Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)

Dual-Regularized Multi-View Outlier Detection*

Handong Zhao[1] **and Yun Fu**[1,2]
[1] Department of Electrical and Computer Engineering, Northeastern University, Boston, USA, 02115
[2] College of Computer and Information Science, Northeastern University, Boston, USA, 02115
{hdzhao,yunfu}@ece.neu.edu

Abstract

Multi-view outlier detection is a challenging problem due to the inconsistent behaviors and complicated distributions of samples across different views. The existing approaches are designed to identify the outlier exhibiting inconsistent characteristics across different views. However, due to the inevitable system errors caused by data-captured sensors or others, there always exists another type of outlier, which consistently behaves abnormally in individual view. Unfortunately, this kind of outlier is neglected by all the existing multi-view outlier detection methods, consequently their outlier detection performances are dramatically harmed. In this paper, we propose a novel Dual-regularized Multi-view Outlier Detection method (**DMOD**) to detect both kinds of anomalies simultaneously. By representing the multi-view data with latent coefficients and sample-specific errors, we characterize each kind of outlier explicitly. Moreover, an outlier measurement criterion is well-designed to quantify the inconsistency. To solve the proposed non-smooth model, a novel optimization algorithm is proposed in an iterative manner. We evaluate our method on five datasets with different outlier settings. The consistent superior results to other state-of-the-art methods demonstrate the effectiveness of our approach.

1 Introduction

Outlier detection (or anomaly detection) is a fundamental data analysis problem in machine learning and data mining fields. With its aim to identify the abnormal samples in the given sample set, it has a wide range of applications, such as image/video surveillance [Krausz and Herpers, 2010], network failure [Ding *et al.*, 2012], email and web spam [Castillo *et al.*, 2007], and many others [Breunig *et al.*, 2000; Angiulli *et al.*, 2003; Koufakou and Georgiopoulos, 2010; Hido *et al.*, 2008; Liu and Lam, 2012]. For more details, we highly recommend readers refer to the surveys [Chandola *et al.*, 2009; Akoglu *et al.*, 2014]. These methods usually analyze the distribution or density of a data set, and identify outliers via some well-defined criteria. However, they only work for single-view data like many other conventional machine learning methods. Nowadays, data are usually collected from diverse domains or obtained from various feature extractors, and each group of features is regarded as a particular view [Xu *et al.*, 2013]. Due to the complicated organization and distribution of data, outlier detection from multi-view data is very challenging.

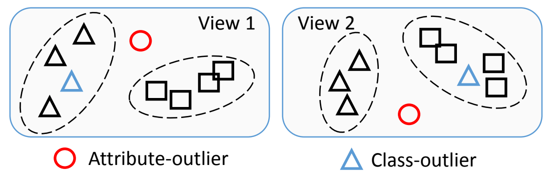


Figure 1: Illustration of attribute-/class- outliers. Both views derive from the same original objects. Here, red circle represents attribute-outlier, and blue triangle denotes class-outlier, respectively.

To date, there are a few methods designed to detect outliers for multi-view data. Das *et al.* [Das *et al.*, 2010] proposed a heterogeneous outlier detection method using multiple kernel learning. Janeja *et al.* developed a multi-domain anomaly detection method to find outliers from spatial datasets [Janeja and Palanisamy, 2013]. Muller *et al.* presented an outlier ranking algorithm for multi-view data by leveraging subspace analysis [Müller *et al.*, 2012]. The most related literatures to our proposed method are cluster-based multi-view outlier detection approaches [Gao *et al.*, 2011; 2013] and [Alvarez *et al.*, 2013].

Although a number of methods have been proposed in either single-view or multi-view category, they can only deal with certain patterns of outliers respectively. We claim that by representing the multi-view data with latent coefficients and sample-specific errors, our proposed model DMOD can identify all the outliers simultaneously. Before we make a further comparison, we first define two kinds of outliers as follows:

*This research is supported in part by the NSF CNS award 1314484, ONR award N00014-12-1-1028, ONR Young Investigator Award N00014-14-1-0484, and U.S. Army Research Office Young Investigator Award W911NF-14-1-0218.

Definition 1. ***Class-outlier*** *is an outlier that exhibits inconsistent characteristics (e.g., cluster membership) across different views, as the blue triangle shown in Figure 1.*

Definition 2. ***Attribute-outlier*** *is an outlier that exhibits consistent abnormal behaviors in each view, as the red circle shown in Figure 1.*

We argue that, on one hand, the existing multi-view methods [Gao *et al.*, 2011; 2013; Alvarez *et al.*, 2013] are only designed for class-outlier outliers. One the other hand, the existing single-view outlier detection methods can only handle attribute-outliers [Liu *et al.*, 2012; Xiong *et al.*, 2011]. However, by representing both kinds of outliers in two different spaces, i.e. latent space and original feature space, our approach can detect both class- and attribute- outliers jointly. In sum, the contributions of our method are summarized as:

- We propose a novel dual-regularized multi-view outlier detection method. To the best of our knowledge, this is the pioneer work to achieve identifying both class- and attribute- outliers simultaneously for multi-view data.
- Outliers are identified from the perspective of data representation, i.e. the coefficient in latent space and sample-specific error in original feature space with a novel cross-view outlier measurement criterion.
- The consistent superior results on five benchmark datasets demonstrate the effectiveness of our method. Specifically, in database *letter*, we raise the performance bar by around 26.13%.

2 Related Works

In this section, we only focus on the most relevant works: multi-view outlier detection methods.

Identifying the abnormal behaviours with multi-view data is a relatively new topic in the outlier detection field. Only a few methods are recently proposed to deal with it [Gao *et al.*, 2011; 2013; Alvarez *et al.*, 2013]. Gao *et al.* presented a multi-view anomaly detection algorithm named horizontal anomaly detection (HOAD) [Gao *et al.*, 2011; 2013]. Several different data sources are exploited to identify outliers from the dataset. The intuition of HOAD is to detect sample whose behavior is inconsistent among different sources, and treat it as anomaly. An ensemble similarity matrix is firstly constructed by the similarity matrices from multiple views, and then spectral embedding is computed for the samples. The anomalous score is computed based on the cosine distance between different embeddings. However, it is worth noticing that HOAD is only designed to deal with the class-outlier, i.e. identify inconsistent behaviors across different views.

Most recently, Alvarez *et al.* proposed a cluster-based multi-view anomaly detection algorithm [Alvarez *et al.*, 2013]. By measuring the differences between each sample and its neighborhoods in different views, the outlier is detected. Four kinds of strategies are provided for anomaly scores estimation. Specifically, for each view, clustering is firstly performed. Then cluster-based affinity vectors are calculated for each sample. Similar to [Gao *et al.*, 2011; 2013], this algorithm only detects the class-outliers, while our approach is able to detect both class- and attribute- outliers simultaneously by virtue of data representation.

3 DMOD: Dual-regularized Multi-view Outlier Detection Method

In this section, before proposing our method, we introduce the preliminary knowledge of clustering indicator based formulation first. Based on that, we propose our novel **D**ual-regularized **M**ulti-view **O**utlier **D**etection (DMOD) method.

3.1 Preliminary Knowledge

The effectiveness of k-means clustering method has been well demonstrated in data representation with its objective using the clustering indicators as:

$$\begin{aligned} &\min_{H,G} \|X - HG\|_{\mathrm{F}}^2, \\ &s.t.\ G_{kl} \in \{0,1\}, \sum_{k=1}^{K} G_{kl} = 1, \forall l = 1,2,\ldots,n \end{aligned} \tag{1}$$

where $X \in \mathbb{R}^{d\times n}$ is the input data with n samples and d dimensional features. Here, $H \in \mathbb{R}^{d\times K}$ is known as the cluster centroid matrix, and $G \in \mathbb{R}^{K\times n}$ is cluster assignment matrix in latent space. Note that the sum of each column of G should equal one, because each data sample x_l has to be assigned to one single cluster. Specifically, if x_l is assigned to k-th cluster, then $G_{kl} = 1$, otherwise $G_{kl} = 0$, which is known as *1-of-K* coding scheme. Although traditional k-means has a wide range of applications, it suffers from the vulnerability to outliers, especially for multi-view data [Bickel and Scheffer, 2004]. This derives one of our motivations to propose a robust outlier detection method for multi-view data.

3.2 The proposed DMOD

Inspired by the success of $\ell_{2,1}$-norm in feature selection [Nie *et al.*, 2010] and error modeling [Liu *et al.*, 2010], we propose a new dual-regularized multi-view outlier detection method for heterogeneous source data. We denote the sample set $X = \{X^{(1)}, \ldots, X^{(i)}, \ldots, X^{(V)}\}$, where V is the number of views and $X^{(i)} \in \mathbb{R}^{d_i\times n}$. $H^{(i)} \in \mathbb{R}^{d_i\times K}$ is the centroid matrix for i-th view. $G^{(i)} \in \mathbb{R}^{K\times n}$ is the clustering indicator matrix for i-th view.

Then our model is formulated as:

$$\begin{aligned} &\min_{H^{(i)},G^{(i)},S^{(i)}} \sum_{i}^{V} \|S^{(i)}\|_{2,1} + \beta \sum_{i}^{V}\sum_{i\neq j}^{V} \|G^{(i)} - M_{ij}G^{(j)}\|_{\mathrm{F}}^2, \\ &\text{s.t. } X^{(i)} = H^{(i)}G^{(i)} + S^{(i)}, \\ &\quad G_{kl} \in \{0,1\}, \sum_{k=1}^{K} G_{kl} = 1, \forall l = 1,2,\ldots,n \end{aligned} \tag{2}$$

where β is a trade-off parameter, M_{ij} denotes the alignment matrix between two different views, and $S^{(i)}$ is the construction error for i-th each view.

Remark 1: Due to the heterogeneous data $X^{(i)}$, different $G^{(i)}$ should be similar. Accordingly, a dual-regularization term $\|G^{(i)} - M_{ij}G^{(j)}\|_{\mathrm{F}}^2, (i \neq j)$ is employed so as to align the indicator matrixes $G^{(i)}$ and $G^{(j)}$ in two different views. Recall that G_i and G_j are orderless clusters, which means even though they are exactly the same, $\|G^{(i)} - G^{(j)}\|_{\mathrm{F}}^2$ cannot be zero without the alignment matrix M_{ij}.

Remark 2: The $\ell_{2,1}$-norm is defined as $\|S\|_{2,1} = \sum_{q=1}^{n}\sqrt{\sum_{p=1}^{d}|S_{pq}|^2}$, where $|S_{pq}|$ is the element of S in p-th row and q-th column. Note that $\ell_{2,1}$-norm has the power to ensure the matrix sparse in row, making it particularly suitable for sample-specific anomaly detection. This robust representation solves the outlier sensitivity problem in Eq. (1). Consequently, the regularization term $\|S^{(i)}\|_{2,1}$ is able to identify the attribute-outliers for i-th view.

Remark 3: In order to detect the class-outlier, i.e. sample inconsistent behaviour across different views, the inconsistency with respect to each view needs to be measured. We argue that the following representation $\sum_{k=1}^{n} G_{kl}^{(i)} G_{kl}^{(j)}$ in latent space can well quantify the inconsistency of sample l across different views i and j. The detailed illustration can be found in the following sub-section.

3.3 Outlier Measurement Criterion

We have discussed the ability of our method in characterizing two kinds of outliers. In order to make the quantitative estimation of inconsistency, we propose a novel outlier measurement function $\varphi(l)$ for sample l as

$$\varphi(l) = \sum_{i}^{V}\sum_{j\neq i}^{V}\left(\sum_{k=1}^{n} G_{kl}^{(i)} G_{kl}^{(j)} - \gamma\|S_l^{(i)}\|\|S_l^{(j)}\|\right), \tag{3}$$

where $\|\cdot\|$ denotes ℓ_2-norm and γ is a trade-off parameter.

The criterion Eq. (3) helps us identify attribute-/class- outliers jointly. Take two-view as an example, the first term measures the anomaly of the l-th sample across view 1 and view 2. When the l-th sample behaves normally in both views, the coefficients in $G^{(1)}$ and $G^{(2)}$ should be consistent. Consequently, $\sum_{k=1}^{n} G_{kl}^{(1)} G_{kl}^{(2)}$ should be relatively large. In contrast, if the l-th sample behaves inconsistently, the coefficients $G^{(1)}$ and $G^{(2)}$ would result in a small value of the first term, which means it is a class-outlier.

The second term $\gamma\|S_l^{(i)}\|\|S_l^{(j)}\|$ identifies the attribute-outliers. If the l-th sample behaves normally in at least one view, $\gamma\|S_l^{(1)}\|\|S_l^{(2)}\|$ is close to zero, which means the overall score $\varphi(l)$ will not decrease much by the second term. On the contrary, if the l-th sample is an attribute-outlier behaves abnormally in both views, the value of the second term increases, which leads to a decreased outlier score $\varphi(l)$.

4 Optimization

So far we have proposed the dual-regularized outlier detection model with a quantitative outlier measurement criterion. In this section, we illustrate the optimization solution to problem (2). Obviously, it is hard to find the global optimizers, since it is not jointly convex with respect to all the variables. Thus, we employ inexact augmented Lagrange method (ALM) [Lin *et al.*, 2009] to optimize each variable iteratively.

4.1 Algorithm Derivation

There are two difficulties to solve the proposed objective. First, $\ell_{2,1}$-norm is non-smooth. Second, each element of the indicator matrix $G^{(i)}$ is a binary integer, and each column vector has to satisfy *1-of-K* coding scheme.

By introducing the Lagrange multiplier $Y^{(i)}$ for each view, the augmented Lagrange function for problem (2) is written as:

$$\begin{aligned}\mathcal{L} &= \sum_{i}^{V}\big(\|S^{(i)}\|_{2,1} + \beta\sum_{i\neq j}^{V}\|G^{(i)} - M_{ij}G^{(j)}\|_{\mathrm{F}}^2 \\ &\quad + \langle Y^{(i)}, X^{(i)} - H^{(i)}G^{(i)} - S^{(i)}\rangle \\ &\quad + \frac{\mu}{2}\|X^{(i)} - H^{(i)}G^{(i)} - S^{(i)}\|_{\mathrm{F}}^2\big),\end{aligned} \tag{4}$$

where $\mu > 0$ is the penalty parameter, and $\langle\cdot\rangle$ denotes the inner product of two matrices, i.e. $\langle A, B\rangle = \mathrm{tr}(A^{\mathrm{T}}B)$. Then we optimize the variables independently in an iterative manner. Specifically, the variables $S^{(i)}$, $H^{(i)}$, $G^{(i)}$, and M_{ij} are updated as follows:

Update $\mathbf{S^{(i)}}$: Fix $H^{(i)}$, $G^{(i)}$, M_{ij}, the Lagrange function with respect to $S^{(i)}$ is written as:

$$\begin{aligned}&\|S^{(i)}\|_{2,1} + \langle Y^{(i)}, X^{(i)} - H^{(i)}G^{(i)} - S^{(i)}\rangle \\ &\quad + \frac{\mu}{2}\|X^{(i)} - H^{(i)}G^{(i)} - S^{(i)}\|_{\mathrm{F}}^2,\end{aligned} \tag{5}$$

which equalizes the following equation:

$$S^{(i)} = \arg\min_{S^{(i)}} \frac{1}{\mu}\|S^{(i)}\|_{2,1} + \frac{1}{2}\|S^{(i)} - \widehat{S}^{(i)}\|_{\mathrm{F}}^2. \tag{6}$$

Here, $\widehat{S}^{(i)} = X^{(i)} - H^{(i)}G^{(i)} + \frac{Y^{(i)}}{\mu}$. This term $S^{(i)}$ can be solved by the shrinkage operator [Yang *et al.*, 2009].

Update $\mathbf{H^{(i)}}$: Fix $S^{(i)}$, $G^{(i)}$, and M_{ij}, and take the derivative $\mathcal{L}$ with respect to $H^{(i)}$, we get

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial H^{(i)}} &= -Y^{(i)}{G^{(i)}}^{\mathrm{T}} + \\ &\quad \mu(-X^{(i)}{G^{(i)}}^{\mathrm{T}} H^{(i)}G^{(i)}{G^{(i)}}^{\mathrm{T}} + S^{(i)}{G^{(i)}}^{\mathrm{T}}).\end{aligned} \tag{7}$$

Setting Eq. (7) as zero, we can update $H^{(i)}$:

$$H^{(i)} = \frac{1}{\mu}\{Y^{(i)} + \mu(X^{(i)} - S^{(i)})\}{G^{(i)}}^{\dagger}, \tag{8}$$

where ${G^{(i)}}^{\dagger}$ denotes the pseudo inverse of $G^{(i)}$.

Update $\mathbf{G^{(i)}}$: Fix $S^{(i)}$, $H^{(i)}$, and M_{ij}, update the cluster indicator matrix $G^{(i)}$, we have

$$\begin{aligned}\mathcal{L} &= \sum_{i}^{V}\big(\beta\sum_{i\neq j}^{V}\|G^{(i)} - M_{ij}G^{(j)}\|_{\mathrm{F}}^2 + \langle Y^{(i)}, X^{(i)} \\ &\quad - H^{(i)}G^{(i)} - S^{(i)}\rangle + \frac{\mu}{2}\|X^{(i)} - H^{(i)}G^{(i)} - S^{(i)}\|_{\mathrm{F}}^2\big).\end{aligned} \tag{9}$$

As mentioned above, $G^{(i)}$ satisfies *1-of-K* coding scheme. We can solve the above problem by decoupling the data and determine each column $\mathbf{g}_m^{(i)} \in \mathbb{R}^{K\times 1}$ one by one, where m is the specified column index and $G^{(i)} = [\mathbf{g}_1^{(i)}, \ldots, \mathbf{g}_m^{(i)}, \ldots, \mathbf{g}_n^{(i)}]$. Thus for each $\mathbf{g}_m^{(i)}$, it satisfies the following equation:

$$\min_{\mathbf{g}_m^{(i)}} \sum_{i}^{V} \big(\beta \sum_{i \neq j}^{V} \|\mathbf{g}_m^{(i)} - M_{ij}\mathbf{g}_m^{(j)}\|_{\mathrm{F}}^2 + \langle \mathbf{y}_m^{(i)}, \mathbf{x}_m^{(i)} - H^{(i)}\mathbf{g}_m^{(i)} - \mathbf{s}_m^{(i)} \rangle + \frac{\mu}{2}\|\mathbf{x}_m^{(i)} - H^{(i)}\mathbf{g}_m^{(i)} - \mathbf{s}_m^{(i)}\|_{\mathrm{F}}^2\big),$$
$$s.t.\ \mathbf{g}_m^{(i)} \in \{0,1\},\ \sum_{m=1}^{K} \mathbf{g}_m^{(i)} = 1, \tag{10}$$

where $\mathbf{g}_m^{(j)}$, $\mathbf{y}_m^{(i)}$, $\mathbf{s}_m^{(i)}$ and $\mathbf{x}_m^{(i)}$ are the m-th column of matrix $G^{(j)}$, $Y^{(i)}$, $S^{(i)}$ and $X^{(i)}$, respectively.

To find the solution of Eq. (10), we do an exhaustive search in the feasible solution set, which is composed of all the columns of identity matrix $I_K = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_K]$.

Update $\mathbf{M_{ij}}$: Fix $S^{(i)}$, $H^{(i)}$ and $G^{(i)}$, update the alignment matrix M_{ij} between $G^{(i)}$ and $G^{(j)}$ as

$$M_{ij} = G^{(i)}{G^{(j)}}^{\dagger}. \tag{11}$$

Finally, the complete optimization algorithm to solve the problem in Eq. (2) is summarized in **Algorithm 1**. The initializations for each variable is also shown in the algorithm. The entire DMOD algorithm for multi-view outlier detection is outlined in **Algorithm 2**.

Algorithm 1. *Optimization Solution of Problem* (2)

Input: multi-view data $X = \{X^{(1)}, \dots, X^{(K)}\}$,
parameter β, the expected number of classes K.
Initialize: Set iteration time $t = 0$
$\mu_0 = 10^{-6}, \rho = 1.2, \mu_{\max} = 10^6$,
$\epsilon = 10^{-6}, S_0^{(i)} = 0, H_0^{(i)} = 0$,
$G_0^{(i)}$ using k-means algorithm.
while not converged **do**
1. Fix the others and update $S^{(i)}$ via Eq. (6).
2. Fix the others and update $H^{(i)}$ via Eq. (8).
3. Fix the others and update each vector $\mathbf{g}_m^{(i)}$ of $G(i)$ using Eq. (9) and (10).
4. Fix the others and update M_{ij} using Eq. (11).
5. Update the multiplier $Y^{(i)}$ via $Y^{(i)} = Y^{(i)} + \mu(X^{(i)} - H^{(i)}G^{(i)} - S^{(i)})$.
6. Update the parameter μ by $\mu = \min(\rho\mu, \mu_{\max})$.
7. Check the convergence condition by $\|X^{(i)} - H^{(i)}G^{(i)} - S^{(i)}\|_\infty < \epsilon$.
8. $t = t + 1$.

end while
Output: $S^{(i)}, H^{(i)}, G^{(i)}$

Algorithm 2. *DMOD for Multi-view Outlier Detection*

Input: Multi-view data X, parameter τ
1. Normalize data $x_i^{(v)}$ by $x_i^{(v)} = x_i^{(v)}/\|x_i^{(v)}\|$.
2. Solve problem (2) by **Algorithm 1**, and get the optimal $G^{(v)}$ and $S^{(v)}$.
3. Compute the outlier scores for all samples by Eq. (3).
4. Generate the binary outlier label L, if $\varphi(i) > \tau$, $L(i) = 0$; otherwise, $L(i) = 1$.

Output: Binary outlier label vector L.

4.2 Complexity Analysis

In this section, we make the time complexity analysis of our model. The most time-consuming parts of **Algorithm 1** are the matrix multiplication and pseudo inverse operations in Step 2, 3 and 4. For each view and each iteration, the pseudo inverse operations in Eq. (8) and Eq. (11) take $O(K^2n + K^3)$ in the worst case. Usually $K \ll n$, then the asymptotic upper-bound for pseudo inverse operation can be expressed as $O(K^2n)$. The multiplication operations take $O(dnK)$. Suppose L is the iteration time, V is the number of views. In general, the time complexity of our algorithm is $O(LVK^2n + LVKdn)$. It is worth noticing that L and V are usually much smaller than n. Thus we claim that our proposed method is linear time complexity with respect to the number of samples n.

5 Experiments

In this section, we collect five benchmark datasets to evaluate the performance. Among them, four are from UCI Machine Learning Repository[1], i.e. *iris*, *breast*, *ionosphere*, and *letter*. The fifth one *VisNir* is from BUAA database [Di Huang and Wang, 2012]. Important statistics are tabulated in Table 1. To generate both types of outliers, we do data pre-processing as follows: for class-outlier, we follow the strategy in [Gao *et al.*, 2011]: (a) split the object feature representation into two subsets, where each subset is considered as one view of the data; (b) take two objects from two different classes and swap the subsets in one view but not in the other. In order to generate attribute-outlier, we randomly select a sample, and replace its features in all views by random values.

Table 1: Databases Statistics

	iris	breast	ionosphere	letter	VisNir
# class	3	2	2	26	150
# sample	150	569	351	20000	1350
# feature	4	32	34	16	200

We compare the proposed method with both the single-view and multi-view outlier detection baselines as follows:

- Direct Robust Matrix Factorization (DRMF) [Xiong *et al.*, 2011] is a single-view outlier detection method which has demonstrated its superiority to several other single-view baselines, i.e. robust PCA [Candès *et al.*, 2011], Stable Principal Component Pursuit [Zhou *et al.*, 2010], and Outlier Pursuit [Xu *et al.*, 2010].
- Low-Rank Representation (LRR) [Liu *et al.*, 2012] is a representative outlier detection method for single-view data. There is a trade-off parameter balancing the low-rank term and error term, which we fine-tune in the range of [0.01, 1] and report the best result.
- HOrizontal Anomaly Detection (HOAD) [Gao *et al.*, 2013] is a cluster-based outlier detection method identifying the inconsistency among multiple sources. Two parameters, i.e. edge-weight m and the number of classes k are fine-tuned to get the best performance.

[1] http://archive.ics.uci.edu/ml/

Table 2: AUC values (mean ± standard deviation) on four UCI datasets with different settings. The setting is formatted as "DatasetName – Class-outlier Ratio (%) – Attribute-outlier Ratio (%)".

Datasets	DRMF [Xiong *et al.*, 2011]	LRR [Liu *et al.*, 2012]	HOAD [Gao *et al.*, 2013]	AP [Alvarez *et al.*, 2013]	DMOD (Ours)
iris-2-8	0.749 ± 0.044	0.779 ± 0.062	0.167 ± 0.057	0.326 ± 0.027	$\mathbf{0.868 \pm 0.036}$
iris-5-5	0.714 ± 0.038	0.762 ± 0.107	0.309 ± 0.062	0.630 ± 0.021	$\mathbf{0.865 \pm 0.047}$
iris-8-2	0.651 ± 0.037	0.740 ± 0.100	0.430 ± 0.055	0.840 ± 0.021	$\mathbf{0.882 \pm 0.043}$
breast-2-8	0.764 ± 0.013	0.586 ± 0.037	0.555 ± 0.072	0.293 ± 0.012	$\mathbf{0.816 \pm 0.038}$
breast-5-5	0.708 ± 0.034	0.493 ± 0.017	0.586 ± 0.061	0.532 ± 0.024	$\mathbf{0.809 \pm 0.020}$
breast-8-2	0.648 ± 0.024	0.508 ± 0.043	0.634 ± 0.046	0.693 ± 0.023	$\mathbf{0.778 \pm 0.019}$
ionosphere-2-8	0.705 ± 0.029	0.699 ± 0.025	0.446 ± 0.074	0.623 ± 0.033	$\mathbf{0.810 \pm 0.044}$
ionosphere-5-5	0.676 ± 0.040	0.627 ± 0.029	0.422 ± 0.051	0.761 ± 0.025	$\mathbf{0.773 \pm 0.041}$
ionosphere-8-2	0.634 ± 0.023	0.511 ± 0.014	0.448 ± 0.041	0.822 ± 0.030	$\mathbf{0.824 \pm 0.029}$
letter-2-8	0.315 ± 0.030	0.503 ± 0.011	0.536 ± 0.046	0.372 ± 0.057	$\mathbf{0.687 \pm 0.041}$
letter-5-5	0.375 ± 0.023	0.499 ± 0.012	0.663 ± 0.057	0.550 ± 0.043	$\mathbf{0.691 \pm 0.037}$
letter-8-2	0.490 ± 0.062	0.499 ± 0.016	0.569 ± 0.049	0.621 ± 0.051	$\mathbf{0.852 \pm 0.037}$

- Anomaly detection using Affinity Propagation (AP) [Alvarez *et al.*, 2013]. AP is the most recent outlier detection approach for multi-view data. Two affinity matrices and four anomaly measurement strategies are presented. In this paper, ℓ-2 distance and Hilbert-Schmidt Independence Criterion (HSIC) are used, since this combination usually performs better than others.

As suggested in [Alvarez *et al.*, 2013; Liu *et al.*, 2012], AUC is reported (area under Receiver Operating Characteristic (ROC) curve) as the evaluation metric. We also adopt ROC curve, representing the trade-off between hit rate and false alarm rate. The hit rate (TPR) and false alarm rate (FPR) are defined as:

$$TPR = \frac{TP}{TP + FN}, \quad FPR = \frac{FP}{TP + TN}, \tag{12}$$

where TP, FN, TN and FP represent true positives, false negatives, true negatives, and false positives, respectively.

5.1 UCI Databases

For each dataset, we strictly follow [Alvarez *et al.*, 2013; Gao *et al.*, 2011; 2013]. The outliers are firstly generated randomly for 50 times. Then the performance of each method is evaluated on those 50 sets. Finally the average results are reported. To simulate the real-world applications happening in different circumstances, we conduct three settings by mixing both outliers with different ratios: (1) 2% class-outlier of the total sample number + 8% attribute-outlier of the total sample number, represented in format "DatasetName-2-8"; (2) 5% class-outlier + 5% attribute-outlier in format "DatasetName-5-5"; (3) 8% class-outlier + 2% attribute-outlier in format "DatasetName-8-2".

Table 2 reports the AUC values (mean ± standard deviations) on four datasets with different outlier settings. Based on Table 2, we have the following observations and discussions.

- Our proposed method DMOD consistently outperforms all the other baselines in all settings.

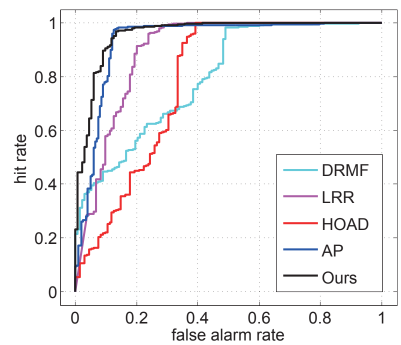


Figure 2: ROC Curves of all the methods on BUAA VisNir database with both outlier levels of 5%.

- In most cases, single-view based methods have superior performance to multi-view based methods in outlier setting "DatasetName-2-8".
- In the experiments with setting "DatasetName-8-2", multi-view based methods perform better than single-view methods in most cases.

Discussion: All the observations are expected, as multi-view based methods are strong in dealing with the class-outliers, while single-view based methods are designed to identify the sample inconsistency in all views. By virtue of the dual-regularized multi-view data representation, both class- and attribute- outliers are well characterized within the proposed DMOD model. Thus a stable and encouraging performance on UCI datasets has been observed. Specifically, in letter dataset, we raise the performance by around 26.13%.

5.2 BUAA-VisNir Database

We make another evaluation on the BUAA VisNir database, which consists of two types of data captured from visual spectral (VIS) and near infrared (NIR) sensors. There are 150 subjects, with the original image size of 287×287 pixels. In order to fasten the computation and keep the key features, we

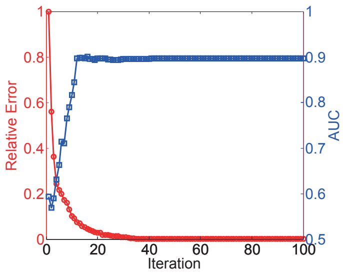


Figure 3: Convergence (red line) and AUC (blue line) curves with respect to iteration time on iris database with parameters β and γ setting as 0.5 and 0.1, respectively.

Table 3: AUC values (mean $\pm$ standard deviation) on BUAA VisNir database with both outlier levels of 5%.

Methods	AUC (mean $\pm$ std)
DRMF [Xiong *et al.*, 2011]	0.7878 ± 0.0112
LRR [Liu *et al.*, 2012]	0.8702 ± 0.0484
HOAD [Gao *et al.*, 2013]	0.7821 ± 0.0182
AP [Alvarez *et al.*, 2013]	0.9041 ± 0.0220
DMOD (Ours)	$\mathbf{0.9296 \pm 0.0147}$

vectorize the images, and project the data matrix into 100-dimension for each view by PCA. It is worth noticing that this pre-processing also helps remove the noise.

To generate 5% class-outliers and 5% attribute-outliers, the same strategies are employed as UCI datasets. Figure 2 and Table 3 show the ROC curves and the corresponding AUC values (mean $\pm$ standard deviation). It is observed that our approach also outperforms all other single-view and multi-view outlier detection algorithms.

5.3 Convergence and Parameter Analysis

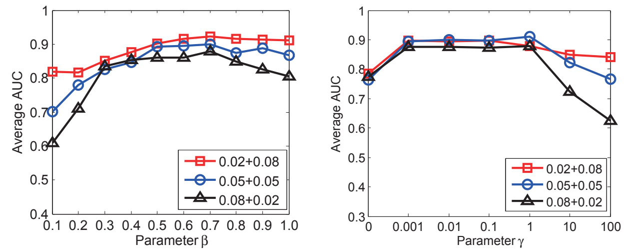
To testify the robustness and stability, we conduct four experiments to study the detection performance in terms of convergence and model parameters. Without explicit specification, all the experiments are conducted on iris dataset with the setting of 5% class-outliers and 5% attribute-outliers. Three parameters β, γ and K are set to 0.5, 0.1 and 3, respectively.

Convergence analysis. To show the convergence property, we compute the relative error of stop criterion $\|X^{(v)} - H^{(v)}G^{(v)} - S^{(v)}\|_\infty$ in each iteration, the convergence curve of our model is drawn in red as shown in Figure 3. It is observed that the relative error drops steadily, and then meets the convergency at around #30 iteration. We also plot the average AUC during each iteration. From the observation, there are three stages before converging: the first stage (from #1 to #15), the AUC goes up steadily; second stage (from #16 to #30), the AUC bumps in a small range; the final stage (from #31 to the end), the AUC achieves the best at the convergence point. Note that our method might converge to the local minimum as k-means does, we employ the strategy to run each set of data 10 times to find the best optimizer.

Parameter analysis. There are three major parameters in our approach, i.e. β, γ and K. Figure 4(a) shows the experiment of outlier detection accuracy with respect to the parameter β under different outlier settings. We set the parameter β in the range of [0.1, 1.0] with the step of 0.1. It is observed that our method reaches the best when β equals 0.7 under different outlier settings, and generally the performance is quite stable in the range of [0.4, 0.9]. Thus in our experiments, we set parameter $\beta = 0.5$ as default.

The experiment shown in Figure 4(b) is designed to testify the robustness of our model in terms of parameter γ. Due to possible amplitude variations of two terms in Eq. 3, we evaluate γ within the following set $\{0, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}\}$. As we observe, the average AUCs in three different settings are relatively steady when $\gamma = \{10^{-3}, 10^{-2}, 10^{-1}, 10^{0}\}$. In practical, we choose $\gamma = 0.1$ as default for all experiments.

Another important parameter in our proposed model is the intrinsic dimension K in matrix factorization step. The intrinsic dimension K of iris dataset is 3 since it has three classes. Therefore we make the evaluation in the range of [1, 8]. The boxplot of average AUC is shown in Figure 5. It is easily observed that our proposed method works well when K is around the true intrinsic dimension, i.e. K is in the range of [2, 5]. However, when K is too small, the performance drops dramatically due to the information lost in the matrix decomposition step. When K is too large, i.e. $K > 5$, the AUC also drops because of introducing more noisy redundant information. Note that, same with [Gao *et al.*, 2011], K is essential and varies depending on data. While we argue that instead of manually searching the best K for each dataset, we can utilize several off-the-shelf methods to predict K [Tibshirani *et al.*, 2000]. A steady and robust performance has been verified



(a) (b)

Figure 4: AUC curves with respect to parameters (a) β and (b) γ. Both experiments are conducted on iris database with three settings (*class-outlier level*+*attribute-outlier level*).

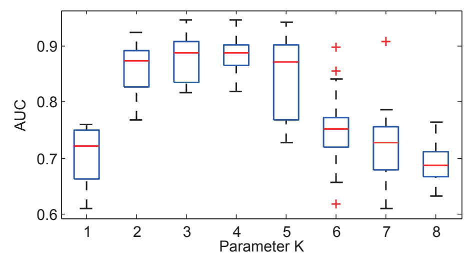


Figure 5: Analysis on matrix decomposition dimension parameter K on iris database with both outliers level of 5%.

as long as the predicted K is not far from the ground-truth.

6 Conclusion

In this paper, we proposed a novel dual-regularized multi-view outlier detection method from the perspective of data representation, named DMOD. An outlier estimation criterion was also presented to measure the inconsistency of each data sample. We introduced an optimization algorithm to effectively solve the proposed objective based on *1-of-K* coding scheme. Extensive experiments on four UCI datasets and one BUAA VisNir database with various outlier settings were conducted. The consistently superior results to four state-of-the-arts demonstrated the effectiveness of our method.

References

[Akoglu *et al.*, 2014] Leman Akoglu, Hanghang Tong, and Danai Koutra. Graph-based anomaly detection and description: A survey. *CoRR*, abs/1404.4679, 2014.

[Alvarez *et al.*, 2013] Alejandro Marcos Alvarez, Makoto Yamada, Akisato Kimura, and Tomoharu Iwata. Clustering-based anomaly detection in multi-view data. In *CIKM*, pages 1545–1548, 2013.

[Angiulli *et al.*, 2003] Fabrizio Angiulli, Rachel Ben-Eliyahu-Zohary, and Luigi Palopoli. Outlier detection using default logic. In *IJCAI*, pages 833–838, 2003.

[Bickel and Scheffer, 2004] Steffen Bickel and Tobias Scheffer. Multi-view clustering. In *ICDM*, pages 19–26, 2004.

[Breunig *et al.*, 2000] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jrg Sander. Lof: Identifying density-based local outliers. In *SIGMOD Conference*, pages 93–104, 2000.

[Candès *et al.*, 2011] Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *J. ACM*, 58(3):11, 2011.

[Castillo *et al.*, 2007] Carlos Castillo, Debora Donato, Aristides Gionis, Vanessa Murdock, and Fabrizio Silvestri. Know your neighbors: web spam detection using the web topology. In *SIGIR*, pages 423–430, 2007.

[Chandola *et al.*, 2009] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Comput. Surv.*, 41(3), 2009.

[Das *et al.*, 2010] Santanu Das, Bryan L. Matthews, Ashok N. Srivastava, and Nikunj C. Oza. Multiple kernel learning for heterogeneous anomaly detection: algorithm and aviation safety case study. In *SIGKDD*, pages 47–56, 2010.

[Di Huang and Wang, 2012] Jia Sun Di Huang and Yunhong Wang. The buaa-visnir face database instructions. *IRIP-TR-12-FR-001*, 2012.

[Ding *et al.*, 2012] Qi Ding, Natallia Katenka, Paul Barford, Eric D. Kolaczyk, and Mark Crovella. Intrusion as (anti)social communication: characterization and detection. In *KDD*, pages 886–894, 2012.

[Gao *et al.*, 2011] Jing Gao, Wei Fan, Deepak S. Turaga, Srinivasan Parthasarathy, and Jiawei Han. A spectral framework for detecting inconsistency across multi-source object relationships. In *ICDM*, pages 1050–1055, 2011.

[Gao *et al.*, 2013] Jing Gao, Nan Du, Wei Fan, Deepak Turaga, Srinivasan Parthasarathy, and Jiawei Han. A multi-graph spectral framework for mining multi-source anomalies. pages 205–228, 2013.

[Hido *et al.*, 2008] Shohei Hido, Yuta Tsuboi, Hisashi Kashima, Masashi Sugiyama, and Takafumi Kanamori. Inlier-based outlier detection via direct density ratio estimation. In *ICDM*, pages 223–232, 2008.

[Janeja and Palanisamy, 2013] Vandana Pursnani Janeja and Revathi Palanisamy. Multi-domain anomaly detection in spatial datasets. *Knowl. Inf. Syst.*, 36(3):749–788, 2013.

[Koufakou and Georgiopoulos, 2010] Anna Koufakou and Michael Georgiopoulos. A fast outlier detection strategy for distributed high-dimensional data sets with mixed attributes. *Data Min. Knowl. Discov.*, 20(2):259–289, 2010.

[Krausz and Herpers, 2010] Barbara Krausz and Rainer Herpers. *MetroSurv*: detecting events in subway stations. *Multimedia Tools Appl.*, 50(1):123–147, 2010.

[Lin *et al.*, 2009] Zhouchen Lin, Minming Chen, and Yi Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. In *Technical Report, UIUC*, 2009.

[Liu and Lam, 2012] Alexander Liu and Dung N. Lam. Using consensus clustering for multi-view anomaly detection. In *IEEE Symposium on Security and Privacy Workshops*, pages 117–124, 2012.

[Liu *et al.*, 2010] Guangcan Liu, Zhouchen Lin, and Yong Yu. Robust subspace segmentation by low-rank representation. In *ICML*, pages 663–670, 2010.

[Liu *et al.*, 2012] Guangcan Liu, Huan Xu, and Shuicheng Yan. Exact subspace segmentation and outlier detection by low-rank representation. In *AISTATS*, pages 703–711, 2012.

[Müller *et al.*, 2012] Emmanuel Müller, Ira Assent, Patricia Iglesias Sanchez, Yvonne Mülle, and Klemens Böhm. Outlier ranking via subspace analysis in multiple views of the data. In *ICDM*, pages 529–538, 2012.

[Nie *et al.*, 2010] Feiping Nie, Heng Huang, Xiao Cai, and Chris H. Q. Ding. Efficient and robust feature selection via joint l2, 1–norms minimization. In *NIPS*, pages 1813–1821, 2010.

[Tibshirani *et al.*, 2000] Robert Tibshirani, Guenther Walther, and Trevor Hastie. Estimating the number of clusters in a dataset via the gap statistic. *Journal of the Royal Statistical Society: Series B*, 63:411–423, 2000.

[Xiong *et al.*, 2011] Liang Xiong, Xi Chen, and Jeff G. Schneider. Direct robust matrix factorizatoin for anomaly detection. In *ICDM*, pages 844–853, 2011.

[Xu *et al.*, 2010] Huan Xu, Constantine Caramanis, and Sujay Sanghavi. Robust PCA via outlier pursuit. In *NIPS*, pages 2496–2504, 2010.

[Xu *et al.*, 2013] Chang Xu, Dacheng Tao, and Chao Xu. A survey on multi-view learning. *CoRR*, abs/1304.5634, 2013.

[Yang *et al.*, 2009] Junfeng Yang, Wotao Yin, Yin Zhang, and Yilun Wang. A fast algorithm for edge-preserving variational multichannel image restoration. *SIAM J. Imaging Sciences*, 2(2):569–592, 2009.

[Zhou *et al.*, 2010] Zihan Zhou, Xiaodong Li, John Wright, Emmanuel J. Candès, and Yi Ma. Stable principal component pursuit. In *ISIT*, pages 1518–1522, 2010.