Anomaly detection of multicariate time series

A summary of detection method of multivariate time series abnormality under the CDN network

Wang Zhong, Gu Rui, Pan Zhongli
LanZhou University, IC@PS

Lanzhou University Annual meeting paper

Abstract

This paper summarizes CDN KPI anomaly detection algorithms and their evaluation methods. Different websites' KPIs exhibit distinct structures and non-stationary relationships over time, posing a challenge for deep learning approaches. To address this, we propose an improved method.

Introduction

With the growth of Internet companies, traditional methods face challenges in detecting KPI exceptions, and machine learning-based CDN KPI abnormal detection is essential[1]. CDN operators collect various KPIs, and the Fig.1 shows an example. Deep anomaly detection uses RNN for feature extraction, but current methods cannot handle non-stationary dependencies and multi-website anomalies effectively.

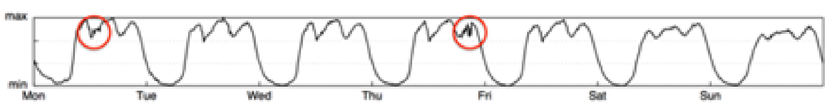

Figure 1: The amount of data visited on the network page of an Internet company was abnormal

Challenge

Challenge 1: Non-stationary dependencies between time periods degrade deep anomaly detection models. Fig.2(a) and Fig.2(d) show different user request behavior on weekdays vs. weekends. KPI changes during scheduled node sets, as shown in Fig.2(b), resulting in non-stationary temporal characteristics. Current methods struggle to capture expected patterns, leading to inferior performance in anomaly detection for CDNs.

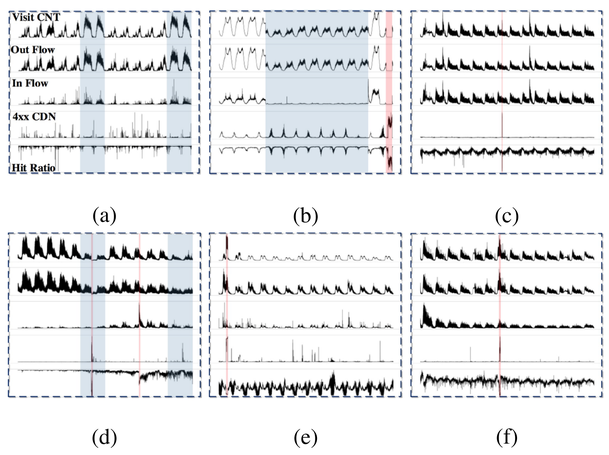

Figure 2: 2-weeks real world typical multivariate CDN KPIs of 6-websites. Periods in light blue show the change points in KPIs; Regions highlighted in red represent the ground-truth anomaly segments.

Challenge 2: CDNs provide services for a diverse set of websites, but current deep anomaly detection models struggle to capture the dynamic complexity of these websites in one model. Training an individual model for each website consumes a lot of resources and raises maintenance costs. Additionally, KPIs of different websites may exhibit similar characteristics, rendering individual models for each website unnecessary.

Methods

CDN KPIs are denoted by x_n, where $n = 1, ..., N$ and N is the number of KPI time series. The observation at time t is $x_{t,n} \in \mathbb{R}^V$, a V-dimensional vector, thus $x_n \in \mathbb{R}^{T \times V}$, where T is the duration of x_n. Anomaly detection on CDN KPIs aims to determine whether $x_{t,n}$ is anomalous or not for a given website. To achieve unsupervised anomaly detection, robust representations of the input data need to be learned using an efficient method.

Variational RNN (VRNN)

The VRNN[4] has VAE[5] at each timestep, conditioned on the last timestep's latent states. The prior on latent variables is not standard Gaussian as assumed in VAE, but follows a specific distribution.

$$\begin{aligned} z_t &\sim \mathcal{N}\left(\mu_{0,t}, \operatorname{diag}(\sigma_{0,t}^2)\right), \\ \left[\mu_{0,t}, \sigma_{0,t}\right] &= \varphi_\tau^{\text{dec}}\left(\varphi_\tau^z(z_t), h_{t-1}\right) \end{aligned} \tag{1}$$

The generating distribution of the VRNN is conditioned on both the latent variable z_t and the hidden state variable h_{t-1} as

$$\begin{aligned} x_t|z_t &\sim \mathcal{N}\left(\mu_{x,t}, \operatorname{diag}(\sigma_{x,t}^2)\right), \\ \left[\mu_{x,t}, \sigma_{x,t}\right] &= \varphi_\tau^{\text{dec}}\left(\varphi_\tau^z(z_t), h_{t-1}\right) \end{aligned} \tag{2}$$

To infer z_t, like standard VAE, a Gaussian distributed variational distribution $q(z_t|x_t)$ is used to approximate its true posterior. In a similar fashion with the generative process of VRNN, the variational distribution is a function of both x_t and h_{t-1} as

$$\begin{aligned} z_t|x_t &\sim \mathcal{N}\left(\mu_{x,t}, \operatorname{diag}(\sigma_{z,t}^2)\right), \\ \left[\mu_{z,t}, \sigma_{z,t}\right] &= \varphi_\tau^{\text{enc}}\left(\varphi_\tau^x(x_t), h_{t-1}\right) \end{aligned} \tag{3}$$

Our Methods

To model the data diversity at different timesteps of multivariate CDN KPIs and solve the non-stationary temporal dependence between them, we extend VRNN into SGmVRNN. Specifically, unlike the standard VRNN, we introduce latent discrete variable $c_{t,n}$ and assign the form of the prior of the latent random variable as

$$z_{t,n}|c_{t,n} \sim \prod_{k=1}^{K} \mathcal{N}\left(z_{t,n}|\mu_k, \operatorname{diag}(\sigma_k)\right)^{c_{t,n,k}} \tag{4}$$

In this way, we assign a different Gaussian distributed prior conditioned on $c_{t,n}$ for latent variable $z_{t,n}$ at different timesteps, which indicates the diverse distribution characteristics of input. Moreover, marginalizing $c_{t,n}$, we can achieve

$$z_{t,n} \sim \sum_{c_{t,n}} p\left(c_{t,n}|\pi_{t.n}\right) p\left(z_{t,n}|c_{t,n}\right) = \sum_{k=1}^{K} \pi_{t,n,k} \mathcal{N}\left(\mu_k, \operatorname{diag}(\sigma_k)\right) \tag{5}$$

Clearly, it is mixture Gaussian distribution with higher representation power than a Gaussian distribution, and thus is just ideal for characterizing the complex structure and temporal characteristics within multivariate CDN KPIs.

Model Training

Our training objective is to minimize the distance between the variational distribution of latent variables $q\left(z_{t,n}, c_{t,n}\right)$ and their true posterior distribution $p(z_{t,n}, c_{t,n}|-)$, which can be quantified by the Kullback-Leibler (KL) divergence, $\operatorname{KL}\left(q(x)||p(x)\right) = \int q(x) \log \frac{q(x)}{p(x)} dx$. So, during the inference of objective, we aim to minimize $\operatorname{KL}\left(q(z_{t,n}, c_{t,n})||p(z_{t,n}, c_{t,n}|-)\right)$. Following the usual strategy of VAE, we can transfer this KL divergence as

$$\begin{aligned} &\operatorname{KL}\left(q(z_{t,n}, c_{t,n})||p\left(z_{t,n}, c_{t,n}|-\right)\right) \\ &= \log p(x_{t,n}) - \mathbb{E}_{q(z_{t,n})q(c_{t,n})}\left(\log p(x_{t,n}|z_{t,n}, c_{t,n}) - \log \frac{q(z_{t,n})q(c_{t,n})}{p(z_{t,n}|c_{t,n})p(c_{t,n})}\right) \\ &= \log p(x_{t,n}) - \text{ELBO} \end{aligned} \tag{6}$$

Anomaly Detection

We apply the reconstruction probability of xt as the anomaly score to determine whether an observed variable is anomalous or not[3], and it is computed as

$$S_{t,n} = \log p(x_{t,n}|z_{t,n}) \tag{7}$$

An observation x_t will be classified as anomalous if S_t is below a specific threshold.

Results

Evaluate our method with "one model for one website" and "one model fits all" experiments. Results in Table1, highlighting best F1-score for each dataset.

Method	CDN			SMD		
	P	R	F1	P	R	F1
VRNN	0.9814	0.8317	0.9003	0.9825	0.8383	0.9047
DOMI	0.9665	0.8348	0.8958	0.9770	0.8036	0.8819
OmniAnomaly	0.8385	0.8757	0.8567	0.9801	0.7843	0.8713
SDFVAE	0.9675	0.8615	0.9115	0.9810	0.8498	0.9107
SGmVRNN	0.9667	0.9204	**0.9430**	0.9607	0.9123	**0.9356**

Table 1: Performance of "one model fits all websites".

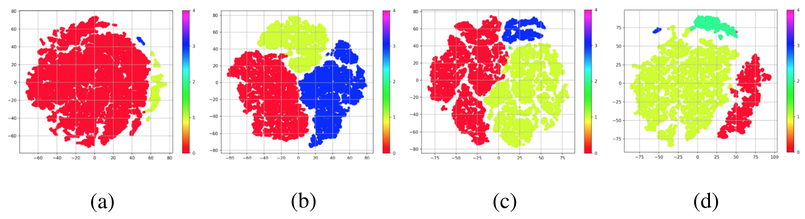

Figure 3: The visualization of latent variables $z_{t,n}$

In Fig.3, each dot represents a latent variable of an observation at a timestep and colors represent clustering groups. Our model can separate KPIs with different structural characteristics into different groups, which confirms the two challenges mentioned and validates our model's effectiveness.

Conclusion

- we propose a method to cope with the anomaly detection challenges that brought by the natural characteristics of multivariate CDN KPIs of diverse websites.

References

[1] Behrouz Zolfaghari, Gautam Srivastava, Swapnoneel Roy, Hamid R. Nemati, Fatemeh Afghah, Takeshi Koshiba, Abolfazl Razi, Khodakhast Bibak, Pinaki Mitra, and Brijesh Kumar Rai. 2020. Content Delivery Networks: State of the Art, Trends, and Future Roadmap. ACM Comput. Surv. 53, 2, Article 34 (March 2021), 34 pages. https://doi.org/10.1145/3380613

[2] D. Li, D. Chen, B. Jin, L. Shi, J. Goh, and S. Ng, "MAD-GAN: multivariate anomaly detection for time series data with generative adversarial networks," in Artificial Neural Networks and Machine Learning - ICANN, 2019, pp. 703–716

[3] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. Pei, "Robust anomaly detection for multivariate time series through stochastic recurrent neural network," in ACM SIGKDD, 2019, pp. 2828–2837.

[4] C. Junyoung, K. Kyle, D. Laurent, G. Kratarth, C. C. Aaron, and B. Yoshua, "A recurrent latent variable model for sequential data," in NeurIPS, 2015, pp. 2980–2988.

[5] P. K. Diederik and W. Max, "Auto-encoding variational bayes," in ICLR, 2014.