

A Non-parametric Bayesian Model for Multi-view Group Anomaly Detection

Hongtao Wang and Hongmei Wang

¹(North China Electric Power University)

Corresponding Author: Hongmei Wang

ABSTRACT: Multi-view anomaly detection is an important problem in machine learning. Existing approaches can generally deal with the multi-view point anomaly, namely the outlier with abnormal behaviors in each view, or inconsistent cluster memberships across different views. However, traditional point anomaly detection methods are not effective for scenarios where individual instances are normal, but their collective behavior is abnormal. Therefore it is crucial to study the collective behavior and discover multi-view group anomalies. Unfortunately, few previous literatures have discussed this kind of anomaly, consequently no existing methods can be applied directly. In this paper, we propose a novel non-parametric Bayesian model, Enhanced Multi-view Group Anomaly Detection (EMGAD), to address this issue. By representing the multi-view data with different latent group structures and topic distributions, EMGAD automatically infers the groups and detects group anomalies simultaneously. To construct the model, we proposed to use the collapsed Gibbs sampling algorithm with four steps in each iteration. The model is then evaluated on both synthetic and real world datasets with different anomaly settings, and the experimental results demonstrate the effectiveness of our approach on detecting multi-view group anomalies.

KEYWORDS: Multi-view; Group anomaly; Anomaly detection

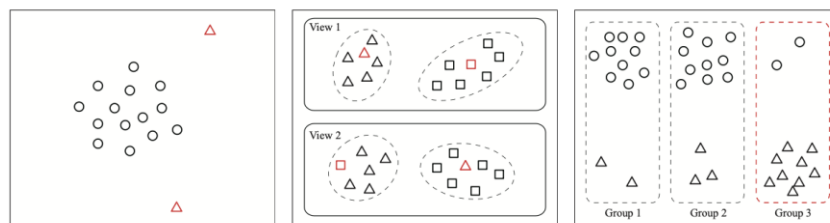
Received 14 December, 2020; Accepted 28 December, 2020 © The author(s) 2020.

Published with open access at www.questjournals.org

I. INTRODUCTION

The recent big data wave has fostered many new technologies that help digest these petabytes to discover hidden knowledge. Among them, anomaly detection has been one of the foci of data science research community. Anomaly detection aims to detect suspicious individuals, patterns, events and groups which do not conform to the expected modality. It can be applied to datasets in almost all domains, such as video surveillance [28], spam detection [10,11,15], malicious insider attack detection [27] and human behavior analytics [18].

Over the past decade, vast amounts of significant work have spawned in this area. According to the perspectives and the types of anomaly, existing work on anomaly detection can be classified into the following three categories: (1) single-view point anomaly detection, (2) multi-view point anomaly detection, and (3) single-view group anomaly detection.



(a) single-view point anomaly (b) multi-view point anomaly (c) single-view group anomaly

Figure 1. Anomalies in Previous Works.

Point anomaly detection, also known as outlier detection, aims to discover those individual data instances whose behavior deviates significantly from the others. Most single-view point anomaly detection methods, including local outlier factor (LOF) [4], angle-based outlier factor (ABOF) [17], one class SVM [24], one-class CRF [25] and isolation forest [20], explore the point anomaly by analyzing the distance of instances from single-view datasets, as shown in Figure 1(a). In recent years, the profusion of diverse data sources has

ushered large amounts of multi-view data such as social media, sensor readings, scientific publication networks or physiological measurements. The goal of multi-view point anomaly detection is to identify those instances with inconsistent behavior among multiple views as anomalies (see Figure 1(b)) [1,12,16,29,33].

However, point anomaly detection methods are incapable for scenarios where individuals are normal, but their collective behavior is abnormal, as shown in Figure 1(c). For example, a group of insiders may collude to leak the confidential information in large organizations; or a group of reviewers may collectively rate a product low. This type of anomaly is called group anomaly [8]. To mitigate this issue, Xiong et al proposed two probabilistic graphical models, MGM [31] and FGM [30], to discover anomalous aggregated group behaviors. Muandet and Schölkopf [21] proposed one-class support measure machines (OCSMMs), to find group anomalies in the space of group feature distributions. However, all of these methods are single-view group anomaly detection, where the groups need to be pre-determined.

In this paper, we investigate a new category of anomaly detection: the multi-view group anomaly detection, which has not been adequately explored. Multi-view group anomaly is a group that can be defined as: given a multi-view dataset, the pattern (or topic distribution) in a number of views does not conform to the majority of other groups. Detecting multi-view group anomaly detection is with promising applicability. For instance, in a scientific community, the highly novel research topics of some groups (usually considered as anomaly) may lead the advances in technologies. However, considering the diverse data generation mechanisms of multi-view data, identifying the multi-view group anomaly is a highly challenging task. First, it is crucial though nontrivial to cluster all the inconsistent groups under different views. Traditionally researchers believe that groups among diverse views always have consistent relations. In this paper, however, we argue that for multi-view data, diverse views may be with completely different structure of clusters. This phenomenon can be observed in many practical domains. For example, a twitter user may have demography view, location view, friends view and tweets view, and so forth. For users in the same group under the friend view, their posted topics in tweets view may vary dramatically. Second, the number of groups under different perspectives is not specified before the clustering process. The isolated but coupled nature of multiple views, call for a novel data driven mechanism that can automatically determine the suitable number of groups. Third, identifying groups with abnormal behavior is also nontrivial, and no criterion in previous work has been defined for multi-view data.

Along this line of research, the similar work is [32], which detects group anomaly in social media under two different views: social relationships and instance features. However, this method only deals with pairwise data to form the social community and requires a predetermined number of groups, which is in general not available for all perspectives. [34] proposed the first general model to deal with this problem. In this paper, we propose an enhanced model (EMGAD) that relaxes the assumption of the predefined topic number under topic views. Extensive experiments on both synthetic and real world datasets show the effectiveness of proposed model. The major contributions of this paper can be summarized as follows:

- We deal with the Multi-view Group Anomaly Detection problem in a fully non-parametric manner.
- We develop an enhanced model called EMGAD that extends MGAD and requires no prior knowledge on the numbers of both groups or topics.
- To evaluate the proposed model, we conduct extensive experiments on both synthetic and real-world datasets.

The rest of the paper is organized as follows. In section 2 we briefly review the related work, and then formally define the multi-view group anomaly detection problem and propose the enhanced MGAD model to discover group anomaly in Section 3. In Section 4, we evaluate the proposed EMGAD model using both synthetic and real world datasets, and report the analysis results. Finally in Section 5 we conclude the paper.

II. RELATED WORKS

Multi-view data learning has been ubiquitous in recent years since multiple view often provides coherent or complimentary information. Studying multi-view Data is with great challenges compared to single-view data due to different generation mechanisms under different views. In this section we review the related models on multi-view data learning and group anomaly detection.

The most popular research on multi-view data is multi-view clustering or co-clustering. Multi-view clustering aims to divide objects into clusters based on multiple views of entities. One class of methods is to utilize traditional single-view clustering algorithms. For example, the straightforward methods incorporate attributes of multiple views into the classical clustering process directly [2]. In contrast, late integration method [5] derives results from each individual view and then fuses them base on consensus. [7] projects multi-view data into a common lower dimensional subspace and then applies k-means to get the partitions. Another category of multi-view clustering is relying on matrix factorization technique (MF). [13] proposed a non-negative matrix factorization (NMF) based approach that gives compatible clustering solutions across multiple views. To find out the key subset of features in each view that are associated with the clusters, [26] simultaneously

decomposes multiple data matrices into sparse row and columns vectors. However, all above methods could not identify anomalies or outliers hidden in multi-view datasets.

To tackle the problem of multi-view anomaly detection, several efforts have been made. Most multi-view anomalies studied by existing work can be classified into two types: the attribute anomaly and the class anomaly. The attribute anomaly refers to an outlier with abnormal behavior in each view. An outlier ranking based method [22] can be applied to evaluate the degree of outlierness for all instances. Iwata and Yamada propose a probabilistic latent variable model [16] under the assumption that all views of a normal instance are generated from a single latent vector. In contrast, the class anomaly refers to instances with inconsistent cluster memberships across different views. Gao et al. proposed the horizontal anomaly detection (HOAD) [12] via spectral clustering on multiple views simultaneously. A more effective method over HOAD is presented based on affinity propagation (AP) [1], and detects anomalies by comparing the neighborhoods in different views. Recently, dual-regularized multi-view outlier detection method (DMOD) [33] and multi-view low-Rank analysis (MLRA) method [19] are proposed respectively to discover both types of multi-view anomalies simultaneously. However, these methods can only deal with point anomalies, but with limitations on group anomalies.

Existing approaches to group anomaly detection can be classified into two general categories. One category investigates the problem following the paradigm of probabilistic framework. The Multinomial Genre Model (MGM) proposed in [31] treats each group as a mixture of topics, and first finds out anomalous groups with rare mixture rates compared to other normal groups. Then a more flexible model Flexible Genre Model (FGM) [30] is proposed to further generalize the generation of group mixture rates. Another category takes a discriminative approach such as one-class support measure machines (OCSMMs) [21]. It maps the distributions into a probability measure space with kernel methods, and recognizes group anomalies via a kernel embedding of the probabilistic distributions. However, these methods are single-view group anomaly detection, where the groups need to be pre-determined. The only work to address multi-view group anomaly detection is [32], which detects group anomaly in social media. It only deals with pair-wise data to form social community and requires a predetermined number of groups, which is not always available in practice. Our approach studies the generalized multi-view group anomaly detection problem, and to the best of our knowledge, the first approach that without any assumptions on the groups.

III. THE ENHANCED MGAD MODEL

3.1. Problem Definition. Motivated by document classification model like Latent Dirichlet Allocation (LDA) [3], we consider the groups as topics or roles, under the perspective of topic views with semantic meanings. Thus under the perspective of topic view, groups are formed with mixture of topics. We assume all groups in one view share the same set of topics in another view but possibly with different topic mixing rates. Normal groups follow the same pattern with respect to their topic mixing rates, but anomalous groups have topic mixing rates that deviate from the normal patterns [32]. Accordingly, we can formalize the problem of multi-view anomaly detection as follows:

Definition 1 (Multi-view Anomaly Detection [34]). Given a dataset with U group views and V topic views, Multi-view Anomaly Detection aims to discover groups from group views and topics from topic views, and to identify abnormal groups whose topic mixing rates do not conform to other groups.

Modelling multi-view data is a highly complicated task since it allows observed attributes under different views to be generated from diverse mechanisms, that is, to form different clusters in different views. To solve this problem, we start from a Bayesian approach to constructing a generative model for multi-view data. In our EMGAD model, consistent group views share the same group partitions and independently generate data features from their own distributions such as Gaussian distributions. On the other hand, given a fixed group partition, instance features under topic view are generated by different latent topics. We treat each group in topic view as a ‘bag-of-words’, and assume that the data instances in each group are exchangeable. According to the De Finetti theorem, the joint distribution of every infinitely exchangeable sequence of random variables can be represented with mixture models, thus, we apply LDA to model topic view data features. Combining those two aspects, the proposed EMGAD model can simultaneously build the relationships among groups under multiple views through the group partition.

To address the second challenge of determining the number of groups, we introduce the Chinese Restaurant Process (CRP) as a prior that can dynamically adjust the group partition.

$$p(C_i = c | C_{-i}, \alpha) \propto \begin{cases} \frac{n_c}{N-1+\alpha} & \text{if } c \leq J \\ \frac{\alpha}{N-1+\alpha} & \text{if } c = J+1 \end{cases}$$

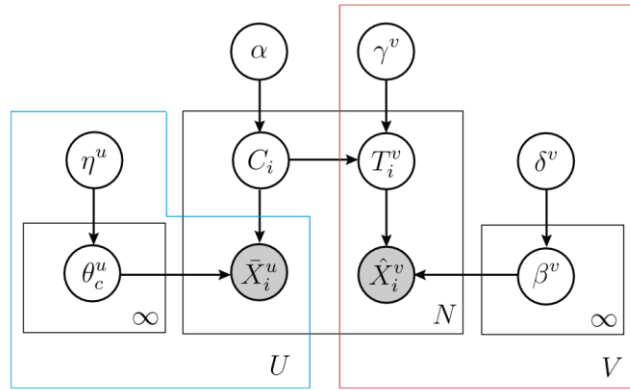


Figure 2. Graphical Model Representation of EMGAD.

MGAD[34] assume that the number of topics under topic view is predetermined. For scenarios where it is impractical to predetermine the number, we propose an enhanced MGAD model that relaxes this assumption and incorporate another CRP to adjust the number of topics automatically as well. Then we flag anomalies as unconformable to other groups by both models. We generalize the MGAD model into the enhanced MGAD (EMGAD), which supports learning not only the number of groups, but also the number of topics under the topic views. We first introduce the EMGAD model and then demonstrate the model inference algorithm.

3.2. Model Specification. In EMGAD, we relax the assumption that topic assignment T_i^v be sampled from a topic mixing rate variable π_{cv} with a fixed number of topics. Instead, we treat the topics in each group as a ‘bag-of-words’, and assume that the topics in each group are exchangeable. Thus we can also apply Chinese Restaurant Process as a prior on the space of latent variable T_i^v to specify a distribution over partitions (topic assignments). In other words, we directly model $T_i^v \sim \text{CRP}(\gamma^v)$, see Figure 2. The concentration parameter γ^v controls the probability to sample a new topic. Different topic view v may have different γ^v according to their context. The generative process of EMGAD is provided in Algorithm 1.

```

foreach instance  $i = 1, \dots, N$  do
  Sample a group  $c$  from  $C_i \sim \text{CRP}(\alpha)$ 
  foreach group view  $u = 1, \dots, U$  do
    Draw  $\theta_c^u$  from a conjugate prior  $\theta_c^u \sim P(\theta_c^u | \eta^u)$ 
    Draw  $\bar{X}_i^u \sim P(\bar{X}_i^u | \theta_c^u)$ 
  end
  foreach topic view  $v = 1, \dots, V$  do
    Draw a topic  $T_i^v \sim \text{CRP}(T_i^v | T_{-i}^v, C_{-i}, \gamma^v)$ 
    Draw  $\hat{X}_i^v \sim P(\hat{X}_i^v | \beta^v, T_i^v)$ 
  end
end

```

Algorithm 1: The Generative Process of EMGAD

3.3. Model Inference. The goal of model inference is to learn the posterior distribution of latent variables, which can further be extended for the anomaly detection. In the proposed MGAD model, The complete joint likelihood of data and latent variables can be written as follows:

$$\begin{aligned}
 p(\mathbf{C}, \mathbf{T}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\pi}, \bar{\mathbf{X}}, \hat{\mathbf{X}} | \Delta) &= \prod_i \text{CRP}(C_i | \alpha) \prod_u \mathcal{N}(\bar{X}_i^u | \theta_{c_i}^u) \text{GIW}(\theta_{c_i}^u | \eta^u) \\
 &\times \prod_v \text{Multinomial}(T_i^v | \pi_{c_i}^v) \text{Dirichlet}(\pi_{c_i}^v | \gamma^v) \mathcal{N}(\hat{X}_i^v | \beta_{T_i^v}^v)
 \end{aligned}$$

Given the likelihood, directly deriving the posterior distribution of latent variables is difficult, since the CRP process is non-parametric with infinite mixtures. Therefore, we adopt the collapsed Gibbs sampling [14], one Markov chain Monte Carlo (MCMC) sampling algorithm, for inference. Compared with the standard Gibbs sampling method, the collapsed Gibbs sampling is fast by alternatively integrating out some variables while sampling another variable. In each iteration, the inference procedure consists of four steps as follows.

Step 1: Sampling group assignment C .

For each instance i , according to the conditional independence principle, the posterior distribution of group assignment C_i can be given by marginalizing parameters of both θ and π as follows:

$$\begin{aligned}
 p(C_i = c | \sim) &\propto p(C_i = c | C_{-i}, \mathbf{T}, \bar{\mathbf{X}}; \Delta) \\
 &\propto p(C_i = c | C_{-i}; \alpha) p(\mathbf{T}, \bar{\mathbf{X}} | C_i = c, C_{-i}; \gamma, \eta) \\
 &\propto CRP(\alpha) p(\bar{\mathbf{X}} | C_i = c, C_{-i}; \eta) p(\mathbf{T} | C_i = c, C_{-i}; \gamma) \\
 &\propto CRP(\alpha) \prod_u p(\bar{X}_i^u | \bar{X}_{-i}^u, C_{-i}; \eta^u) \prod_v p(T_i^v | T_{-i}^v, C_{-i}; \gamma^v)
 \end{aligned}$$

Step 2: Sampling topic assignment \mathbf{T} .

We also perform a collapsed Gibbs Sampling method to infer latent variables. The inference procedure of EMGAD is similar to MGAD model in which four steps are needed. There are same forms for the inference of posterior distribution of group assignments \mathbf{C} and topic assignments \mathbf{T} . The difference is that the posterior of T_i^v do not follow a Dirichlet-Multinomial distribution, but a CRP prior.

$$T_i^v | T_{-i}^v, C_{-i}; \gamma^v \sim CRP(T_i^v | T_{-i}^v, C_{-i}, \gamma^v)$$

Step 3: Update parameters β by new \mathbf{T} .

We consider the features in view v as observations and the aim is to estimate parameter β^v for the topic-feature distribution. By assuming a Gaussian distribution for $\hat{\mathbf{X}}$ in view v , β^v can be estimated by maximizing the likelihood

$$p(\hat{\mathbf{X}}^v | \beta^v, \mathbf{T}^v) = \prod_i \mathcal{N}(\hat{X}_i^v | \beta_k^v, T_i^v = k)$$

Step 4: Update hyper-parameter in Δ .

With regard to the CRP concentration hyper-parameter α , the conditional posterior distribution only depends on J , the number of distinct groups. It can be re-sampled by Metropolis-Hastings methods. In this paper, we adopt a standard technique [9].

3.4. Group Anomaly Detection Criterion. In this section we present the group anomaly detection criterion, based on the learned MGAD model. After the process of model inference, latent variables including groups and topics, are allocated with specific values. According to Def.(1), group anomalies are those with rare topic mixing rates. Since there are multiple topic views, we examine every group in the perspective of all topic views and give them anomaly scores.

$$S_c = \sum w_v E_{\pi_c^v} [-\ln p(\pi^v; \Delta)]$$

The abnormal group will have a higher value. Since it is not straightforward to perform the integral calculation, we use the Importance Sampling, a Monte Carlo method using samples drawn from an approximate distribution. By alternatively computing anomaly score S_c^v in every topic view, we rank groups by their anomaly score S_c in the perspective of all topic views.

IV. PERFORMANCE EVALUATION

4.1. Baselines. To the best of our knowledge, few existing models can be directly adopted to multi-view group anomaly detection problem. Thus to evaluate the performance of our model, We design several two-stages approaches that: (i) identify groups on groups views; and (ii) discover topics on topic views and detection anomalous groups. We compare the EMGAD model with the following three approaches[34]: MGAD, KM-LDA and DP-LDA.

4.2. Synthetic Datasets. In this section, we compare the EMGAD model and other baselines on synthetic datasets. We firstly introduce the generating process of synthetic data, as well as the anomaly injection strategy. Secondly we report the performance results on these datasets.

4.2.1. Synthetic Data Preparation. We generate the synthetic datasets according to different type of views. Without loss of generality, we set the number of both group view and topic view to be one, i.e., $U = 1$ and $V = 1$. Let J be the ground truth number of groups, K be the number of topics under topic view and IAP be the Injected Anomaly Proportion. Suppose the features under group view and topic view are both two-dimensional vectors.

We first generate the normal instances. Each normal group in the group view is generated from a Gaussian distribution. We randomly choose a mean vector between interval $[0, 10] \times [0, 10]$, and the covariance are $\Sigma = aI_2$, where I_2 denotes the identity matrix and a ranges from 0.1 to 0.3. We sample instances for each group by different means and covariance and every group with at least 20 instances. We adopt the anomaly injection technique to derive the anomalous groups as in [30,32,34]. We define IAP as the Injection Anomaly Proportion, and then the number of anomalous groups is $J \times \text{IAP}$. We set a different topic mixing rate and generate features under the topic view for all $J \times \text{IAP}$ groups according to that rate. We evaluate the performance of different methods with three metrics, including Accuracy, False Positive Rate and AUC.

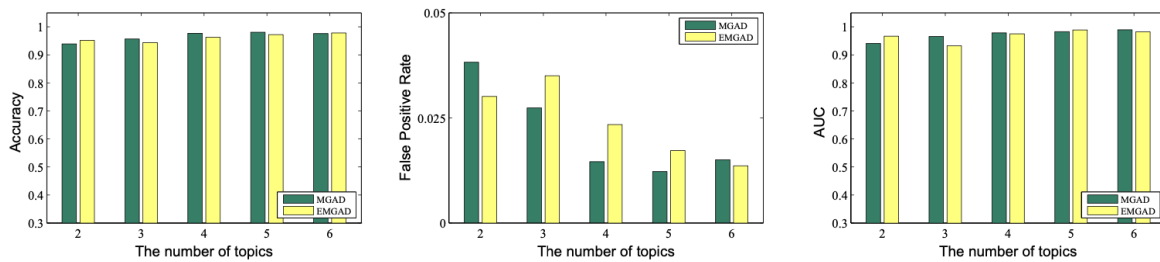


Figure 3. Performance of EMGAD vs. MGAD.

4.2.2 Performance of EMGAD. Finally in the experiments of synthetic data, we evaluate the effectiveness of our proposed EMGAD model. EMGAD model can be applied to scenarios when the number of topics under topic view is not explicitly known. We fix the number of groups J at 20, the injected anomaly proportion IAP at 0.2, and test its performance by varying the number of topics K from 2 to 6. We also generate 100 synthetic datasets for each K and demonstrate the results by average Accuracy, False Positive Rate and AUC in Figure.3. Note we only compare EMGAD model with our MGAD model since MGAD performs distinctly better than the baselines. We set initial value of concentration parameter $\gamma = 0.1$ for the CRP process under topic view. From Figure.3 we can see that the EMGAD model achieves a similar performance with MGAD, namely, their difference is very small and both outperform the baselines. The reason is that the mechanism of CRP process has the ability to find new topics and the rest of the model is similar with MGAD. Note that the false positive rate of EMGAD is slightly higher than MGAD($K = 3, 4, 5$), because once the number of topics are estimated inconsistent with the true number, some groups with influenced topic mixing rate may interfere the anomaly score calculation. We recommend EMGAD model when the number of topics of some topic views is not known before the anomaly detection process.

4.3. Real World Datasets. This section of experiments is to study the performance and utility of proposed models on real world datasets. We first implement tests anomaly detection performance on two public datasets by anomaly injection technique.

There are lots of real-world multi-view datasets that have two views: a correlation view and a content view, such as the scientific publications and web pages, to name two. Thus we study our models on two representative public datasets of such type. The first dataset is Cora, which consists of 2708 Machine Learning papers. Every paper is from one of seven topics, namely Case Based (CB), Genetic Algorithms (CA), Neural Networks (NN), Probabilistic Methods (PM), Reinforcement Learning (REL), Rule Learning (PUL) and Theory (TH). Each publication in Cora is described by a bag-of-word vector with 0/1-value indicating the absence/presence of the corresponding word from the dictionary. For the correlation view it consists of 5429 links indicating the publication citations. The second dataset is called WebKB, which is a hypertext dataset collected from four universities in the USA. For each university there are pages of five topics: Course, Student, Faculty, Project and Staff. There are totally 877 web pages including 1608 links in this corpus. Each page is also described by a bag-of-word vector with 0/1-value.

The goal of evaluation is to pick out the ‘anomalous’ groups from the prepared corpus. Then we apply different models to infer the group and topic assignments distributions, as well as computing the anomaly score for all detected groups. For every sub-dataset, we rank the groups with respect to their anomaly scores, and by their anomaly proportion, treat the top groups as the detected anomalies. Table 1 and Table 2 show the average accuracy and false positive rate of four different models respectively.

4.3.1. Anomaly Detection Performance.

Table 1. Group Anomaly Average Accuracy.

| Sub-dataset | MGAD | EMGAD | KM-LDA | DP-LDA |
|-----------------|-----------------------|-----------------------|---------------|---------------|
| Cora: CB/RUL | 0.6923 ±0.0032 | 0.6769±0.0044 | 0.6857±0.0131 | 0.6744±0.0072 |
| Cora: NN/REL | 0.7099 ±0.0057 | 0.6847±0.0040 | 0.6533±0.0081 | 0.6445±0.0039 |
| Cora: NN+PM/REL | 0.6684±0.0029 | 0.6935 ±0.0043 | 0.6500±0.0079 | 0.6505±0.0057 |
| Cora: GA+OM/TH | 0.6886 ±0.0053 | 0.6727±0.0048 | 0.6467±0.0074 | 0.6806±0.0106 |
| WebKB-1 | 0.6500±0.0493 | 0.6622 ±0.0520 | 0.5500±0.0233 | 0.5767±0.0549 |
| WebKB-2 | 0.6267 ±0.0448 | 0.6200±0.0389 | 0.5500±0.0233 | 0.5733±0.0489 |

Table 2. Group Anomaly False Positive Rate.

| Sub-dataset | MGAD | EMGAD | KM-LDA | DP-LDA |
|-----------------|-----------------------|-----------------------|---------------|---------------|
| Cora: CB/RUL | 0.1818 ±0.0013 | 0.2222±0.0018 | 0.1964±0.0051 | 0.2002±0.0028 |
| Cora: NN/REL | 0.1744 ±0.0021 | 0.1901±0.0016 | 0.2167±0.0032 | 0.2188±0.0015 |
| Cora: NN+PM/REL | 0.2003±0.0014 | 0.1839 ±0.0016 | 0.2188±0.0031 | 0.2157±0.0023 |
| Cora: GA+OM/TH | 0.1879 ±0.0021 | 0.1990±0.0020 | 0.2208±0.0029 | 0.1968±0.0041 |
| WebKB-1 | 0.2333±0.0219 | 0.2233 ±0.0231 | 0.3000±0.0103 | 0.2839±0.0242 |
| WebKB-2 | 0.2472 ±0.0198 | 0.2500±0.0172 | 0.3000±0.0103 | 0.2838±0.0214 |

From Table 1 and Table 2 we can see that the performance of all models degrades heavily compared to the results of synthetic datasets. This is because of the complex and high dimension of features in both group and topic views. The performance of WebKB is worse than Cora since there are less normal groups to differentiate with anomalous groups. But EMGAD is almost superior to other baselines models for different sub-datasets. For example, EMGAD promotes the average accuracy of all tested sub-datasets by approximately 5% and 4%, compared to KM-LDA and DP-LDA.

V. CONCLUSION

In this paper, we propose a non-parametric generative multi-view group anomaly detection model, named EMGAD. With this model, inconsistent groups or topics in different views are connected and learned simultaneously. To endow our model with robustness of group generation, we introduce the CRP prior into EMGAD, and propose an effective algorithm for model inference. Finally, we devise a novel anomaly detection criterion to the inference results based on cross entropy principle.

REFERENCES

- [1]. Alejandro Marcos Alvarez, Makoto Yamada, Akisato Kimura, and Tomoharu Iwata. Clustering-based anomaly detection in multi-view data. In 22nd ACM International Conference on Information and Knowledge Management, CIKM’13, pages 1545–1548, 2013.
- [2]. Steffen Bickel and Tobias Scheffer. Multi-view clustering. In Proceedings of the 4th IEEE International Conference on Data Mining (ICDM 2004), pages 19–26, 2004.
- [3]. David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.
- [4]. Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. LOF: identifying density-based local outliers. In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, pages 93–104, 2000.
- [5]. Eric Bruno and Stéphane Marchand-Maillet. Multiview clustering: a late fusion approach using latent models. In Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 736–737, 2009.
- [6]. Jie Cao, Zhiang Wu, Bo Mao, and Yanchun Zhang. Shilling attack detection utilizing semi-supervised learning method for collaborative recommender system. *World Wide Web*, 16(5-6):729–748, 2013.
- [7]. Kamalika Chaudhuri, Sham M. Kakade, Karen Livescu, and Karthik Sridharan. Multi-view clustering via canonical correlation analysis. In Proceedings of the 26th Annual International Conference on Machine Learning, ICML 2009, pages 129–136, 2009.
- [8]. Kaustav Das, Jeff Schneider, and Daniel B Neill. Detecting anomalous groups in categorical datasets. Carnegie Mellon University, School of Computer Science, Machine Learning Department, 2009.
- [9]. Michael D Escobar and Mike West. Bayesian density estimation and inference using mixtures. *Journal of the american statistical association*, 90(430):577–588, 1995.
- [10]. Shobeir Fakhraei, James Foulds, Madhusudana Shashanka, and Lise Getoor. Collective spammer detection in evolving multi-relational social networks. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 1769–1778, 2015.
- [11]. Amir Fayazi, Kyumin Lee, James Caverlee, and Anna Squicciarini. Uncovering crowdsourced manipulation of online reviews. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 233–242, 2015.
- [12]. Jing Gao, Wei Fan, Deepak Turaga, Srinivasan Parthasarathy, and Jiawei Han. A spectral framework for detecting inconsistency across multi-source object relationships. In 2011 IEEE 11th International Conference on Data Mining, pages 1050–1055, 2011.
- [13]. Jing Gao, Jiawei Han, Jialu Liu, and Chi Wang. Multi-view clustering via joint nonnegative matrix factorization. In Proceedings of the 13th SIAM International Conference on Data Mining, May 2–4, 2013. Austin, Texas, USA., pages 252–260, 2013.
- [14]. Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. Bayesian data analysis, volume 2. Chapman & Hall/CRC Boca Raton, FL, USA, 2014.
- [15]. Stephan Günnemann, Nikou Günnemann, and Christos Faloutsos. Detecting anomalies in dynamic rating data: A robust probabilistic model for rating evolution. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 841–850, 2014.
- [16]. Tomoharu Iwata and Makoto Yamada. Multi-view anomaly detection via robust probabilistic latent variable models. In Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, pages 1136–1144, 2016.
- [17]. Hans-Peter Kriegel, Arthur Zimek, et al. Angle-based outlier detection in high-dimensional data. In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 444–452, 2008.

- [18]. Rikard Laxhammar and Göran Falkman. Online learning and sequential anomaly detection in trajectories. *IEEE Trans. Pattern Anal. Mach. Intell.*, 36(6):1158–1173, 2014.
- [19]. Sheng Li, Ming Shao, and Yun Fu. Multi-view low-rank analysis for outlier detection. In *Proceedings of the 2015 SIAM International Conference on Data Mining*, pages 748–756, 2015.
- [20]. Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 6(1):3, 2012.
- [21]. Krikamol Muandet and Bernhard Schölkopf. One-class support measure machines for group anomaly detection. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, 2013.
- [22]. Emmanuel Müller, Ira Assent, Patricia Iglesias Sánchez, Yvonne Müller, and Klemens Böhm. Outlier ranking via subspace analysis in multiple views of the data. In *12th IEEE International Conference on Data Mining, ICDM 2012*, pages 529–538, 2012.
- [23]. Kevin P Murphy. Conjugate bayesian analysis of the gaussian distribution. *def*, 1(2σ2):16, 2007.
- [24]. Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7):1443–1471, 2001.
- [25]. Yale Song, Zhen Wen, Ching-Yung Lin, and Randall Davis. One-class conditional random fields for sequential anomaly detection. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pages 1685–1691, 2013.
- [26]. Jiangwen Sun, Jin Lu, Tingyang Xu, and Jinbo Bi. Multi-view sparse co-clustering via proximal alternating linearized minimization. In *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015*, pages 757–766, 2015.
- [27]. E Ted, Henry G Goldberg, Alex Memory, William T Young, Brad Rees, Robert Pierce, Daniel Huang, Matthew Reardon, David A Bader, Edmond Chow, et al. Detecting insider threats in a real corporate database of computer usage activity. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1393–1401, 2013.
- [28]. Hui Wen, Shiming Ge, Shuixian Chen, Hongtao Wang, and Limin Sun. Abnormal event detection via adaptive cascade dictionary learning. In *2015 IEEE International Conference on Image Processing*, pages 847–851, 2015.
- [29]. Ou Wu, Jun Gao, Weiming Hu, Bing Li, and Mingliang Zhu. Identifying multi-instance outliers. In *Proceedings of the SIAM International Conference on Data Mining, SDM 2010*, pages 430–441, 2010.
- [30]. Liang Xiong, Barnabás Póczos, and Jeff G. Schneider. Group anomaly detection using flexible genre models. In *25th Annual Conference on Neural Information Processing Systems*, pages 1071–1079, 2011.
- [31]. Liang Xiong, Barnabás Póczos, Jeff G. Schneider, Andrew J. Connolly, and Jake VanderPlas. Hierarchical probabilistic models for group anomaly detection. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2011*, pages 789–797, 2011.
- [32]. Qi Rose Yu, Xinran He, and Yan Liu. GLAD: group anomaly detection in social media analysis. In *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 372–381, 2014.
- [33]. Handong Zhao and Yun Fu. Dual-regularized multi-view outlier detection. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, pages 4077–4083, 2015.
- [34]. Hongtao Wang, Pan Su, Miao Zhao, Hongmei Wang and Gang Li. Multi-View Group Anomaly Detection . In *CIKM*, October 22-26, 2018, Torino, Italy.