Review of Anomaly Detection of Time Series with Smoothness-Inducing Sequential Variational Auto-Encoder

Shihao Zhang

December 2023

1 Key concepts

1.1 Time Series anomaly detection

Matrix $\mathbf{X} \in \mathbb{R}^{M \times T}$, where M correlated streams with T time steps for each stream.

- Subsequence or whole sequence level: $\mathbf{X}_{m,t_1:t_2}$.
- Point Level: $x_{m,t}$.

Anomaly detection recovers "normal" patterns in the presence of anomalies. Then anomalies can be detected via a certain criterion by comparison of the reconstructed data and the raw data.

1.2 AE, VAE, SVAE, SISVAE

1. AE $\rightarrow$ VAE: VAE maps input data to distributions instead of fixed vectors of AE. (Suitable for generating new data points and modelling uncertainty).
2. VAE $\rightarrow$ SVAE: SVAE incorporate the temporal structure to handle the time series data. (RNN)
3. SVAE $\rightarrow$ SISVAE: SISVAE induce smoothness prior (regularizer) to the generative model to account for the biases caused by anomalies.

1.3 Main Contributions

The article focuses on unsupervised point level anomaly detection for multidimensional correlated time series by proposing SISVAE.

1. Smoothness prior (regularizer): incorporate the smoothness prior to the learning of a deep generative model.

2. Liberate the constant noise assumption: parameterize mean and variance of the distributions individually for each time-stamp.

3. Good performance both on synthetic data and real-world data: use reconstruction probability and reconstruction error as decision criteria.

2 Method

2.1 Problem Formulation

The goal is to recover the true signal $f(t)$ and calculate the anomaly score matrix A, then compare the anomaly score with the threshold to get anomaly indication matrix I from $x_{1:T}$.

2.2 Deep Generative Models and SGVB

$$\tilde{L} = -D_{KL}(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})||p_{\boldsymbol{\theta}}(\mathbf{z})) + \frac{1}{L}\sum_{l=1}^{L}(\log p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z}^l))$$

, where $\mathbf{z}^l = g_{\boldsymbol{\phi}}(\boldsymbol{\epsilon}^l, \mathbf{x})$, $\boldsymbol{\epsilon}^l \sim p(\boldsymbol{\epsilon})$, $\boldsymbol{\theta}$ is generative model parameters, and $\boldsymbol{\phi}$ is inference model parameters.

2.3 SVAE

1. Inference Model: Approximated posterior: $\mathbf{z}_t \sim q(\mathbf{z}_t|\mathbf{x}_t) = N(\boldsymbol{\mu}_{z,t}, diag(\boldsymbol{\sigma}^2_{z,t}))$, where $[\boldsymbol{\mu}_{z,t}, \boldsymbol{\sigma}_{z,t}] = \varphi^{enc}_{\phi}(\varphi^{\mathbf{x}}_{\theta}(\mathbf{x}_t), \mathbf{h}_{t-1})$

2. Generative Model:

   - Data reconstruction: $\hat{\mathbf{x}} \sim p(\hat{\mathbf{x}}|\mathbf{z}_t) = N(\boldsymbol{\mu}_{x,t}, diag(\boldsymbol{\sigma}^2_{x,t}))$, where $[\boldsymbol{\mu}_{x,t}, \boldsymbol{\sigma}_{x,t}] = \varphi^{dec}_{\theta}(\varphi^{\mathbf{z}}_{\theta}(\mathbf{z}_t), \mathbf{h}_{t-1})$ (Bernoulli for binary data, multivariate Gaussian for continuous data)
   - Prior of the latent state: $\mathbf{z}_t \sim p_{\theta}(\mathbf{z}_t|\mathbf{x}_{<t}, \mathbf{z}_{<t}) = N(\boldsymbol{\mu}_{0,t}, diag(\boldsymbol{\sigma}^2_{0,t}))$, where $[\boldsymbol{\mu}_{0,t}, \boldsymbol{\sigma}_{0,t}] = \varphi^{prior}_{\theta}(\mathbf{h}_{t-1})$
   - Hidden states of RNN: $\mathbf{h}_t = f_{\theta}(\varphi^{\mathbf{z}}_{\theta}(\mathbf{z}_t), \mathbf{h}_{t-1})$ (Gated Recurrent unit (GRU))

3. Reparameterization of $\mathbf{z}$: $\mathbf{z}_t = \boldsymbol{\mu}_{z,t} + \boldsymbol{\sigma}_{z,t} \odot \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim N(0, \mathbf{I})$

2.4 Smoothness prior

It assumes that the probability density function over time would vary smoothly over time. Among all possible reconstructed series, the true signal should have a smooth transition of distribution.

$$-\lambda D_{KL}(p_{\boldsymbol{\theta}}(\hat{\mathbf{x}}_{t-1}|\mathbf{z}_{\leq t-1}, \mathbf{x}_{<t-1})||p_{\boldsymbol{\theta}}(\hat{\mathbf{x}}_t|\mathbf{z}_{\leq t}, \mathbf{x}_{<t}))$$

2.5 Training

Divide the long time series into short chunks, using sliding windows, which slide over multiple time series synchronously.

3 Anomaly Scores

Given an input $\{\mathbf{x}_t\}_{1:W}$, the model encode it into $\mathbf{z}_{1:W}$, then decode to $\{\hat{\mathbf{x}}\}_{1:W}$.

3.1 Reconstruction error

$$e_{m,t} = ||x_{m,t} - \mu_{m,t}||$$

, where $e_{m,t}$ is anomaly score for observation $x_{m,t}$. $\mu_{m,t}$ is the reconstructed mean.

3.2 Reconstruction probability

$$-\frac{1}{L}\sum_{l=1}^{L}\log(p_{\boldsymbol{\theta}}(\mathbf{x}_t|\mathbf{z}_t^{(l)}))$$

4 Performance

Imbalanced number of anomaly data and normal data. AUPRC is more discriminate than AUROC.

4.1 Synthetic Data

- 100×200 sequence from a multioutput Gaussian process.
- Sequences are correlated through a random sampled positive definite coregionalization matrix.
- Random nonstationary Guassian noise for all data point.
- Bernoulli for different fractions of anomalies.
- Poisson noise for anomalies with intensity λ_t of each anomaly is proportional to the amplitude of inserted place x_t.
- Plus and minus sigh is sampled from Bernoulli with $p = 0.5$.

1. Effects of Anomaly Proportion
2. Effects of Regularization Hyper-parameter
3. Convergence Study

4.2 Real-World Data

- Yahoo's S5 Webscope Data
- Open μPMU Power Network Data

1. Threshold-Based Anomaly Detection
2. Model Characteristics
3. Ranking-Based Anomaly Detection
4. Tradeoff Between Recall and Precision

4.3 Case Study

1. False Alarms on Normal Time Series
2. Detection of Point-Level Anomalies
3. Detection of Subsequence Level Anomalies

4.4 Comparision with Donut

Simulation Study on Univariate Time Series

5 Useful tricks

- Sliding windows.
- Changing multidimensional data to univariate data.
- precision@k
- Reparameterization Trick

6 Questions

- SGVB and AEVB
- SMC
- precision, recall, f1-score, accuracy
- Reparameterization Trick