: AutoEncoder

Anomaly Detection with AutoEncoder

- Anomaly Score = Reconstruction Error
- Base anomaly detection procedure with AutoEncoder
    1. Train the AutoEncoder with only normal training dataset.
    2. If abnormal data comes in the model, then reconstruction error will be large.
    3. Predict data as an anomaly when reconstruction-error is larger than specific threshold.

Problem

- Some problems in anomaly detection with AutoEncoder
  - noise in normal training dataset
  - abnormal data is similar with normal data
  - reconstructs normal data not well
  - reconstructs abnormal data well
  - low performance

Algorithm1: Adaptive-AE

- Update(remake) the training dataset with some hyperparameters during training epochs
- Forces the AutoEncoders to focus on inliers than outliers

Algorithm1: Adaptive-AE

for *epoch* $= 1, 2, \ldots$ **do**

    if epoch $>$ initialEpoch(hp) and epoch $=$ samplingTerm(hp) **then**

        Get recon-error of the original training dataset through AutoEncoder prediction

        Replace high recon-error instances with low recon-error instances at a specific samplingRatio(hp) from the original training dataset

        Update the training dataset using the above result

        Train model

    else

        Train model

    end if

end for

Algorithm1: update the training dataset

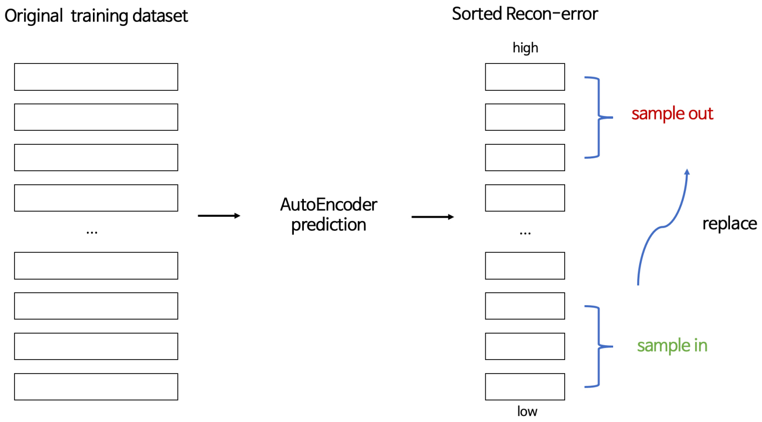

Figure: how to update the training dataset

Algorithm1: sample in & out instances

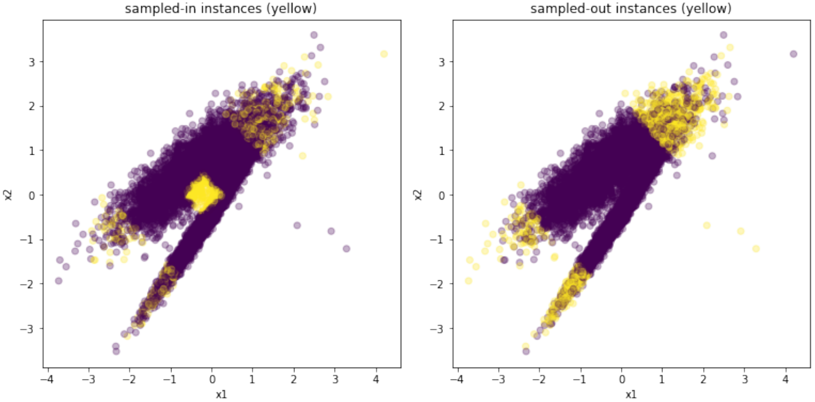

Figure: forces the AutoEncoders to focus on inliers than outliers

Algorithm1: reconstuction-error changes

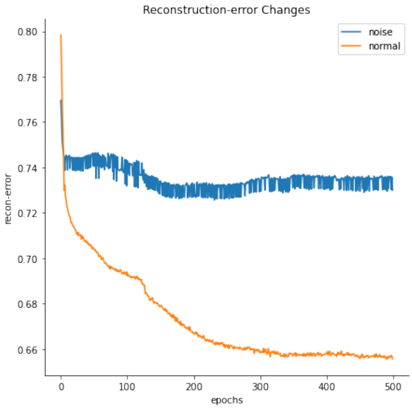

Figure: reconstuction-error changes in well-trained model

Algorithm2: Ensemble

- robust than the individual components and improved performance
- less sensitive to hyperparameters, so can complement algorithm1
- can ensemble various models by algorithm1 (diverse hyperparameters combinations)
- usefullness in anomaly detection
  - can filter out spurious findings of individual learners
  - consider in various perspectives
- Ensemble method
  - mean, median, maximum, minimum
  - different combination functions work better for different dataset

Algorithm2: Ensemble

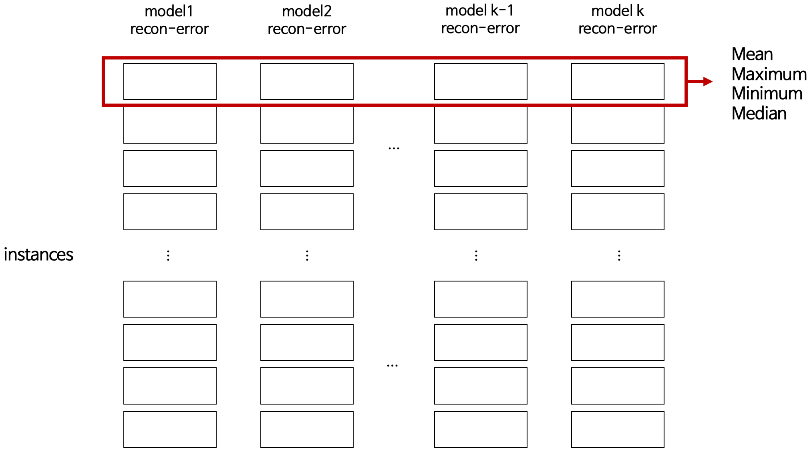

Figure: ensemble using reconstruction-error

Experiments setting

- model
  - AE with 6 fully-connected layers (node: [8,4,2])
- train
  - 500 epochs (select the lowest validation loss epoch model)
- hyperparameters (8 combinations, grid search)
  - initialEpoch: [5, 20]
  - samplingTerm: [1, 5]
  - samplingRatio: [0.01, 0.1]

Data

- Unsupervised Learning data: Generated data
  - 1. Normal data
  - 2. Anomalies in Training dataset
    - ratio: [0.01, 0.1]
    - weight: [0.1, 0.9]
  - 3. Anomalies in Test dataset (ratio: 0.1)
    - weight: [0.1, 0.9]
- Semi-Supervised Learning data: Benchmark data
  - 8 data

Data: unsupervised case

Generated data	shape	IR
Normal	(20000, 16)	.
TrainNoise1	(20000, 16)	4
TrainNoise2	(20000, 16)	3
TrainNoise3	(20000, 16)	4
TrainNoise4	(20000, 16)	3
TestNoise1	(20000, 16)	3
TestNoise2	(20000, 16)	3

Table: dataset description

Data: semi-supervised case

Benchmark data	shape	IR
abalone	(731,9)	16
shuttle	(1829,10)	13
vowel	(988,14)	9
wine	(4898,12)	25
yeast1	(947,9)	30
yeast2	(1484,9)	28
yeast3	(1484,9)	32
yeast4	(1484,9)	41

Table: dataset description

Single Result: unsupervised learning case

- score: Average of the 10 models in each dataset (Adaptive-AE is hyperparameter tuned with grid search)

	AUCROC		AUCPR	
	Base	Adaptive-AE	Base	Adaptive-AE
Normal	0.8277	**0.8605**	0.7944	**0.8299**
TrainNoise1	0.8439	**0.8722**	0.8145	**0.8581**
TrainNoise2	0.8242	**0.8539**	0.7855	**0.8215**
TrainNoise3	0.8268	**0.8424**	0.7877	**0.8064**
TrainNoise4	0.7824	**0.7913**	0.7259	**0.7408**
TestNoise1	0.8135	**0.8506**	0.7757	**0.8348**
TestNoise2	0.7892	**0.8235**	0.7601	**0.8163**

Table: generated data result in unsupervised learning (no label) case

Single Result: semi-supervised learning case

	AUCROC		AUCPR	
	Base	Adaptive-AE	Base	Adaptive-AE
abalone	0.8104	**0.8313**	0.6478	**0.6679**
shuttle	1.0	1.0	1.0	1.0
vowel	**0.6990**	0.6774	**0.6992**	0.6630
wine	0.7003	**0.7233**	0.4411	**0.4667**
yeast1	0.4674	**0.5517**	0.1542	**0.1888**
yeast2	0.6876	**0.6979**	0.2803	**0.2817**
yeast3	0.6472	**0.9050**	0.2182	**0.5189**
yeast4	0.6241	**0.7567**	0.1941	**0.2808**

Table: benchmark data result in semi-supervised learning (know only normal) case

Ensemble Result: unsupervised learning case

- ensemble score
  - ensemble with 8 single models in each dataset
    - each single model is not hyperparameter tuned
  - best score method from maximum, minimum, mean, median

Ensemble Result: unsupervised learning case

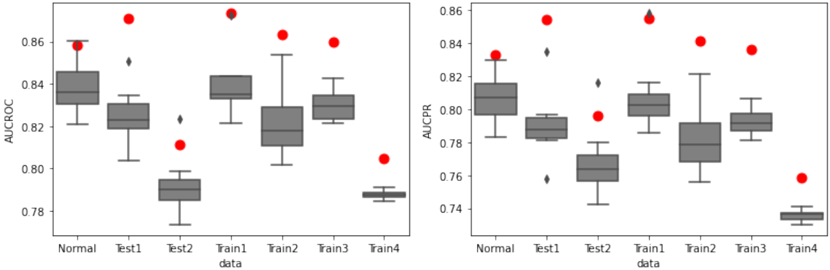

Figure: (Left) AUCROC, (Right) AUCPR, boxplot is made with 8 adaptive-AE models and red dot is ensemble result

Ensemble Result: semi-supervised learning case

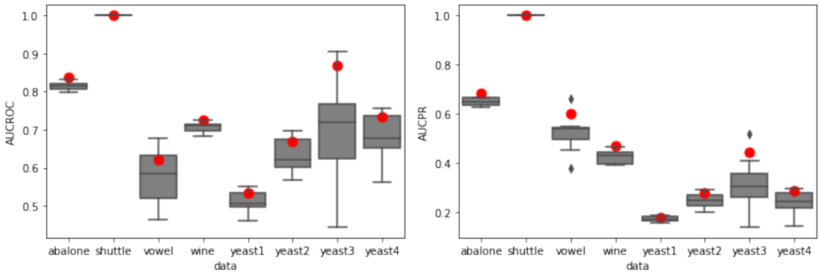

Figure: (Left) AUCROC, (Right) AUCPR, boxplot is made with 8 adaptive-AE models and red dot is ensemble result

Conclusion

- Enables the autoencoders to learn better representations of the inliers
- With proper hyperparameters, Adaptive-AE can get better result than base method
- Adaptive-AE can make diverse autoencoder models, so ensemble approach could be effective
- Adaptive-AE ensemble can get robust and improved result (less necessary hyperparamter tuning)

Future work

- enhance the process of subtracting anomalies in training