2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust

Non-linear Dimensionality Reduction for Privacy-Preserving Data Classification

Khaled Alotaibi, V. J. Rayward-Smith, Wenjia Wang and Beatriz de la Iglesia
School of Computing Sciences,
University of East Anglia, Norwich, NR4 7TJ, UK
Email: {K.Alotaibi, vjrs, Wenjia.Wang, B.Iglesia}@uea.ac.uk

***Abstract*—Many techniques have been proposed to protect the privacy of data outsourced for analysis by external parties. However, most of these techniques distort the underlying data properties, and therefore, hinder data mining algorithms from discovering patterns. The aim of Privacy-Preserving Data Mining (PPDM) is to generate a data-friendly transformation that maintains both the privacy and the utility of the data. We have proposed a novel privacy-preserving framework based on non-linear dimensionality reduction (i.e. non-metric multidimensional scaling) to perturb the original data. The perturbed data exhibited good utility in terms of distance-preservation between objects. This was tested on a clustering task with good results. In this paper, we test our novel PPDM approach on a classification task using a k-Nearest Neighbour (k-NN) classification algorithm. We compare the classification results obtained from both the original and the perturbed data and find them to be much same particularly for the few lower dimensions. We show that, for distance-based classification, our approach preserves the utility of the data while hiding the private details.**

I. Introduction

In the real world, there are increased concerns about individual privacy, especially when data are outsourced or shared with external parties. This privacy issue was firstly addressed by the statistics community in order to protect the individual's identity within a statistical database (i.e. microdata), using methods known as inference control in statistical databases or Statistical Disclosure Control (SDC) [1]. However, the main weakness of this method is that some data properties, such as distance between data points and correlations between variables, are not adequately considered, although, in practice, most data mining algorithms rely on these properties to ensure accurate results.

The term "privacy-preserving data mining" (PPDM) can be simply defined as obtaining valid data mining results without revealing the underlying data values. Generally, it aims to achieve two fundamental objectives — data privacy and data utility; that is, producing accurate mining results without disclosing "private" information. However, this is a challenging task for any data mining technique, as obtaining highly accurate results often depends on the use of the original values. Most approaches to PPDM usually make a trade-off between the two objectives instead of bringing them together.

Multidimensional scaling (MDS) is a non-linear dimensionality reduction technique used to project the data into a lower dimensional space in order to mainly achieve two objectives. The first is to eliminate irrelevant features and reduce noise which may affect the analysis. The second is to easily visualise data using only two or three dimensions, so that a better interpretation of "hidden" structures within the data can be gained. The basic idea of the MDS technique is as follows: given a matrix of proximities that express the similarities or dissimilarities between data objects, find a configuration of data points whose distances fit these proximities best. In non-metric MDS, the inter-point distances between data points in the new space approximate a non-linear transformation from the proximities. Hence, the generated data are subject to high uncertainty, where each data object has been represented by possibly completely different data values, and each data variable has a different pdf.

We have proposed a non-metric MDS [2] as a data perturbation method for PPDM, and shown it preserves the utility for data clustering. In this paper we show how our proposed method produces data that maintains utility for classification analysis too, while providing no information about the original data to the analyser that could be exploited maliciously to disclose the real data values. We use the perturbed data to carry out a classification task using the k-NN algorithm, and show that the results are similar to those obtained from the original data.

This paper is organised as follows: Section II presents some related work. Section III describes some of the detail of our proposed method and, in particular, how to evaluate the effectiveness of the perturbed data in terms of data utility and privacy. Section IV presents and discusses our experimental results. Finally, Section V contains our conclusions.

II. Related Work

Most proposed methods for PPDM sanitise data through a linear transformation of the data values. Data randomisation methods attempt to disguise the original data by randomly modifying the data values, often using additive or multiplicative noise. Compared with other data perturbation methods, the randomisation methods are relatively simple, as the transformation process for the data values is performed in a data-independent way. In additive perturbation, a random matrix R, which has a normal distribution with mean $\mu = 0$ and standard deviation σ, is added to the original data matrix X to produce perturbed data Y [3], [4], [5]. For instance, Agrawal and Srikant [6] reconstruct aggregate distribution that approximates the original data distribution, from data perturbed using

978-0-7695-4848-7/12 $26.00 © 2012 IEEE
DOI 10.1109/SocialCom-PASSAT.2012.76

additive noise. Then, the reconstructed distribution was used to build a decision tree. The drawback of this method is that the added noise will distort the distances between the data points, and therefore poor results may be obtained. Another drawback is that the additive noise can be filtered out and the privacy can then be compromised [7], [8].

Multiplicative perturbation is a further enhancement to additive perturbation, to provide more data utility [9], [10], [11], [12]. It can be described by multiplying X by R in order to generate the new matrix Y; this can then be released to the data miner. In other words, the original data are either rotated around their centre or projected into lower dimensions. However, this method is vulnerable to some privacy attacks. Attackers may exploit some theoretical properties of the random matrices, which usually have a predictable structure, to disclose the original data values [13], [14], [15].

The focus of our PPDM effort is to transform or perturb data so that they can be outsourced for analysis without disclosing *any* of the original values of the attributes. Data anomymisation and cryptography-based approaches are not suitable for our purpose as they tackle other challenges within PPDM. Therefore, in the experiment section, we compare our results with other perturbation approaches.

III. Proposed Approach for data perturbation

MDS attempts to position a set of data points (configurations) in some lower space while retaining the pair-wise distances between these points as much as possible. Given a set of objects $x_1, x_2, \cdots, x_n \in \mathbb{R}^m$ with dissimilarities $\delta_{ij}, 1 \leq i \leq j \leq n$, MDS aims to map these objects to a configuration or set of points $y_1, y_2, \cdots, y_n \in \mathbb{R}^p$, $p < m$, where each point represents one of the objects, and the distance between points $d_{ij} = ||y_i - y_j||$ are such that $d_{ij} = f(\delta_{ij})$, where f is a function chosen in some optimal way (also known as the *representation function*) that relates the dissimilarities in the original space to distances in the new configuration. In non-metric MDS, f is a non-decreasing monotonic function that maintains a monotone relationship between the dissimilarities and the distances in the configuration. For convenience, we assume throughout this paper that the dissimilarities are calculated using the Euclidean metric,

$$\delta(x_i, x_j) = \delta_{ij} = \sqrt{\sum_{k=1}^{m}(x_{ik} - x_{jk})^2}, \tag{1}$$

where m is the number of dimensions, and x_{ik} and x_{jk} are the k^{th} attributes of x_i and x_j, respectively.

The main purpose of using non-metric MDS in PPDM is to transform the entire original data, X, in the high dimensional space, m, into new data, Y, in a lower dimensional space, p, so that it becomes harder (if not impossible) to disclose the real values of the original data variables. The final configuration resulting from this transformation (i.e. Y) is called perturbed data; the general perturbation model is defined as

$$Y = T(\Delta), \tag{2}$$

where Δ is the dissimilarity matrix of X and $T : \mathbb{R}^m \rightarrow \mathbb{R}^p$ is non-metric MDS transformation that satisfies

$$||x_i - x_j|| = ||T(x_i) - T(x_j)|| + e_{ij}, \tag{3}$$

for any two points $x_i, x_j \in \mathbb{R}^m$. The error e theoretically represents the sum of squared deviations caused by the transformation T. The simplest form of e can be defined by

$$e = \sum_{i,j}^{n}(||x_i - x_j|| - ||y_i - y_j||)^2, \tag{4}$$

where $||x_i - x_j||$ is the Euclidean distance between data points x_i and x_i in the high-dimensional space, X, and $||y_i - y_j||$ is the Euclidean distance between their maps in the low-dimensional space, Y.

Non-metric MDS is quite similar to non-parametric procedures that are based on ranked data. Given two pairs of points, (x_i, x_j) and (x_k, x_l), with dissimilarity rank-order $\delta_{ij} \leq \delta_{kl}$, the corresponding distances in the perturbed data must ideally satisfy $d_{ij} \leq d_{kl}$. In practice, this is not always achievable for all pairs of points and we seek a set of points Y that achieves this as best as possible.

Let the set $\{\delta_{ij} \mid i < j\}$ represents $M = n(n-1)/2$ ordered elements of the upper triangle of dissimilarity matrix Δ,

$$\delta_{i_1 j_1} \leq \delta_{i_2 j_2} \leq \ldots \leq \delta_{i_M j_M}. \tag{5}$$

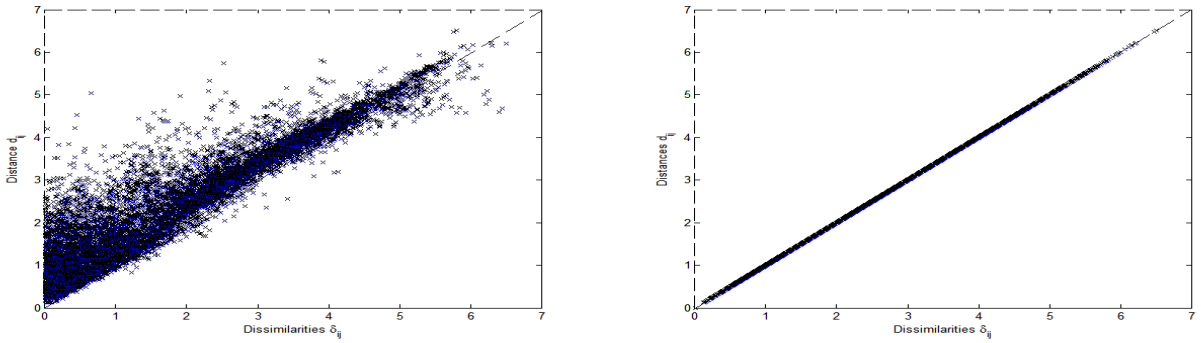
The corresponding distances in Y are $d_{i_1 j_1}, d_{i_2 j_2}, \ldots, d_{i_M j_M}$, which should be in ascending order to achieve so-called *monotonicity*. Assume that the initial configuration in the lower space, Y^0, has been generated, and the initial distances, $d^0_{i_1 j_1}, d^0_{i_2 j_2}, \ldots, d^0_{i_M j_M}$, have also been calculated, a monotone regression algorithm was proposed by [16], [17] to compute a non-decreasing sequence, which are called *disparities*,

$$\hat{d}_{i_1 j_1} \leq \hat{d}_{i_2 j_2} \leq \ldots \leq \hat{d}_{i_M j_M}. \tag{6}$$

That is, $\hat{d_{ij}}$ are numbers generated from the monotone least-square regression of d_{ij} on δ_{ij} to achieve the monotonicity and to obtain a "best-fitting" curve, where the sum of square deviations has to be minimised.

$$\begin{aligned} S^* = \sum_{i,j}^{n} e_{ij} &= \sum_{i,j}^{n} (f(\delta_{ij}) - d_{ij})^2 \\ &= \sum_{i,j}^{n} (\hat{d_{ij}} - d_{ij})^2. \end{aligned} \tag{7}$$

The quantity S^*, called the *raw stress* is often normalised to generate so-called *stress*, S, so it becomes invariant under the change of scale of dissimilarities.



(a) Random projection (b) Non-metric MDS

Fig. 1. Comparing the deviation of distance mapping between points in two solutions: (a) random projection and (b) non-metric MDS.

$$S = \sqrt{\sum_{i,j}^{n} (\hat{d_{ij}} - d_{ij})^2 / \sum_{i,j}^{n} d_{ij}^2}. \quad (8)$$

The non-metric MDS solution is usually found by choosing an initial configuration $Y^0 \subset \mathbb{R}^p$ and moving its points around, in iterative steps, to approximate the best model relation $d_{ij} \approx f(\delta_{ij})$. The initial configuration can be obtained either by decomposing Δ into its eigenvalues and their associated eigenvectors, or by using an arbitrary configuration generated from normally distributed random numbers and scaled by the average squared dissimilarities, in order to obtain average initial distances d_{ij}^0 equal to the average dissimilarities δ_{ij}. The coordinates of each point in $\mathbb{R}^p$ at each iteration are adjusted, using a steepest descent algorithm [17] in the direction that maximally reduces the stress. That is, the computational complexity is considerably high since fitting the monotone regression requires $\mathcal{O}(n^2)$. The algorithm terminates when no further improvements are possible.

A. Data Utility Measures

The measures of data utility are often dependent on the data analysis task to be undertaken. In supervised distance-based classification, data mining algorithms utilise the pair-wise distance between data objects. For example, the k-NN classifier predicts the class label of a test example by computing its distance to the rest of the data points in the training set. A new data item is then classified based on the majority class voting of its nearest neighbours [18]. However, the choice of k can affect the classification accuracy, as every neighbour has the same impact. Other approaches have been suggested to reduce the impact of k using, for example, a distance-weighted constraint [19].

Given a set of training objects $X = \{x_1, x_2, \cdots, x_n\}$ and a predefined set of classes $C = \{c_1, c_2, \cdots, c_s\}$, the k-NN classification rule of a test object x' is defined by

$$g(x', c'_j) = \arg\max_{j=1,\cdots,s} \sum_{i=1}^{n} I(c_j = j), \quad (9)$$

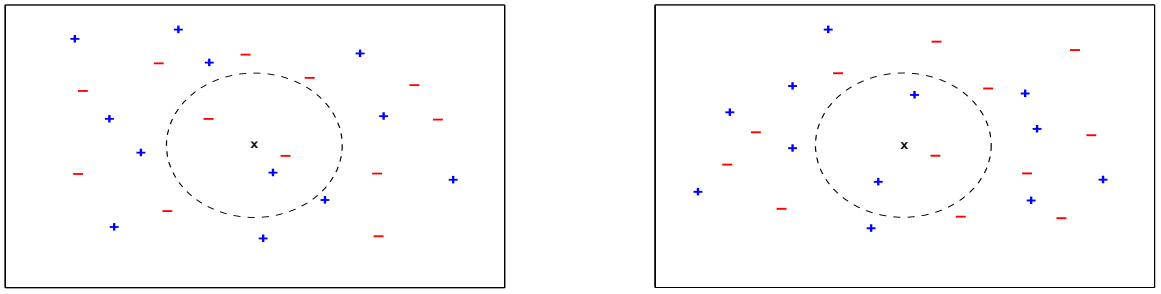
where

$$I(.) = \begin{cases} 1 & \text{if } x_i \text{ belongs to the k-nearest neighbours of } x'. \\ 0 & \text{otherwise.} \end{cases}$$

x_i is said to be the kth nearest neighbour of x' when $||x_i - x'||$ is the kth smallest among $||x_1 - x'||, ||x_2 - x'||, \cdots, ||x_n - x'||$.

We define three utility measures that describe the "goodness" of the perturbed data for distance-based classification analysis. Firstly, we require that the distances between any two data points in the perturbed data are well-preserved with respect to the distances between those points in the original data. Secondly, for each data point in the perturbed data, the k neighbourhood points should be the same as the k neighbourhood points in the original data. Thirdly, for each data point with class c_i, the number of the k neighbourhood points with the same class should be the same before and after the transformation. Measures that correspond to those properties are described as follows:

1) pair-wise Distance Preservation: For the purpose of distance preservation, an error measure of distance deviation should be defined, and then should be minimised to obtain the best mapping. The stress S (8) is usually defined as a weighted sum of differences between distances in the input space, X, and the corresponding distances in the output space, Y. The pair-wise distance between any two points, y_i and y_j, in the p-dimensional space after their transformation, is maintained with small error e_{ij}. The summation of e_{ij} over all pairs (y_i, y_j) yields how well the dissimilarities, δ_{ij}, of the original data, X, or their transformation, $\hat{d_{ij}}$, are fitted by the corresponding distances, d_{ij}, in the perturbed data, Y. That is, the stress indicates the size of information loss between data points before and after the transformation, and can be expressed as a percentage, with 0% stress being equivalent to a perfect configuration, i.e. one that preserves a perfect monotone relationship between dissimilarities and distances. In Figure 1, the Shepard plot shows a comparison between the distances before and after the transformation, which are generated by (a) a linear random projection and (b) non-metric



(a) Original data

(b) Perturbed data

Fig. 2. The impact of the transformation on classifying example $\mathbf{x}$ where the distances of 3-nearest neighbours have been changed. In the original data, the example classified "negative" whereas in the perturbed data it classified "positive".

MDS. For random projection, there is a wide scatter around the regression line, indicating a lack of fit of the distances to the dissimilarities, while the results from non-metric MDS exhibit a linear pattern indicating good fit.

2) Neighbourhood Preservation: Preserving the topological structure of data in the lower dimensional space would demonstrate the usefulness of the data for analysis or visualisation. The neighbourhood is preserved if the distribution of the k-nearest neighbours, in the original space, X, is unchanged or well-approximated in the perturbed space, Y [20]. That is, data points in X are mapped to data points in Y, such that nearby points and faraway points are still nearby and faraway, respectively.

When the pair-wise distances are distorted as a result of the transformation, the accuracy of the classifier is likely to decrease. Consider the example in Figure 2 and let $k = 3$. Unlabelled test object $\mathbf{x}$, located at the centre of the circle, will be classified based on the majority class label of its k-neighbours' training objects, which belong to either a "+" or "−" class. In the original data (2a), the point is classified as a "−" example because the majority class of its neighbours is negative. However, in the perturbed data (2b), the point will be classified as "+" for the same reason. Thus, the quantification of the neighbourhood preservation in the new space can be an indicator of utility for distance-base classification analysis. The quality of neighbourhood preservation [21] can be measured by

$$NP = \frac{1}{k}\sum_{i=1}^{n}\frac{|U_k(x_i) \cap U_k(y_i)|}{n}, \tag{10}$$

where $U_k(x_i)$ and $U_k(y_i)$ are sets of the k-nearest neighbours of point x_i, in the original data, X, and point y_i, in the perturbed data, Y, respectively.

In other words, this measure represents the average number of data points that enter or leave the neighbourhoods in the perturbed data. Hence, it should be maximised. If the maximum value of this measure is 1, the resulting new space is clearly useful for analysis, as the distances of the neighbourhoods are well-preserved.

3) Class Compactness: If members of the same class are close to each other in the original data space, they should also be close to each other in the perturbed data space, i.e. the cluster they belong to should be compact and separable from other clusters [21]. This property indicates that the distance between any two points in different groups should be larger than the distance between any two points within the group.

Given a set of objects $X = \{x_1, x_2, \cdots, x_n\}$ and a predefined set of classes $C = \{c_1, c_2, \cdots, c_s\}$, class compactness for any class $c_j \in C$ is defined by

$$CC_j = \frac{1}{k}\sum_{x_i \in c_j}\frac{|U_k(x_i, c_j)|}{r}, \tag{11}$$

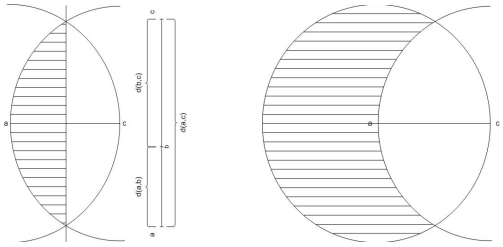
where $U_k(x_i, j)$ is a set of the k-nearest neighbours of point x_i having class label c_j, and r is the number of points in class c_j. The overall class compactness is

$$CC = \frac{1}{r}\sum_{j=1}^{r} CC_j. \tag{12}$$

A low value of this measure would indicate variance of group membership. In contrast, a better preservation of the underlying class structure can be achieved when the value is close to 1.

B. Disclosure Risk Evaluation

In general, the degree of the privacy preserved in any PPDM solution would be described in how much the perturbed data, Y, differs from the original data, X. The higher the probability of estimating the original values, the worse the privacy preservation. The transformation $T : \mathbb{R}^m \rightarrow \mathbb{R}^p$ proposed here satisfies the privacy constraint, as the new dimensions of data Y have different statistics from the original dimensions of data X (i.e. different pdfs and with different ranges). Unlike data randomization methods that are based on a transformation basis (i.e. random noise) in order to derive the new space,



(a) $d_{ab} \leq d_{bc} \leq d_{ac}$ (b) $d_{ab} \leq d_{ac} \leq d_{bc}$

Fig. 3. Representation of all possible positions (shaded area) to place the point b, without violating the constraint: (a) $d_{ab} \leq d_{bc} \leq d_{ac}$ and (b) $d_{ab} \leq d_{ac} \leq d_{bc}$.

in non-metric MDS, the distances are non-linearly mapped to the new space. Even though some attributes are strongly correlated in the original space, their images, in the perturbed space, are irrelevant and not necessary to be correlated as well. Consequently, data Y have no much information that can be exploited by attacks proposed in [22] which use PCA-based technique in order to disclose the original data values. This is also true for attacks in [15], [23] that attempt to relate the eigenvectors of the covariance of a prior known sample with the eigenvectors of the covariance matrix of data Y. Thus, the attacker will not learn anything from Y, and it would be difficult to accurately estimate X. However, data Y still contains much useful information for analysis.

Furthermore, the transformation does not start from the original data matrix X itself, but rather, from the dissimilarity matrix Δ. More precisely, the rank-order of the dissimilarities δ_{ij} (not their magnitude) is assumed to be important for generating the lower dimensional data, so it is often called *ordinal* MDS. This implies that the underlying dimensionality of the original data is unknown and the attributes are irrelevant, which increases the uncertainty about Y. In non-metric MDS, the distances d_{ij} between the points in Y should be, as far as possible, in the same rank-order as the dissimilarities δ_{ij} obtained from X. This can be achieved by an unknown function that monotonically relates the rank-order of the disparities $\hat{d}_{ij}$ resulting from $f : \delta_{ij} \rightarrow d_{ij}$ to the given rank-order of the dissimilarities δ_{ij}.

Moreover, the data, Y, are subject to high uncertainty since the monotone regression geometrically implies that Y are moved iteratively in the direction that minimises the stress S, and therefore, the points are placed within an uncertain area under the restriction of monotonicity. To illustrate how the points are placed in the new space, assume that a,b and c are three data points in Y; their pair-wise distances are d_{ab}, d_{bc} and d_{ac} conforming to the rank-order $d_{ab} \leq d_{bc} \leq d_{ac}$, as illustrated in Figure 3. Assume that the points a and c have been placed and that the distance between them is d_{ac}. Without loss of generality, all possible positions for placing a point b, without violating the constraint $d_{ab} \leq d_{bc} \leq d_{ac}$, are bounded by the shaded area as shown in Figure 3a. Consider another distances order: $d_{ab} \leq d_{ac} \leq d_{bc}$. In this case, the uncertainty about placing b will increase to include a wider area, as shown in Figure 3b. The shaded area represents the uncertainty in placing the points, which can prevent the attacker from exactly determining the position of any point, and hence, privacy can not be compromised.

TABLE I
DATASETS USED IN THE EXPERIMENTS.

Dataset	# Records	# Attributes	# Classes
Wine	178	13	3
Breast Cancer Wisconsin	699	9	2
Handwritten Digits	3823	64	10
Image Segmentation	2100	19	7
Multiple Features	2000	216	10

IV. Experimental Results

To test our model, we conducted experiments on real numeric datasets taken from the UCI machine learning repository. The description of the datasets is given in Table I. To examine the quality of classification, we compare the quality of accuracy obtained from X and Y. If the misclassification error of the classifier that is trained on Y is equal to that error from the classifier on X, then a good perturbation is achieved. This implies that the classifier on Y is invariant to the transformation T.

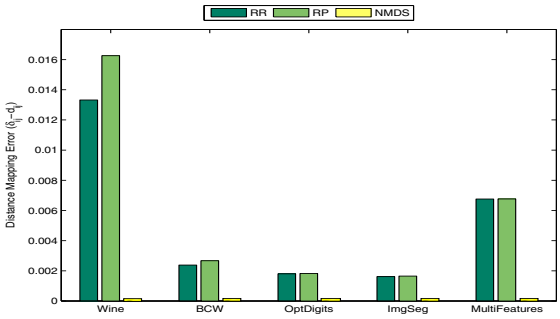


Fig. 4. A comparison between three different perturbation techniques: random rotation (RR), random projection (RP) and non-metric MDS (NMDS). The last exhibits a perfect distance mapping between objects in the original data, X, and the perturbed data, Y.

The original data, X, has been normalised so all variables have zero mean and $\sigma = 1$. Then, the dissimilarities Δ between the records in X were calculated using (1). The perturbed data, Y, are generated in the lower dimension $p = m - 1$ based on Δ and using non-metric MDS. To show how much information is lost as a result of the transformation, we calculate the deviation of distance between objects in the upper and lower spaces, $\delta_{ij} - d_{ij}$, where δ_{ij} and d_{ij} is the distance between the objects in the upper and lower space, respectively. We also compare our method with other methods that are based on data randomisation [9], [12]. The results from experiments on the five datasets are depicted in Figure 4, which indicate that the non-metric MDS outperforms other

methods for this measurement, as the error was negligible and lower than for the other methods.

To estimate the accuracy of the k-NN classifier on both X and Y, we use 10-fold cross-validation and the value of k is set to 4. For all datasets, the accuracy on Y was exactly the same as on X. The accuracies of datasets were as follows: 97% for Wine, 97% for Breast Cancer Wisconsin, 98% for Handwritten Digits, 94% for Image Segmentation and 96% for Multiple Features. This implies that the perturbed data are still useful for classification analysis. We also evaluate the impact of k on the classifier performance to examine if there exists a variation in the accuracy on both data X and Y. The value of k is varied from 3 to 10. For all datasets, as k increases, the classification error increases except for BCW dataset the error decreases for $k \geq 8$. However, the results on both data X and Y were exactly the same. The effect of varying k is shown in Figure 5.

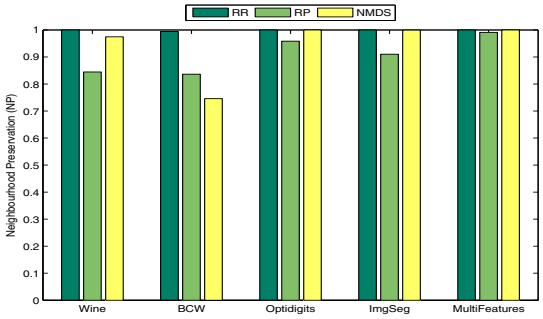


Fig. 6. The average of neighbourhood preservation in the perturbed data, Y, at different number of k neighbours by using random rotation (RR), random projection (RP) and non-metric MDS (NMDS).

The results of the experiments on the quality of neighbourhood preservation in the perturbed data, Y, are presented in Figure 6. The NP measure is calculated at different number of k ($k = 3, 4, \cdots, 10$), and then averaged. We can see that all methods find a configuration that preserves the neighbourhoods for most datasets. This gives the insight that Y can be used in classification analysis and good results can be obtained. However, random rotation exhibits better neighbourhood preservation than other methods. In contrast, for random projection, point neighbourhoods have been slightly affected by the transformation, as the results were low compared with the other methods. Note that the results for the BCW dataset show low neighbourhood preservation for lower values of k. This may be because some of the data points tend to be outliers instead of forming dense clusters.

Finally, we examine class compactness in both data X and Y. Figure 7 shows the results for all datasets using three different methods (i.e. non-metric MDS, random rotation and projection). The class compactness is not much changed before and after the transformation. This indicates good data utility of the perturbed data, specifically for low values of k, where all methods exhibit better class compactness.

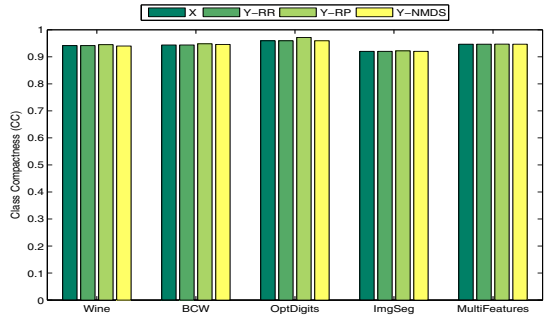


Fig. 7. A comparison of the average of class compactness in the original data, X and the perturbed data, Y, at different number of k and using different transformations.

We also compared the accuracy on Y using our method and random projection. In both cases, the data were transformed to different numbers of dimensions. Based on the results shown in Figure 8, for random projection, the accuracy increases as the number of dimensions increases. However, better accuracy is obtained on data perturbed using non-metric MDS for all dimensions, but especially for the space below 10 dimensions, where the differences are quite remarkable.

One distinctive feature of the non-linear dimensionality reduction method is that better class separation in data would be effectively achieved. Therefore, the k-NN classifier would work better in classifying data objects, as the pair-wise distances within one group are relatively small and between two groups are relatively large. A comparison of class compactness in the Wine dataset before and after the transformation is shown in Figure 9. In data randomization methods (i.e. random rotation and projection), both close and far apart points are mapped close together, whereas in non-metric MDS, clear separation and tight appearance of the clusters were achieved, as the pair-wise distances are well-preserved.

V. Conclusions

In this paper, we have evaluated our previously proposed conceptual privacy-preserving framework for outsourcing data using non-metric MDS. In particular, we used non-metric MDS as a perturbation tool to hide the original data values, and to achieve better PPDM for a distance-based classification task. The accuracy of the classification on both the original data X and the perturbed data Y was compared. The results were exactly the same, which indicates that the non-linear transformation of data into a lower dimension can preserve the distances-related properties, allowing for classification results comparable to those of the original data, and meanwhile, protecting the underlying data values. Our method is independent of both the original data and mining algorithms. It merely takes into account the dissimilarities between data objects, so existing data mining algorithms can be applied directly on the perturbed data without any modifications. Furthermore, our method demonstrates good neighbourhood preservation, good

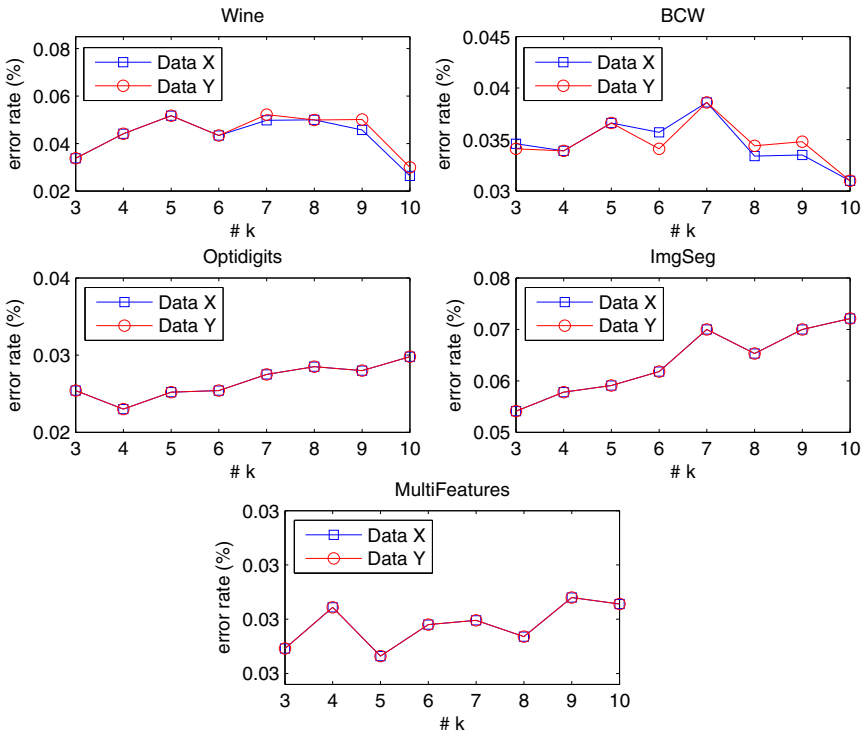
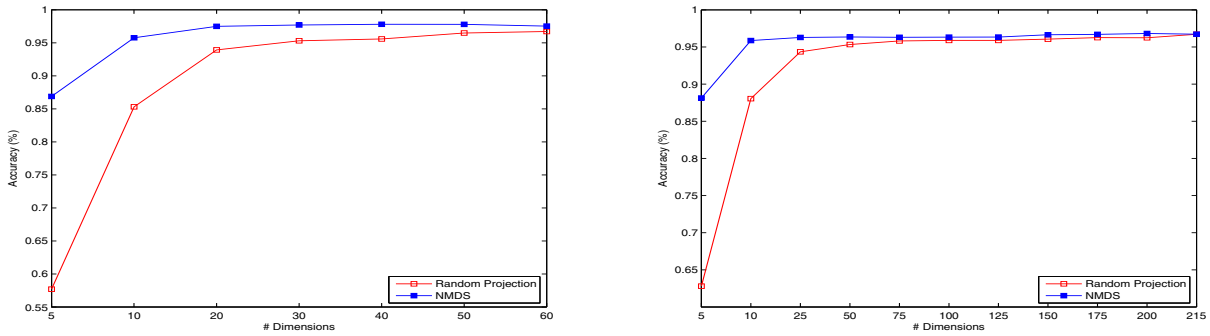


Fig. 5. The classification test error, as function of k, on both the original data, X, and the perturbed data, Y.



(a) Handwritten Digits

(b) Multiple Features

Fig. 8. Classification accuracy of k-NN on two datasets transformed using random projection and non-metric MDS.

class compactness, better class separation, and is more robust in terms of privacy preservation. In summary, the application of non-metric MDS transformation often works efficiently and can produce better results than other perturbation techniques based on data randomisation.

REFERENCES

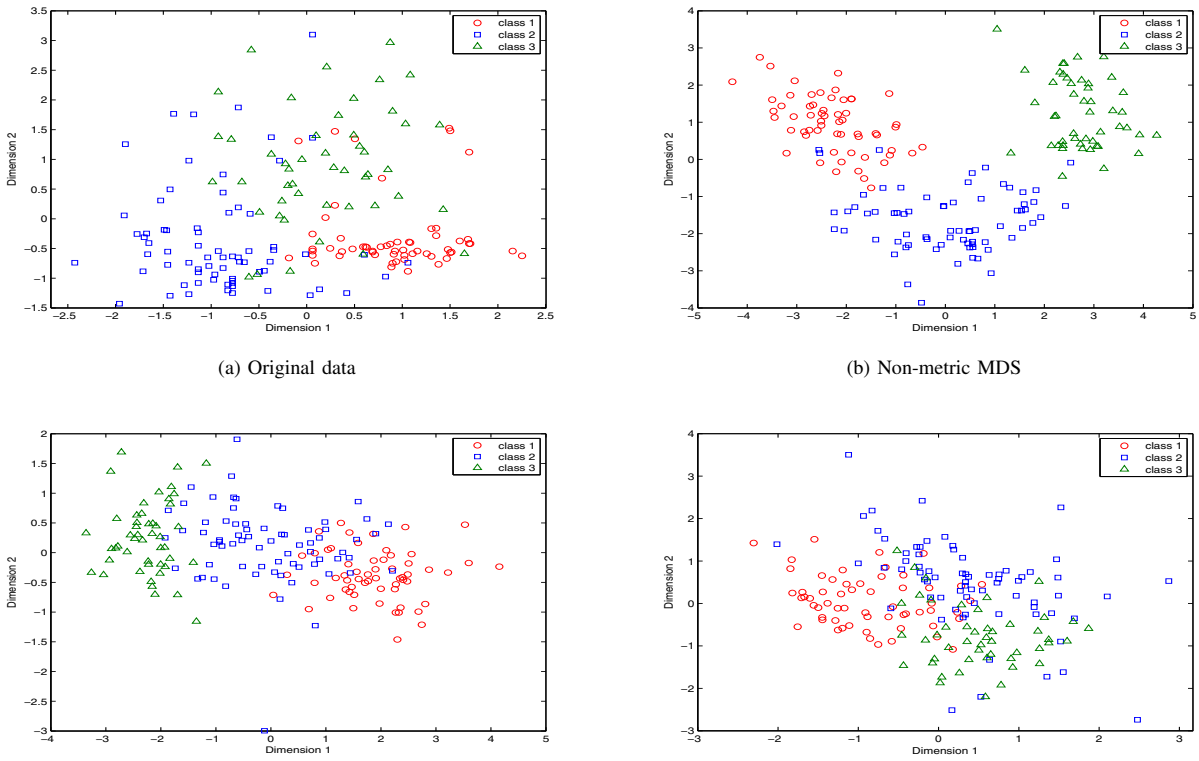
[1] J. Domingo-Ferrer, "A survey of inference control methods for privacy-preserving data mining," in *Privacy-Preserving Data Mining: Models and Algorithms*, C. Aggarwal and P. Yu, Eds. Springer, 2008, ch. 3, pp. 53–80.

[2] K. Alotaibi, V. Rayward-Smith, and B. de la Iglesia, "Non-metric multidimensional scaling for privacy-preserving data clustering," in *Intelligent Data Engineering and Automated Learning-IDEAL 2011*. Springer Berlin, Heidelberg, 2011, pp. 287–298.

[3] M. Kadampur and D. Somayajulu, "A noise addition scheme in decision tree for privacy preserving data mining," *Journal of computing*, vol. 2, pp. 137–144, 2010.

[4] J. Kim, "A method for limiting disclosure in microdata based on random noise and transformation," in *Proceedings of the section on survey research methods*, 1986, pp. 303–308.

[5] R. Little, "Statistical analysis of masked data," *Journal of Official Statistic-Stockholm*, vol. 9, pp. 407–407, 1993.



(a) Original data

(b) Non-metric MDS

(c) Random rotation

(d) Random projection

Fig. 9. A comparison of class compactness between data objects in (a) the original data, X, and the perturbed data, Y, generated by different methods (b),(c),(d). The classes in non-metric MDS solution are reasonably well-separated relative to the classes in others generated by data randomisation.

[6] R. Agrawal and R. Srikant, "Privacy-preserving data mining," *ACM Sigmod Record*, vol. 29, no. 2, pp. 439–450, 2000.

[7] D. Agrawal and C. Aggarwal, "On the design and quantification of privacy preserving data mining algorithms," in *Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. ACM, 2001, p. 255.

[8] H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar, "Random-data perturbation techniques and privacy-preserving data mining," *Knowledge and Information Systems*, vol. 7, no. 4, pp. 387–414, 2005.

[9] K. Chen and L. Liu, "Privacy preserving data classification with rotation perturbation," in *Proceedings of the Fifth IEEE International Conference on Data Mining*, 2005, pp. 589–592.

[10] K. Chen, G. Sun, and L. Liu, "Towards attack-resilient geometric data perturbation," in *Proceedings of the 2007 SIAM Data Mining Conference*. SDM'07, 2007.

[11] K. Liu, H. Kargupta, and J. Ryan, "Random projection-based multiplicative data perturbation for privacy preserving distributed data mining," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 1, pp. 92–106, 2006.

[12] S. Oliveira and O. Zaïane, "Privacy-preserving clustering to uphold business collaboration: A dimensionality reduction based transformation approach," *International Journal of Information Security and Privacy*, vol. 1, no. 2, p. 13, 2007.

[13] S. Guo and X. Wu, "Deriving private information from arbitrarily projected data," *Advances in Knowledge Discovery and Data Mining*, pp. 84–95, 2007.

[14] H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar, "On the Privacy Preserving Properties of Random Data Perturbation Techniques," in *Proceedings of the Third IEEE International Conference on Data Mining*. IEEE Computer Society, 2003, pp. 99–106.

[15] K. Liu, C. Giannella, and H. Kargupta, "An attackers view of distance preserving maps for privacy preserving data mining," *Knowledge Discovery in Databases: PKDD 2006*, pp. 297–308, 2006.

[16] J. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964.

[17] ——, "Nonmetric multidimensional scaling: a numerical method," *Psychometrika*, vol. 29, no. 2, pp. 115–129, 1964.

[18] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[19] S. Dudani, "The distance-weighted k-nearest-neighbor rule," *Systems, Man and Cybernetics, IEEE Transactions on*, no. 4, pp. 325–327, 1976.

[20] J. Venna and S. Kaski, "Neighborhood preservation in nonlinear projection methods: An experimental study," *Artificial Neural Networks-ICANN 2001*, pp. 485–491, 2001.

[21] A. N. Gorban and A. Zinovyev, "Principal manifolds and graphs in practice: From molecular biology to dynamical systems," *International Journal of Neural Systems*, vol. 20, no. 3, pp. 219–232, 2010.

[22] Z. Huang, W. Du, and B. Chen, "Deriving private information from randomized data," in *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*. ACM, 2005, p. 48.

[23] E. Turgay, T. Pedersen, Y. Saygın, E. Savaş, and A. Levi, "Disclosure risks of distance preserving data transformations," in *Scientific and Statistical Database Management*. Springer, 2008, pp. 79–94.