Neighborhood Consensus Networks for Unsupervised Multi-view Outlier Detection

Li Cheng, Yijie Wang*, Xinwang Liu
College of Computer, National University of Defense Technology, Changsha, China
{chengli09, wangyijie, xinwangliu}@nudt.edu.cn

Abstract

Multi-view outlier detection recently attracted rapidly growing attention with the development of multi-view learning. Although promising performance demonstrated, we observe that identifying outliers in multi-view data is still a challenging task due to the complicated characteristics of multi-view data. Specifically, an effective multi-view outlier detection method should be able to handle (1) different types of outliers; (2) two or more views; (3) samples without clusters; (4) high dimensional data. Unfortunately, little is known about how these four issues can be handled simultaneously. In this paper, we propose an unsupervised multi-view outlier detection method to address these issues. Our method is based on the proposed novel neighborhood consensus networks termed NC-Nets, which automatically encodes intrinsic information into a comprehensive latent space for each view (for issue (4)) and uniforms the neighborhood structures among different views (for issue (2)). Accordingly, we propose an outlier score measurement which consists of two parts: the within-view reconstruction score and the cross-view neighborhood consensus score. The measurement is designed based on the characteristics of the different outlier types (for issue (1)) and no cluster assumption is needed (for issue (3)). Experimental results show that our method significantly outperforms state-of-the-art methods. On average, our method achieves $11.2\% \sim 96.2\%$ improvement in term of AUC and $33.5\% \sim 352.7\%$ improvement in term of F1-Score.

Introduction

Unsupervised outlier detection aims at identifying outliers in a given dataset without labels, which has been intensively studied and widely used in various applications, such as medical diagnosis (Wang et al. 2019), fraud detection (Wang et al. 2018), and information security (Kang et al. 2019; Wang and Ma 2014), to name just a few. In recent decades, a number of outlier detection methods have been proposed including distance-based methods, density-based methods, and clustering-based methods (Aggarwal 2015; Wang et al. 2013). These outlier detection algorithms are designed for data from one source, i.e., single-view data.

Nowadays, data are usually collected from diverse sources and features from a particular source are regarded as a particular view. Multi-view learning approaches can utilize the complementary information across various views and explore the consensus property to get better performance than their single-view counterparts (Xu, Wang, and Lai 2016; Tang et al. 2020; Sheng et al. 2019; Ji et al. 2019; Zhao et al. 2017). Although promising performance has been demonstrated in existing methods, detecting outliers from multi-view data is still a challenging problem due to the complicated distribution and inconsistent behavior of samples across different views. Specifically, based on the characteristics of multi-view data and existing literature, we argue that there are four important issues that need to be addressed:

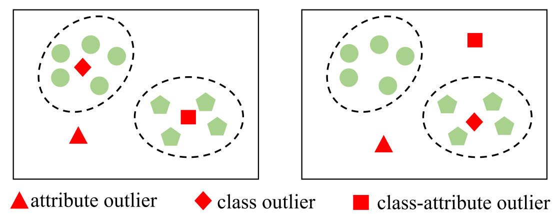

Figure 1: Illustration of three types of outliers: (i) attribute outlier (the red triangle): outlier that exhibits consistent abnormal behaviors in each view; (ii) class outlier(the red diamond): outlier that exhibits inconsistent characteristics (e.g., cluster membership) across different views; (iii) class-attribute outlier(the red square): outlier that exhibits consistent abnormal behaviors in some views, while exhibits inconsistent characteristics across some other views.

- **I1** handle different types of outliers: there are three types of outliers, i.e., attribute outlier, class outlier and class-attribute outlier, as illustrated in Figure 1;
- **I2** handle two or more views: there are usually more than two data sources, which indicates that the detection process could not be conducted in a pairwise manner and should be easily extended to multiple views;
- **I3** handle samples without clusters: there may exist no clusters in complex data, thus the commonly used cluster membership is no longer effective;
- **I4** handle high dimensional data: there are quite a number of features that could be extracted in practical applications. Such data may contain a few abundant features, leading to unsatisfying detection performance.

*Corresponding author

Method	I1	I2	I3	I4
HOAD(Gao et al. 2011)	×	×	×	×
CC(Liu and Lam 2012)	×	√	×	×
AP(Marcos Alvarez et al. 2013)	×	×	×	×
DMOD(Zhao and Fu 2015)	√	×	×	×
CRMOD(Zhao et al. 2017)	√	√	×	×
LDSR(Li and Li 2018)	√	√	×	×
MODDIS(Ji et al. 2019)	√	√	√	×
MUVAD(Sheng et al. 2019)	√	×	√	×
Our method	√	√	√	√

Table 1: The comparisons between existing methods and our proposed method on solving the issues. The value is √ if the method can address this issue and × otherwise.

A number of methods have been proposed for multi-view outlier detection (Gao et al. 2011; Liu and Lam 2012; Marcos Alvarez et al. 2013; Zhao and Fu 2015; Zhao et al. 2017; Li and Li 2018). Early methods usually identify outliers in a pairwise manner, leading to serious complications when faced with three or more views. Most existing methods rely on the clustering assumption: multiple views of inliers share consistent clustering structures while class outliers tend to fall into different clusters and attribute outliers consistently deviate from all clusters w.r.t. different views. Apparently, such methods failed to detect outliers when there are no clusters in data. Recently, a nearest neighbor-based outlier measurement criterion has been proposed in (Sheng et al. 2019) to handle data without clusters. It uses the naive distance calculation on the original features, which is risky due to the high-dimensionality and possible noise involved. The method in (Ji et al. 2019) is the first one to adopt neural networks into multi-view outlier detection. It tries to learn an intact space that preserves the original pairwise distances, which may be biased in high-dimensional data. Besides, its objective is not designed for outlier detection, leading to unsatisfying performance. In summary, to the best of our knowledge, there is still a lack of methods that can simultaneously address the aforementioned four issues (see Table 1), which makes the outlier detection problem on multi-view data more challenging. It motivates us to find out an integral solution to solve all four issues together.

In this paper, we propose a neighborhood consensus networks based multi-view outlier detection method, termed NCMOD. The key advantage of NCMOD lies in the joint consideration of within-view reconstruction and the cross-view neighborhood consensus. Specifically, the major contributions of this paper are summarized as follows:

- We propose a novel Neighborhood Consensus Networks framework for multi-view outlier detection, which can encode intrinsic information into a comprehensive latent space for each view and uniform neighborhood structures among different views. Learning an effective latent space enables NCMOD to handle high-dimensional data (issue **I4**) and such a learning strategy allows our model to conveniently handle two or more views (issue **I2**).
- We propose an outlier score measurement which consists of two parts: the within-view reconstruction score and the cross-view neighborhood consensus score. The measurement is designed according to the characteristics of the three outlier types (issue **I1**) and no cluster assumptions are needed. Consequently, NCMOD can handle data that have no clusters (issue **I3**).
- To the best of our knowledge, NCMOD is the first multi-view outlier detection method to adopt neural networks with direct outlier detection objective. Extensive experimental results verify the superiority of NCMOD.

Related Work

Outlier detection is an important research topic in machine learning and data mining (Aggarwal 2015; Wang and Li 2006). Various methods have been proposed for outlier detection (Aggarwal and Yu 2001; Knorr, Ng, and Tucakov 2000; Breunig et al. 2000; Schölkopf et al. 2001; Liu, Ting, and Zhou 2008; Keller, Muller, and Bohm 2012). However, most existing outlier detection methods are designed for single-view data.

The pioneering work on this topic is horizontal anomaly detection (HOAD) method (Gao et al. 2011). HOAD can be regarded as performing constrained spectral clustering in each view firstly and then finding instances that belong to different clusters in different views. Two similar clustering based methods were proposed in (Liu and Lam 2012; Marcos Alvarez et al. 2013), aiming at detecting class outliers by exploring the inconsistency of clustering results across multiple views. As a result, these methods are only designed to detect class outliers. Zhao and Fu proposed Dual-Regularized Multi-View Outlier Detection (DMOD) (Zhao and Fu 2015), which represents multi-view data with latent coefficients and sample-specific errors and characterize each type of outlier explicitly. The above models have a common disadvantage that they must identify outliers in a pairwise manner, leading to serious complication problem when faced with three or more views. To overcome the pairwise constraints, an enhanced version of DMOD is proposed, name CRMOD (Zhao et al. 2017), which uses a consensus cluster indicator rather than a dual one. Another low-rank subspace learning based method detects the outliers by learning a latent discriminant representation for each view data and defining a novel outlier score function based on the latent discriminant representations of all views. However, all the previous methods are based on cluster assumption and can not effectively handle data with no clusters. There is also another low-rank analysis based method MLRA (Li, Shao, and Fu 2015). We do not detail it since it is a supervised method that makes use of class labels.

Most recently, a nearest neighbor-based outlier measurement criterion has been proposed in (Sheng et al. 2019) to handle data without clusters. Since it uses the naive distance calculation on the original data, the performance may drop on high dimensional data because of the "curse of dimensionality". Our proposed NCMOD fills this gap by jointly considering mapping from heterogeneous views into a comprehensive latent space and uniforming neighborhood structures among different views. The method in (Ji et al. 2019) is the first one to adopt neural networks into multi-view outlier detection. But its objective is to integrate multi-view data

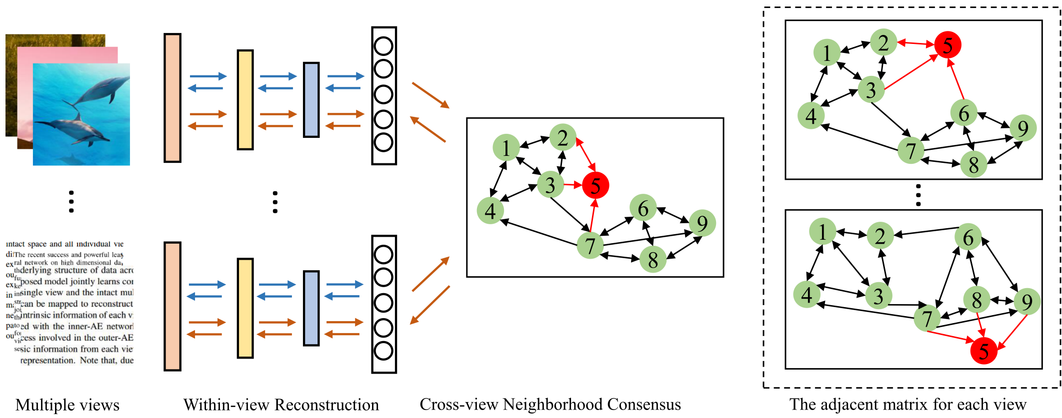

Figure 2: Overview of the proposed NC-Nets. The key components are the within-view reconstruction and the cross-view neighbor consensus. Within-view reconstruction automatically extracts features for each view with the intrinsic information preserved (the blue arrows), while cross-view neighborhood consensus tries to uniform the neighborhood structure of different views (the orange arrows). The joint consideration of reconstruction error and neighborhood distance in the latent space can characterize the multi-view outliers explicitly. For instance, we can easily identify the sample with id 5 as an outlier by comparing the consensus adjacent matrix and the adjacent matrix for each view.

into a latent intact space, which is not designed for outlier detection, leading to unsatisfying performance.

The Proposed Algorithm

In this section, we present the NCMOD for outlier detection with a set of multi-view samples $\mathcal{X} = \{\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^V\}$, where $\mathbf{X}^v \in \mathbb{R}^{d_v \times N}$ is the feature matrix of the v-th view. V, n and d_v are the number of views, number of samples and number of features for the v-th view, respectively. Specifically, $\mathbf{X}^v = \{\mathbf{x}_1^v, \mathbf{x}_2^v, \ldots, \mathbf{x}_N^v\}$ where $\mathbf{x}_i^v$ is the i-th sample of the v-th view.

The recent success and powerful learning capacity of neural networks on high dimensional data naturally motivate the use of neural networks. The key goal of NC-Nets (as presented in Figure 2) is to recover a latent space that can well reveal the neighborhood structure of data across multiple views. The proposed model jointly learns representation for each single view and the consistent neighborhood structure for all the views. Then, the intrinsic information of each view is automatically extracted with the within-view networks, while the neighborhood consensus involved between cross-view networks ensures the neighborhood structures of all views are similar. Based on the analysis of the three types of outliers, we adopt the combination of reconstruction error and neighborhood distance in the latent space as the outlier score measurement.

Neighborhood Consensus Networks

We adopt the autoencoder (AE) networks in the within-view part, the reasons of using AE networks are: (1) since no supervised information is provided to guide the learning process, we employ AE networks instead of general neural networks to ensure the intrinsic information to be preserved in the learned latent space; (2) with variants of AE (e.g., convolutional autoencoder for images), our model has the potential to be extended to multiple real world applications; (3) for existing multi-view outlier detection models, outlier scoring is usually directly based on the given features, which is risky due to the high-dimensionality and possible noise involved. The introduced encoding networks could extract intrinsic information to be encoded into the latent representation instead of the original high-dimensional/noisy features.

For simplicity, the AE network for the v-th view is denoted as $(f^v(\cdot); h^v(\cdot))$, where $f^v(\cdot)$, $h^v(\cdot)$ are the encoder and decoder, respectively. Let $\mathbf{Z} = f(\mathbf{X})$ be the latent space, $\hat{\mathbf{X}} = h(f(\mathbf{X}))$ be the recovered representations of AE. To preserve the intrinsic information in the low dimensional latent space, we should minimize the reconstruction loss:

$$\min_{\{f^v, h^v\}_{v=1}^V} \sum_{v=1}^{V} \sum_{i=1}^{N} ||\mathbf{x}_i^v - \hat{\mathbf{x}}_i^v||^2 \tag{1}$$

where $\hat{\mathbf{x}}_i^v$ is the recovered representation of $\mathbf{x}_i^v$.

Inspired by the works (Sheng et al. 2019; Keller, Muller, and Bohm 2012), we use k nearest neighbors to refer to the neighborhood structure, which indicates that an inlier have similar k nearest neighbors across different views while an outlier does not. We first define the adjacent matrix for the v-th view:

$$\mathbf{G}_{ij}^v = \left\{ \begin{array}{l} 1, \mathbf{z}_j^v \in knn(\mathbf{z}_i^v), \\ 0, otherwise. \end{array} \right. \tag{2}$$

where $knn(\mathbf{z})$ denotes the k nearest neighbors of $\mathbf{z}$.

Accordingly, the goal is to equalize $\mathbf{G}^v$ and $\mathbf{G}^u$ for each pair (u, v). To overcome the pairwise constraints, we change the aim to approximate all the adjacent matrices to a consensus adjacent matrix $\mathbf{G}^*$. In other words, the goal becomes finding a consensus adjacent matrix $\mathbf{G}^*$ which minimizes the neighbors' distances for each view:

$$\min_{\mathbf{G}^*} \sum_{v=1}^{V} \sum_{i,j=1}^{N} \mathbf{G}_{ij}^* ||\mathbf{z}_i^v - \mathbf{z}_j^v||^2 \tag{3}$$

Considering that the elements in the k nearest neighbors may have different affinities, which indicates that they should have different weights in the objective. We use the commonly used Gauss kernel based affinity matrix rather than adjacent matrix:

$$\mathbf{A}_{ij} = \begin{cases} \exp(-\frac{\sum_{v=1}^{V} ||\mathbf{z}_i^v - \mathbf{z}_j^v||^2}{2\delta^2}), \mathbf{G}_{ij}^* = 1, \\ 0, otherwise. \end{cases} \tag{4}$$

where δ is the bandwidth parameter and we set $\delta = 1.0$ by default for simplicity.

Then we can get the final objective of the proposed NC-Nets:

$$\begin{aligned} \min_{\mathbf{G}^*, f, h} L = \sum_{v=1}^{V} \sum_{i=1}^{N} (||\mathbf{x}_i^v - \hat{\mathbf{x}}_i^v||^2 + \alpha \sum_{j=1}^{N} \mathbf{A}_{ij} ||\mathbf{z}_i^v - \mathbf{z}_j^v||^2) \\ \mathbf{s.t.} \quad \mathbf{G}_{ij}^* \in \{0, 1\}, \sum_{j=1}^{N} \mathbf{G}_{ij}^* = k \\ \mathbf{A}_{ij} = \mathbf{G}_{ij}^* * \exp(-\frac{\sum_{v=1}^{V} ||\mathbf{z}_i^v - \mathbf{z}_j^v||^2}{2\delta^2}) \end{aligned} \tag{5}$$

where α is the trade-off parameter.

Outlier Scoring

On one hand, multi-view data may provide abundant information, especially on high-dimensional data which may contain a few abundant features; on the other hand, it should provide consistent information since they are describing the same samples from different perspectives. In particular, for the task of outlier detection, the consistency refers to similar neighborhood structures. Thus it is natural to assume that: if we can remove the abundant information by mapping the original view onto a low-dimensional latent space, the multiple views of an inlier have similar neighborhood structures. With the learned latent representations, we propose a novel outlier scoring strategy:

$$s(\mathbf{x}_i) = s_r(\mathbf{x}_i) + s_n(\mathbf{x}_i) \tag{6}$$

where

$$\begin{aligned} s_r(\mathbf{x}_i) = \sum_{v=1}^{V} ||\mathbf{x}_i^v - \hat{\mathbf{x}}_i^v||^2 \\ s_n(\mathbf{x}_i) = \sum_{v=1}^{V} \sum_{j=1}^{N} \mathbf{G}_{ij}^* ||\mathbf{z}_i^v - \mathbf{z}_j^v||^2 \end{aligned} \tag{7}$$

are respectively the reconstruction error and the sum of k nearest neighbors' distances. This strategy helps to identify all the three types of outliers simultaneously. The analysis is given as follows:

- For an inlier $\mathbf{x}_i$, its reconstruction errors on all the views should be small since inlier is the majority, resulting in a small value of $s_r(\mathbf{x}_i)$; since it is consistent across multiple views, its neighborhood structure in one view should be similar to the consensus one, which would give a small value of $s_n(\mathbf{x}_i)$. Thus the value of $s(\mathbf{x}_i)$ should be small.
- For an attribute outlier $\mathbf{x}_i$, since it is consistently anomalous and dissimilar to the majority of samples in each view, both its reconstruction error and neighborhood distances are large. Consequently, the values of both $s_r(\mathbf{x}_i)$ and $s_n(\mathbf{x}_i)$ are large, leading to a large value of $s(\mathbf{x}_i)$.
- For a class outlier $\mathbf{x}_i$, since it is inconsistent across multiple views, its neighborhood structure in one view could be dissimilar to it in the other view. Thus, it is difficult to find an optimal $\mathbf{G}^*$ for $\mathbf{x}_i$, leading to a large value of $s_n(\mathbf{x}_i)$. Then the value of $s(\mathbf{x}_i)$ should be large, too.
- For a class-attribute outlier $\mathbf{x}_i$, it is easy to derive that the values of both $s_r(\mathbf{x}_i)$ and $s_n(\mathbf{x}_i)$ are large since it contains the characteristics of both attribute outlier and class outlier. Thus the value of $s(\mathbf{x}_i)$ should be large, too.

The above strategy adopts the neighborhood structure of data rather than the cluster structures. Consequently, it can handle data that have no clusters. By taking both the within-view reconstruction and cross-view neighborhood consensus into account, it explicitly captures the characteristics of all the three types of multi-view outliers.

Algorithm 1 Optimization of NCMOD

Input: Multi-view input.
Output: AE Networks $(f^v(\cdot); h^v(\cdot))_{v=1}^V$,
Consensus Adjacent Matrix $\mathbf{G}^*$.

```
Initialize (f^v(·); h^v(·))_{v=1}^V;
repeat
    Update G* by optimizing Equation (8);
    Update A by Equation (4);
    for v = 1 → V do
        Update (f^v(·); h^v(·)) by Equation (9);
    end for
until convergence
return {(f^v(·); h^v(·))}_{v=1}^V, G*.
```

Optimization

The optimization of our proposed NCMOD is summarized in Algorithm 1. We first pre-train the deep AE without the neighborhood consensus part on all multi-view data because the initial value of $\mathbf{G}^*$ is unknown. We then use the pre-trained parameters to initialize the consensus adjacent matrix $\mathbf{G}^*$. There are multiple blocks of variables in our problem, and the objective function of our NCMOD is not jointly convex for all these variables. Therefore, we optimize our objective function by employing Alternating Direction Minimization (ADM) strategy (Lin, Liu, and Su 2011; Zhang, Liu, and Fu 2019). To adopt the ADM strategy, the optimization is cycled over the following three steps: updating the consensus adjacent matrix $\mathbf{G}^*$, updating the affinity matrix $\mathbf{A}$ and updating the AE networks $\{(f^v(\cdot); h^v(\cdot))\}_{v=1}^V$ by fixing the other blocks of variables. The optimization for each step is as follows:

Updating the consensus adjacent matrix $\mathbf{G}^*$. When f and h are fixed, we can obtain the latent representation $\mathbf{Z}$ and the recovered representation $\hat{\mathbf{X}}$, then the optimization of $\mathbf{G}^*$ can be simplified as:

$$\begin{aligned} \min_{\mathbf{G}^*} L = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{v=1}^{V} \mathbf{G}_{ij}^* ||\mathbf{z}_i^v - \mathbf{z}_i^v||^2 \\ \mathbf{s.t.} \quad G_{ij}^* \in \{0, 1\}, \sum_{j=1}^{N} = k \end{aligned} \tag{8}$$

the above problem is a k nearest neighbors searching problem, which can be efficiently solved by several algorithms such as kd-tree, ball-tree, and so on(Bhatia et al. 2010).

Updating the affinity matrix A. When $\mathbf{G}^*$, f and h are given, A can be easily recalculated by Equation (4).

Updating the networks $\{(f^v(\cdot); h^v(\cdot))\}_{v=1}^V$. When A is fixed, the gradients of NCMOD for the back propagation are:

$$\begin{aligned}
\frac{\partial L}{\partial h^v} &= \sum_{i=1}^{N} 2(h^v(f^v(\mathbf{x}_i^v)) - \mathbf{x}_i^v) \\
\frac{\partial L}{\partial f^v} &= \sum_{i=1}^{N} 2(h^v(f^v(\mathbf{x}_i^v)) - \mathbf{x}_i^v)\frac{\partial h^v}{\partial \mathbf{x}}(f^v(\mathbf{x}_i^v)) \\
&\quad + \alpha \sum_{i=1}^{N}\sum_{j=1}^{N} 2A_{ij}(f^v(\mathbf{x}_i^v) - f^v(\mathbf{x}_j^v))
\end{aligned} \tag{9}$$

Experiments

Setup

For experimental purposes, we follow the way in the previous work (Zhao et al. 2017; Sheng et al. 2019) to generate multi-view data with three types of outliers in given outlier ratios and make some reasonable changes to the generation of attribute outliers. Given a standard benchmark with multiple classes, we first randomly choose N samples from two or three classes(”inlier class”) as the original dataset. The remained classes are called “outlier class” here. Then the original features are split into a serial of feature subsets, each corresponding to a view. For instance, to generate two view data, we can cut a D-dimension sample and take features of the first $\lfloor \frac{D}{2} \rfloor$ dimensions as the first view and the other dimensions as the second view. The next step is to generate outliers. For attribute outliers, we randomly select samples and replace features of all D dimensions in all views with an object randomly sampled from “outlier classes”, which is different from existing works that use random values. This is because replacing with random values is too biased when compared to the original pattern and too easy to detect in our empirical tests. For class outliers, we randomly take some pairs of samples and swap feature vectors in $\lfloor \frac{V}{2} \rfloor$ views while keeping feature vectors in the other views unchanged. For class-attribute outliers, we also randomly choose some pairs of samples, swap feature vectors in $\lfloor \frac{V}{2} \rfloor$ views, and replacing features with values of randomly sampled objects from the “outlier classes” in the other views.

Three widely used high dimensional benchmarks are used: MNIST, REUTERS, TTC, and the dimensionality of them are respectively 784, 2000 and 7507. We generate various datasets with a fixed size of $N = 1000$ by different settings of outlier ratios and number of views. The datasets are formatted as ”BenchmarkAbbr+id”. For BenchmarkAbbr, ”M”, ”R” and ”T” denotes MNIST, REUTERS and TTC, respectively. And the ”id” refers to different ratios of attribute outlier(ρ_1), class outlier(ρ_2) and class-attribute outlier(ρ_3), which is shown in Table 2. For instance, ”M1” is generated from MNIST with 2% attribute outlier, 5% class outlier and 8% class-attribute outlier.

id	ρ_1	ρ_2	ρ_3
1	0.02	0.05	0.08
2	0.02	0.08	0.05
3	0.05	0.02	0.08
4	0.05	0.08	0.02
5	0.08	0.02	0.05
6	0.08	0.05	0.02

Table 2: The settings of outlier ratios for different id.

We compare our proposed NCMOD with seven state-of-the-art methods: OCSVM (Schölkopf et al. 2001), HOAD (Gao et al. 2011), DMOD (Zhao and Fu 2015), CRMOD (Zhao et al. 2017), MUVAD (Sheng et al. 2019), LDSR (Li and Li 2018) and MODDIS (Ji et al. 2019). Notably, OCSVM is a representative approach for single-view outlier detection and we include it to investigate the performance of single-view approach on multi-view data. As for OCSVM, multiple views are first concatenated into one single view and then used as input. MUVAD is complex and there is no available source code online. We only implement the naïve version which uses both inliers and outliers to calculate the outlier scores. However, our method also uses both inliers/outliers and the learned latent representations can also be fed into the MUVAD, thus it is still a meaningful comparison. The methods are all implemented in Python 3.4 executed at a PC in a 3.6GHz CPU with 16GB memory. All the source codes, including the proposed NCMOD, the competitors and datasets generation, are provided in the supplementary.

As for evaluation, we adopt the commonly-used Area under the Receiver Operating Characteristic Curve (AUC) as threshold-independent metric, and F1-Score (F1) as threshold-dependent metric (Cheng et al. 2020; Gupta et al. 2013; Wang et al. 2017). The threshold is set as the outlier ratio, i.e., 0.15. The higher the two metrics are, the better the approach performs.

Comparison Results

We compare AUC and F1 values of the NCMOD and its competitors as shown in Table 3 and 5. According to the experimental results, we have the following observations:

- In most cases, our proposed NCMOD performs better than other competitors. And NCMOD significantly outperforms other baselines on average.
- Some multi-view outlier detection methods, i.e. HOAD, DMOD and CRMOD, failed to work on high-dimensional data. The later methods, i.e., MUVAD, LDSR and MODDIS, have relatively good performance.
- In general, the detection performance of 3-view case is slightly worse than that of 2-view case.

The above observations are within expectation. For the 2-view case, NCMOD gets the best performance on 15 and 13 of 18 datasets in terms of AUC and F1 respectively, while the performance on the other ones is close to the best. NCMOD averagely performs better than seven competitors with $11.2\% \sim 96.2\%$ improvement in term of AUC. And in term of F1, the average improvements are $33.5\% \sim 349.3\%$. For the 3-view case, NCMOD gets the best performance

	OCSVM		HOAD		DMOD		CRMOD		MUVAD-s		LDSR		MODDIS		NCMOD	
Data	AUC	F1	AUC	F1	AUC	F1	AUC	F1	AUC	F1	AUC	F1	AUC	F1	AUC	F1
M1	0.572	0.173	0.503	0.133	0.267	0.033	0.311	0.033	0.934	0.707	**0.963**	**0.773**	0.617	0.247	0.929	0.707
M2	0.654	0.280	0.461	0.127	0.365	0.047	0.366	0.053	0.906	0.600	**0.935**	**0.693**	0.687	0.307	0.916	0.607
M3	0.731	0.327	0.471	0.133	0.169	0.000	0.233	0.000	0.920	0.633	0.927	0.620	0.746	0.293	**0.953**	**0.727**
M4	0.578	0.253	0.522	0.167	0.358	0.053	0.388	0.047	0.754	0.453	0.752	**0.533**	0.615	0.267	**0.768**	0.427
M5	0.633	0.247	0.491	0.141	0.269	0.020	0.303	0.020	0.823	0.533	0.824	0.527	0.706	0.340	**0.867**	**0.593**
M6	0.639	0.240	0.564	0.153	0.311	0.020	0.319	0.020	0.871	0.647	0.849	0.580	0.699	0.273	**0.909**	**0.660**
R1	0.595	0.280	0.495	0.154	0.371	0.087	0.380	0.080	0.779	0.380	0.432	0.187	0.624	0.273	**0.913**	**0.713**
R2	0.463	0.187	0.542	0.194	0.514	0.140	0.527	0.200	0.704	0.273	0.533	0.287	0.548	0.233	**0.920**	**0.687**
R3	0.649	0.347	0.544	0.175	0.324	0.113	0.335	0.113	0.736	0.407	0.420	0.187	0.602	0.287	**0.858**	**0.513**
R4	0.610	0.333	0.493	0.135	0.358	0.120	0.377	0.133	0.680	0.267	0.492	0.287	0.580	0.260	**0.890**	**0.613**
R5	0.436	0.120	0.504	0.165	0.527	0.167	0.526	0.173	0.823	0.400	0.446	0.167	0.703	0.320	**0.827**	**0.487**
R6	0.453	0.120	0.475	0.142	0.518	0.173	0.520	0.180	0.684	0.340	0.435	0.220	0.569	0.253	**0.761**	**0.507**
T1	0.456	0.120	0.599	0.160	0.612	0.227	0.613	0.220	0.480	0.120	**0.647**	**0.300**	0.456	0.113	0.602	0.193
T2	0.545	0.167	0.509	0.120	0.553	0.213	0.553	0.213	0.577	0.185	0.650	**0.307**	0.538	0.167	**0.662**	0.260
T3	0.559	0.167	0.533	0.307	0.314	0.078	0.314	0.079	0.598	0.160	0.409	0.240	0.568	0.180	**0.642**	**0.293**
T4	0.531	0.160	0.435	0.133	0.528	0.193	0.533	0.193	0.537	0.185	0.539	0.213	0.531	0.153	**0.655**	**0.293**
T5	0.517	0.173	0.509	0.080	0.603	0.267	0.603	0.267	0.550	0.173	0.601	0.340	0.512	0.173	**0.612**	**0.427**
T6	0.506	0.153	0.506	0.157	0.296	0.044	0.311	0.074	0.450	0.127	0.429	0.253	0.523	0.187	**0.560**	**0.253**
Avg.	0.563	0.214	0.509	0.154	0.403	0.111	0.417	0.117	0.711	0.366	0.627	0.373	0.601	0.240	**0.791**	**0.498**
$\Delta(\%)$	40.6	133.0	55.5	222.9	96.2	349.3	89.5	327.3	11.2	36.0	26.2	33.5	31.5	107.2	-	-

Table 3: Detection performance on on datasets with 2-view split. The best results are in bold. AVG is the average performance of a method over all datasets, and Δ is the improvement of NCMOD compared to the corresponding competitor.

on 13 and 16 of 18 datasets in terms of AUC and F1 respectively, while the performance on the other ones is close to the best. NCMOD averagely performs better than seven competitors with $13.3\% \sim 81.3\%$ improvement in term of AUC. And in term of F1, the average improvements are $49.1\% \sim 352.7\%$. These results are due to the reason that NCMOD effectively captures the characteristics of all the three types of outliers on high dimensional data, resulting in high-quality outlier rankings.

As single-view outlier detection methods can not capture the inconsistency across multiple views, OCSVM obtains poor performance on multi-view datasets. By making full use of the multi-view data representation in the latent space, our proposed NCMOD better characterizes all the three types of outliers across views. Different from HOAD, DMOD, CRMOD and LDSR that adopt inconsistent cluster membership to evaluate the outlierness, our proposed method abandons this clustering assumption. This enables NCMOD to successfully handle data without cluster structures, substantially reduce its detection errors and obtain significant detection improvement. Among these four methods, LDSR achieves relatively better performance on some datasets, e.g., M2, M3, T1. This is because LDSR actually utilizes multi-view subspace clustering technique, which could alleviate the problem caused by high dimensionality. MUVAD adopts a nearest neighbor-based outlier measurement criterion and makes no assumption on the clustering structures. Consequently, it can handle data that have no clusters and achieve relatively good performance. However, the naive distance calculation on the original features may becomes ineffective, especially on high dimensional data. Thus NCMOD performs much better than MUVAD. MODDIS is the first one to introduce neural networks into multi-view outlier detection. It integrates multi-view data into a latent intact space that preserves sufficient irregularities of samples. Unfortunately, the objective of MODDIS is not directly served for outlier detection, leading to unsatisfying performance. In our proposed NC-Nets, the objective is directly designed for outlier detection, which consequently ensures good performance of NCMOD.

	AUC			F1		
	M	R	T	M	R	T
AE	0.629	0.544	0.547	0.247	0.197	0.196
NC	0.692	0.491	0.533	0.341	0.285	0.157
NCMOD	**0.884**	**0.834**	**0.616**	**0.607**	**0.540**	**0.232**

Table 4: Ablation study. Due to space limit, we only report the average performance over the three benchmarks, i.e., MNIST(M), REUTERS(R) and T(TTC).

Furthermore, let us take a closer look at the Table 3 and 5. By averaging the performance of all outlier settings in 2-view case, we get the mean AUC and F1 as 0.791 and 0.498 for the proposed NCMOD. While these numbers are 0.764 and 0.433 for 3-view case. The reasons could be: (1) the more views make the data group structure more disordered than that in 2-view case, which further influence the accuracy of neighborhood consensus; (2) The ratio of abnormal features in 3-view case is lower than that in 2-view case. For instance, $1/3$ features are swapped for the class-outlier in 3-view case. And in 2-view case, the ratio is $1/2$, which makes the anomaly easier to be identified.

Ablation Study

To verify the effectiveness of within-view reconstruction part and cross-view neighborhood consensus part, we conduct the ablation study with respect to the proposed NCMOD. AE represents the proposed model without neighbor-

	OCSVM		HOAD		DMOD		CRMOD		MUVAD-s		LDSR		MODDIS		NCMOD	
Data	AUC	F1	AUC	F1	AUC	F1	AUC	F1	AUC	F1	AUC	F1	AUC	F1	AUC	F1
M1	0.584	0.213	0.493	0.173	0.208	0.007	0.380	0.027	0.851	0.487	0.890	0.593	0.647	0.233	**0.896**	**0.607**
M2	0.619	0.213	0.543	0.220	0.227	0.013	0.272	0.013	0.913	0.673	**0.931**	0.673	0.633	0.213	0.913	**0.687**
M3	0.756	0.380	0.500	0.327	0.280	0.013	0.236	0.027	0.862	0.533	**0.902**	**0.727**	0.779	0.433	0.878	0.587
M4	0.588	0.220	0.479	0.207	0.429	0.080	0.386	0.067	0.839	0.473	0.846	0.527	0.647	0.260	**0.881**	**0.593**
M5	0.637	0.260	0.445	0.133	0.374	0.053	0.370	0.047	0.776	0.420	**0.833**	0.513	0.687	0.313	0.822	**0.527**
M6	0.678	0.353	0.539	0.307	0.340	0.053	0.300	0.067	0.848	0.487	0.838	0.533	0.709	0.413	**0.871**	**0.567**
R1	0.570	0.293	0.506	0.120	0.402	0.153	0.425	0.140	0.653	0.260	0.374	0.107	0.561	0.287	**0.823**	**0.473**
R2	0.497	0.180	0.469	0.153	0.478	0.147	0.484	0.147	0.727	0.347	0.385	0.047	0.600	0.347	**0.888**	**0.607**
R3	0.432	0.173	0.551	0.147	0.586	0.240	0.554	0.180	0.616	0.313	0.313	0.093	0.474	0.213	**0.771**	**0.480**
R4	0.542	0.280	0.503	0.167	0.434	0.147	0.440	0.140	0.645	0.320	0.357	0.087	0.583	0.320	**0.810**	**0.433**
R5	0.450	0.153	0.503	0.153	0.550	0.180	0.544	0.167	0.658	0.347	0.370	0.060	0.569	0.293	**0.729**	**0.473**
R6	0.401	0.133	0.512	0.167	0.612	0.287	0.578	0.207	0.742	0.353	0.269	0.100	0.623	0.300	**0.822**	**0.493**
T1	0.565	0.193	0.525	0.167	0.478	0.140	0.477	0.127	0.580	0.200	**0.603**	0.213	0.570	0.180	0.585	**0.213**
T2	0.518	0.147	0.477	0.127	0.521	0.160	0.519	0.147	0.510	0.160	**0.631**	**0.327**	0.515	0.133	0.584	0.167
T3	0.580	0.147	0.509	0.180	0.514	0.180	0.517	0.167	0.600	0.147	0.453	0.120	0.568	0.147	**0.623**	**0.200**
T4	0.528	0.147	0.522	0.167	0.543	0.146	0.543	0.153	0.543	0.133	0.503	0.160	0.519	0.153	**0.590**	**0.207**
T5	0.609	0.193	0.518	0.220	0.482	0.120	0.474	0.093	0.652	0.233	0.367	0.080	0.601	0.193	**0.646**	**0.233**
T6	0.529	0.113	0.472	0.187	0.396	0.051	0.400	0.064	0.551	0.127	0.342	0.127	0.554	0.147	**0.628**	**0.240**
Avg.	0.560	0.211	0.504	0.185	0.436	0.121	0.439	0.110	0.698	0.334	0.567	0.283	0.602	0.254	**0.764**	**0.433**
Δ(%)	41.2	136.5	57.0	169.8	81.3	313.1	80.3	352.7	13.3	49.1	39.5	76.2	31.4	95.8	-	-

Table 5: Detection performance on on datasets with 3-view split. The best results are in bold. AVG is the average performance of a method over all datasets, and Δ is the improvement of NCMOD compared to the corresponding competitor.

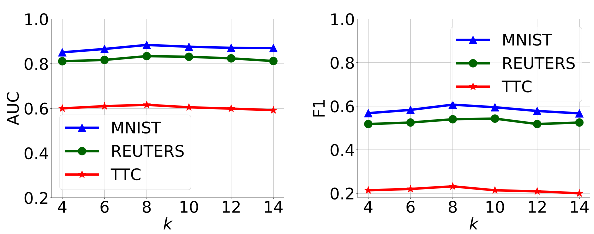

Figure 3: The performance curve w.r.t k.

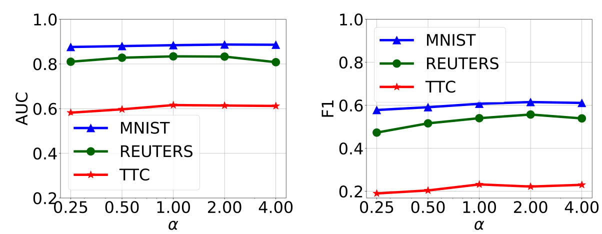

Figure 4: The performance curve w.r.t α.

hood consensus part while NC refers to the one without AE. As presented in Table 4, NCMOD substantially outperforms NC, which numerically indicates that we cannot ignore the intrinsic information. Besides, our proposed method consistently performs better than AE, which validates that the neighborhood consensus can successfully capture the characteristics of multi-view outlier. In sum, both within-view reconstruction part and cross-view neighborhood consensus part contribute to the enhancement of the proposed model in terms of multi-view outlier detection performance.

Parameters Sensitivity

To analyze the impacts of the parameters on the detection performance of NCMOD, We plot both the average AUC and F1 values over the three benchmarks with respect to nearest neighbors number k and trade-off parameter α. As shown in Figure 3 and 4, we present the parameter tuning with different values for k and α. It is observed that the promising performance could be expected when the values of k and α are within a wide range and our method is fairly robust with various values of k and α. We set $k = 8$ and $\alpha = 1.0$ as default.

Conclusion

We propose a novel neighborhood consensus networks based unsupervised multi-view outlier detection method, termed NCMOD. NCMOD automatically encode intrinsic information of each view into a comprehensive latent space with consensus neighborhood structures. Accordingly, we propose an outlier score measurement to characterize the three types of multi-view outliers. Extensive experimental results verify the superiority of the proposed NCMOD.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (Grant No.2016YFB1000101), the National Natural Science Foundation of China (Grant No.61379052), the Science Foundation of Ministry of Education of China (Grant No.2018A02002), the Natural Science Foundation for Distinguished Young Scholars of Hunan Province (Grant No.14JJ1026).

References

Aggarwal, C. C. 2015. Outlier analysis. In *Data mining*, 237–263. Springer.

Aggarwal, C. C.; and Yu, P. S. 2001. Outlier detection for high dimensional data. In *Proceedings of the 2001 ACM SIGMOD international conference on Management of data*, 37–46.

Bhatia, N.; et al. 2010. Survey of nearest neighbor techniques. *arXiv preprint arXiv:1007.0085* .

Breunig, M. M.; Kriegel, H.-P.; Ng, R. T.; and Sander, J. 2000. LOF: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 93–104.

Cheng, L.; Wang, Y.; Liu, X.; and Li, B. 2020. Outlier Detection Ensemble with Embedded Feature Selection. In *AAAI*, 3503–3512.

Gao, J.; Fan, W.; Turaga, D.; Parthasarathy, S.; and Han, J. 2011. A spectral framework for detecting inconsistency across multi-source object relationships. In *2011 IEEE 11th International Conference on Data Mining*, 1050–1055. IEEE.

Gupta, M.; Gao, J.; Aggarwal, C. C.; and Han, J. 2013. Outlier detection for temporal data: A survey. *IEEE Transactions on Knowledge and data Engineering* 26(9): 2250–2267.

Ji, Y.-X.; Huang, L.; He, H.-P.; Wang, C.-D.; Xie, G.; Shi, W.; and Lin, K.-Y. 2019. Multi-view outlier detection in deep intact space. In *2019 IEEE International Conference on Data Mining (ICDM)*, 1132–1137. IEEE.

Kang, Z.; Lu, Y.; Su, Y.; Li, C.; and Xu, Z. 2019. Similarity Learning via Kernel Preserving Embedding. *arXiv preprint arXiv:1903.04235* .

Keller, F.; Muller, E.; and Bohm, K. 2012. HiCS: high contrast subspaces for density-based outlier ranking. In *2012 IEEE 28th international conference on data engineering*, 1037–1048. IEEE.

Knorr, E. M.; Ng, R. T.; and Tucakov, V. 2000. Distance-based outliers: algorithms and applications. *The VLDB Journal* 8(3-4): 237–253.

Li, K.; and Li, S. 2018. Latent discriminant subspace representations for multi-view outlier detection. In *AAAI*.

Li, S.; Shao, M.; and Fu, Y. 2015. Multi-view low-rank analysis for outlier detection. In *Proceedings of the 2015 SIAM International Conference on Data Mining*, 748–756. SIAM.

Lin, Z.; Liu, R.; and Su, Z. 2011. Linearized alternating direction method with adaptive penalty for low-rank representation. In *Advances in neural information processing systems*, 612–620.

Liu, A. Y.; and Lam, D. N. 2012. Using consensus clustering for multi-view anomaly detection. In *2012 IEEE Symposium on Security and Privacy Workshops*, 117–124. IEEE.

Liu, F. T.; Ting, K. M.; and Zhou, Z.-H. 2008. Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, 413–422. IEEE.

Marcos Alvarez, A.; Yamada, M.; Kimura, A.; and Iwata, T. 2013. Clustering-based anomaly detection in multi-view data. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, 1545–1548.

Schölkopf, B.; Platt, J. C.; Shawe-Taylor, J.; Smola, A. J.; and Williamson, R. C. 2001. Estimating the support of a high-dimensional distribution. *Neural computation* 13(7): 1443–1471.

Sheng, X.-R.; Zhan, D.-C.; Lu, S.; and Jiang, Y. 2019. Multi-View Anomaly Detection: Neighborhood in Locality Matters. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 4894–4901.

Tang, C.; Liu, X.; Zhu, X.; Zhu, E.; Luo, Z.; Wang, L.; and Gao, W. 2020. CGD: Multi-View Clustering via Cross-View Graph Diffusion. In *AAAI*, 5924–5931.

Wang, C.; Deng, Z.; Lai, J.; and Philip, S. Y. 2018. Serendipitous recommendation in e-commerce using innovator-based collaborative filtering. *IEEE transactions on cybernetics* 49(7): 2678–2692.

Wang, H.; Yang, Y.; Liu, B.; and Fujita, H. 2019. A study of graph-based system for multi-view clustering. *Knowledge-Based Systems* 163: 1009–1019.

Wang, Y.; and Li, S. 2006. Research and performance evaluation of data replication technology in distributed storage systems. *Computers & Mathematics with Applications* 51(11): 1625–1632.

Wang, Y.; Li, X.; Li, X.; and Wang, Y. 2013. A survey of queries over uncertain data. *Knowledge and information systems* 37(3): 485–530.

Wang, Y.; and Ma, X. 2014. A general scalable and elastic content-based publish/subscribe service. *IEEE Transactions on Parallel and Distributed Systems* 26(8): 2100–2113.

Wang, Y.; Pei, X.; Ma, X.; and Xu, F. 2017. Ta-update: An adaptive update scheme with tree-structured transmission in erasure-coded storage systems. *IEEE Transactions on Parallel and Distributed Systems* 29(8): 1893–1906.

Xu, Y.; Wang, C.; and Lai, J. 2016. Weighted multi-view clustering with feature selection. *Pattern Recognition* 53: 25–35.

Zhang, C.; Liu, Y.; and Fu, H. 2019. Ae2-nets: Autoencoder in autoencoder networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2577–2585.

Zhao, H.; and Fu, Y. 2015. Dual-regularized multi-view outlier detection. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.

Zhao, H.; Liu, H.; Ding, Z.; and Fu, Y. 2017. Consensus regularized multi-view outlier detection. *IEEE Transactions on Image Processing* 27(1): 236–248.