Multi-View Low-Rank Analysis for Outlier Detection

Sheng Li* Ming Shao* Yun Fu†

Abstract

Outlier detection is a fundamental problem in data mining. Unlike most existing methods that are designed for single-view data, we propose a multi-view outlier detection approach in this paper. Multi-view data can provide plentiful information of samples, however, detecting outliers from multi-view data is still a challenging problem due to the complicated distribution and inconsistent behavior of samples across different views. We address this problem through robust data representation, by building a Multi-view Low-Rank Analysis (MLRA) framework. Our framework contains two major components. First, it performs cross-view low-rank analysis for revealing the intrinsic structures of data. Second, it identifies outliers by estimating the outlier score for each test sample. Specifically, we formulate the cross-view low-rank analysis as a constrained rank-minimization problem, and present an efficient optimization algorithm to solve it. Different from the existing multi-view outlier detection methods, our framework is able to detect two different types of outliers from multiple views simultaneously. To this end, we design a criterion to estimate the outlier scores by analyzing the obtained representation coefficients. Experimental results on seven UCI datasets and the USPS-MNIST dataset demonstrate that our approach outperforms several state-of-the-art single-view and multi-view outlier detection methods in most cases.

1 Introduction

Outlier detection (or anomaly detection) is a data mining technique that identifies the abnormal samples in a data set, and it has been widely used in many safety-critical applications such as fraud detection, network intrusion identification and system health monitoring. Many outlier detection methods have been proposed over the past decades [19, 44, 35, 9], including the reference-based approach [33], inductive logic programming based algorithm [2], information-theoretic algorithm [42], and isolation based algorithm [27]. Recently, some outlier detection methods have been developed to deal with the high-dimensional data. Pham *et al.* designed an efficient algorithm for angle-based outlier detection in high-dimensional data [34]. Zimek *et al.* studied the subsampling problem in statistical outlier detection, and provided effective solutions [45]. Schubert *et al.* presents a generalization of density-based outlier detection methods using kernel density estimation [35]. Generally, these existing methods usually analyze the distribution or density of a dataset, and identify outliers by using some well-defined criteria. Moreover, these methods were designed for single-view data like many other conventional data mining methods.

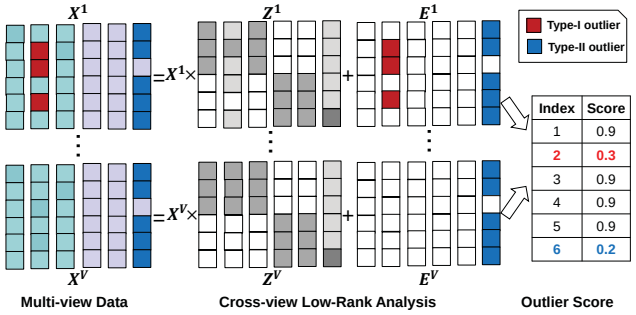

Figure 1: Flowchart of the proposed MLRA framework. Given a multi-view sample set $X = \{X^1, X^2, \cdots, X^V\}$, MLRA first performs cross-view low-rank analysis to reveal the reconstruction relationship of samples, and then calculate outlier scores. Finally, it identifies the outliers contained in data (i.e., the second and sixth column in data matrix). Z^v are representation coefficient matrices with low-rank structures, and E^v are sparse matrices.

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 [43]. Many algorithms have been designed in the multi-view settings, by considering the complementary information from different data views. Moreover, some machine learning and data mining problems, such as clustering [40] and subspace learning [41], have been greatly benefitted from the multi-view data. Nevertheless, detecting outliers from multi-view data is still a challenging problem, due to the complicated distribution and inconsistent behavior of samples across different views.

*Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA. {shengli, mingshao}@ece.neu.edu

†Department of Electrical and Computer Engineering and College of Computer and Information Science, Northeastern University, Boston, MA, USA. yunfu@ece.neu.edu

1.1 Motivation and Contribution In this paper, we tackle the multi-view outlier detection problem from the perspective of data representation. We would argue that, by leveraging the representation relationship of samples, the outliers contained in data set can be correctly identified.

Recently, low-rank matrix recovery has been extensively studied to exploit the intrinsic structure of data [8]. Many applications have been benefited from such structural information, such as subspace clustering [30], multi-task learning [10], subspace learning [23], transfer learning [36] and semi-supervised classification [22]. In low-rank subspace clustering [30], the sample set is served as bases (or dictionary) to reconstruct itself, which inspires us to explore the representation relationship of samples. Our intuition is that *a normal sample usually serves as a good contributor in representing the other normal samples, while the outliers do not.* Therefore, it is reasonable to identify outliers from the representation coefficients in low-rank matrix recovery.

Based on the assumptions above, we propose a novel outlier detection framework named Multi-view Low-Rank Analysis (MLRA). Figure 1 shows the flowchart of our framework. It contains two successive components: (1) robust data representation by cross-view low-rank analysis, and (2) the calculation of outlier scores. In particular, two types of outliers are considered in our framework. The Type-I outliers are samples that show inconsistent clustering results across different views, and the Type-II outliers have abnormal behaviors in each view. In Figure 1, the second column in X^1 (marked by red color) is an outlier in view 1, but it's a normal sample in other views. So it's a Type-I outlier. Moreover, if the last column (marked by blue color) is abnormal in all of the V views, it's a Type-II outlier.

By far, only a few methods have been proposed to detect outliers in multi-view data. Das *et al.* presented a heterogeneous anomaly detection method using multiple kernel learning [12]. Muller *et al.* proposed a multi-view outlier ranking algorithm using subspace analysis [32]. Hsiao *et al.* utilized the pareto depth analysis to develop a multi-criteria anomaly detection algorithm [15]. The most relevant works to our approach are clustering based multi-view outlier detection methods, horizontal anomaly detection (HOAD) [14] and anomaly detection using affinity propagation (AP) [1]. Unlike the existing methods, our approach tackles the multi-view outlier detection problem from a different perspective, i.e., robust data representation.

Furthermore, although the two types of outliers discussed above exist in many real-world applications, traditional single-view and multi-view outlier detection methods cannot handle them simultaneously. For example, the multi-view methods proposed in [14] and [1] are only designed for the Type-I outliers. However, our approach can detect both Type-I and Type-II outliers.

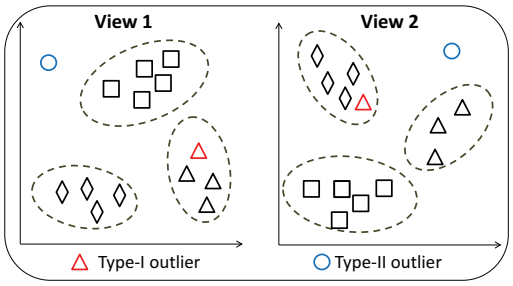

Figure 2: Illustration of Type-I outliers (red triangles) and Type-II outliers (blue circles) in two-view data.

We formulate the cross-view low-rank analysis in our framework as a constrained rank-minimization problem, and present an efficient optimization algorithm to solve it. After that, we devise a criterion to estimate the outlier score for each sample, considering two types of outliers in multiple views. Extensive results on several benchmark datasets are reported.

In summary, the major contributions of this paper are as follows:

- We design a multi-view outlier detection framework. To the best of our knowledge, this paper is the first attempt to detect two types of outliers in a joint framework.
- Outliers are identified from the perspective of data representation. To this end, we develop a cross-view low-rank analysis model, and present an efficient optimization algorithm to solve it.
- Extensive results on seven UCI datasets, and the USPS-MNIST dataset demonstrate the effectiveness of our approach.

2 Preliminary

Given a single-view sample set $\bar{X} = \{x_1, x_2, \cdots, x_n\} \in \mathbb{R}^{d\times n}$ that contains a small amount of outliers, traditional *single-view outlier detection* methods aim at identifying those outliers automatically. These methods usually utilize the distance or density information from sample set [3], and identify outliers using decision boundaries or outlier scores.

When data are collected from multiple views, we have a collection of sample sets, $X = \{X^{(1)}, X^{(2)}, \cdots, X^{(V)}\}$, where V is the total number of views. For the sample set observed in view v, we also have $X^{(v)} = \{x_1^{(v)}, x_2^{(v)}, \cdots, x_n^{(v)}\}$, where n is the number of samples in each view. Generally, *multi-view outlier detection* is more difficult than *single-view outli-*

er detection, as outliers may behave completely different across multiple views.

In this paper, we focus on detecting outliers from multi-view data. In particular, we aim to identify two types of outliers that are defined below.

Definition 1 *Type-I Outlier* is an outlier that exhibit inconsistent characteristics (e.g., cluster membership) across different views.

Figure 2 illustrates two data views, and each view contains three clusters and several outliers. The red triangle belongs to different clusters in view 1 and view 2. Thus, it's a Type-I outlier. Note that the existing multi-view outlier detection algorithms [14, 1] are designed for the Type-I outlier.

Definition 2 *Type-II Outlier* is an outlier that exhibits consistent characteristics across different views, but it shows abnormal behavior in each view.

In Figure 2, the blue circle is a Type-II outlier because it does not belong to any cluster in both views. We also notice that this type of outliers are ignored by existing multi-view outlier detection methods.

3 Multi-View Low-Rank Analysis (MLRA)

In this section, we describe the proposed multi-view low-rank analysis (MLRA) framework. Our goal is to detect two types of outliers simultaneously. As shown in Figure 1, our MLRA framework contains two successive components, which are cross-view low-rank analysis and the calculation of outlier scores.

3.1 Cross-View Low-Rank Analysis

We formulate the cross-view low-rank analysis as a constrained rank minimization problem, and then present an optimization algorithm to solve it.

Problem Formulation

Unlike clustering based methods presented in [14, 1], we tackle the multi-view outlier detection problem from the perspective of data representation. In particular, for the sample set $X^{(v)} \in \mathbb{R}^{d\times n}$ observed in the v-th view, we can represent it as

$$X^{(v)} = X^{(v)}Z^{(v)} + E^{(v)}, \tag{3.1}$$

where $Z^{(v)} \in \mathbb{R}^{n\times n}$ is a coefficient matrix and $E^{(v)} \in \mathbb{R}^{d\times n}$ is a noise matrix.

Like other outlier detection algorithms, we assume that the (normal) samples came from K clusters. The samples in the same cluster could be drawn from the same subspace. Therefore, $Z^{(v)}$ should be a low-rank coefficient matrix that has the block-diagonal structure. The coefficient vectors in $Z^{(v)}$ belong to the same cluster tend to have high correlations.

On the other hand, outliers are actually "sample-specific" noises in data matrix. It's reasonable to use $l_{2,1}$ norm to measure the noise matrix, as $l_{2,1}$ norm makes the column of the matrix to be zero. Moreover, we consider the cross-view relationships between coefficient matrices $Z^{(v)}$. Our intuition is that the representation coefficients should be consistent for normal data in different views, but should be inconsistent for outliers.

Based on the observations above, we present the objective function as follows

$$\begin{aligned}\min_{Z^{(v)},E^{(v)}} \quad & \sum_{v=1}^{V}(\mathrm{rank}(Z^{(v)}) + \alpha||E^{(v)}||_{2,1}) \\ & +\beta\sum_{v=1}^{V-1}\sum_{p=v+1}^{V}||Z^{(v)} - Z^{(p)}||_{2,1} \\ \text{s.t.} \quad & X^{(v)} = X^{(v)}Z^{(v)} + E^{(v)},\ v = 1, 2, \cdots, V,\end{aligned} \tag{3.2}$$

where $\|E\|_{2,1}$ denotes the $l_{2,1}$ norm, and $\|E\|_{2,1} = \sum_{i=1}^{n}\sqrt{\sum_{j=1}^{d}([E]_{ji})^2}$, α and β are trade-off parameters to balance different terms.

In (3.2), the first two terms $\sum_{v=1}^{V}(\mathrm{rank}(Z^{(v)}) + \alpha||E^{(v)}||_{2,1})$ represent the low-rank and sparse constraints on each view, respectively. The last term $\sum_{v=1}^{V-1}\sum_{p=v+1}^{V}||Z^{(v)} - Z^{(p)}||_{2,1}$ indicates the summation of pair-wise error of coefficient matrices $Z^{(v)}$. Considering the inconsistency columns in $Z^{(v)}$ and $Z^{(p)}$, we utilize the $l_{2,1}$ norm on $(Z^{(v)} - Z^{(p)})$. This term ensures robust data representations. If the coefficient matrix in a specific view (e.g., $Z^{(2)}$) are unreliable or corrupted, it would be fixed by virtue of the last term.

For the sake of simplicity, we only provide detailed derivations and solutions for the two-view case. They can be extended easily to multi-view cases. In the two-view case, we have $X = \{X^{(1)}, X^{(2)}\}$, and $x_i^{(v)}$ $(v = 1, 2)$ denote the i-th sample in view v. Then we can modify the object function in (3.2) for two-views. However, the optimization problem in (3.2) is hard to solve, as rank$(\cdot)$ function is neither convex nor continuous. Trace norm is a commonly-used approximation of the non-convex function rank$(\cdot)$ [8, 18]. Then, (3.2) for the two-view case is formulated as

$$\begin{aligned}\min_{Z^{(v)},E^{(v)}} \quad & \sum_{v=1}^{2}(\|Z^{(v)}\|_* + \alpha||E^{(v)}||_{2,1}) + \beta||Z^{(1)} - Z^{(2)}||_{2,1} \\ \text{s.t.} \quad & X^{(v)} = X^{(v)}Z^{(v)} + E^{(v)},\ v = 1, 2,\end{aligned} \tag{3.3}$$

where $\|\cdot\|_*$ represents the trace norm [8].

Optimization

To solve (3.3), we employ an efficient optimization technique, the inexact augmented Lagrange multiplier (ALM) algorithm [25]. First, we introduce relaxation

variables $J^{(v)}$ and S to (3.3), and obtain

(3.4)
$$\begin{aligned}\min_{Z^{(v)},J^{(v)},E^{(v)},S} \quad & \sum_{v=1}^{2}(\|J^{(v)}\|_* + \alpha||E^{(v)}||_{2,1}) + \beta||S||_{2,1} \\ \text{s.t.} \quad & X^{(v)} = X^{(v)}Z^{(v)} + E^{(v)}, \\ & Z^{(v)} = J^{(v)},\ v = 1, 2, \\ & S = Z^{(1)} - Z^{(2)}.\end{aligned}$$

Furthermore, the augmented Lagrangian function of (3.4) is

(3.5)
$$\begin{aligned}\mathcal{L} = & \sum_{v=1}^{2}(\|J^{(v)}\|_* + \alpha||E^{(v)}||_{2,1}) + \beta||S||_{2,1} + \\ & \sum_{v=1}^{2}(\langle W^{(v)}, X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)}\rangle + \frac{\mu}{2}||Z^{(v)} - J^{(v)}||_{\mathrm{F}}^2 \\ & + \langle P^{(v)}, Z^{(v)} - J^{(v)}\rangle + \frac{\mu}{2}||X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)}||_{\mathrm{F}}^2) \\ & + \langle Q, S - (Z^{(1)} - Z^{(2)})\rangle + \frac{\mu}{2}||S - (Z^{(1)} - Z^{(2)})||_{\mathrm{F}}^2,\end{aligned}$$

where $W^{(v)}$, $P^{(v)}$ and Q are Lagrange multipliers, and $\mu > 0$ is a penalty parameter.

The objective function is not jointly convex to all the variables, but it is convex to each of them when fixing the others. Therefore, we update each variable as follows.

Update $J^{(v)}$

By ignoring the irrelevant terms w.r.t. $J^{(v)}$ in (3.5), we have the objective as follows

(3.6)
$$J^{(v)} = \arg\min_{J^{(v)}} \sum_{v=1}^{2}(\frac{1}{\mu}||J^{(v)}||_* + \frac{1}{2}||J^{(v)} - (Z^{(v)} + \frac{P^{(v)}}{\mu})||_{\mathrm{F}}^2).$$

The optimal solution to (3.6) can be obtained by using the singular value thresholding (SVT) algorithm [7]. In detail, we have $\Delta_J = Z^{(v)} + (P^{(v)}/\mu)$. The SVD of Δ_J is written as $\Delta_J = U_J\Sigma_J V_J$, where $\Sigma_J = diag(\{\sigma_i\}_{1\le i\le r})$, r denotes the rank, and σ_i denote the singular values. The solution is $J^{(v)} = U_J\Omega_{(1/\mu)}(\Sigma_J)V_J$, where $\Omega_{(1/\mu)}(\Sigma_J) = diag(\{\sigma_i - (1/\mu)\}_+)$, and $(a)_+$ indicates the positive portion of a.

Update Z^v

We ignore the terms independent of $Z^{(v)}$ in (3.5), and obtain

(3.7)
$$\begin{aligned}& \sum_{v=1}^{2}(\langle W^{(v)}, X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)}\rangle + \frac{\mu}{2}||Z^{(v)} - J^{(v)}||_{\mathrm{F}}^2 \\ & + \langle P^{(v)}, Z^{(v)} - J^{(v)}\rangle + \frac{\mu}{2}||X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)}||_{\mathrm{F}}^2) \\ & + \langle Q, S - (Z^{(1)} - Z^{(2)})\rangle + \frac{\mu}{2}||S - (Z^{(1)} - Z^{(2)})||_{\mathrm{F}}^2,\end{aligned}$$

By setting the derivative w.r.t. $Z^{(1)}$ and $Z^{(2)}$ respectively , we obtain the solutions as follows

(3.8)
$$\begin{aligned}Z^{(1)} = \ & (2\mathrm{I} + X^{(1)\top}X^{(1)})^{-1}(X^{(1)\top}(X^{(1)} - E^{(1)}) \\ & + J^{(1)} + S + Z^{(2)} + \frac{X^{(1)\top}W^{(1)} - P^{(1)} + Q}{\mu}).\end{aligned}$$

(3.9)
$$\begin{aligned}Z^{(2)} = \ & (2\mathrm{I} + X^{(2)\top}X^{(2)})^{-1}(X^{2\top}(X^{(2)} - E^{(2)}) \\ & + J^{(2)} - S + Z^{(1)} + \frac{X^{(2)\top}W^{(2)} - P^{(2)} - Q}{\mu}).\end{aligned}$$

Update S

By dropping the terms irrelevant to S, the equation (3.5) is reduced to

(3.10)
$$\begin{aligned}S = \arg\min_{S} \quad & \frac{\beta}{\mu}||S||_{2,1} \\ & + \frac{1}{2}||S - (Z^{(1)} - Z^{(2)} + \frac{Q}{\mu})||_{\mathrm{F}}^2.\end{aligned}$$

Update $E^{(v)}$

Similarly, after dropping terms independent of $E^{(v)}$, we have

(3.11)
$$\begin{aligned}E^{(v)} = \arg\min_{E^{(v)}} & \sum_{v=1}^{2}(\frac{\alpha}{\mu}||E^{(v)}||_{2,1} \\ & + \frac{1}{2}||E^{(v)} - (X^{(v)} - X^{(v)}Z^{(v)} + \frac{W^{(v)}}{\mu})||_{\mathrm{F}}^2).\end{aligned}$$

The solution to problems like (3.10) and (3.11) is discussed in [30]. Take (3.10) as an example and let $\Psi = Z^{(1)} - Z^{(2)} + \frac{Q}{\mu}$, the i-th column of S is

(3.12)
$$S(:,i) = \begin{cases} \frac{\|\Psi_i\| - \beta}{\|\Psi_i\|}\Psi_i, & if\ \beta < \|\Psi_i\|, \\ 0, & \text{otherwise.}\end{cases}$$

Finally, the complete optimization algorithm for solving (3.5) is outlined in Algorithm 1. We also show the initializations for each variable in the algorithm.

Discussion and Complexity Analysis

The Inexact ALM is a mature optimization technique. It usually converges well in practice, although proving the convergence in theory is still an open issue [11]. In the experiments, we will show the convergence property of our algorithm.

Steps 2-4 are the most time-consuming parts in Algorithm 1. Let n denote the sample size. In Step 2, the SVD of $n \times n$ matrices is required by the SVT operator, which costs $O(n^3)$. Steps 3-4 involve the matrix inversion and matrix multiplication, which usually cost $O(n^3)$. If the number of iterations is t, the overall time complexity of Algorithm 1 becomes $O(tn^3)$.

3.2 Outlier Score Estimation With the optimal solutions $Z^{(v)}$ and $E^{(v)}$, we design a criterion to estimate the outlier score of each sample. To calculate the outlier score vector o, our criterion (for the two-view case) is formulated as

(3.13)
$$o(i) = \sum_{k=1}^{n}(u_k^{(i)}Z_{ik}^{(1)}Z_{ik}^{(2)}) - \lambda\sum_{k=1}^{n}(E_{ik}^{(1)}E_{ik}^{(2)}),$$

where $o(i)$ denotes the outlier score of the i-th sample, $u^{(i)} \in \mathbb{R}^{n\times 1}$ is a constant indictor vector. In detail, the

Algorithm 1. *Solving* (3.5) *using Inexact ALM*

Input: data set $X = \{X^{(1)}, X^{(2)}\}$, parameters α, β, $Z^{(v)} = J^{(v)} = 0$, $E^{(v)} = 0$, $W^{(v)} = 0$, $P^{(v)} = 0$, $Q = 0$, $\rho = 1.2$, $\mu = 0.1$, $\mu_{\max} = 10^{10}$, $\epsilon = 10^{-8}$

1:**while** *not converged* **do**
2: Fix the others and update $J^{(1)}$ and $J^{(2)}$ using (3.6).
3: Fix the others and update $Z^{(1)}$ using (3.8).
4: Fix the others and update $Z^{(2)}$ using (3.9).
5: Fix the others and update S using (3.10).
6: Fix the others and update $E^{(v)}$ using (3.11).
7: Update the multipliers $W^{(v)}$, $P^{(v)}$ and Q
$W^{(v)} = W^{(v)} + \mu(X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)})$,
$P^{(v)} = P^{(v)} + \mu(Z^{(v)} - J^{(v)})$,
$Q = Q + \mu(S - (Z^{(1)} - Z^{(2)}))$.
8: Update the penalty parameter μ by
$\mu = \min(\mu_{\max}, \rho\mu)$
9: Examine the conditions for convergence
$||X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)}||_\infty < \epsilon$ and
$||Z^{(v)} - J^{(v)}||_\infty < \epsilon$ and
$||S - (Z^{(1)} - Z^{(2)})||_\infty < \epsilon$
10: **end while**

Output: $Z^{(v)}, E^{(v)}$

Algorithm 2. *MLRA for Outlier Detection*

Input: Multi-view sample set X, threshold γ

Output: Binary outlier label vector L

1: Normalize each sample $x_i^{(v)}$,
$x_i^{(v)} = x_i^{(v)}/||x_i^{(v)}||$.
2: Solve objective (3.5) using Algorithm 1 and obtain optimal solution Z^v, E^v.
3: Calculate outlier score for each sample using (3.14).
4: Generate binary label vector L
If $o(i) < \gamma$, $L(i) = 1$*; otherwise,* $L(i) = 0$.

k-th element in $u^{(i)}$ corresponds to the k-th sample in X. If samples x_i and x_k belong to the same class, then $u_k^{(i)} = 1$; otherwise, $u_k^{(i)} = 0$. λ is a trade-off parameter (we set $\lambda = 0.5$ in the experiments).

The criterion (3.13) helps us detect two types of outliers simultaneously. The first term in (3.13) measures the inconsistency of the i-th sample across two views. From the point of view of data representation, a sample is mainly represented by those samples came from the same cluster. Therefore, we evaluate the inner-class representation coefficients by virtue of $u^{(i)}$. For instance, if the i-th sample is normal in both views, the coefficients in $Z_i^{(1)}$ and $Z_i^{(2)}$ should be consistent. As a result, the value of $\sum_{k=1}^{n}(u_k^{(i)} Z_{ik}^{(1)} Z_{ik}^{(2)})$ should be relatively large. On the contrary, if the i-th sample is an outlier that exhibits diverse characteristics in different views, the inconsistent coefficients $Z_i^{(1)}$ and $Z_i^{(2)}$ would lead to a small value. Therefore, this term is suitable for detecting the Type-I outliers.

The second term in (3.13) contributes to identifying the Type-II outliers. Each column in $E^{(1)}$ and $E^{(2)}$ corresponds to the reconstruction error vectors in view 1 and view 2, respectively. If the i-th sample is normal in at least one of the views, the value of $\sum_{k=1}^{n}(E_{ik}^{(1)} E_{ik}^{(2)})$ tends to be zero, and then this term won't affect the outlier score $o(i)$ too much. However, if the i-th sample is a Type-II outlier which shows abnormal behavior in both views, the summation in the second term will be increased. Finally, the outlier score $o(i)$ will be further decreased.

Further, a general criterion for the V-view case is (3.14)

$$o(i) = \sum_{p=1}^{V-1}\sum_{q=p+1}^{V}(\sum_{k=1}^{n}(u_k^{(i)} Z_{ik}^{(p)} Z_{ik}^{(q)}) - \lambda\sum_{k=1}^{n}(E_{ik}^{(p)} E_{ik}^{(q)})),$$

After calculating the outlier scores for all the samples, the sample x_i is marked as an outlier if the score $o(i)$ is smaller than the threshold γ. The entire MLRA algorithm is summarized in Algorithm 2.

4 Experiments

The performance of our MLRA framework is evaluated on seven UCI datasets [5] and the USPS-MNIST dataset [16, 20]. The details of datasets, settings, and some additional results are provided in the supplementary document[1] due to space limitation.

4.1 Settings Our approach is compared with several state-of-the-art single-view and multi-view outlier detection methods in the presence of two types of outliers. The compared methods are listed as follows:

- Low-Rank Representations (LRR) [29]. LRR is a representative outlier detection method for single-view data. Thus, we testify its performance on two views separately.
- HOrizontal Anomaly Detection (HOAD) [14]. HOAD is a clustering-based multi-view outlier detection method. Two parameters m and k in HOAD have been fine tuned to obtain its best performance.
- Anomaly detection using Affinity Propagation (AP) [1]. AP is the state-of-the-art multi-view outlier detection method. The authors employed two affinity measurements and four anomaly score calculation strategies. In this paper, we use the L-2 distance and Hilbert-Schmidt Independence Criterion (HSIC), as they usually yield better performance than others.

[1] `https://sites.google.com/site/lisheng1989/home/SDM15_Sup.pdf`

Table 1: Average AUC values (± standard deviations) on seven UCI datasets with only Type-I outliers.

Datasets	Single-view Methods		Multi-view Methods		
	LRR (view 1) [29]	LRR (view 2) [29]	HOAD [14]	AP [1]	Ours
iris	0.5051±0.0818	0.4933±0.0740	0.8308±0.0582	**0.9551±0.0293**	0.8416±0.0194
Letter	0.4937±0.0171	0.4816±0.0183	0.5307±0.0398	0.8483±0.0127	**0.8755±0.0154**
Ionosphere	0.4641±0.0432	0.4633±0.0451	0.5009±0.0571	**0.9380±0.0255**	0.8673±0.0295
zoo	0.6020±0.0842	0.5860±0.0961	0.5486±0.0948	**0.9149±0.0483**	0.8993±0.0506
waveform	0.5012±0.0315	0.5021±0.0296	0.7525±0.0358	0.6177±0.0192	**0.7678±0.0247**
pima	0.4898±0.0339	0.5257±0.0274	0.5582±0.0304	0.6682±0.0351	**0.7390±0.0253**
wdbc	0.5281±0.0403	0.5473±0.0395	0.4493±0.0613	0.9188±0.0251	**0.9311±0.0149**

Table 2: Average AUC values (± standard deviations) on seven UCI datasets with Type-I and Type-II outliers.

Datasets	Single-view Methods		Multi-view Methods		
	LRR (view 1) [29]	LRR (view 2) [29]	HOAD [14]	AP [1]	Ours
iris	0.4036±0.0642	0.3845±0.0556	0.3704±0.0401	0.6950±0.0209	**0.8361±0.0475**
Letter	0.3434±0.0120	0.3307±0.0134	0.3437±0.0122	0.6864±0.0125	**0.7785±0.0093**
Ionosphere	0.4258±0.0321	0.4401±0.0363	0.4990±0.0456	0.7597±0.0215	**0.7913±0.0284**
zoo	0.4262±0.0517	0.4018±0.0646	0.5846±0.0660	0.7692±0.0731	**0.8528±0.0422**
waveform	0.4300±0.0189	0.3894±0.0254	0.7708±0.0263	0.4222±0.0131	**0.8323±0.0201**
pima	0.3325±0.0240	0.3430±0.0242	0.3747±0.0203	0.4637±0.0216	**0.7718±0.0271**
wdbc	0.3275±0.0281	0.3422±0.0229	0.3323±0.0693	0.4768±0.0282	**0.7913±0.0125**

As suggested in [14, 1], we adopt the receiver operating characteristic (ROC) curves as the evaluation metric, which represents the trade-off between detection rate and false alarm rate. We also report the area under ROC curve (AUC). The false positive rate (FPR) and true positive rate (TPR) used for generating ROC curves are defined as follows:

$$FPR = \frac{FP}{FP+TN}, \quad TPR = \frac{TP}{TP+FN}, \tag{4.15}$$

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

4.2 UCI Datasets We employ seven benchmark datasets, namely "Iris", "Letter", "Waveform", "Zoo", "Ionosphere", "Pima" and "Wdbc" from the UCI machine learning repository [5]. For each dataset, we generate two types of outliers.

To illustrate the convergence property of our algorithm, we show in Figure 3(a) the relative error on the Iris dataset. The relative error in each iteration is calculated by $\max(||X^{(1)} - X^{(1)}Z^{(1)} - E^{(1)}\|_{\mathrm{F}}/||X^{(1)}||_{\mathrm{F}}, ||X^{(2)} - X^{(2)}Z^{(2)} - E^{(2)}\|_{\mathrm{F}}/||X^{(2)}||_{\mathrm{F}})$. Figure 3(a) shows that our algorithms converges quickly, which ensures the less computational cost of our approach.

As for the parameter selection, we adopted a coarse-to-fine strategy to find the proper range for parameters. There are two major parameters in our approach, α and β. We tuned their values in the range of $\{10^{-2}, 10^{-1}, \cdots, 10^{2}\}$. Figure 3(b) shows the AUC

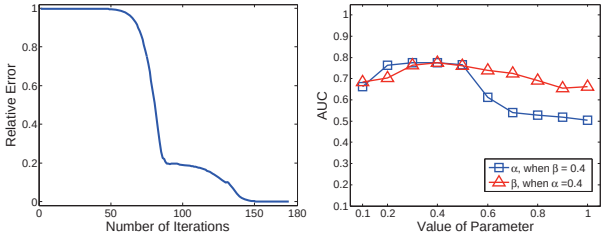

Figure 3: (a) Convergence curve of our algorithm on Iris dataset. (b) AUC of our approach on Piam dataset by varying the values of α and β.

of our approach on the Pima dataset, varying the values of α and β. Note that we obtained similar results on other datasets. We can observe from Figure 3(b) that, as the dataset contain sample-specific noise, the two parameters usually tend to be small values around 0.04. Also, as we chose ROC and AUC as the evaluation metrics, we do not need to specify the threshold γ in Algorithm 2. In fact, different values of γ were employed to generate the ROC curves.

For each dataset, we repeat the random outlier generation procedures for 50 times, evaluate the performance of each compared method on those 50 sets, and report the average results. We conduct two settings for each method: (1) Type-I outliers only; (2) Type-I and Type-II outliers. In this way, we can observe the strengths and limitations of different methods.

Table 1 reports the average area under ROC curve (AUC) values (with standard deviations) on seven datasets with Type-I outliers. From Table 1, we have

the following observations. First, the results of single-view method LRR on two views are quite close, and they are much lower than the multi-view methods. Secondly, AP performs better than LRR and HOAD in most cases, and it achieves the best results on the Iris, Ionosphere and Zoo datasets. Thirdly, our approach outperforms the other compared methods on four datasets, and also obtains competitive results on the Zoo dataset. In all, it shows that AP and our approach work very well in detecting the Type-I outliers, and our approach obtains the best results in most cases.

Table 2 shows the average AUC values on seven datasets with both Type-I and Type-II outliers. We can observe from Table 2 that our approach significantly outperforms other competitors in all the cases. In addition, LRR-v1 and LRR-v2 are very close to each other, and AP still performs better than HOAD in most datasets except waveform. The results demonstrate that our approach can detect two types of outliers simultaneously.

4.3 USPS-MNIST Dataset We construct a two-view dataset by combining two popular handwritten datasets, USPS [16] and MINST [20]. The USPS dataset contains 9298 handwritten digit images, and the MNIST handwritten contains 70000 digit images. The same digits in two datasets can be considered as two different views, as they were collected under different scenarios.

In the experiments, we randomly select 50 images per digit from each dataset. Thus, there are 500 samples in each view. We employed the same strategies as in the UCI datasets to generate 5% Type-I outliers and 5% Type-II outliers. This procedure was repeated 20 times, and we then evaluated each compared method on these 20 sample sets.

Figure 4 shows the ROC curves, and Table 3 lists the average AUC values with standard deviations. We can observe that our approach significantly outperforms other single-view and multi-view outlier detection algorithms. It's very interesting that LRR (view 1) attains even better performance than two multi-view methods, HOAD and AP. The reason is that, in this dataset we have the same amount of Type-I and Type-II outliers. LRR is suitable for handling the Type-II outliers. Two multi-view methods, HOAD and AP, are only capable of detecting Type-I outliers. However, our approach is able to detect two types of outliers effectively.

5 Related Works

In general, our work is closely related to the following topics: multi-view learning, outlier detection, and low-rank learning.

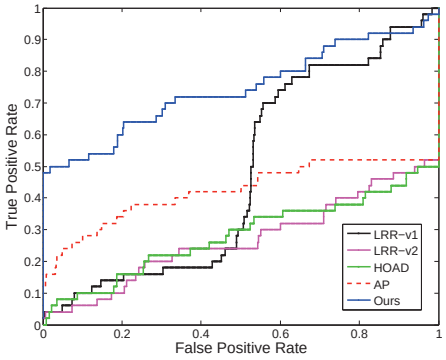

Figure 4: ROC Curves of all compared methods on USPS-MNIST dataset.

Table 3: Average AUC values with standard deviations of compared methods on USPS-MNIST dataset.

Method	AUC (± standard deviation)
LRR (view 1) [29]	0.4758±0.0291
LRR (view 2) [29]	0.2686±0.0316
HOAD [14]	0.3046±0.0233
AP [1]	0.4281±0.0154
Ours	**0.7444±0.0105**

5.1 Multi-View Learning Multi-view learning has been receiving increasing attention in recent years. One implicit assumption is that either view alone has sufficient information about the samples, but the complexity of learning problems can be reduced by eliminating hypotheses from each view that tend not to agree with each other [38]. One of the representative work in multi-view learning is co-training [6], which learns from the samples that are described by two distinct views. Some recent multi-view learning algorithms include manifold co-regularization [37], and multi-view feature learning [31].

The basic idea of these methods is to exploit the *consistency* among multiple views to enhance the learning performance. In our MLRA framework, however, we exploit the *inconsistency* information among different views to identify outliers.

5.2 Outlier Detection As introduced in Section 1, many *single-view* outlier detection algorithms have been developed over the past decade, such as [45, 39, 27, 13, 24]. Tong *et al.* proposed a non-negative residual matrix factorization (NrMF) method for anomaly detection in graph data. It estimates outlier scores from the residuals, but it is only designed for single-view data. To date, only a few methods have been developed to handle the *multi-view* outlier detection problem [17, 26]. The most relevant multi-view methods to our approach are clustering based multi-view outlier detection methods, horizontal anomaly detection (HOAD) [14] and anomaly

detection using affinity propagation (AP) [1].

HOAD aims to detect outliers from several different data sources that can be considered as multi-view data. In HOAD, the samples that have inconsistent behavior among different data sources are marked as anomalies. HOAD first constructs a combined similarity graph based on the similarity matrices in multiple views and computes spectral embeddings for the samples. Then it calculates the anomalous score of each sample using the cosine distance between different spectral embeddings. However, HOAD is only designed for the outliers that show inconsistent behavior across different views (i.e., the Type-I outlier defined in this paper). In addition, the graph constructed in HOAD will be dramatically expanded for multi-view data, which increases considerable computational cost.

Most recently, Alvarez *et al.* proposed an affinity propagation (AP) based multi-view anomaly detection algorithm [1]. This algorithm identifies anomalies by analyzing the neighborhoods of each sample in different views, and it adopts four different strategies to calculate anomaly scores. Specifically, it performs clustering in different view separately. The clustering-based affinity vectors are then calculated for each sample. There are significant differences between our approach and Alvarez's algorithm. First, like HOAD, Alvarez's algorithm is a clustering based method that analyze clustering results in different views to detect outliers. However, our approach models the multi-view outlier detection problem from the perspective of data reconstruction, and performs low-rank analysis to identify outliers. Second, Alvarez's algorithm was only designed for detecting the Type-I outliers. However, our approach can detect both Type-I and Type-II outliers jointly.

5.3 Low-Rank Learning Our approach is also related to low-rank matrix learning, which has attracted increasing attention in recent years [4, 21]. Robust PCA (RPCA) [8] and Low-Rank Representation (LRR) are two representative low-rank learning methods. RPCA can recover noisy data from one single space, while LRR is able to recover multiple subspaces in the presence of noise [30, 28]. The most successful application of LRR is subspace clustering. It can correctly recover the subspace membership of samples, even if the samples are heavily corrupted.

In addition, LRR shows promising performance in outlier detection [29]. However, it can only deal with single-view data. Different from LRR, our approach performs multi-view low-rank analysis for outlier detection. To the best of our knowledge, MLRA is the first multi-view low-rank learning approach.

6 Conclusions

We propose a multi-view low-rank analysis (MLRA) framework in this paper for outlier detection. Our framework performs cross-view low-rank analysis, and employs a well designed criterion to calculate the outlier score for each sample. We formulate it as a rank-minimization problem, and adopt Inexact ALM algorithm to solve it. By analyzing the representation coefficients in different views, our framework is able to detect two different types of outliers simultaneously. Experimental results on seven UCI datasets and the USPS-MNIST dataset show that the proposed approach outperforms the state-of-the-art single-view and multi-view outlier detection methods under various settings. Especially when the datasets contain both Type-I and Type-II outliers, our approach can significantly boost the performance of outlier detection.

Acknowledgements

This research is supported in part by the NSF CNS award 1314484 and NSF IIS award 1449266, ONR award N00014-12-1-1028, ONR Young Investigator Award N00014-14-1-0484, U.S. Army Research Office Young Investigator Award W911NF-14-1-0218, and Air Force Office of Scientific Research award FA9550-12-1-0201.

References

[1] A. M. Alvarez, M. Yamada, A. Kimura, and T. Iwata. Clustering-based anomaly detection in multi-view data. In *CIKM*, pages 1545–1548, 2013.

[2] F. Angiulli and F. Fassetti. Outlier detection using inductive logic programming. In *ICDM*, pages 693–698, 2009.

[3] I. Assent, X. H. Dang, B. Micenková, and R. T. Ng. Outlier detection with space transformation and spectral analysis. In *SDM*, pages 225–233, 2013.

[4] F. R. Bach. Consistency of trace norm minimization. *Journal of Machine Learning Research*, 9:1019–1048, 2008.

[5] K. Bache and M. Lichman. UCI machine learning repository, 2013.

[6] A. Blum and T. M. Mitchell. Combining labeled and unlabeled sata with co-training. In *COLT*, pages 92–100, 1998.

[7] J. F. Cai, E. J. Candes, and Z. W. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

[8] E. J. Candès, X. D. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of ACM*, 58(3):11, 2011.

[9] S. Chawla and A. Gionis. k-means-: A unified approach to clustering and outlier detection. In *SDM*, pages 189–197, 2013.

[10] J. Chen, J. Zhou, and J. Ye. Integrating low-rank and group-sparse structures for robust multi-task learning. In *KDD*, pages 42–50, 2011.

[11] B. Cheng, G. Liu, J. Wang, Z. Huang, and S. Yan. Multi-task low-rank affinity pursuit for image segmentation. In *ICCV*, pages 2439–2446, 2011.

[12] S. Das, B. L. Matthews, A. N. Srivastava, and N. C. Oza. Multiple kernel learning for heterogeneous anomaly detection: algorithm and aviation safety case study. In *KDD*, pages 47–56, 2010.

[13] A. F. Emmott, S. Das, T. Dietterich, A. Fern, and cWeng Keen Wong. Systematic construction of anomaly detection benchmarks from real data. In *KDD Workshop on Outlier Detection and Description*, pages 16–21, 2013.

[14] J. Gao, W. Fan, D. S. Turaga, S. Parthasarathy, and J. Han. A spectral framework for detecting inconsistency across multi-source object relationships. In *ICDM*, pages 1050–1055, 2011.

[15] K.-J. Hsiao, K. S. Xu, J. Calder, and A. O. H. III. Multi-criteria anomaly detection using pareto depth analysis. In *NIPS*, pages 854–862, 2012.

[16] J. Hull. A database for handwritten text recognition research. *IEEE TPAMI*, 16(5):550–554, 1994.

[17] V. P. Janeja and R. Palanisamy. Multi-domain anomaly detection in spatial datasets. *Knowledge and Information Systems*, 36(3):749–788, 2013.

[18] R. H. Keshavan, A. Montanari, and S. Oh. Matrix completion from noisy entries. In *NIPS*, pages 952–960, 2009.

[19] A. Koufakou and M. Georgiopoulos. A fast outlier detection strategy for distributed high-dimensional data sets with mixed attributes. *Data Mining and Knowledge Discovery*, 20(2):259–289, 2010.

[20] Y. LeCun, L. Bottou, Y. Bengio, and P. Haaffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[21] L. Li, S. Li, and Y. Fu. Learning low-rank and discriminative dictionary for image classification. *Image Vision Comput.*, 32(10):814–823, 2014.

[22] S. Li and Y. Fu. Low-rank coding with b-matching constraint for semi-supervised classification. In *IJCAI*, pages 1472–1478, 2013.

[23] S. Li and Y. Fu. Robust subspace discovery through supervised low-rank constraints. In *SDM*, pages 163–171, 2014.

[24] S. Li, M. Shao, and Y. Fu. Locality linear fitting one-class SVM with low-rank constraints for outlier detection. In *IJCNN*, pages 676–683, 2014.

[25] Z. C. Lin, M. M. Chen, L. Q. Wu, and Y. Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *Technique Report, UIUC*, 2009.

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

[27] F. T. Liu, K. M. Ting, and Z. Zhou. Isolation-based anomaly detection. *TKDD*, 6(1):3, 2012.

[28] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma. Robust recovery of subspace structures by low-rank representation. *IEEE PAMI*, 35(1):171–184, 2013.

[29] G. Liu, H. Xu, and S. Yan. Exact subspace segmentation and outlier detection by low-rank representation. In *AISTATS*, pages 703–711, 2012.

[30] G. C. Liu, Z. C. Lin, and Y. Yu. Robust subspace segmentation by low-rank representation. In *ICML*, pages 663–670, 2010.

[31] R. Memisevic. On multi-view feature learning. In *ICML*, 2012.

[32] E. Müller, I. Assent, P. I. Sanchez, Y. Mülle, and K. Böhm. Outlier ranking via subspace analysis in multiple views of the data. In *ICDM*, pages 529–538, 2012.

[33] Y. Pei, O. R. Zaïane, and Y. Gao. An efficient reference-based approach to outlier detection in large datasets. In *ICDM*, pages 478–487, 2006.

[34] N. Pham and R. Pagh. A near-linear time approximation algorithm for angle-based outlier detection in high-dimensional data. In *KDD*, pages 877–885, 2012.

[35] E. Schubert, A. Zimek, and H. Kriegel. Generalized outlier detection with flexible kernel density estimates. In *SDM*, pages 542–550, 2014.

[36] M. Shao, D. Kit, and Y. Fu. Generalized transfer subspace learning through low-rank constraint. *International Journal of Computer Vision*, 109(1-2):74–93, 2014.

[37] V. Sindhwani and D. S. Rosenberg. An rkhs for multi-view learning and manifold co-regularization. In *ICML*, pages 976–983, 2008.

[38] K. Sridharan and S. M. Kakade. An information theoretic framework for multi-view learning. In *COLT*, pages 403–414, 2008.

[39] H. Tong and C. Lin. Non-negative residual matrix factorization with application to graph anomaly detection. In *SDM*, pages 143–153, 2011.

[40] G. Tzortzis and A. Likas. Kernel-based weighted multi-view clustering. In *ICDM*, pages 675–684, 2012.

[41] M. White, Y. Yu, X. Zhang, and D. Schuurmans. Convex multi-view subspace learning. In *NIPS*, pages 1682–1690, 2012.

[42] S. Wu and S. Wang. Information-theoretic outlier detection for large-scale categorical data. *IEEE TKDE*, 25(3):589–602, 2013.

[43] C. Xu, D. Tao, and C. Xu. A survey on multi-view learning. *CoRR*, abs/1304.5634, 2013.

[44] X. Zhou, C. Yang, and W. Yu. Automatic mitral leaflet tracking in echocardiography by outlier detection in the low-rank representation. In *CVPR*, pages 972–979, 2012.

[45] A. Zimek, M. Gaudet, R. J. G. B. Campello, and J. Sander. Subsampling for efficient and effective unsupervised outlier detection ensembles. In *KDD*, pages 428–436, 2013.