Multi-View Low-Rank Analysis with Applications to Outlier Detection

SHENG LI, Adobe Research
MING SHAO, University of Massachusetts Dartmouth
YUN FU, Northeastern University

Detecting outliers or anomalies is a fundamental problem in various machine learning and data mining applications. Conventional outlier detection algorithms are mainly designed for single-view data. Nowadays, data can be easily collected from multiple views, and many learning tasks such as clustering and classification have benefited from multi-view data. However, outlier detection from multi-view data is still a very challenging problem, as the data in multiple views usually have more complicated distributions and exhibit inconsistent behaviors. To address this problem, we propose a multi-view low-rank analysis (MLRA) framework for outlier detection in this article. MLRA pursuits outliers from a new perspective, robust data representation. It contains two major components. First, the cross-view low-rank coding is performed to reveal the intrinsic structures of data. In particular, we formulate a regularized rank-minimization problem, which is solved by an efficient optimization algorithm. Second, the outliers are identified through an outlier score estimation procedure. Different from the existing multi-view outlier detection methods, MLRA 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. Moreover, we extend MLRA to tackle the multi-view group outlier detection problem. Extensive evaluations on seven UCI datasets, the MovieLens, the USPS-MNIST, and the WebKB datasets demon strate that our approach outperforms several state-of-the-art outlier detection methods.

CCS Concepts: • **Information systems** → **Data mining**; **Data cleaning**; • **Computing methodologies** → *Regularization*;

Additional Key Words and Phrases: Multi-view learning, low-rank matrix recovery, outlier detection

ACM Reference format:
Sheng Li, Ming Shao, and Yun Fu. 2018. Multi-View Low-Rank Analysis with Applications to Outlier Detection. *ACM Trans. Knowl. Discov. Data.* 12, 3, Article 32 (March 2018), 22 pages.
https://doi.org/10.1145/3168363

1 INTRODUCTION

As a fundamental data mining technique, outlier detection (or anomaly detection) identifies the abnormal samples in a dataset. Many effective outlier detection algorithms have been developed

This research is supported in part by the NSF IIS award 1651902, ONR Young Investigator Award N00014-14-1-0484, and U.S. Army Research Office Award W911NF-17-1-0367.
Authors' addresses: S. Li, Adobe Research, 345 Park Avenue, San Jose, CA, USA, 95110; email: sheli@adobe.com; M. Shao, Department of Computer and Information Science, University of Massachusetts Dartmouth, 285 Old Westport Road Dartmouth, MA, USA, 02747; email: mshao@umassd.edu; Y. Fu, Northeastern University, 403 Dana Research Center, 360 Huntington Avenue Boston, MA USA, 02115; email: yunfu@ece.neu.edu.

https://doi.org/10.1145/3168363

during the past decades, and they have been extensively applied to many safety-critical applications, such as fraud detection, network intrusion identification and system health monitoring [43, 47, 60]. The representative outlier detection methods include the reference-based approach [44], inductive logic programming based algorithm [2], information-theoretic algorithm [54], and isolation based algorithm [34]. 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 [46]. Zimek et al. studied the subsampling problem in statistical outlier detection, and provided effective solutions [61]. Schubert et al. presented a generalization of density-based outlier detection methods using kernel density estimation [47]. In general, these existing methods 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.

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 [57]. Multi-view data provide plentiful information to characterize the properties of objects. 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 [52] and subspace learning [53], have been greatly benefitted from the multi-view data. Nevertheless, detecting outliers from multi-view data is still a challenging problem for two reasons: (1) the multi-view data usually have more complicated distributions than the single-view data; (2) the data points may exhibit inconsistent behaviors in different views. In other words, outliers may be easily observed as normal data in one or more views.

1.1 Motivation and Contribution

In this article, 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 dataset can be correctly identified.

Recently, low-rank matrix recovery has been extensively studied to exploit the intrinsic structure of data [8, 36]. Many applications have been benefited from such structural information, such as subspace clustering [39], multi-task learning [9], subspace learning [25], transfer learning [48], and semi-supervised classification [24]. In low-rank subspace clustering [39], 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 1 outliers are samples that show inconsistent clustering results across different views, and the Type 2 outliers have abnormal behaviors in each view. For example, an animal dataset might contain two data views, including the image view and text view. The features extracted from a horse image might be very similar to these from a deer image, as these two species have similar limbs. But they have quite different features in text view. Thus, they could be the Type 1 outliers. In addition, if some natural scene images are accidentally included in the animal dataset, they are considered as Type 2 outliers. To build effective learning systems, it is crucial to identify such outliers in advance. In Figure 1, the second column in X^1 (marked by red color) is an outlier

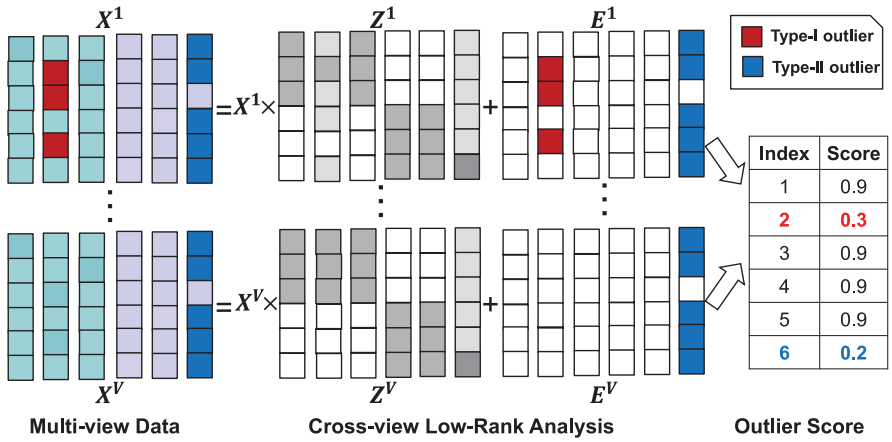


Fig. 1. Flowchart of the proposed MLRA framework. Given a multi-view sample set $X = \{X^1, X^2, \ldots, X^V\}$, MLRA first performs cross-view low-rank analysis to reveal the reconstruction relationship of samples, and then calculate outlier scores. Finally, it identifies the outliers contained in data (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.

in view 1, but it is a normal sample in other views. So it is a Type 1 outlier. Moreover, if the last column (marked by blue color) is abnormal in all of the V views, it is a Type 2 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 [11]. Muller et al. proposed a multi-view outlier ranking algorithm using subspace analysis [42]. Hsiao et al. utilized the pareto depth analysis to develop a multi-criteria anomaly detection algorithm [16]. 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]. These methods obtained promising results in various applications. They detect outliers from the perspective of ensemble learning [11] or clustering [14]. 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 1 outliers. However, our approach can detect both Type 1 and Type 2 outliers.

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. Moreover, we extend the MLRA framework to a new problem, multi-view group anomaly detection. Extensive results on ten benchmark datasets are reported.

This article is an extension of our previous work [26, 27]. In summary, the major contributions of this article are as follows:

—We design a multi-view outlier detection framework, MLRA. To the best of our knowledge, our work is the first attempt to detect two types of outliers in a joint framework.
—We identify the outliers 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.

— We extend MLRA to multi-view group outlier detection, by refining the objective function and devising a criterion for estimating outlier scores.
— We evaluate our approach and related methods on seven UCI datasets, the MovieLens, the USPS-MNIST, and the WebKB datasets. Extensive results demonstrated the effectiveness of our approach.

1.2 Organization

The rest of the article is organized as follows. In Section 2, we review the related works and discuss how they differ from our approach. In Section 3, we introduce the preliminary knowledge of outlier detection, and defines two types of outliers in the multi-view case. In Section 4, we present the MLRA framework, including the problem formulation, optimization, and outlier score estimation. In Section 5, we extend MLRA framework for multi-view group outlier detection. The experimental results and discussions are reported in Section 6. Section 7 concludes this article.

2 RELATED WORKS

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

2.1 Multi-View Learning

Multi-view learning has been receiving increasing attention in recent years [57]. 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 [50]. 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. The representative multi-view learning algorithms include manifold co-regularization [49], and multi-view feature learning [40].

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.

2.2 Outlier Detection

Many *single-view* outlier detection algorithms have been developed over the past decade, such as [13, 34, 51, 61]. 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. Lee et al. designed an anomaly detection algorithm via online oversampling PCA [22]. Liu et al. studied a specific scenario when data have imperfect labels [33]. Du et al. presented a discriminative metric learning for outlier detection [12]. Perozzi et al. proposed the focused clustering and outlier detection method in large attributed graphs [45]. Liu et al. designed the support vector data description (SVDD) based outlier detection method [32].

To date, only a few methods have been developed to handle the *multi-view* outlier detection problem [19, 31]. The most relevant multi-view methods to our approach are clustering based multi-view outlier detection methods, HOAD [14] and anomaly detection using 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 1 outlier defined in this article). 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 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 1 outliers. However, our approach can detect both Type 1 and Type 2 outliers jointly.

Group anomaly detection is a relatively new topic. Several effective algorithms have been presented to address this problem [41, 56, 59]. However, these algorithms can only handle the single-view data. To the best of our knowledge, our work is the first attempt to address the multi-view group outlier detection problem.

2.3 Low-Rank Learning

Our approach is also related to low-rank matrix learning, which has attracted increasing attention in recent years [4, 23]. 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 [35, 39]. 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 [37, 38]. Liu et al. applied the LRR model to outlier detection, and achieved promising results [37]. Li et al. incorporated the low-rank constraint into the SVDD model, and detected outliers from image datasets [28]. However, these methods can only deal with single-view data.

Different from LRR and its variants, our approach performs MLRA for outlier detection. To the best of our knowledge, MLRA is the first multi-view low-rank learning approach.

3 PRELIMINARY

Given a single-view sample set $\bar{X} = \{x_1, x_2, \ldots, x_n\} \in \mathbb{R}^{d \times n}$ 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. In addition, the notations summarized in Table 1 will be used throughout this article.

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

In this article, 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 1 Outlier is an outlier that exhibit inconsistent characteristics (e.g., cluster membership) across different views.

Table 1. Summary of Notations

Notation	Description
X	Multi-view sample set
$\bar{X}$	Single-view sample set
n	The number of samples in X
V	Number of views
$X^{(v)}$	Sample set in view v
$Z^{(v)}$	Representation coefficients in view v
$D^{(v)}$	Dictionary in view v
$E^{(v)}$	Error matrices in view v
o_i	Outlier score for the ith sample in X
$\|\cdot\|_*$	Trace norm (i.e., nuclear norm)
$\|\cdot\|_F$	Frobenius norm
$\|\cdot\|_{2,1}$	$l_{2,1}$ norm

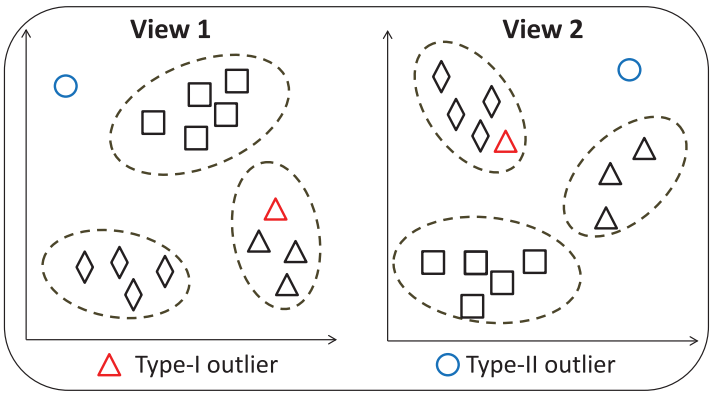


Fig. 2. Illustration of Type 1 outliers (red triangles) and Type 2 outliers (blue circles) in two-view data.

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 is a Type 2 outlier. Note that the existing multi-view outlier detection algorithms [1, 14] are designed for the Type 1 outlier.

Definition 2. Type 2 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 2 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.

4 MULTI-VIEW LOW-RANK ANALYSIS (MLRA)

In this section, we describe the proposed 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.

4.1 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 [1, 14], 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 vth view, we can represent it as

$$X^{(v)} = X^{(v)}Z^{(v)} + E^{(v)}, \tag{1}$$

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 is 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:

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

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 tradeoff parameters to balance different terms.

In (2), the first two terms $\sum_{v=1}^{V}(\text{rank}(Z^{(v)}) + \alpha||E^{(v)}||_{2,1})$ 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 pairwise 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 ith sample in view v. Then, we can modify the object function in (2) for two-views. However, the optimization problem in (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, 20]. Then, (2) for the two-view case is formulated as

$$\begin{aligned} \min_{Z^{(v)},E^{(v)}} \quad & \sum_{v=1}^{2}\left(\left\|Z^{(v)}\right\|_{*} + \alpha\left\|E^{(v)}\right\|_{2,1}\right) + \beta\left\|Z^{(1)} - Z^{(2)}\right\|_{2,1} \\ \text{s.t.} \quad & X^{(v)} = X^{(v)}Z^{(v)} + E^{(v)},\ v = 1, 2, \end{aligned} \tag{3}$$

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

Optimization

To solve (3), we employ an efficient optimization technique, the inexact augmented Lagrange multiplier (ALM) algorithm [30]. First, we introduce relaxation variables $J^{(v)}$ and S to (3), and obtain

$$\begin{aligned} \min_{Z^{(v)},J^{(v)},E^{(v)},S} \quad & \sum_{v=1}^{2}(\|J^{(v)}\|_* + \alpha||E^{(v)}||_{2,1}) + \beta||S||_{2,1} \\ \text{s.t.} \quad & X^{(v)} = X^{(v)}Z^{(v)} + E^{(v)}, \\ & Z^{(v)} = J^{(v)},\ v = 1, 2, \\ & S = Z^{(1)} - Z^{(2)}. \end{aligned} \tag{4}$$

Furthermore, the augmented Lagrangian function of (4) is

$$\begin{aligned} \mathcal{L} = & \sum_{v=1}^{2}(\|J^{(v)}\|_* + \alpha||E^{(v)}||_{2,1}) + \beta||S||_{2,1} \\ + \sum_{v=1}^{2}(&\langle W^{(v)}, X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)}\rangle + \frac{\mu}{2}||Z^{(v)} - J^{(v)}||_{\mathrm{F}}^2 \\ &+ \langle P^{(v)}, Z^{(v)} - J^{(v)}\rangle + \tfrac{\mu}{2}||X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)}||_{\mathrm{F}}^2) \\ &+ \langle Q, S - (Z^{(1)} - Z^{(2)})\rangle + \tfrac{\mu}{2}||S - (Z^{(1)} - Z^{(2)})||_{\mathrm{F}}^2, \end{aligned} \tag{5}$$

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 (5), we have the objective as follows:

$$J^{(v)} = \arg\min_{J^{(v)}} \sum_{v=1}^{2}\left(\frac{1}{\mu}\left\|J^{(v)}\right\|_* + \frac{1}{2}\left\|J^{(v)} - \left(Z^{(v)} + \frac{P^{(v)}}{\mu}\right)\right\|_{\mathrm{F}}^2\right). \tag{6}$$

The optimal solution to (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 $\Sigma_J = \mathrm{diag}(\{\sigma_i\}_{1\le i\le r})$, 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 $\Omega_{(1/\mu)}(\Sigma_J) = \mathrm{diag}(\{\sigma_i - (1/\mu)\}_+)$, and $(a)_+$ indicates the positive portion of a.

Update Z^{v}

We ignore the terms independent of $Z^{(v)}$ in (5), and obtain

$$\begin{aligned} \mathcal{L}(Z^{(v)}) = \sum_{v=1}^{2}(&\langle W^{(v)}, X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)}\rangle + \frac{\mu}{2}||Z^{(v)} - J^{(v)}||_{\mathrm{F}}^2 \\ &+ \langle P^{(v)}, Z^{(v)} - J^{(v)}\rangle + \tfrac{\mu}{2}||X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)}||_{\mathrm{F}}^2) \\ &+ \langle Q, S - (Z^{(1)} - Z^{(2)})\rangle + \tfrac{\mu}{2}||S - (Z^{(1)} - Z^{(2)})||_{\mathrm{F}}^2, \end{aligned} \tag{7}$$

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

$$\begin{aligned} Z^{(1)} = (2\mathrm{I} + X^{(1)\top}X^{(1)})^{-1}&(X^{(1)\top}(X^{(1)} - E^{(1)}) \\ &+ J^{(1)} + S + Z^{(2)} + \frac{X^{(1)\top}W^{(1)} - P^{(1)} + Q}{\mu}). \end{aligned} \tag{8}$$

$$Z^{(2)} = (2\mathrm{I} + X^{(2)\top}X^{(2)})^{-1}(X^{2\top}(X^{(2)} - E^{(2)}) + J^{(2)} - S + Z^{(1)} + \frac{X^{(2)\top}W^{(2)} - P^{(2)} - Q}{\mu}). \tag{9}$$

Update S

By dropping the terms irrelevant to S, Equation (5) is reduced to

$$S = \arg\min_{S} \quad \frac{\beta}{\mu}||S||_{2,1} + \frac{1}{2}\left\|S - \left(Z^{(1)} - Z^{(2)} + \frac{Q}{\mu}\right)\right\|_{\mathrm{F}}^2. \tag{10}$$

Update $E^{(v)}$

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

$$E^{(v)} = \arg\min_{E^{(v)}} \sum_{v=1}^{2}\left(\frac{\alpha}{\mu}||E^{(v)}||_{2,1} + \frac{1}{2}\left\|E^{(v)} - \left(X^{(v)} - X^{(v)}Z^{(v)} + \frac{W^{(v)}}{\mu}\right)\right\|_{\mathrm{F}}^2\right). \tag{11}$$

The solution to problems like (10) and (11) is discussed in [39]. Take (10) as an example and let $\Psi = Z^{(1)} - Z^{(2)} + \frac{Q}{\mu}$, the ith column of S is

$$S(:,i) = \begin{cases} \frac{\|\Psi_i\| - \beta}{\|\Psi_i\|}\Psi_i, & if\ \beta < \|\Psi_i\|, \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

Finally, the complete optimization algorithm for solving (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 [10]. 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 $O(n^3)$. As a result, the time complexity of one iteration in Algorithm 1 is $O(n^3)$. We will show the running time of our algorithm and its competitors in the experiments.

4.2 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

$$o(i) = \sum_{k=1}^{n}\left(u_k^{(i)}Z_{ik}^{(1)}Z_{ik}^{(2)}\right) - \lambda\sum_{k=1}^{n}\left(E_{ik}^{(1)}E_{ik}^{(2)}\right), \tag{13}$$

where $o(i)$ denotes the outlier score of the ith sample, $u^{(i)} \in \mathbb{R}^{n\times 1}$ is a constant indictor vector. In detail, the kth element in $u^{(i)}$ corresponds to the kth sample in X. We consider two settings for $u^{(i)}$. When class labels are available, if samples x_i and x_k belong to the same class, then $u_k^{(i)} = 1$; otherwise, $u_k^{(i)} = 0$. When class labels are unknown, we simply set all $u^{(i)}$ to 1. λ is a tradeoff parameter (we set $\lambda = 0.5$ in the experiments).

ALGORITHM 1: Solving (5) using Inexact ALM

Input: dataset $X = \{X^{(1)}, X^{(2)}\}$, 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 (6).
3: Fix the others and update $Z^{(1)}$ using (8).
4: Fix the others and update $Z^{(2)}$ using (9).
5: Fix the others and update S using (10).
6: Fix the others and update $E^{(v)}$ using (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: $Z^{(v)}, E^{(v)}$

As discussed above, our objective function in (2) considers both view-specific reconstruction and the cross-view consistency, which enables us to discover both of the Type 1 and Type 2 outliers by analyzing the learned representation coefficients and noise matrices. In particular, the criterion (13) could detect two types of outliers simultaneously. The first term in (13) measures the inconsistency of the ith sample across two views. From the perspective of data reconstruction, 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 ith sample is a normal sample 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 ith 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 1 outliers. The second term in (13) contributes to identifying the Type 2 outliers. Each column in $E^{(1)}$ and $E^{(2)}$ corresponds to the reconstruction error vectors in view 1 and view 2, respectively. If the ith 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 would not affect the outlier score $o(i)$ too much. However, if the ith sample is a Type 2 outlier, which shows abnormal behavior in both views, the summation in the second term will be increased, and then the outlier score $o(i)$ will be further decreased. Therefore, the proposed criterion (13) is able to detect both Type 1 outliers and Type 2 outliers.

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

$$o(i) = \sum_{p=1}^{V-1} \sum_{q=p+1}^{V} \left(\sum_{k=1}^{n} \left(u_k^{(i)} Z_{ik}^{(p)} Z_{ik}^{(q)} \right) - \lambda \sum_{k=1}^{n} \left(E_{ik}^{(p)} E_{ik}^{(q)} \right) \right), \tag{14}$$

ALGORITHM 2: MLRA for Outlier Detection

Input: Multi-view sample set X, threshold γ

1: Normalize each sample $x_i^{(v)}$,
$x_i^{(v)} = x_i^{(v)}/||x_i^{(v)}||$.
2: Solve objective (5) using Algorithm 1 and obtain
optimal solution Z^v, E^v.
3: Calculate outlier score for each sample using (14).
4: Generate binary label vector L
If $o(i) < \gamma$, $L(i) = 1$; *otherwise,* $L(i) = 0$.

Output: Binary outlier label vector L

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 complete MLRA algorithm is summarized in Algorithm 2.

5 MLRA FOR MULTI-VIEW GROUP OUTLIER DETECTION

In this section, we extend the MLRA framework to a novel application, multi-view group outlier detection. Different from original outlier detection tasks that identify individual abnormal data points, we aim to detect a group of abnormal data points across different views.

5.1 Motivation

In practice, outlier may not only appear as an individual point, but also as a group. For example, a group of people collude to create false product reviews in social media websites [59]. The most challenging part in group anomaly detection is that the outliers appear to be normal at the individual level. Existing works on group anomaly detection mainly deal with the single-view data [41, 56, 59]. In this article, we propose a new problem, multi-view group outlier detection.

The group outlier detection problem becomes more complicated in the multi-view settings. Our assumption is that the dataset contains several groups, and each group is considered as a cluster. In other words, we can observe several clusters in each view. Ideally, the cluster structure in all the views should be consistent. In the multi-view group outlier detection problem, one cluster might be identified as an outlier group, if it exhibits inconsistent behavior in different views.

Formally, an outlier group in the multi-view setting is defined as follows.

Definition 3. An *Outlier Group* is a set of data points that form as a cluster in each view, but show inconsistent behavior across different views.

The group outlier is actually a special case of Type 1 outlier we defined in Section 3. Our goal is to identify such outlier groups from the perspective of data representation.

5.2 Formulation and Algorithm

We extend our MLRA framework for the multi-view group outlier detection problem. As before, we provide detailed derivations and solutions for the two-view case. As the individual outlier points are not considered in this problem, we can simplify (3) as

$$\min_{Z^{(v)}} \sum_{v=1}^{2} (\|Z^{(v)}\|_* + \alpha||X^{(v)} - X^{(v)}Z^{(v)}||_{\mathrm{F}}) + \beta||Z^{(1)} - Z^{(2)}||_{2,1}. \tag{15}$$

In (15), we drop the $l_{2,1}$ constraints on the reconstruction errors, and utilize the Frobenius norm. The reason is that we assume the data only contain group outliers. The group outliers are the Type 1 outliers, which show inconsistent behavior across different views. As discussed in Section 4, the $l_{2,1}$ norm is very suitable to detect the Type 2 outliers. Therefore, it is not very necessary to use $l_{2,1}$ norm in this scenario. However, if the data contain both individual-level outliers and group-level outliers, we suggest using the $l_{2,1}$ norm to model reconstruction errors. The problem (15) can be solved using the same optimization technique described in Section 4.

Using the optimal solutions $Z^{(v)}$, we design a criterion to calculate the outlier score vector o_g. The criterion for group outlier detection is formulated as

$$o_g(i) = \sum_{k=1}^{n} - \left| \left(\frac{U \circ (Z^{(1)\top} Z^{(1)})}{\|Z^{(1)}\|_{\mathrm{F}}^2} - \frac{U \circ (Z^{(2)\top} Z^{(2)})}{\|Z^{(2)}\|_{\mathrm{F}}^2} \right)_{ik} \right|, \tag{16}$$

where $U \in \mathbb{R}^{n \times n}$ is a pairwise cluster membership indicator matrix. If two samples are from the sample cluster, the corresponding element in U will be 1. $|a|$ denotes the absolute value of a, and $A \circ B$ denotes the element-wise product of matrices A and B.

In (16), $(Z^{(v)\top} Z^{(v)}) \in \mathbb{R}^{n \times n}$ is actually an affinity matrix derived from the LRRs, which can be considered a robust estimation of the cluster structure in view v. Particularly, each element in this affinity matrix is the inner product of two LRR vectors. The benefits of employing such an affinity measurement has also been discussed in [24] and [17]. $U \circ (Z^{(v)\top} Z^{(v)})$ means that we only count the block diagonal parts in coefficients matrices $Z^{(v)}$. In practice, U can be obtained by performing spectral clustering on $Z^{(v)}$. We normalize the estimations in each view, and measure the differences in two views. For instance, if each view contains two groups, and $U \circ (Z^{(v)\top} Z^{(v)})$ should have two clear block diagonals. If one group is an outlier group, the corresponding block diagonal part in $|\frac{U \circ (Z^{(1)\top} Z^{(1)})}{\|Z^{(1)}\|_{\mathrm{F}}^2} - \frac{U \circ (Z^{(2)\top} Z^{(2)})}{\|Z^{(2)}\|_{\mathrm{F}}^2}|$ should be enlarged, as this group has inconsistent characteristics in two views.

The sample x_i is marked as a member of the outlier group if the score $o_g(i)$ is smaller than the threshold γ. The MLRA based group outlier detection algorithm is summarized in Algorithm 3.

ALGORITHM 3: MLRA for Group Outlier Detection

Input: Multi-view sample set X, threshold γ

1: Normalize each sample $x_i^{(v)}$,
$x_i^{(v)} = x_i^{(v)} / ||x_i^{(v)}||$.
2: Solve objective (15) and obtain optimal solutions Z^v.
3: Calculate outlier score for each sample using (16).
4: Generate binary label vector L
If $o(i) < \gamma$, $L(i) = 1$; *otherwise,* $L(i) = 0$.

Output: Binary outlier label vector L_g

6 EXPERIMENTS

The performance of our MLRA framework is evaluated on seven UCI datasets [5], MovieLens-1M dataset,[1] the USPS-MNIST dataset [18, 21], and the WebKB dataset [6].[2]

[1] http://grouplens.org/datasets/movielens/.
[2] http://lig-membres.imag.fr/grimal/data.html.

6.1 Baselines and Evaluation Metrics

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) [37]. LRR is a representative outlier detection method for single-view data. Thus, we testify its performance on two views separately.
- —Direct Robust Matrix Factorization (DRMF) [55]. DRMF formulates robust factorization as a matrix approximation problem with constraints on the cardinality of the outlier set.
- —Outlier Pursuit (OP) [58]. OP is able to recover the optimal low-dimensional space and identifies outliers. It is also a single-view method.
- —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 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 article, we use the L-2 distance and Hilbert-Schmidt Independence Criterion (HSIC), as they usually yield better performance than others.

As suggested in [1, 14], we adopt the receiver operating characteristic (ROC) curves as the evaluation metric, which represents the tradeoff 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:

$$FPR = \frac{FP}{FP + TN}, \quad TPR = \frac{TP}{TP + FN}, \tag{17}$$

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

6.2 Synthetic Multi-View Settings on Real Data

UCI Datasets

We employ seven benchmark datasets, namely "Iris," "Letter," "Waveform," "Zoo," "Ionosphere," "Pima," and "Wdbc" from the UCI machine learning repository [5]. To conduct fair comparisons, we follow the sample settings in [1]. Since all the seven datasets are not multi-view datasets, we simulate two views as suggested in [14]. In particular, the feature representations of each dataset are divided into two subsets, where each subset is considered as one view of the data. In order to generate a Type 1 outlier, we take two objects from two different classes and swap the subsets in one view but not in the other. To generate a Type 2 outlier, we randomly select a sample, and replace its features in two views as random values. In total, 15% data are preprocessed and labeled as outliers. Table 2 summarizes the detailed information of all the UCI datasets used in this article.

To illustrate the convergence property of our algorithm, we show in Figure 4(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 4(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 $\{10^{-2}, 10^{-1}, \ldots, 10^{2}\}$. Figure 4(b) shows the AUC 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 4(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

Table 2. Summary of Seven UCI Datasets (n = Number of Samples, m_1 = Number of Type 1 Outliers, m_2 = Number of Type 2 Outliers, d = Number of Dimensions)

Datasets	n	m_1	m_2	d
Iris	150	16	8	4
Letter	1300	130	65	16
Ionosphere	351	36	18	34
Zoo	101	10	5	16
Waveform	1200	120	60	21
Pima	768	76	38	8
Wdbc	569	56	28	30

Table 3. Average AUC Values (±Standard Deviations) on Seven UCI Datasets with only Type 1 Outliers

	Single-View Methods			Multi-View Methods		
Datasets	OP	DRMF	LRR	HOAD	AP	Ours
Iris	0.42±0.08	0.46±0.07	0.50±0.08	0.83±0.06	**0.96±0.03**	0.84±0.02
Letter	0.39±0.05	0.43±0.03	0.49±0.02	0.53±0.04	0.85±0.01	**0.88±0.02**
Ionosphere	0.41±0.06	0.42±0.03	0.46±0.05	0.50±0.06	**0.94±0.03**	0.87±0.03
Zoo	0.49±0.08	0.53±0.08	0.60±0.09	0.55±0.10	**0.91±0.05**	0.90±0.05
Waveform	0.40±0.04	0.45±0.03	0.51±0.03	0.75±0.04	0.62±0.02	**0.77±0.02**
Pima	0.45±0.03	0.48±0.04	0.51±0.04	0.56±0.03	0.67±0.04	**0.74±0.03**
Wdbc	0.47±0.04	0.50±0.05	0.54±0.04	0.45±0.06	0.92±0.03	**0.93±0.01**

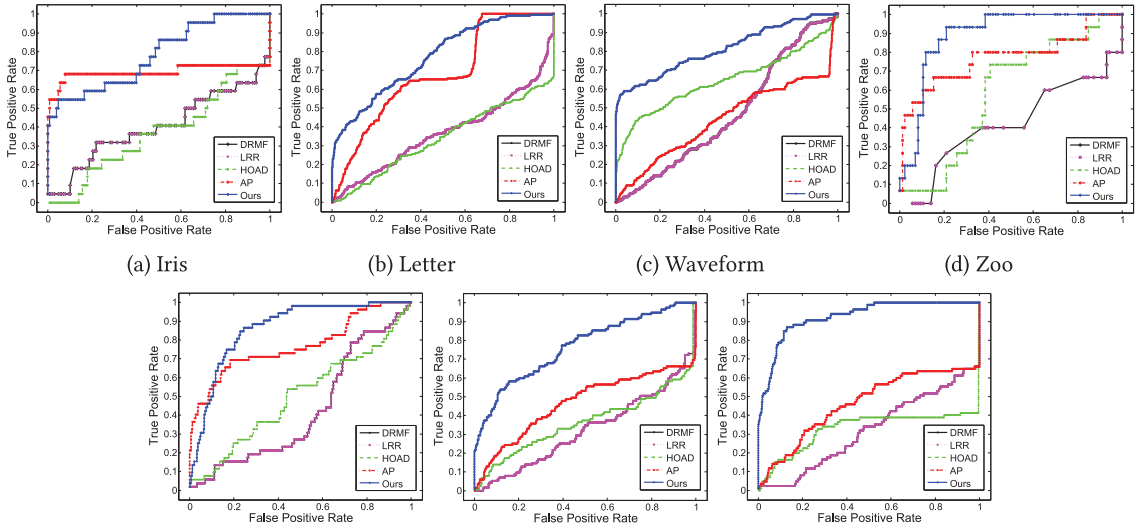
Note: The bold fonts denote the best results on each dataset.

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 1 outliers only; (2) Type 1 and Type 2 outliers. In this way, we can observe the strengths and limitations of different methods.

Table 3 reports the average AUC values (with standard deviations) on seven datasets with Type 1 outliers. From Table 3, we have the following observations. First, the results of single-view method like LRR are much lower than the multi-view methods. Second, the multi-view method AP performs better than single-view methods and HOAD in most cases, and it achieves the best results on the Iris, Ionosphere, and Zoo datasets. Third, 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 1 outliers, and our approach obtains the best results in most cases.

Figure 3 shows the detailed ROC curves of compared methods on UCI datasets. It shows that our approach obtains the best performance in most cases. Table 4 shows the average AUC values on seven datasets with both Type 1 and Type 2 outliers. We can observe from Table 4 that our approach significantly outperforms other competitors in all the cases. In addition, AP still performs better than HOAD in most datasets except waveform. The results demonstrate that our approach can detect two types of outliers simultaneously.



(a) Iris (b) Letter (c) Waveform (d) Zoo

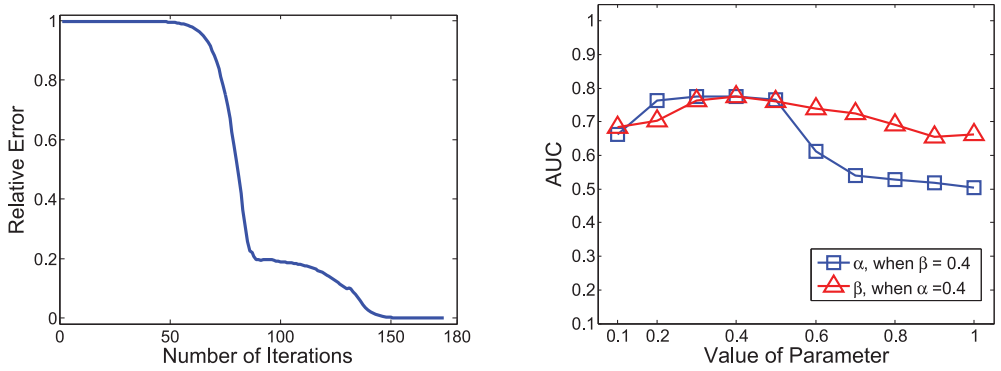
(e) Ionosphere (f) Pima (g) Wdbc

Fig. 3. ROC curves of outlier detection on seven UCI datasets.

Table 4. Average AUC Values (±Standard Deviations) on Seven UCI Datasets with Type 1 and Type 2 Outliers

Datasets	Single-View Methods			Multi-View Methods		
	OP	DRMF	LRR	HOAD	AP	Ours
Iris	0.36±0.05	0.38±0.04	0.39±0.06	0.37±0.04	0.70±0.02	**0.84±0.05**
Letter	0.32±0.02	0.34±0.02	0.34±0.01	0.34±0.01	0.67±0.01	**0.78±0.01**
Ionosphere	0.39±0.03	0.46±0.03	0.43±0.04	0.50±0.05	0.76±0.02	**0.79±0.03**
Zoo	0.35±0.06	0.37±0.04	0.41±0.06	0.58±0.07	0.77±0.07	**0.85±0.04**
Waveform	0.40±0.02	0.43±0.03	0.42±0.02	0.77±0.03	0.42±0.01	**0.83±0.02**
Pima	0.32±0.03	0.35±0.02	0.34±0.02	0.37±0.02	0.46±0.02	**0.77±0.03**
Wdbc	0.31±0.02	0.33±0.02	0.33±0.03	0.33±0.07	0.48±0.03	**0.79±0.01**

Note: The bold fonts denote the best results on each dataset.



(a) (b)

Fig. 4. (a) Convergence curve of our algorithm on Iris dataset. (b) AUC of our approach on Pima dataset by varying the values of α and β.

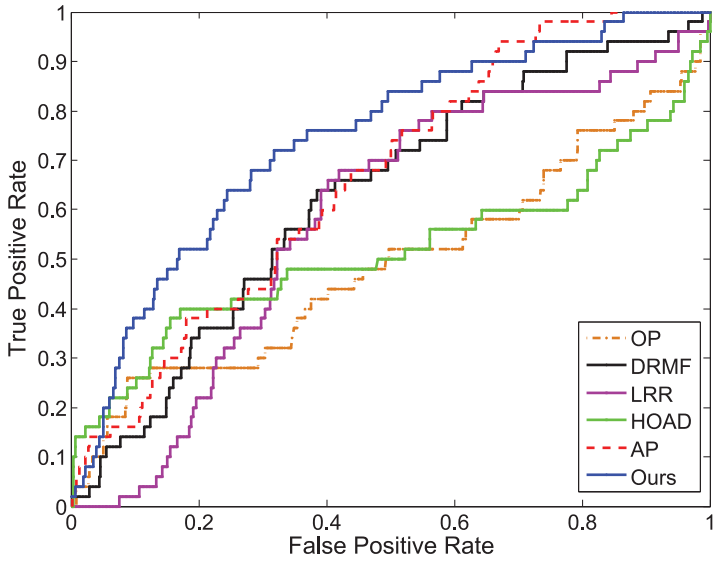


Fig. 5. ROC Curves of all compared methods on two-view USPS dataset.

Table 5. Average AUC Values with Standard Deviations of Compared Methods on two-View USPS Dataset

Method	AUC (±Standard Deviation)
OP [58]	0.4892±0.0746
DRMF [55]	0.6412±0.0673
LRR [37]	0.5960±0.0461
HOAD [14]	0.5193±0.0429
AP [1]	0.6745±0.0848
Ours	**0.7381±0.0702**

Note: The bold fonts denote the best results on each dataset.

USPS Digit Dataset

We construct a two-view dataset by using the USPS dataset [18], which contains 9,298 handwritten digit images. We extract two types of features from each image as two data views, including pixel values and Fourier coefficients.

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 1 outliers and 5% Type 2 outliers. This process was repeated 20 times, and we evaluated the performance of each method on these 20 sample sets.

Figure 5 shows the ROC curves, and Table 5 lists the average AUC values with standard deviations. For single-view outlier detection methods OP, DRMF, and LRR, we simply concatenate the features from two views together as inputs. From Figure 5 and Table 5, we observe that the single-view methods DRMF and LRR attain even better performance than the multi-view method HOAD. Since DRMF and LRR are low-rank based methods, they are capable of detecting the Type 2 outliers. Moreover, as our approach can detect two types of outliers effectively, it outperforms all of the single-view and multi-view outlier detection baselines.

Table 6. Average AUC Values with Standard Deviations of Compared Methods on WebKB Dataset

Method	AUC (±Standard Deviation)
OP [58]	0.4219±0.0611
DRMF [55]	0.4624±0.0603
LRR [37]	0.4805±0.0530
HOAD [14]	0.5027±0.0643
AP [1]	0.4965±0.0655
Ours	**0.5532±0.0475**

Note: The bold fonts denote the best results on each dataset.

6.3 Real-World Multi-View Data with Synthetic Outliers

The WebKB dataset [6] has been widely used for evaluating multi-view learning algorithms [15, 29]. It contains webpages collected from four universities, including *Cornell, Texas, Washington, and Wisconsin*. The webpages can be categorized into five classes: student, course, project, faculty, and staff. Each webpage is described by two views, the *content* view and the *citation* view. In the content view, each webpage is represented by a word vector of length 1,703. The citation view characterizes the number of citation links across pages. In our experiments, we use the *Cornell* subset, which contains 195 webpages. We follow the procedures described in Section 6.2 to generate two types of outliers, and evaluate the performance of our approach and baselines. Table 6 shows the average AUC with standard deviations of all compared methods on the WebKB dataset. In addition, the improvement of our approach is not very significant. We can observe that the TPR of all compared methods are very low, as it is a challenging task in real world. In addition, the improvement of our approach is not very significant. Clearly, our MLRA approach achieves higher AUC than its competitors.

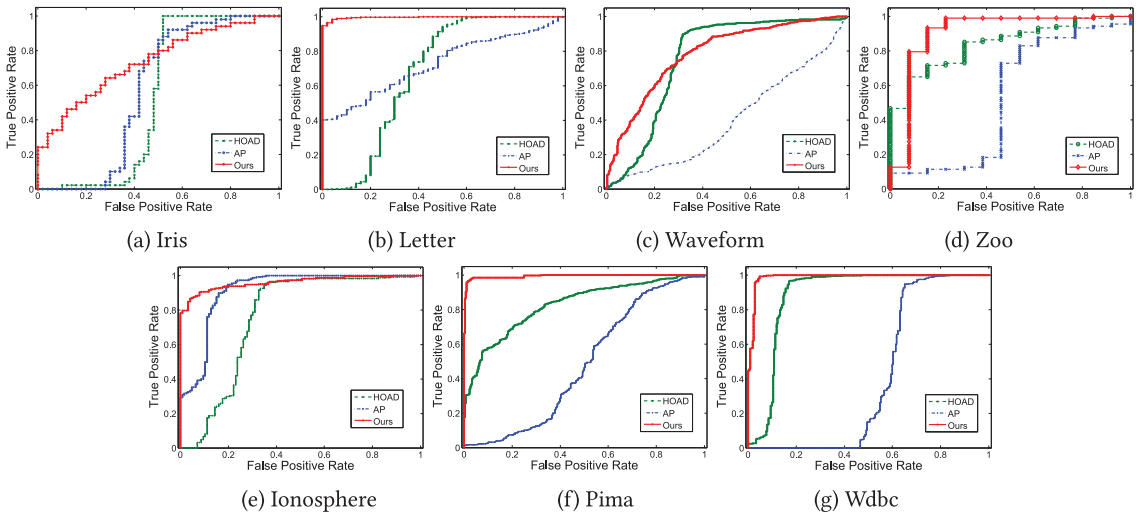
6.4 Real-World Multi-View Data with Real Outliers

We employ the popular MovieLens-1M dataset, which contains 1 million ratings for 3,883 movies by 6,040 users. We consider the movies as samples, and exploit two perspectives of movies: (1) Genre. There are 18 genres in this dataset, such as Action, Comedy, and Horror. Each movie was classified as one or more of these 18 genres, which can be converted as a binary vector. (2) User's feedback. Each movie was rated by one or more users, which can also be represented as binary vectors (across all users). As the ground truth information of outliers in this real-world dataset are unknown, we mainly perform qualitative analysis to show the performance of our approach in detecting outliers.

We sample 500 movies and 600 users from the dataset, and perform our MLRA approach to assign an outlier score to each movie. Table 7 shows some movies with high outlier scores and low outlier scores. The movie "Quiz Show" belongs to the "Drama" genre. It was considered as an outlier as it receives much more ratings than other movies in the "Drama" genre. In other words, this movie exhibits inconsistent behavior in the genre view and the rating view. On the other hand, the movies "Toy Story" and "Jumanji" are categorized to three different genres, and they share the same genre of "Children's." Meanwhile, both of them received a large number of ratings, as many other movies belonging to the same genre. Therefore, they have very low outlier scores, and can be labeled as normal samples. In a word, the qualitative analysis on the MovieLens-1M shows that our approach is able to produce meaningful outlier detection results on real-world data.

Table 7. Movies with High and Low Outlier Scores Calculated by the Proposed Approach

Movie Title	Score	Movie Title	Score
Quiz Show	0.98	Wings of Courage	0.15
Dumb & Dumber	0.96	Balto	0.13
Forget Paris	0.95	GoldenEye	0.09
While You Were Sleeping	0.95	Jumanji	0.07
Speed	0.93	Toy Story	0.02



(a) Iris (b) Letter (c) Waveform (d) Zoo

(e) Ionosphere (f) Pima (g) Wdbc

Fig. 6. ROC curves of group outlier detection on seven UCI datasets.

6.5 Group Outlier Detection

In this experiment, we evaluate the performance of our method on detecting group outliers. We use seven UCI datasets and the USPS digit dataset in the experiments. First, as described in previous sections, the feature vectors of each dataset are divided into two subsets to construct the multi-view dataset. Each subset is considered as one view of the data. In order to generate an outlier group, we randomly select a class and a view, remove the data in this class and the specific view, and fill in random vectors drawn from a multivariate standard normal distribution. We generate one outlier group for each dataset. In this way, we can ensure that the outlier group has a cluster structure, but they are quite different from the normal samples.

As multi-view methods usually perform better than the single-view ones, we compare our MLRA framework with two multi-view outlier detection methods, HOAD and AP, in this section. In addition to the group outlier detection model named as MLRA-Group, we also evaluate the performance of our individual-level outlier detection model (as described in Section 4), which is named as MLRA-Individual. Figure 6 shows the ROC curves on seven UCI datasets, and Table 8 shows the average AUC results. We can observe that (1) AP performs better than HOAD on the Iris, Letter and Ionosphere datasets, but HOAD performs much better than AP on the rest datasets; (2) both of our MLRA-Individual and MLRA-Group methods outperform HOAD and AP on all the datasets. As MLRA-Individual does not consider the group prior information, it obtains lower AUC values

Table 8. Average AUC Values on UCI and USPS Datasets with Group Outliers

Datasets	HOAD	AP	MLRA-Individual	MLRA-Group
Iris	0.5324	0.5720	0.6832	**0.7372**
Letter	0.6749	0.7247	0.7582	**0.9963**
Ionosphere	0.7563	0.9127	0.9235	**0.9606**
Zoo	0.8400	0.5271	0.8522	**0.9047**
Waveform	0.7568	0.3979	0.7629	**0.7846**
Pima	0.8274	0.5231	0.8732	**0.9934**
Wdbc	0.8864	0.4096	0.9129	**0.9870**
USPS	0.5422	0.7653	0.9015	**0.9244**

Note: The bold fonts denote the best results on each dataset.

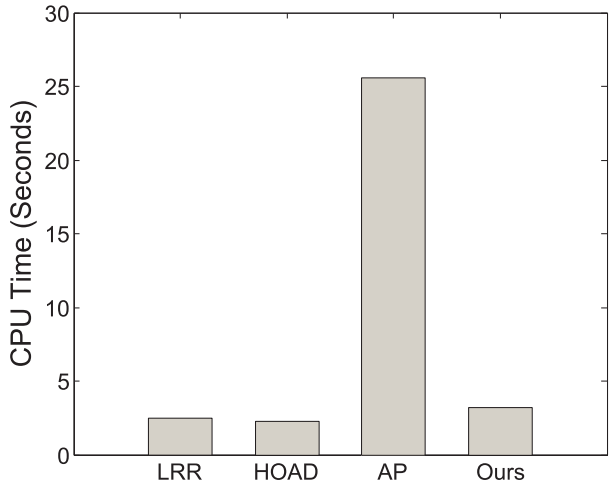


Fig. 7. CPU time (seconds) of all compared methods on UCI-Letter dataset.

than MLRA-Group. In particular, the AUC of MLRA-Group is quite close to 1.0 on the Letter, Pima and Wdbc datasets, which demonstrates the strength of our framework.[3]

6.6 Discussions

We evaluate the computational cost of different methods on the Letter dataset. The machine used in our experiments installs 24GB RAM and Intel Xeon W3350 CPU. Figure 7 shows the average running time over 50 runs of each compared method. We can observe that AP took much more computing time than other methods, due to its AP procedure. LRR, HOAD, and our approach have similar computational costs, as they are all matrix factorization based methods with similar time complexities.

To perform an in-depth analysis of the outlier detection results, Figure 8 shows the number of detected outliers for each type on the USPS-MNIST dataset, when the FPR is equal to 0.8. It shows that two multi-view methods, HOAD and AP, are only capable of detecting Type 1 outliers. However, our approach is able to detect two types of outliers effectively.

[3]AUC=1.0 implies that the outlier detector is perfect.

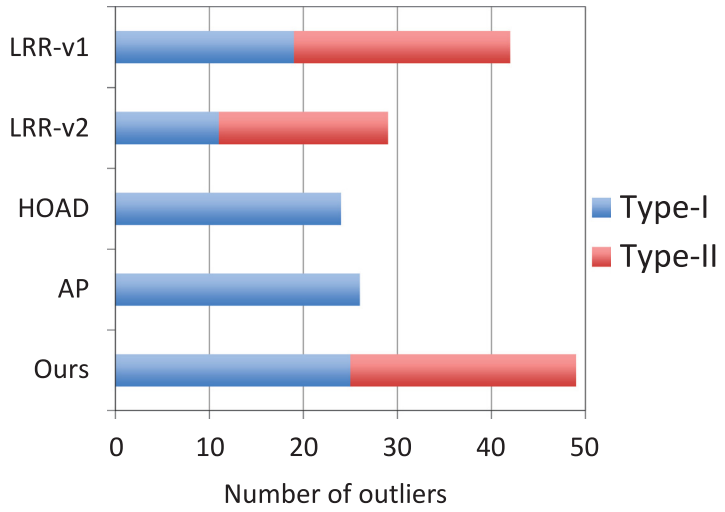


Fig. 8. Number of detected outliers (two types) when $FPR = 0.8$ on USPS-MINST dataset.

7 CONCLUSIONS

We have proposed a MLRA framework in this article for outlier detection. Our framework performed cross-view low-rank analysis, and employed a well designed criterion to calculate the outlier score for each sample. We formulated it as a rank-minimization problem, and adopted the Inexact ALM algorithm to solve it. By analyzing the representation coefficients in different views, our framework was able to detect two different types of outliers simultaneously. Moreover, MLRA has been extended to multi-view group outlier detection. Experimental results on seven UCI datasets, USPS-MNIST, MovieLens, and WebKB datasets showed 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 1 and Type 2 outliers, our approach can significantly boost the performance of outlier detection.

In our future work, we will apply MLRA framework to more outlier detection applications, and we would also like to develop a divide-and-conquer version of MLRA to make it more suitable for large scale datasets and further improve its performance.

REFERENCES

[1] Alejandro Marcos Alvarez, Makoto Yamada, Akisato Kimura, and Tomoharu Iwata. 2013. Clustering-based anomaly detection in multi-view data. In *CIKM*. 1545–1548.
[2] Fabrizio Angiulli and Fabio Fassetti. 2009. Outlier detection using inductive logic programming. In *ICDM*. 693–698.
[3] Ira Assent, Xuan Hong Dang, Barbora Micenková, and Raymond T. Ng. 2013. Outlier detection with space transformation and spectral analysis. In *SDM*. 225–233.
[4] F. R. Bach. 2008. Consistency of trace norm minimization. *Journal of Machine Learning Research* 9 (2008), 1019–1048.
[5] K. Bache and M. Lichman. 2013. UCI Machine Learning Repository. (2013). Retrieved from http://archive.ics.uci.edu/ml.
[6] Avrim Blum and Tom M. Mitchell. 1998. Combining labeled and unlabeled data with co-training. In *COLT*. ACM, 92–100.
[7] J. F. Cai, E. J. Candes, and Z. W. Shen. 2010. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization* 20, 4 (2010), 1956–1982.
[8] E. J. Candès, X. D. Li, Y. Ma, and J. Wright. 2011. Robust principal component analysis?*Journal of ACM* 58, 3 (2011), 11.
[9] Jianhui Chen, Jiayu Zhou, and Jieping Ye. 2011. Integrating low-rank and group-sparse structures for robust multi-task learning. In *KDD*. 42–50.

[10] Bin Cheng, Guangcan Liu, Jingdong Wang, ZhongYang Huang, and Shuicheng Yan. 2011. Multi-task low-rank affinity pursuit for image segmentation. In *ICCV*. 2439–2446.
[11] Santanu Das, Bryan L. Matthews, Ashok N. Srivastava, and Nikunj C. Oza. 2010. Multiple kernel learning for heterogeneous anomaly detection: Algorithm and aviation safety case study. In *KDD*. 47–56.
[12] Bo Du and Liangpei Zhang. 2014. A discriminative metric learning based anomaly detection method. *IEEE Transactions on Geoscience and Remote Sensing* 52, 11 (2014), 6844–6857. DOI: http://dx.doi.org/10.1109/TGRS.2014.2303895
[13] Andrew F. Emmott, Shubhomoy Das, Thomas Dietterich, Alan Fern, and Weng-Keen Wong. 2013. Systematic construction of anomaly detection benchmarks from real data. In *KDD Workshop on Outlier Detection and Description*. 16–21.
[14] Jing Gao, Wei Fan, Deepak S. Turaga, Srinivasan Parthasarathy, and Jiawei Han. 2011. A spectral framework for detecting inconsistency across multi-source object relationships. In *ICDM*. 1050–1055.
[15] Yuhong Guo. 2013. Convex subspace representation learning from multi-view data. In *AAAI*. Vol. 1, 2.
[16] Ko-Jen Hsiao, Kevin S. Xu, Jeff Calder, and Alfred O. Hero III. 2012. Multi-criteria anomaly detection using pareto depth analysis. In *NIPS*. 854–862.
[17] Han Hu, Zhouchen Lin, Jianjiang Feng, and Jie Zhou. 2014. Smooth representation clustering. In *CVPR*. 3834–3841.
[18] Jonathan Hull. 1994. A database for handwritten text recognition research. *IEEE Transactions on Pattern Analysis and Machine* 16, 5 (1994), 550–554.
[19] Vandana Pursnani Janeja and Revathi Palanisamy. 2013. Multi-domain anomaly detection in spatial datasets. *Knowledge and Information Systems* 36, 3 (2013), 749–788.
[20] R. H. Keshavan, A. Montanari, and S. Oh. 2009. Matrix completion from noisy entries. In *NIPS*. 952–960.
[21] Yann LeCun, Leon Bottou, Yoshua Bengio, and Patrick Haaffner. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86, 11 (1998), 2278–2324.
[22] Yuh-Jye Lee, Yi-Ren Yeh, and Yu-Chiang Frank Wang. 2013. Anomaly detection via online oversampling principal component analysis. *IEEE Transactions on Knowledge and Data Engineering* 25, 7 (2013), 1460–1470. DOI: http://dx.doi.org/10.1109/TKDE.2012.99
[23] Liangyue Li, Sheng Li, and Yun Fu. 2014. Learning low-rank and discriminative dictionary for image classification. *Image and Vision Computing* 32, 10 (2014), 814–823. DOI: http://dx.doi.org/10.1016/j.imavis.2014.02.007
[24] Sheng Li and Yun Fu. 2013. Low-rank coding with b-matching constraint for semi-supervised classification. In *IJCAI*. 1472–1478.
[25] Sheng Li and Yun Fu. 2014. Robust subspace discovery through supervised low-rank constraints. In *SDM*. 163–171. DOI: http://dx.doi.org/10.1137/1.9781611973440.19
[26] Sheng Li and Yun Fu. 2015. Multi-view low-rank analysis for outlier detection. In *SDM*.
[27] Sheng Li and Yun Fu. 2017. *Robust Representation for Data Analytics*. Springer.
[28] Sheng Li, Ming Shao, and Yun Fu. 2014. Locality linear fitting one-class SVM with low-rank constraints for outlier detection. In *IJCNN*. 676–683. DOI: http://dx.doi.org/10.1109/IJCNN.2014.6889446
[29] Shao-Yuan Li, Yuan Jiang, and Zhi-Hua Zhou. 2014. Partial multi-view clustering. In *AAAI*. Citeseer, 1968–1974.
[30] Z. C. Lin, M. M. Chen, L. Q. Wu, and Y. Ma. 2009. *The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices*. Technical Report, University of Illinois at Urbana-Champaign.
[31] Alexander Liu and Dung N. Lam. 2012. Using consensus clustering for multi-view anomaly detection. In *IEEE Symposium on Security and Privacy Workshops*. 117–124.
[32] Bo Liu, Yanshan Xiao, Longbing Cao, Zhifeng Hao, and Feiqi Deng. 2013. SVDD-based outlier detection on uncertain data. *Knowledge and Information Systems* 34, 3 (2013), 597–618. DOI: http://dx.doi.org/10.1007/s10115-012-0484-y
[33] Bo Liu, Yanshan Xiao, Philip S. Yu, Zhifeng Hao, and Longbing Cao. 2014. An efficient approach for outlier detection with imperfect data labels. *IEEE Transactions on Knowledge and Data Engineering* 26, 7 (2014), 1602–1616. DOI: http://dx.doi.org/10.1109/TKDE.2013.108
[34] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. 2012. Isolation-based anomaly detection. *TKDD* 6, 1 (2012), 3. DOI: http://dx.doi.org/10.1145/2133360.2133363
[35] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma. 2013. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine* 35, 1 (2013), 171–184.
[36] Guangcan Liu, Qingshan Liu, and Ping Li. 2017. Blessing of dimensionality: Recovering mixture data via dictionary pursuit. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 1 (2017), 47–60.
[37] Guangcan Liu, Huan Xu, Jinhui Tang, Qingshan Liu, and Shuicheng Yan. 2016. A deterministic analysis for LRR. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38, 3 (2016), 417–430.
[38] Guangcan Liu, Huan Xu, and Shuicheng Yan. 2012. Exact subspace segmentation and outlier detection by low-rank representation. In *AISTATS*. 703–711.
[39] G. C. Liu, Z. C. Lin, and Y. Yu. 2010. Robust subspace segmentation by low-rank representation. In *ICML*. 663–670.
[40] Roland Memisevic. 2012. On multi-view feature learning. In *ICML*.

[41] Krikamol Muandet and Bernhard Schölkopf. 2013. One-class support measure machines for group anomaly detection. In *UAI*. DOI: https://dslpitt.org/uai/displayArticleDetails.jsp?mmnu=1&smnu=2&article_id=2406&proceeding_id=29.
[42] Emmanuel Müller, Ira Assent, Patricia Iglesias Sanchez, Yvonne Mülle, and Klemens Böhm. 2012. Outlier ranking via subspace analysis in multiple views of the data. In *ICDM*. 529–538.
[43] Colin O'Reilly, Alexander Gluhak, and Muhammad Ali Imran. 2015. Adaptive anomaly detection with kernel eigenspace splitting and merging. *IEEE Transactions on Knowledge and Data Engineering* 27, 1 (2015), 3–16. DOI: http://dx.doi.org/10.1109/TKDE.2014.2324594
[44] Yaling Pei, Osmar R. Zaïane, and Yong Gao. 2006. An efficient reference-based approach to outlier detection in large datasets. In *ICDM*. 478–487.
[45] Bryan Perozzi, Leman Akoglu, Patricia Iglesias Sanchez, and Emmanuel Müller. 2014. Focused clustering and outlier detection in large attributed graphs. In *KDD*. 1346–1355. DOI: http://dx.doi.org/10.1145/2623330.2623682
[46] Ninh Pham and Rasmus Pagh. 2012. A near-linear time approximation algorithm for angle-based outlier detection in high-dimensional data. In *KDD*. 877–885.
[47] Erich Schubert, Arthur Zimek, and Hans-Peter Kriegel. 2014. Generalized outlier detection with flexible kernel density estimates. In *SDM*. 542–550. DOI: http://dx.doi.org/10.1137/1.9781611973440.63
[48] Ming Shao, Dmitry Kit, and Yun Fu. 2014. Generalized transfer subspace learning through low-rank constraint. *International Journal of Computer Vision* 109, 1–2 (2014), 74–93. DOI: http://dx.doi.org/10.1007/s11263-014-0696-6
[49] Vikas Sindhwani and David S. Rosenberg. 2008. An RKHS for multi-view learning and manifold co-regularization. In *ICML*. 976–983.
[50] Karthik Sridharan and Sham M. Kakade. 2008. An information theoretic framework for multi-view learning. In *COLT*. 403–414.
[51] Hanghang Tong and Ching-Yung Lin. 2011. Non-negative residual matrix factorization with application to graph anomaly detection. In *SDM*. 143–153. DOI: http://dx.doi.org/10.1137/1.9781611972818.13
[52] Grigorios Tzortzis and Aristidis Likas. 2012. Kernel-based weighted multi-view clustering. In *ICDM*. 675–684.
[53] Martha White, Yaoliang Yu, Xinhua Zhang, and Dale Schuurmans. 2012. Convex multi-view subspace learning. In *NIPS*. 1682–1690.
[54] Shu Wu and Shengrui Wang. 2013. Information-theoretic outlier detection for large-scale categorical data. *IEEE Transactions on Knowledge and Data Engineering* 25, 3 (2013), 589–602.
[55] Liang Xiong, Xi Chen, and Jeff Schneider. 2011. Direct robust matrix factorization for anomaly detection. In *ICDM*. IEEE, 844–853.
[56] Liang Xiong, Barnabás Póczos, and Jeff G. Schneider. 2011. Group anomaly detection using flexible genre models. In *NIPS*. 1071–1079.
[57] Chang Xu, Dacheng Tao, and Chao Xu. 2013. A survey on multi-view learning. *CoRR* abs/1304.5634 (2013).
[58] Huan Xu, Constantine Caramanis, and Sujay Sanghavi. 2010. Robust PCA via outlier pursuit. In *NIPS*. 2496–2504.
[59] Qi Rose Yu, Xinran He, and Yan Liu. 2014. GLAD: Group anomaly detection in social media analysis. In *KDD*. 372–381. DOI: http://dx.doi.org/10.1145/2623330.2623719
[60] Xiaowei Zhou, Can Yang, and Weichuan Yu. 2012. Automatic mitral leaflet tracking in echocardiography by outlier detection in the low-rank representation. In *CVPR*. 972–979.
[61] Arthur Zimek, Matthew Gaudet, Ricardo J. G. B. Campello, and Jörg Sander. 2013. Subsampling for efficient and effective unsupervised outlier detection ensembles. In *KDD*. 428–436.

Received September 2016; revised April 2017; accepted November 2017