Semi-supervised Variational Multi-view Anomaly Detection

Shaoshen Wang, Ling Chen, Farookh Hussain, and Chengqi Zhang

University of Technology Sydney, Sydney, Australia
shaoshenwang@gmail.com,{ling.chen,farookh.hussain,Chengqi.Zhang}@uts.edu.au

Abstract. Multi-view anomaly detection (Multi-view AD) is a challenging problem due to the inconsistent behaviors across multiple views. Meanwhile, learning useful representations with little or no supervision has attracted much attention in machine learning. There are a large amount of recent advances in representation learning focusing on deep generative models, such as Variational Auto Encoder (VAE). In this study, by utilizing the representation learning ability of VAE and manipulating the latent variables properly, we propose a novel Bayesian generative model as a semi-supervised multi-view anomaly detector, called MultiVAE. We conduct experiments to evaluate the performance of MultiVAE on multi-view data. The experimental results demonstrate that MultiVAE outperforms the state-of-the-art competitors across popular datasets for semi-supervised multi-view AD. As far as we know, this is the first work that applies VAE-based deep models on multi-view AD.

Keywords: Multi-view Anomaly Detection· VAE · Semi-supervised

1 Introduction

Anomaly Detection (AD) algorithms [3] aim to identify data points that are significantly different from the remaining data. While the traditional problem setting focuses on data of single-view, we have to deal with data of multiple views in many practical scenarios. For example, social media content usually contains texts, images and user behavior features that provide complementary information of the same items from different perspectives. Consequently, multi-view anomaly detection emerges as a crucial research problem that finds many real-world applications such as purchase behavior analysis [6], malicious insider detection [14], and disparity management [5].

Following the terminology proposed in [19] and [13], multi-view anomalies can be grouped into three categories: 1) *Attribute anomalies*, which refer to instances that exhibit abnormal behaviours in each view. For example, given the data of two views in Figure 1, the yellow triangles represent an attribute anomaly because it behaves differently from other instances in each view. 2) *Class anomalies*, which refer to instances that exhibit inconsistent characteristics across different views. Such instances behave normally in each view. However, if multiple views are considered collectively, the instances show inconsistent properties. For example, the red squares in Figure 1 represent a class anomaly. Even though it appears to be normal by falling into a cluster in each view, it is similar to different sets of instances in different views. 3) *Class-attribute*

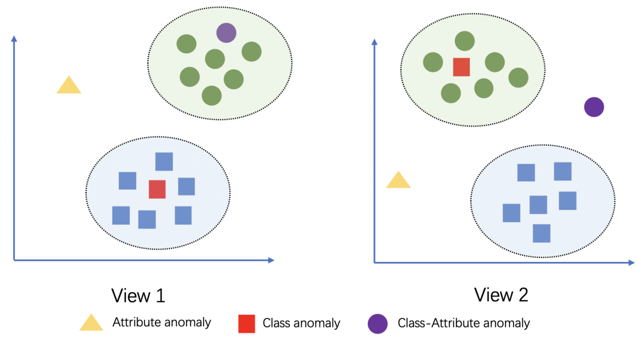


Fig. 1. Three classes of Multi-view anomaly.

anomalies, which are instances having characteristics of attribute anomalies in some views and properties of class anomalies in the other views. The purple circles in Figure 1 correspond to a class-attribute anomaly.

Existing multi-view AD algorithms have their respective limitations in detecting the three types of anomalies simultaneously. For example, clustering-based algorithms, such as HOrizontal Anomaly Detection (HOAD) [6] and Affinity Propagation (AP) [15], become less effective when there is no clear tendency of clusters in the data; while other modals, such as Probabilistic Latent Variable Model (PLVM) [8], Latent Discriminant Subspace Representatio (LDSR) and the recent Hierarchical Bayesian Model (HBM) [17], assume there exists a linear transformation between instances and a common latent variable shared by all views. However, it fails to capture the nonlinearities and impairs the detecting capability for complex data distributions. Moreover, it is risky to assume all the views share a same latent variable, especially for complicated data distribution. Therefore, more advanced algorithms are desired for multi-view AD.

Most existing anomaly detection methods are unsupervised, having access to unlabelled data including both normal and anomalous instances. In real practices, however, it is easy to obtain labeled normal samples because anomalies are defined to be rare [17]. Therefore, it is practically meaningful to develop semi-supervised anomaly detection algorithms trained on labeled normal data, which is supposed to have better performance as the trained model can capture characteristics of normal data better than those trained on polluted data.

In this paper, we propose a Bayesian generative model for semi-supervised multi-view anomaly detection, called MultiVAE. In order to learn the correlation between views directly and detect the inconsistency existing in aforementioned three types of anomalies, we leverage the representation learning ability of VAE to learn a latent distribution for each view, which is then used to reconstruct other views of the same instance. Three types of multi-view anomalies are detected by an anomaly score based on cross-view reconstruction losses. Further, we propose an importance weighted version of reconstruction loss to achieve higher accuracy and stability. Noticing that many real applications of multi-view AD involve data of discrete values, we introduce a Categorical distribution as the likelihood in the decoder of MultiVAE. To enable gradient-based optimization of the model, the Gumbel-softmax [10] technique is utilized to approximate the Categorical distribution.

2 Preliminaries

As a directed probabilistic graphical model (DPGM), a Variational Auto-encoder (VAE) [12] aims to learn a Bayesian latent variable model by maximizing the log-likelihood of the training data $\{x^{(i)}\}_{i=1}^{N}$ via variational inference. It introduces a distribution $q_{\phi}(z|x)$ to approximate the intractable true posterior $p(z|x)$. Mean and variance vectors μ_v and σ_v are estimated by the encoder, and the latent variable z_v is sampled via reparameterization trick [12]. Then, the decoder takes z_v as input to generate X'_v as a reconstruction for X_v. VAE is trained by maximizing the following Evidence Lower Bound (ELBO):

$$\mathcal{L} = \mathbb{E}_{z \sim q_{\phi}(z|x^{(i)})}[\log\, p_{\theta}(x^{(i)}|z)] - D_{\mathrm{KL}}(q_{\phi}(z|x^{(i)})||p(z))$$

The importance weighted autoencoder (IWAE) [2] is an important variant of the vanilla VAE. IWAE computes a tighter lower bound through appropriate weighting of a multi-sample estimator, as

$$\mathcal{L}_{IWAE} = \mathbb{E}_{z^{1:K} \sim q_{\phi}(z|x)}[\log\, \Sigma_{k=1}^{K} \frac{1}{K} p_{\theta}(x|z^{k})] - D_{\mathrm{KL}}(q_{\phi}(z|x)||p_{\theta}(z))$$

3 Methodology

3.1 Problem Setting and Proposed Framework

Suppose that we are given N instances $\{X_1, ..., X_n\}$ with D views. $X_n = (\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(D)})$ is a set of multi-view observation vectors for the n-th instance, and $\boldsymbol{x}^{(d)} \in \mathbb{R}^{M_d}$ is the observation vector of the d-th view where M_d is the corresponding dimensionality. The objective of semi-supervised multi-view AD is to find anomalous instances that have inconsistent characteristics or behaviors across multiple views (i.e., the three types of anomalies discussed in Section 1). The training set includes only normal instances while the testing set contains both normal and anomalous instances. For simplicity, we focus on the situation when $D = 2$. However, as will be discussed, our model can be extended to handle more views straightforwardly.

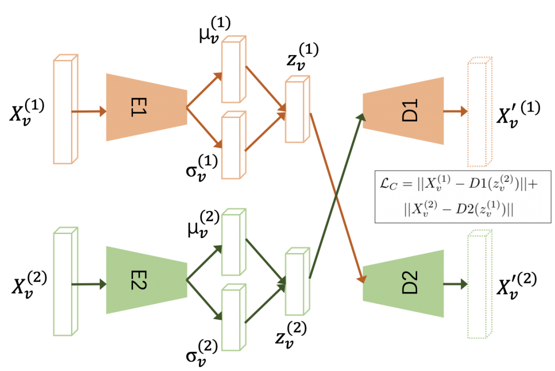


Fig. 2. Architecture of MultiVAE. The model aims to learn cross-view dependency by cross reconstruction.

As mentioned above, multi-view anomalies show inconsistencies between views. That is, if a data point is anomalous, it is likely that its cross-view dependency is lower than that of normal data points. Given a data instance X of two views, we model the cross-view dependency using the conditional distribution between two views, $p(X^{(2)}|X^{(1)})$ and $p(X^{(1)}|X^{(2)})$. Then, given a normal instance $X_i = [X_i^{(1)}, X_i^{(2)}]$ and an anomalous instance $X_j = [X_j^{(1)}, X_j^{(2)}]$ (being a class anomaly, an attribute anomaly, or a class-attribute anomaly), we have the following assumption:

$$p(X_i^{(2)}|X_i^{(1)}) + p(X_i^{(1)}|X_i^{(2)}) > p(X_j^{(2)}|X_j^{(1)}) + p(X_j^{(1)}|X_j^{(2)})$$

Based on the cross-view dependency assumption, we propose MultiVAE which generates one view from the other. For example, as shown in Fig 2, the model utilizes the input of view_1 to estimate the latent variable $z_v^{(1)}$ and generate output for view_2. In the same time, view_2 is used to generate view_1. The model can be expressed as follows:

$$\mu^{(1)} = f_{\phi_1^1}(X^{(1)}), \sigma^{(1)} = f_{\phi_2^1}(X^{(1)}); \mu^{(2)} = f_{\phi_1^2}(X^{(2)}), \sigma^{(2)} = f_{\phi_2^2}(X^{(2)})$$
$$z_v^{(1)} \sim Gaussian(\mu^{(1)}, \sigma^{(1)}), z_v^{(2)} \sim Gaussian(\mu^{(2)}, \sigma^{(2)})$$
$$X_v'^{(2)} \sim p_{\theta_2}(X|z_v^{(1)}), X_v'^{(1)} \sim p_{\theta_1}(X|z_v^{(2)})$$

The parameters of model is set $\{\phi_1^1, \phi_2^1, \phi_1^2, \phi_2^2, \theta_1, \theta_2\}$. f is a function parameterized by NN serving as the encoder, p is a predefined distribution serving as the decoder. Note that, similar to VAE, we can seek help from amortized variational inference to estimate the parameters of the model.

Our framework can be extended straightforwardly to handle more views with linear time complexity with respect to the number of views. Suppose there are n views where $n > 2$, we can concatenate all the $n-1$ views except view i to generate view i. Regarding the time complexity, suppose each encoder or decoder is implemented by Multilayer perceptron that has time consumption $O(1)$ for feed forward of one instance. Given n training instances, D encoders and D decoders, the overall time complexity is $O(nD)$ for training phase, $O(D)$ for detecting a single instance in testing phase.

3.2 Loss function deduction

For simplification, we denote $X_v^{(1)} = X_1$, $X_v^{(2)} = X_2$. Under the maximum likelihood estimation (MLE) framework, it would be complicated to directly maximize $p(X_1, X_2)$. Hence, we infer the parameters of the model in the way similar to the variational inference used in VAE. We derive an Evidence Lower Bound (ELBO) to approximate the log likelihood of $p(X_1, X_2)$, which can be formulated as follows:

$$\begin{aligned}
&\log\ p(X_1, X_2) \\
&\geq \mathbb{E}_{q(z_1,z_2)}\left[\log\ \frac{p(X_1, X_2, z_1, z_2)}{q(z_1, z_2)}\right] \textit{ (Jensen's inequality)} \\
&\approx \mathbb{E}_{q(z_1|X_1)q(z_2|X_2)}\left[\log\ \frac{p(X_2|z_1)p(X_1|z_2)p(z_1)p(z_2)}{q(z_1|X_1)q(z_2|X_2)}\right] \\
&\quad \textit{(mean-field approximation)} \\
&= -\mathcal{L}_{\mathcal{C}} - D_{\mathrm{KL}}(q_\phi(z_1|X_1)||p(z_1)) - D_{\mathrm{KL}}(q_\phi(z_2|X_2)||p(z_2))
\end{aligned}$$

where $\mathcal{L}_{\mathcal{C}}$ represents the reconstruction error, D_{KL} represents Kullback-Leibler divergence between two distributions, $p(z_1)$ and $p(z_2)$ can be isotropic multivariate Gaussian. In practice, we adopt the importance weighted ELBO as the training objective to obtain tighter lower bound and less variance:

$$\mathcal{L}_{MultiVAE} = \mathbb{E}_{z_1^{1:K} \sim q_\phi(z_1|X_1)}[\log \Sigma_{k=1}^{K} \frac{1}{K} p_\theta(X_2|z_1^k)] - D_{\mathrm{KL}}(q_\phi(z_1|X_1)||p_\theta(z_1)) + \mathrm{E}_{z_2^{1:K} \sim q_\phi(z_2|X_2)}[\log \Sigma_{k=1}^{K} \frac{1}{K} p_\theta(X_1|z_2^k)] - D_{\mathrm{KL}}(q_\phi(z_2|X_2)||p_\theta(z_2))$$

3.3 Categorical ditribution for discrete data

Discrete data (e.g., categorical data) frequently appears in many real AD applications. VAE-based models usually use Gaussian or Bernoulli distribution as $p(x|z)$ for generating instances, which performs poorly on discrete data. In order to alleviate this issue, we assume that discrete data are generated from categorical distribution. We introduce a K dimensional categorical distribution $Cat(\pi_1, \pi_2 ... \pi_k)$ as $p(x|z)$ and assume K dimensional one-hot instance vector $\boldsymbol{x}$ sampled from the generative process: $\boldsymbol{x} \sim p(x|z) = Cat(\pi_1, \pi_2 ... \pi_k)$ where $Cat(x = i|\pi_1, \pi_2 ... \pi_k) = \pi_i$ and $\sum_{i=1}^{K} \pi_i = 1$. However, this sampling process is not differentiable and the model cannot be optimized by gradient-based methods. We thus apply the Gumbel-Softmax re-parameterization trick [10] to approximate the sampling process and make it to be differentiated. The element i of vector $\boldsymbol{x}$ can be approximated as:

$$\boldsymbol{x}_i = \frac{exp((\log(\pi_i) + g_i)/\tau)}{\Sigma_{j=1}^{K} exp((\log(\pi_j) + g_j)/\tau)} \; for\, i = 1, 2 ... K$$

where π_i is estimated by neural network in decoder, $\tau > 0$ is an adjustable hyperparameter which controls how closely the samples approximate discrete values. A smaller τ means a closer approximation. $g \sim Gumbel(0, 1)$ are i.i.d. samples from standard Gumbel distribution.

3.4 Semi-supervised Multi-view Anomaly Detection Score Design

To detect the three types of multi-view anomalies, we compute an anomaly score for each instance. We analyze the behaviour of normal and abnormal data during the testing phase as follows:

For *normal instances*, since the correlation between views accords with normal training data, it's easy to reconstruct one view from another view. This gives rise to a smaller cross reconstruction error $\mathcal{L}_{\mathcal{C}}$. For *Class anomalies*, since different view belongs to different class, it is inconsistent cross multiple views, making it difficult to reconstruct one view from another. The cross reconstruction error $\mathcal{L}_{\mathcal{C}}$ should be high. For *Attribute anomalies*, since views belong to distribution different from that of the training data, the correlation between views is different from what the model has learned during training, which also leads to a high $\mathcal{L}_{\mathcal{C}}$. For *Class-Attribute anomalies*, the situation is similar to *attribute anomalies* so that this type of anomalies can be detected by high $\mathcal{L}_{\mathcal{C}}$ as well.

Overall, it is expected that the cross-view reconstruction error $\mathcal{L}_{\mathcal{C}}$ for a normal instance is less than that of an anomalous instance, including *Class anomaly*, *Attribute* *anomaly* and *Class-attribute anomaly*. Therefore, we use $\mathcal{L}_{\mathcal{C}}$ as the anomaly score for an instance $\boldsymbol{x}_i$ as follows (K is the number of importance weighted sampling):

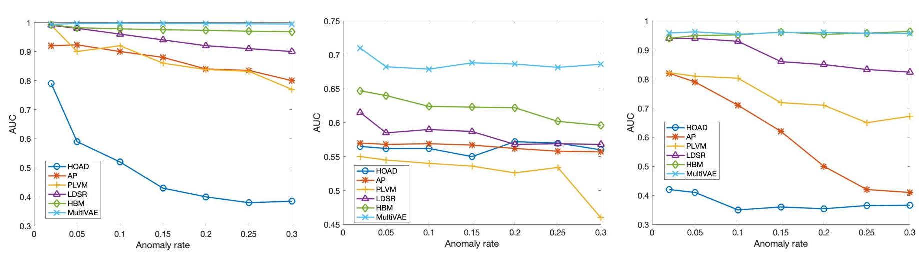
$$S_{\text{MultiVAE}}(\boldsymbol{x}_i) = \mathcal{L}_{\mathcal{C}} = -(\mathbb{E}_{z_1^{1:K} \sim q_\phi(z_1|\boldsymbol{x}_i^{(1)})}[\log \Sigma_{k=1}^{K} \frac{1}{K} p(\boldsymbol{x}_i^{(2)}|z_1^k)] + \mathbb{E}_{z_2^{1:K} \sim q_\phi(z_2|\boldsymbol{x}_i^{(2)})}[\log \Sigma_{k=1}^{K} \frac{1}{K} p(\boldsymbol{x}_i^{(1)}|z_2^k)]) \quad (1)$$

As S_{MultiVAE} measures the reconstruction error, a higher score indicates a larger probability of an instance being anomalous.

4 Experiments

We compare our method with the HOrizontal Anomaly Detection (HOAD) [6], Affinity Propagation (AP) [15], the Probabilistic Latent Variable Model (PLVM) [8], the Latent Discriminant Subspace Representation (LDSR) [13] and the state-of-the-art Hierarchical Bayesian Model (HBM) [17] to evaluate performance of MultiVAE. Seven datasets, including Thyroid, Annthyroid, Forestcover, Vowels, Pima, Wine and Glass, from the ODDS library [16] are used. We follow the same experimental setting in [17] for fair comparison. Three types of multi-view anomalies are generated in the same way described in previous works [17,13]. We use the area under the ROC curve (AUC) as the evaluation measure. The higher the AUC is, the better the approach performs.

Table 1 shows the average AUCs achieved by the comparing models on the 7 datsets of the training data (Some results of the baselines are from [17]). It demonstrates the advantage of our proposed approach for multi-view AD clearly, where MultiVAE consistently outperforms the other models on most datasets. This can be explained by the capability of capturing the correlation among multiple views via multiple latent vectors, supported by the non-linear learning ability of neural networks together with cross-view reconstruction.



(a) Attribute anomaly (b) Class anomaly (c) Class-attribute anomaly

Fig. 3. The curves of AUC W.R.T anomaly rate in Pima dataset.

To investigate how the anomaly rate affects the performance of different models, we experiment on data polluted by an increasing percentage of outliers. Fig 3 shows the variation of AUCs on data set *pima* with outlier ratio of 2%, 5%, 10%, 15%, 20%,

	Model	Thyroid	Annthyroid	ForestCover	Vowels	Pima	Wine	Glass
A	HOAD	.5202±.0864	.5078±.0724	.6801±.0866	.8540±.0691	.5921±.0768	.6503±.1574	.7083±.1410
	AP	.6737±.1164	.5747±.0669	.6774±.0739	.7062±.1125	.9376±.0293	.6947±.1078	.7497±.1117
	PLVM	.8989±.0091	.8904±.0363	.4870±.0126	.5481±.0067	.9086±.0083	.4058±.0481	.4087±.0246
	LDSR	.9751±.0074	.9876±.0022	.9983±.0005	.9181±.0153	.9858±.0057	**.9932±.0009**	**.9940±.0040**
	HBM	.9877±.0056	.9979±.001	.9995±.0027	.9875±.0071	.9877±.0044	.9417±.0450	.9530±.0292
	MultiVAE	**.9991±.0005**	**.9996 ±.0001**	**.9998±.0001**	**.9988±.0010**	**.9964±.0030**	.9658±.0316	.9824±.0165
C	HOAD	.5393±.0303	.5849±.0348	.6872±.0337	.3818±.0384	.5557±.0310	.7124±.0638	.4277±.0932
	AP	.5847±.0227	.5265±.0350	.7906±.0332	.7520±.0513	.5659±.0365	.5629±.0933	.5576±.0518
	PLVM	.5676±.0090	.4087±.0176	.6035±.0044	.5479±.0282	.5425±.0138	.4860±.0040	.5433±.0104
	LDSR	.8631±.0217	.7128±.0418	.7551±.0293	.9245±.0173	.5924±.0543	.5889±.0916	.7098±.0498
	HBM	.8744±.0205	.7383±.0450	.8672±.0197	.9360±.0158	.6354±.0400	.8373±.0424 .	.7613±.0570
	MultiVAE	**.9678±.0144**	**.7891±.0819**	**.9814±.0118**	**.9711±.0132**	**.6562±.0670**	**.9068±.0794**	**.7698±.1490**
C-A	HOAD	.4934±.0270	.4976±.0311	.4342±.0468	.5994±.1342	.4181±.0260	.5798±.0615	.5598±.0652
	AP	.6380±.0723	.5647±.0819	.8054±.0373	.8511±.0713	.7916±.0555	.5481±.1173	.7308±.0676
	PLVM	.7122±.0191	.8933±.0134	.8184±.0087	.6390±.0223	.8249±.0063	.7094±.0145	.9555±.0092
	LDSR	.9344±.0179	.9122±.0220	.9845±.0049	.9642±.0064	.9315±.0146	**1±0**	.9900±.0026
	HBM	.9863 ±.0075	.9842±.0076	.9857±.0095	.9757±.0082	.9510±.0169	.9201±.0470	**.9984±.0023**
	MultiVAE	**.9930±.0057**	**.9943±.0054**	**.9989±.0012**	**.9937±.0045**	**.9571±.0232**	.9018±.0872	.9456±.0300

Table 1. AUC values (mean±std) of semi-supervised multi-view AD on seven datasets with anomaly rate = 0.05. (A:Attribute anomaly; C:Class anomaly; C-A:Class-Attribute anomaly)

25% and 30% for three types of outliers. We see that, in general, as the anomaly rate increases, the performance decreases. And the proposed method is comparatively robust compared with the other methods.

To evaluate the situation when there are more than two views of the data (i.e., $n > 2$), we run the comparative methods on the WebKB dataset [1], which has been widely used for evaluating multi-view learning algorithms. We use its Cornell subset in our experiment, which contains 195 webpages over 5 labels. Each webpage is described by four views: content, inbound link, outbound link and cites. Table 2 shows the AUC values of all compared methods on the dataset with outlier ratio of 5% and 10%. It can be observed that MultiVAE again achieves higher AUC than its competitors, which demonstrates the strength of our Bayesian detector.

	HOAD	AP	LDSR	HBM	MultiVAE
5 % anomalies	0.811	0.755	0.672	0.930	**0.942**
10 % anomalies	0.769	0.715	0.647	0.922	**0.934**

Table 2. AUC values of competitors on 4-view WebKB dataset.

5 Conclusion

We propose a VAE-based deep framework for multi-view AD. Under the framework, MultiVAE is developed to model the dependency between views for semi-supervised AD. Our experimental results on benchmark data and real-world data demonstrate the effectiveness of MultiVAE as the first effort that leverages variational auto-encoder in multi-view anomaly detection, by exploiting the cross-view reconstruction loss.

Acknowledgement. This work has been partially supported by ARC DP180100966.

References

1. Blum, A., Mitchell, T.: Combining labeled and unlabeled data with co-training. In: Proceedings of the eleventh annual conference on Computational learning theory. pp. 92–100 (1998)
2. Burda, Y., Grosse, R., Salakhutdinov, R.: Importance weighted autoencoders. arXiv preprint arXiv:1509.00519 (2015)
3. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. ACM Comput. Surv. **41**, 15:1–15:58 (2009)
4. Das, S., Matthews, B.L., Srivastava, A.N., Oza, N.C.: Multiple kernel learning for heterogeneous anomaly detection: algorithm and aviation safety case study. In: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 47–56. ACM (2010)
5. Duh, K., man Au Yeung, C., Iwata, T., Nagata, M.: Managing information disparity in multilingual document collections. ACM Trans. Speech Lang. Process. **10**, 1:1–1:28 (2013)
6. Gao, J., Fan, W., Turaga, D., Parthasarathy, S., Han, J.: A spectral framework for detecting inconsistency across multi-source object relationships. In: 2011 IEEE 11th International Conference on Data Mining. pp. 1050–1055. IEEE (2011)
7. Huang, F., Zhang, X., Li, C., Li, Z., He, Y., Zhao, Z.: Multimodal network embedding via attention based multi-view variational autoencoder. In: Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval. pp. 108–116. ACM (2018)
8. Iwata, T., Yamada, M.: Multi-view anomaly detection via robust probabilistic latent variable models. In: Advances In Neural Information Processing Systems. pp. 1136–1144 (2016)
9. Janeja, V.P., Palanisamy, R.: Multi-domain anomaly detection in spatial datasets. Knowledge and information systems **36**(3), 749–788 (2013)
10. Jang, E., Gu, S., Poole, B.: Categorical reparameterization with gumbel-softmax. arXiv preprint arXiv:1611.01144 (2016)
11. Kawachi, Y., Koizumi, Y., Harada, N.: Complementary set variational autoencoder for supervised anomaly detection. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 2366–2370. IEEE (2018)
12. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
13. Li, K., Li, S., Ding, Z., Zhang, W., Fu, Y.: Latent discriminant subspace representations for multi-view outlier detection. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)
14. Liu, A.Y., Lam, D.N.: Using consensus clustering for multi-view anomaly detection. In: 2012 IEEE Symposium on Security and Privacy Workshops. pp. 117–124. IEEE (2012)
15. Marcos Alvarez, A., Yamada, M., Kimura, A., Iwata, T.: Clustering-based anomaly detection in multi-view data. In: Proceedings of the 22nd ACM international conference on Information & Knowledge Management. pp. 1545–1548. ACM (2013)
16. Rayana, S.: Odds library (2016)
17. Wang, Z., Lan, C.: Towards a hierarchical bayesian model of multi-view anomaly detection. In: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (2020)
18. Xu, H., Chen, W., Zhao, N., Li, Z., Bu, J., Li, Z., Liu, Y., Zhao, Y., Pei, D., Feng, Y., et al.: Unsupervised anomaly detection via variational auto-encoder for seasonal kpis in web applications. In: Proceedings of the 2018 World Wide Web Conference. pp. 187–196 (2018)
19. Zhao, H., Fu, Y.: Dual-regularized multi-view outlier detection. In: Twenty-Fourth International Joint Conference on Artificial Intelligence (2015)