

Anomaly Detection in Transportation Corridors using Manifold Embedding

Amrudin Agovic Arindam Banerjee

Dept of Computer Science & Engineering
University of Minnesota, Twin Cities
{aagovic,banerjee}@cs.umn.edu

Auroop Ganguly Vladimir Protopopescu

Computational Sciences & Engineering
Oak Ridge National Laboratory
{gangulayar,protopopesva}@ornl.gov

ABSTRACT

The formation of secure transportation corridors, where cargoes and shipments from points of entry can be dispatched safely to highly sensitive and secure locations, is a high national priority. One of the key tasks of the program is the detection of anomalous cargo based on sensor readings in truck weigh stations. Due to the high variability, dimensionality, and/or noise content of sensor data in transportation corridors, appropriate feature representation is crucial to the success of anomaly detection methods in this domain. In this paper, we empirically investigate the usefulness of manifold embedding methods for feature representation in anomaly detection problems in the domain of transportation corridors. We focus on both linear methods, such as multi-dimensional scaling (MDS), as well as nonlinear methods, such as locally linear embedding (LLE) and isometric feature mapping (ISOMAP). Our study indicates that such embedding methods provide a natural mechanism for keeping anomalous points away from the dense/normal regions in the embedding of the data. We illustrate the efficacy of manifold embedding methods for anomaly detection through experiments on simulated data as well as real truck data from weigh stations.

1. INTRODUCTION

Anomaly detection has remained one of the most difficult tasks in data mining due to the inherent difficulty in precisely defining and quantifying the notion of anomaly. Unlike other data mining tasks such as classification, clustering, and association analysis, anomaly detection has to be typically customized to the application domain, since its definition is domain-dependent. Nevertheless, with several emerging application domains (particularly in the realm of national and homeland security) that rely heavily on anomaly detection, the need for a careful study of such methods is more urgent than ever before.

Over the past several years, significant research effort has

gone in the design of anomaly detection methods that are appropriate in unsupervised [3], semi-supervised [20][32], and fully supervised [31][22] settings. A comprehensive survey of existing methods can be found in [30]. Several of these methods implicitly assume that the input data has a representation appropriate for anomaly detection. In reality, there is little or no control over the features, and the features, in their original representation, may not be appropriate for the anomaly detection algorithm. Thus, a natural question to ask is: Can the existing anomaly detection methods benefit from feature extraction? Unlike the classification/prediction literature, where feature extraction is the norm, few results exist on appropriate feature representation for anomaly detection problems. In this paper, we study the application of manifold embedding methods for feature representation in anomaly detection problems, specifically in the context of secure transportation corridors, where the goal is to be able to safely dispatch cargoes and shipments from points of entry to highly sensitive and secure locations. The formation of transportation corridors is a high national priority, and, at this stage, one of the key tasks of the program is to be able to detect anomalous cargo based on sensor readings in truck weigh stations. While (nonlinear) manifold embedding methods have been around for over a decade now [4][24], the novelty of our study stems from the fact that the (nonlinear) embedding methods have not been applied to the emerging anomaly detection problems in the domain of transportation corridors.

We argue that manifold embeddings, particularly the more recent non-linear approaches such as ISOMAP [4] and LLE [24], are surprisingly natural for effective representations of data for anomaly detection purposes. In proximity based anomaly detection methods [18][6][17], one typically makes use of the intuition that normal points are usually close to and anomalous points are far away from the “normal” points. In particular, one often looks at what fraction of a point’s k -nearest neighbors view the point under consideration to be their k -nearest neighbor [6]. On the other hand, ISOMAP approximates geodesic distances using the k -nearest neighbors. Outlier points will have larger geodesic distances to all other points, and hence, will lie far away from the normal points in the manifold embedding. An intuition for the usefulness of LLE embedding can be similarly obtained. Thus, manifold embeddings can result in an appropriate representation of the data, which ensures outliers to stay away from the normal points. Through extensive experimentation, we illustrate that this simple intuition is

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

KDD 2007, San Jose, CA

Copyright 200X ACM X-XXXXX-XX-X/XX/XX ...\$5.00.

useful in practice, and has the potential of significantly improving the effectiveness of existing methods for anomaly detection.

Anomaly detection problems are typically not solved in a fully automated way, and often involve human experts in the loop. In fact, it is often desirable to be able to tune the performance of an anomaly detection method based on the false-positive and false-negative rates. In light of such desiderata, finding appropriate low-dimensional feature representations using manifold embedding (i) can give valuable information about the structure of the data, which can be used for visualization and tuning purposes, and (ii) can be directly fed into any standard anomaly detection method, such as thresholded Parzen window density estimators [11] and one-class support vector machines [25, 28]. In fact, visualization of the embedding can give valuable clues about potential anomalies in an off-the-shelf anomaly detection method may not detect.

The main contribution of the paper is the application of manifold embedding based anomaly detection methodology to transportation corridors. We apply this methodology to real weigh station sensor data collected over several months and uncover important structure in such data, including potential anomalous behavior as well as group structure among normal trucks. The domain specific insights obtained from the analysis is proving valuable for planning the next steps of the project.

The rest of the paper is organized as follows. In Section 2 we review linear and nonlinear manifold embedding methods, and discuss the rationale behind anomaly detection with manifold embedding. We present experimental results on simulated datasets in Section 3. Section 4 discusses the anomaly detection problem in transportation corridor and presents experimental results on real life data. A summary of the results is given in section Section 5.

2. ANOMALY ANALYSIS USING MANIFOLD EMBEDDING

In this section, we review three widely used manifold embedding methods, viz, multidimensional scaling (MDS), locally linear embedding (LLE), and isometric feature mapping (ISOMAP), and discuss their applicability for anomaly detection purposes. In particular, we argue that while the methods were originally designed for obtaining low dimensional manifold structure of high-dimensional observations, all the methods provide a natural way to ensure that points that are away from the dense regions in the high-dimensional observations stay away from the dense regions even in the embedding. As a result, such embedding methods could be very useful for anomaly detection.

For the purpose of our discussion of embedding methods, we consider a set of high-dimensional observations $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, where $\mathbf{x}_i \in \mathbb{R}^d, i = 1, \dots, n$. Further, we assume that the data is centered at the origin. The primary goal of manifold embedding methods is to compute n corresponding data points $\psi_i \in \mathbb{R}^m$ where $m < d$, while preserving important “structure” in the data. The structure to be preserved determines to a certain extent the choice of the dimensionality reduction approach. We consider three popular methods—MDS, LLE, and ISOMAP. MDS is a linear embedding method which has been studied over several decades now, whereas LLE and ISOMAP are more recent nonlinear

embedding methods. Choosing an appropriate target dimension m can be a challenge. In our case we restrict m such that the results can be visualized.

2.1 Metric Multidimensional Scaling (MDS)

Given a $n \times n$ dissimilarity matrix D and a distance measure, the goal of MDS is to perform dimensionality reduction in a way that will preserve dot products between data points as closely as possible [5]. We consider a particular form of MDS called classical scaling. In classical scaling, the Euclidean distance measure is used and the following objective function is minimized:

$$E_{MDS} = \sum_{i,j,i \neq j} (\mathbf{x}_i^T \mathbf{x}_j - \psi_i^T \psi_j)^2 = \sum_{i,j,i \neq j} D_{ij}^2. \quad (1)$$

The first step of the method is to construct the Gram matrix XX^T from D . This can be accomplished by double-centering D^2 [1]:

$$\mathbf{x}_i^T \mathbf{x}_j = -\frac{1}{2} [D_{ij}^2 - D_{i.}^2 - D_{.j}^2 + D_{..}^2], \quad (2)$$

where

$$D_{i.}^2 = \frac{1}{n} \sum_{a=1}^n D_{ia}^2, \quad D_{.j}^2 = \frac{1}{n} \sum_{b=1}^n D_{bj}^2, \quad D_{..}^2 = \frac{1}{n^2} \sum_{c=1}^n \sum_{d=1}^n D_{cd}^2.$$

The minimizer of the objective function is computed from the spectral decomposition of the Gram matrix. Let V denote the matrix formed with the first m eigenvectors of $X^T X$ with corresponding eigenvalue matrix Λ that has positive diagonal entries $\{\lambda_i\}_{i=1}^m$. The projected data point in the lower dimensional space are the rows of $V\sqrt{\Lambda}$, i.e.,

$$\sqrt{\Lambda}V^T = [\psi_1 \dots \psi_n].$$

The output of classical scaling maximizes the variance in the data set while reducing dimensionality. Distances that are far apart in the original data set will tend to be far apart in the projected data set. Since Euclidean distances are used, the output of the above algorithm is equivalent to the output of PCA [15, 29, 1]. However, other variants of metric MDS are also possible where, for example, non-Euclidean distance measures or different objective functions are used. Further, in recent years, PCA has been extended to work with exponential family distributions [8] and their corresponding Bregman divergences [13, 2]. In recent literature, [21] has looked at applying distributed nonlinear PCA for visualization of network data, and have suggested the possibility of using such a methodology for anomaly detection.

2.2 Locally Linear Embedding (LLE)

LLE is a graph-based dimensionality reduction method which attempts to preserve the local linear structure [24]. Given a graph, LLE linearly approximates each point on the manifold with its closest neighbors. This is done by solving a least squares regression problem on local neighborhoods. A lower dimensional representation is obtained by reconstructing each point based on its neighbors.

The first step of LLE is to compute k -nearest neighbors. In the second step a weight matrix W is computed which allows the representation of each point as a linear combination of its neighbors. LLE treats each point as being sampled from a locally linear region. The weight matrix W is

computed by minimizing the reconstruction error

$$E_W = \sum_i \|\mathbf{x}_i - \sum_j W_{ij} \mathbf{x}_j\|^2 \quad (3)$$

subject to the constraints: $W_{ij} = 0$ when \mathbf{x}_i and \mathbf{x}_j are not neighbors, and $\sum_j W_{ij} = 1$.

The last step of the LLE algorithm is to compute a lower dimensional representation of the data. The output of the algorithm is obtained by minimizing:

$$E_\Psi = \sum_i \|\psi_i - \sum_j W_{ij} \psi_j\|^2 \quad (4)$$

subject to the constraints: $\sum_{i=1}^n \psi_i = 0$ and $\Psi^T \Psi = I_{m \times m}$.

While the computation of W is carried out locally, the reconstruction of the points is computed globally in one step. As a result, data points with overlapping neighborhoods are coupled. This way LLE can uncover global structure as well. The constraints on the optimization problems in the last two steps force the embedding to be scale and rotation invariant. LLE is a widely used method that has been successfully used on certain applications [24, 27] and has motivated several methods including supervised [9] and semi-supervised [33] extensions, as well as other embedding methods [10]. One can use it to uncover nonlinearities which cannot be detected with MDS. LLE has been used in conjunction with k-means in [16] to detect anomalies in hyperspectral images. Unlike [16] we use dimensionality reduction alone for preprocessing of features and apply it to transportation corridors.

2.3 Isometric Feature Mapping (ISOMAP)

ISOMAP is another graph-based embedding method [26, 4]. The idea behind ISOMAP is to embed points by preserving geodesic distances between data points. The method attempts to preserve the global structure in the data as closely as possible. Given a graph, geodesic distances are measured in terms of shortest paths between points. Once geodesic distances are computed MDS is used to obtain an embedding.

The algorithm consists of three steps. The first step is to construct a graph by computing k -nearest neighbors. In the second step, one computes pairwise distances D_{ij} between any two points. This can be done using Dijkstra's shortest path algorithm. The last step of ISOMAP is to run the metric MDS algorithm with D_{ij} as input. The resulting embedding will give $\|\psi_i - \psi_j\|^2$ approximately equal to D_{ij}^2 for any two points. By using a local neighborhood graph and geodesic distances, the ISOMAP method exploits both local and global information. In practice, this method works fairly well on a range of problems. One could prove [4] that as the density of data points is increased the graph distances converge to the geodesic distances. ISOMAP has been used in a wide variety of applications [19, 23], and has motivated several extensions in the recent past [34, 14, 33].

The success of anomaly detection methods often depends on the representation of the data. Existence of irrelevant features makes it hard to understand the true structure of the data, and can make the task of anomaly detection significantly harder. In a general sense, manifold embedding methods uncover the true structure in the data in low dimensions. A low dimensional representation is often desirable since (i) one can do visualization based analysis, which

can be very effective in practice, and (ii) certain anomaly detection techniques, e.g., based on non-parametric density estimates [12], are more effective in low dimensions. In addition to the general benefits, embedding methods can be particularly effective for anomaly detection in that they have a natural mechanism to keep the anomalous points away from the dense/normal regions even in the embedding of the data.

3. EXPERIMENTS ON ARTIFICIAL DATA

To illustrate the advantage of transformation/embedding methods in anomaly detection we create two high dimensional simulated data sets. Both data sets are embedded onto a two dimensional space using ISOMAP and MDS. We compare the performance of anomaly detection on the original data set against the embedded data sets. For anomaly detection we use the popular One-class SVM algorithm. One-class SVM [25] computes a hyperplane in the feature space such that a predefined fraction of the training data lies on one side of the separating hyperplane. Outlier points end up on the side facing the origin. The objective function of the method maximizes the margin between the hyperplane and the origin. For our experiments we use the LIBSVM [7] implementation of the algorithm. We ran experiments on two sets of artificial data:

DataSet1: Normal data is sampled from three non-overlapping Gaussians in \mathbb{R}^{50} , with each Gaussian having an identity covariance matrix. The anomalous points are randomly generated. The coordinates of the first two dimensions are scaled such that they lie within the same range as the Gaussian data. The remaining 48 dimensions are scaled to be further away from the normal points. DataSet1 contains 3300 points, out of which 300 are anomalies.

DataSet2: In this case all data is sampled from a Gaussian in \mathbb{R}^{50} . Anomalous points are generated by taking one half of the points, and setting the first three dimensions such that they form a swiss roll. Data sets resembling a swiss roll are common in dimensionality reduction literature [24, 26]. The size of DataSet2 is 1200.

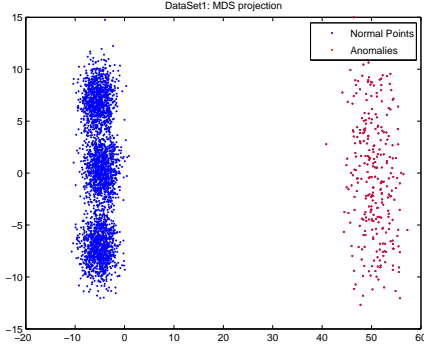
We ran a five-fold cross validation on both data sets. During cross-validation training was performed only on normal points, while testing was done on both the current fold and the anomalous points.

3.1 Results

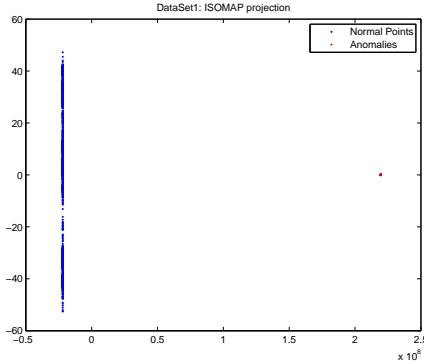
For DataSet1, both MDS and ISOMAP embeddings keep outliers away from the normal points (Figure 1).¹ Thus, in the embedding, anomalies and normal data points are not difficult to detect. However, in the original data set the separation appears to be a more difficult task. The error rate for one-class SVM is 12.3% on the original data set, while it is 6.1 % on the MDS embedding and 3.6 % on the ISOMAP embedding.

DataSet2 illustrates clearly what happens when anomalies adhere to some structure (Figure 2). In this case the structure is a swiss roll in the first three dimensions. The error rate of 16.7 % on the original data set is the highest. Running one-class SVM on the embedded data sets results in an improved performance. The error rate on the MDS embedding is with 5.5 % the lowest. On the ISOMAP embedding it is somewhat higher with 11.3 %, but still better in comparison to the original data set. While both embed-

¹All plots are best viewed in color.



(a)



(b)

Figure 1: DataSet1 2-dimensional embedding using (a) MDS, and (b) ISOMAP.

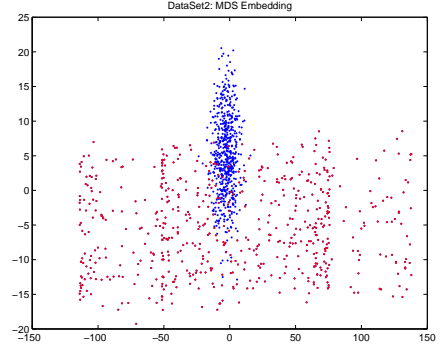
dings appear to keep most anomalies as outliers, ISOMAP seems to be more effective in uncovering the structure of the data. While MDS shows the swiss roll from a top view, in ISOMAP the swiss roll appears to be unfolded. The unfolding allows one to analyze anomalies at different ends of the swiss roll.

The projected points for Dataset1 appear rather separable, yet the error rates for it are not zero. In addition one can also notice a better performance on Dataset2 for MDS-projected data. These observations can be explained by the way the experiments were conducted. One-class SVM had to be run on both projected data and high dimensional data. To be fair to all scenarios four sets of svm parameters were used when running cross validation. The reported results are based on the best performing parameter set for each scenario. Given the same conditions, our objective was to evaluate how the one-class svm algorithm would perform on each scenario.

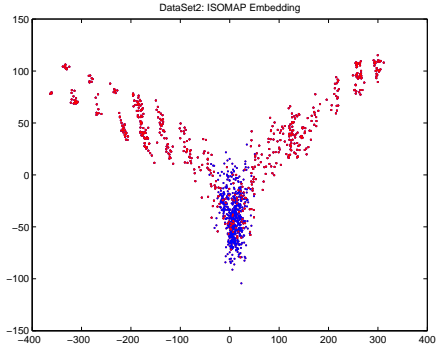
The summary of results is shown in Table 1. Our experiments illustrate that one can indeed benefit from using embedding methods prior to performing anomaly detection. In particular, using (nonlinear) embedding methods can reveal structure within anomalies, which can make a significant difference in real-life anomaly detection tasks.

4. APPLICATION TO TRANSPORTATION CORRIDORS

The formation of secure transportation corridors, where



(a)



(b)

Figure 2: DataSet2 2-dimensional embedding using (a) MDS, and (b) ISOMAP.

Embedding Method	DataSet1	DataSet2
None (original data)	12.3 %	16.7 %
MDS	6.1%	5.5 %
ISOMAP	3.6%	11.3 %

Table 1: Error-rate on anomaly detection using manifold embedding.

cargoes and shipments from points of entry can be dispatched safely to highly sensitive and secure locations, is a high national priority. The primary objective is to ensure rapid intermodal cargo movement, specifically focusing on large trucks along the nation's highways, in a manner that ensures supply chain security without disrupting commerce. This could be achieved through a network of truck weigh stations equipped with state-of-the-art sensor infrastructures. In this context, sensors are defined to include weigh-in-motion scales, static scales, radiation sensors, RFID scanners, cameras, video, text scanners for truck manifests, as well as OCR-based scanners of truck license plates and DOT number plates. The massive volumes of disparate data gathered from sensors are useful in the context of their end-use, which is to detect cargoes that represent plausible safety or security hazards.

The automated discovery of hazards is especially challenging as such events, i.e., trucks carrying contraband or dangerous cargo, are extremely rare. The current practice is

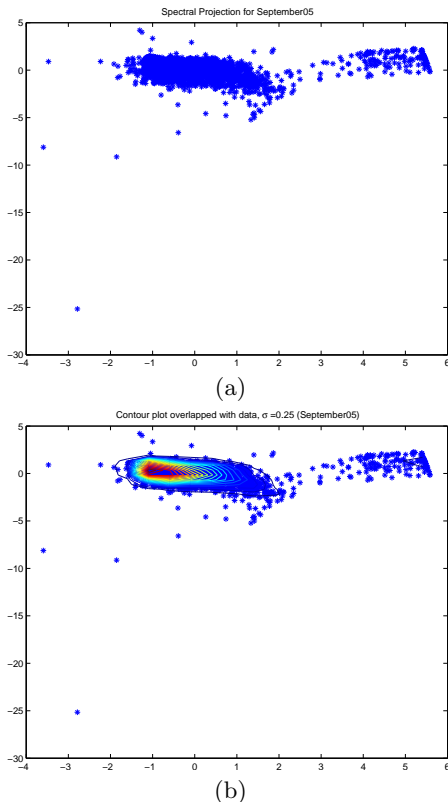


Figure 3: Results on September'05: (a) Spectral projection of data, (b) Parzen estimator for anomaly detection.

to detain a truck for manual inspection when certain alarm thresholds are exceeded. In general, this approach tends to err on the side of caution, which can sometimes generate false or nuisance alarms. Nuisance alarms can, in turn, be rather expensive in terms of human resources, for example, in the time spent by the law enforcement agent in manual inspection of trucks, and may lead to traffic delays. On the other hand, given the high cost of missed detections, the alarm thresholds cannot be set too low lest suspicious trucks pass by unchecked. The critical need to set an optimal threshold, and hence balance false alarm rates with the probability of detection, is a motivating factor for the development of more advanced anomaly detection methods.

As a part of the SensorNet program, the Oak Ridge National Laboratory (ORNL) is currently performing pilot studies at a few weigh stations, one of which is located near Watt Road along the I-40 highway in eastern Tennessee near Knoxville. In this paper, we analyze what is called “static scale data” from the Watt Road weigh stations over several months. This data includes truck lengths, weights at three locations, number of axles, vehicle speeds at the weigh station and vehicle on road distance, i.e., the distance of the vehicle from the sensor. Such static scale data is the focus of our analysis since they were generated from the initial pilot studies in a usable form, and because they happen to be unclassified and generally available once specific truck information has been abstracted.

4.1 Experimental Results

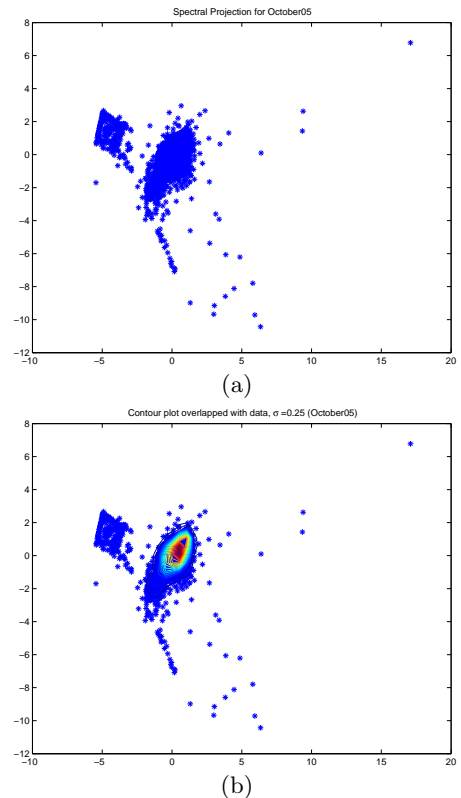


Figure 4: Results on October'05: (a) Spectral projection of data, (b) Parzen estimator for anomaly detection.

We analyzed the weigh-station data using manifold embedding, followed by a simple parzen density estimator [11] to get density contours. We display the results for each of the three months: September'05 (Figure 3), October'05 (Figure 4), and November'05 (Figure 5). For each month, we plot the embedding in 2 dimensions and the parzen density contours on the projected data. As a simple approach, any data point below a certain density contour level can be declared as an outlier or anomaly. The plots clearly show that for the features we used, the normal set is almost always a central big blob, with the anomalous points either scattered around individually or even forming small clusters in some cases. Studying the properties of the small anomalous groups will be instructive in differentiating normal from anomalous.

Several additional experiments were run to study the relationship between the projections of MDS, ISOMAP and LLE. Figure 6 shows a direct comparison between the MDS and ISOMAP projections. It is clear from the color-coding that potential outliers detected by MDS (Figure 6(a)) are also detected by ISOMAP (Figure 6(b)). Similar plots comparing MDS with LLE is shown in Figure 7.

The potential value of nonlinear dimensionality reduction methods over MDS is demonstrated in Figure 8. We consider a particular truck that has one of the highest axle counts and a somewhat uncommon weight distribution on its axles. LLE detects this particular truck as a potential outlier as it is far away from the rest of the group (Figure 8(a)). On the contrary, MDS projects the truck among

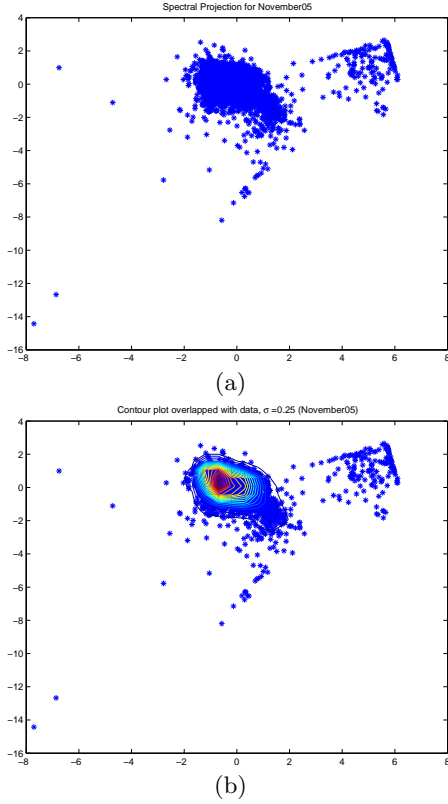


Figure 5: Results on November'05: (a) Spectral projection of data, (b) Parzen estimator for anomaly detection.

the main group of “normal” points (Figure 8(b)). While its not clear if this truck can be called anomalous in the application domain, the example demonstrates the capability of the nonlinear methods capturing subtle oddities in the observed measurements/features, that can be valuable in the context of anomaly detection.

Since some features, such as vehicle speed, have obvious outliers (e.g., a truck moving at 70 mph through the weigh station), we applied box constraints on each feature to get rid of the obvious outliers. After visualizing each feature, threshold values were selected to create the box constraints. As shown in Figure 11(d), the MDS projection of the boxed data does not have any obvious outliers. Further, MDS cannot reveal any interesting structure in the data after the removal of the obvious outliers using box constraints. We applied ISOMAP and LLE to the original as well as the boxed data. As shown in Figure 11, both ISOMAP and LLE show that there are broadly three groups of trucks in the boxed data. A similar structure in the data is revealed for the other months as well. Detailed investigation of the properties of these groups, possibly using additional data, will be an important item for future investigation.

A more detailed comparison between ISOMAP and MDS applied to the boxed data is presented in Figure 12. We locate points in each of the three groups found by ISOMAP in the MDS projection. Clearly, MDS is unable to find this subtle structure. It is interesting to note that the three groups are actually located in three parts of the big point cloud projection of MDS. However, most anomaly detection

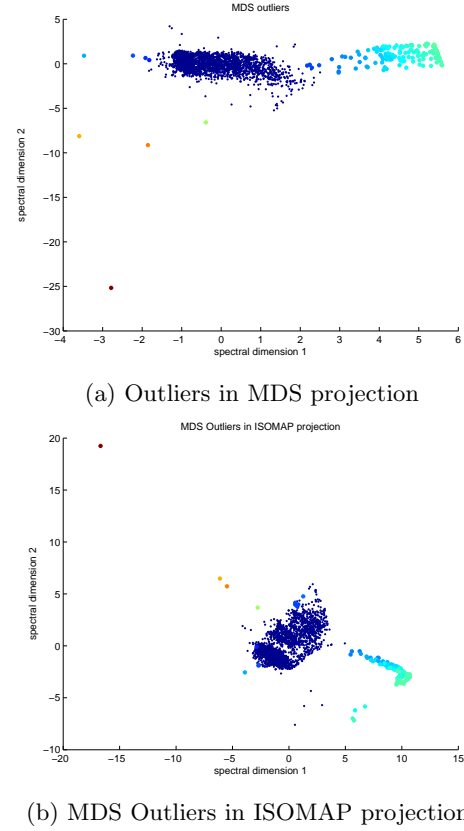
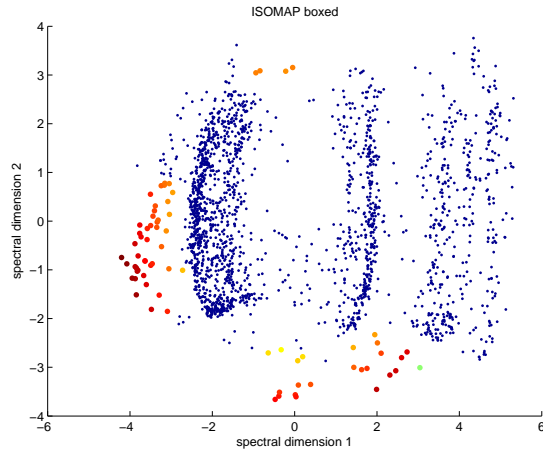


Figure 6: Comparison between MDS and ISOMAP projections. One can see that all outliers given by MDS are also outliers with ISOMAP

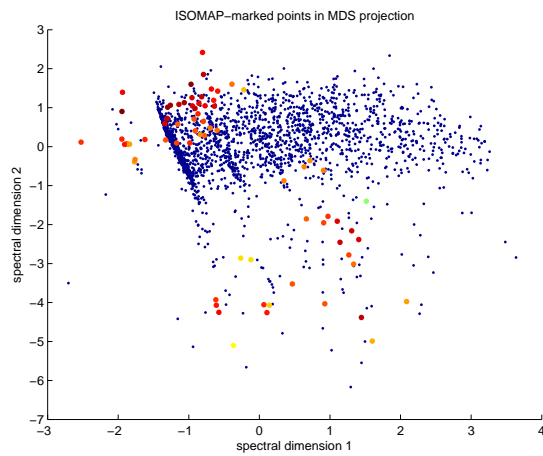
methods applied on the MDS projection will unlikely be able to differentiate between the groups, as all are part of the same big point cloud. In Figure 9, we compare the potential outliers in the boxed ISOMAP projection and corresponding points in the boxed MDS projection. The potential outliers detected by ISOMAP are spread all over the MDS projection. As a result, detecting such potentially anomalous points from the MDS representation can be significantly more difficult.

We also take a simplistic first-cut look at the three groups detected by the nonlinear algorithms. In Figure 10, we show the ISOMAP embedding of the boxed points based on only three features: axle count, vehicle on road distance, and vehicle length. We wanted to see which set of features are causing the formation of three groups in the original ISOMAP embedding. Any single feature, or pair of features did not explain the structure in the data. It seems that the relationships between Axle Count, Vehicle on road distance, and Vehicle length are responsible for the three groups. More detailed experiments will be needed to determine the nature of these relationships.

While our preliminary data analysis has revealed interesting structure in the weigh station data, incorporating additional data about the trucks as well as domain knowledge (say, in terms of rules, box-constraints, etc.) would make our current data mining methodology significantly more effective in practice.



(a) Outliers in boxed ISOMAP projection



(b) Corresponding points in MDS projection

Figure 9: Comparison between outliers in boxed ISOMAP and corresponding points in boxed MDS. The potential outliers detected by ISOMAP are spread all over the MDS projection.

expected insights. The first two variables which dominate the data patterns, namely, axle counts and vehicle length, do seem to be surrogates for the vehicle type and hence appear to validate the previous suppositions. However, the first surprise is that weights or weight profiles are not among these variables. One possible reason may be that the weight profile may be captured on the average by a combination of the other two variables, length and axle counts, although this does not appear too likely. One other possibility may be that the weights and loading profiles can vary significantly even when all other variables remain the same, and hence do not convey meaningful information when the inter-relationships among variables are considered. Further studies may be needed to understand this matter in depth. The influence of the variable called "vehicle on road distance" is the most unexpected. The measurement of this variable is supposed to be among the (if not the) least accurate, and the variable itself is not thought to be much relevant other than perhaps as a potential indicator for the accuracy of the readings. The fact that this apparently unimportant and in-

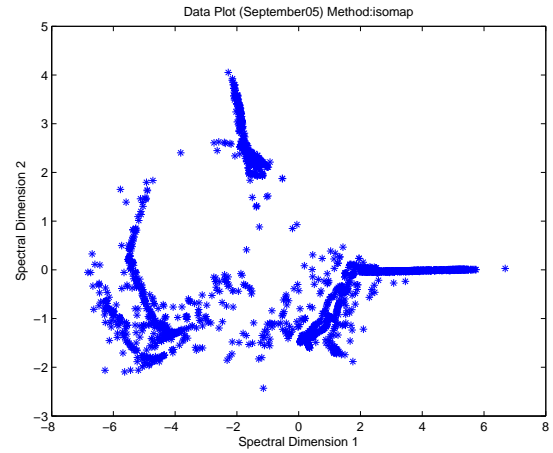


Figure 10: ISOMAP embedding using only three features: Axle Count, Vehicle on road distance, and Vehicle length.

accurate variable is a necessary condition for the patterns, and also happens to be one of three variables that sufficiently or dominantly explain the features in the data, is surprising. Further studies are needed to explore either the domain significance or the spurious nature of this specific insight.

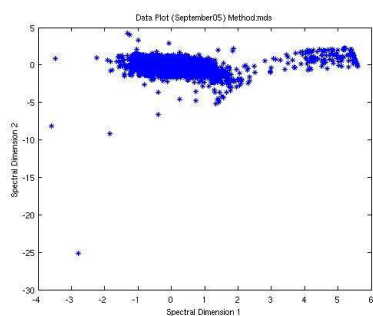
Acknowledgements: This research was sponsored by the Laboratory Directed Research and Development (LDRD) Program of Oak Ridge National Laboratory (ORNL), managed by UT-Battelle, LLC for the U. S. Department of Energy under Contract No. DE-AC05-00OR22725. Any opinions, findings, conclusions, and recommendations expressed in the paper are those of the authors and do not necessarily reflect those of the sponsor.

6. REFERENCES

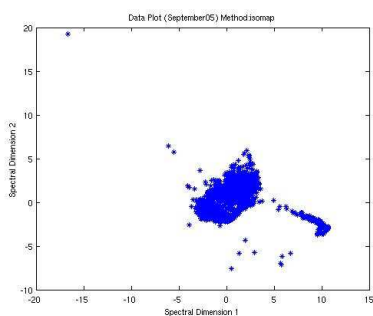
- [1] A. M. Andrew. Statistical pattern recognition. *Robotica*, 18(2):219–223, 2000.
- [2] A. Banerjee, S. Merugu, I. Dhillon, and J. Ghosh. Clustering with Bregman divergences. *Journal of Machine Learning Research*, 6:1705–1749, 2005.
- [3] D. Barbara, Y. Li, J. Couto, J.-L. Lin, and S. Jajodia. Bootstrapping a data mining intrusion detection system. In *Proceedings of the 2003 ACM Symposium on Applied Computing*, pages 421–425. ACM Press, 2003.
- [4] M. Bernstein, V. de Silva, J. Langford, and J. Tenenbaum. Graph approximations to geodesics on embedded manifolds. Technical report, Stanford University, 2000.
- [5] I. Borg and P. Groenen. *Modern Multi-dimensional Scaling*. Springer, 1996.
- [6] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and Jörg Sander. LOF: identifying density-based local outliers. In *ACM International Conference on Management of Data (SIGMOD)*, pages 93–104, 2000.
- [7] C. Chang and C. Lin. Libsvm: a library for support vector machines, 2001.
- [8] M. Collins, S. Dasgupta, and R. Schapire. A generalization of principal component analysis to the exponential family. In *Proc. of the 14th Annual*

Conference on Neural Information Processing Systems (NIPS), 2001.

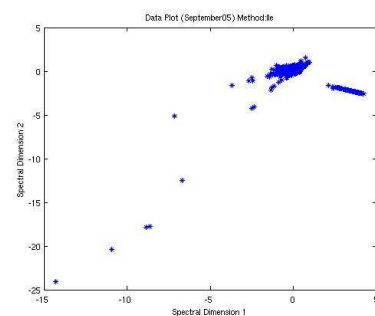
- [9] D. de Ridder and R. Duin. Locally linear embedding for classification. Technical Report PH-2002-01, Delft University of Technology, 2002.
- [10] D. Donoho and C. Grimes. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proceedings of the National Academy of Science*, 100(10), 2003.
- [11] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, 2001.
- [12] S. Forrest, S. A. Hofmeyr, A. Somayaji, and T. A. Longstaff. A sense of self for unix processes. In *Proceedings of the IEEE Symposium on Research in Security and Privacy*, pages 120–128. IEEE Computer Society Press, 1996.
- [13] J. Forster and M. K. Warmuth. Relative expected instantaneous loss bounds. In *Proc. of the 13th Annual Conference on Computational Learning Theory (COLT)*, pages 90–99, 2000.
- [14] O. C. Jenkins and M. J. Mataric. A spatio-temporal extension to Isomap nonlinear dimensionality reduction. In *Proceedings of the 21st International Conference on Machine Learning*, 2004.
- [15] I. Joliffe. *Principal Component Analysis*. Springer-Verlag, 1996.
- [16] L.H. Kim, D.H. Finkel. Hyperspectral image processing using locally linear embedding. In *Conference Proceedings. First International IEEE EMBS Conference on Neural Engineering*, pages 316–319, 2003.
- [17] E. M. Knorr and R. T. Ng. A unified notion of outliers: Properties and computation. In *Knowledge Discovery and Data Mining*, pages 219–222, 1997.
- [18] E. M. Knorr and R. T. Ng. Algorithms for mining distance-based outliers in large datasets. In *Proc. 24th International Conference on Very Large Data Bases, VLDB*, pages 392–403, 24–27 1998.
- [19] D. Kulpinski. Lle and isomap analysis of spectral and color images. Master’s thesis, Simon Fraser University, 2002.
- [20] D. Marchette. A statistical method for profiling network traffic. In *First USENIX Workshop on Intrusion Detection and Network Monitoring*, pages 119–128, Santa Clara, CA, April 9–12, 1999.
- [21] N. Patwari, A. O. Hero, and A. Pacholski. Manifold learning visualization of network traffic data. In *SIGCOMM Workshop on Mining Network Data*, 2005.
- [22] C. Phua, D. Alahakoon, and V. Lee. Minority report in fraud detection: classification of skewed data. *SIGKDD Explor. Newsl.*, 6(1):50–59, 2004.
- [23] R. Pless. Image spaces and video trajectories: Using Isomap to explore video sequences. In *Proceedings of IEEE International Conference on Computer Vision*, 2003.
- [24] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, 2000.
- [25] B. Schoelkopf, J. C. Platt, J. C. Shawe-Taylor, A. J. Smola, and R. C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Comput.*, 13(7):1443–1471, 2001.
- [26] J. Tanenbaum, V. de Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.
- [27] T. Tangkuampien and T.-J. Chin. Locally linear embedding for markerless human motion capture using multiple cameras. In *Proceedings of Digital Image Computing: Techniques and Applications*, 2005.
- [28] D. Tax and R. Duin. Data domain description by support vectors. In *Proceedings of ESANN*, pages 251–256, 1999.
- [29] Introduction to Statistical Pattern Recognition. K. Fukunaga. Academic Press, 1990.
- [30] V. Chandola, A. Banerjee, and V. Kumar. Outlier detection - A survey. In *Preparation*, 2007.
- [31] R. Vilalta and S. Ma. Predicting rare events in temporal domains. In *Proceedings of ICDM’02*, page 474, Washington, DC, USA, 2002. IEEE Computer Society.
- [32] N. Wu and J. Zhang. Factor analysis based anomaly detection. In *IEEE Workshop on Information Assurance*, West Point, NY, USA, June 2003. United States Military Academy.
- [33] X. Yang, H. Fu, H. Zha, and J. Barlow. Semi-supervised nonlinear dimensionality reduction. In *ICML*, 2006.
- [34] H. Zha and Z. Zhang. Isometric embedding and continuum ISOMAP. In *ICML*, 2003.



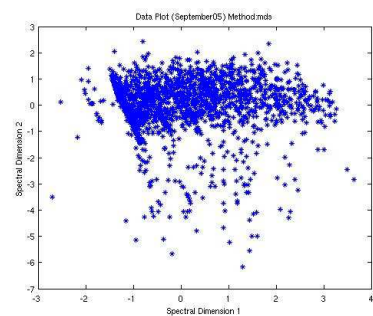
(a) MDS on normal data



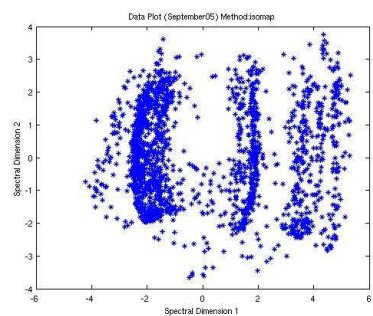
(b) ISOMAP on normal data



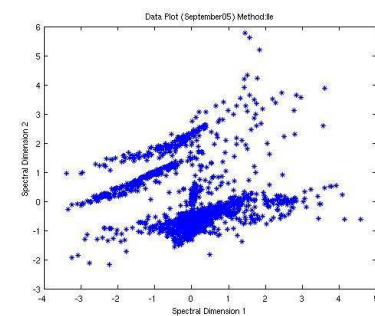
(c) LLE on normal data



(d) MDS on boxed data

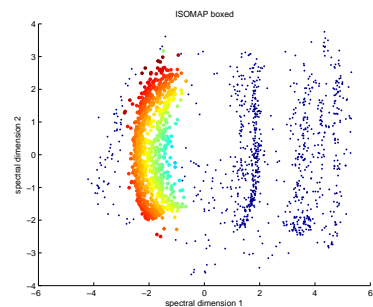


(e) ISOMAP on boxed data

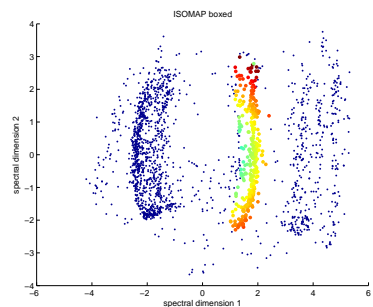


(f) LLE on boxed data

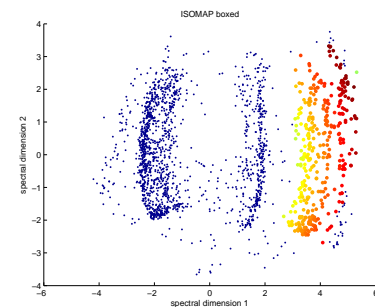
Figure 11: Normal and Boxed data for September'05 projected using MDS, ISOMAP and LLE.



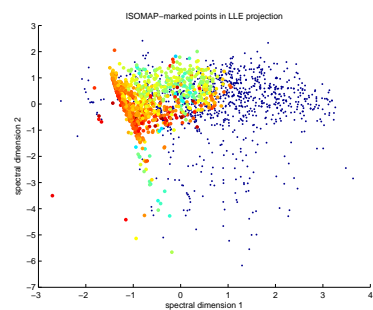
(a) Grp 1 in ISOMAP prj.



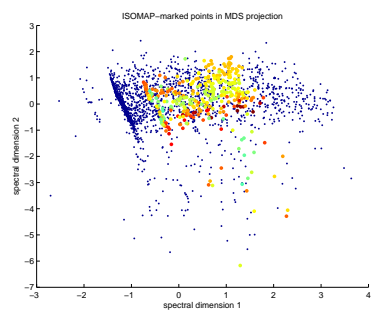
(b) Grp 2 in ISOMAP prj.



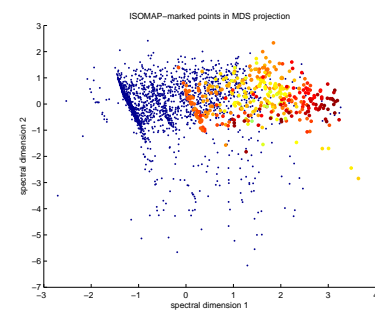
(c) Grp 3 in ISOMAP prj.



(d) Grp 1 points in MDS prj.



(e) Grp 2 points in MDS prj.



(f) Grp 3 points in MDS prj.

Figure 12: Plots directly comparing three point clouds detected by ISOMAP to boxed MDS projection.