On the Privacy Preserving Properties of Random Data Perturbation Techniques

Hillol Kargupta and Souptik Datta
Computer Science and
Electrical Engineering Department
University of Maryland Baltimore County
Baltimore, Maryland 21250, USA
{hillol, souptik1}@cs.umbc.edu

Qi Wang and Krishnamoorthy Sivakumar
School of Electrical Engineering
and Computer Science
Washington State University
Pullman, Washington 99164-2752, USA
{qwang, siva}@eecs.wsu.edu

Abstract

Privacy is becoming an increasingly important issue in many data mining applications. This has triggered the development of many privacy-preserving data mining techniques. A large fraction of them use randomized data distortion techniques to mask the data for preserving the privacy of sensitive data. This methodology attempts to hide the sensitive data by randomly modifying the data values often using additive noise. This paper questions the utility of the random value distortion technique in privacy preservation. The paper notes that random objects (particularly random matrices) have "predictable" structures in the spectral domain and it develops a random matrix-based spectral filtering technique to retrieve original data from the dataset distorted by adding random values. The paper presents the theoretical foundation of this filtering method and extensive experimental results to demonstrate that in many cases random data distortion preserve very little data privacy.

1. Introduction

Many data mining applications deal with privacy-sensitive data. Financial transactions, health-care records, and network communication traffic are some examples. Data mining in such privacy-sensitive domains is facing growing concerns. Therefore, we need to develop data mining techniques that are sensitive to the privacy issue. This has fostered the development of a class of data mining algorithms [2, 9] that try to extract the data patterns without directly accessing the original data and guarantees that the mining process does not get sufficient information to reconstruct the original data.

This paper considers a class of techniques for privacy-preserving data mining by randomly perturbing the data while preserving the underlying probabilistic properties. It explores the random value perturbation-based approach [2], a well-known technique for masking the data using random noise. This approach tries to preserve data privacy by adding random noise, while making sure that the random noise still preserves the "signal" from the data so that the patterns can still be accurately estimated. This paper questions the privacy-preserving capability of the random value perturbation-based approach. It shows that in many cases, the original data (sometimes called "signal" in this paper) can be accurately estimated from the perturbed data using a spectral filter that exploits some theoretical properties of random matrices. It presents the theoretical foundation and provides experimental results to support this claim.

Section 2 offers an overview of the related literature on privacy preserving data mining. Section 3 presents the motivation behind the framework presented in this paper. Section 4 describes the random data perturbation method proposed in [2]. Section 5 presents a discussion on the eigenvalues of random matrices. Section 6 presents the intuition behind the thoery to separate out random component from a mixture of non-random and random component. Section 7 describes the proposed random matrix-based filtering technique . Section 8 applies the proposed technique and reports its performance for various data sets. Finally, Section 9 concludes this paper.

2. Related Work

There exists a growing body of literature on privacy-sensitive data mining. These algorithms can be divided into several different groups. One approach adopts a distributed framework. This approach supports computation of data mining models and extraction of "patterns" at a given node by exchanging only the minimal necessary information among the participating nodes without transmitting the raw data. Privacy preserving association rule mining from homogeneous [9] and heterogeneous [19] distributed data sets are few examples. The second approach is based on

data-swapping which works by swapping data values within same feature [3].

There is also an approach which works by adding random noise to the data in such a way that the individual data values are distorted preserving the underlying distribution properties at a macroscopic level. The algorithms belonging to this group works by first perturbing the data using randomized techniques. The perturbed data is then used to extract the patterns and models. The randomized value distortion technique for learning decision trees [2] and association rule learning [6] are examples of this approach. Additional work on randomized masking of data can be found elsewhere [18].

This paper explores the third approach [2]. It points out that in many cases the noise can be separated from the perturbed data by studying the spectral properties of the data and as a result its privacy can be seriously compromised. Agrawal and Aggarwal [1] have also considered the approach in [2] and have provided a expectation-maximization (EM) algorithm for reconstructing the distribution of the original data from perturbed observations. They also provide information theoretic measures (mutual information) to quantify the amount of privacy provided by a randomization approach. Agrawal and Aggarwal [1] remark that the method suggested in [2] does not take into account the distribution of the original data (which could be used to guess the data value to a higher level of accuracy). However, [1] provides no explicit procedure to reconstruct the original data values. Evfimievski et al. [5, 4] and Rizvi [15] have also considered the approach in [2] in the context of association rule mining and suggest techniques for limiting privacy breaches. Our primary contribution is to provide an explicit filtering procedure, based on random matrix theory, that can be used to estimate the original data values.

3. Motivation

As noted in the previous section, a growing body of privacy preserving data mining techniques are adopting randomization as a primary tool to "hide" information. While randomization is an important tool, it must be used very carefully in a privacy-preserving application.

Randomness may not necessarily imply uncertainty. Random events can often be analyzed and their properties can be explained using probabilistic frameworks. Statistics, randomized computation, and many other related fields are full of theorems, laws, and algorithms that rely on probabilistic characterization of random processes that often work quite accurately. The signal processing literature [12] offers many filters to remove white noise from data and they often work reasonably well. Randomly generated structures like graphs demonstrate interesting properties [7]. In short, randomness does seem to have "structure" and this structure may be used to compromise privacy issues unless we pay careful attention. The rest of this paper illustrates this challenge in the context of a well-known privacy preserving technique that works using random additive noise.

4. Random Value Perturbation Technique: A Brief Review

For the sake of completeness, we now briefly review the random data perturbation method suggested in [2] for hiding the data (i.e. guaranteeing protection against the reconstruction of the data) while still being able to estimate the underlying distribution.

4.1. Perturbing the Data

The random value perturbation method attempts to preserve privacy of the data by modifying values of the sensitive attributes using a randomized process [2]. The authors explore two possible approaches — Value-Class Membership and Value Distortion — and emphasize the Value Distortion approach. In this approach, the owner of a dataset returns a value $u_i + v$, where u_i is the original data, and v is a random value drawn from a certain distribution. Most commonly used distributions are the uniform distribution over an interval $[-\alpha, \alpha]$ and Gaussian distribution with mean $\mu = 0$ and standard deviation σ. The n original data values $u_1, u_2, \ldots, u_n$ are viewed as realizations of n independent and identically distributed (i.i.d.) random variables U_i, $i = 1, 2, \ldots, n$, each with the same distribution as that of a random variable U. In order to perturb the data, n independent samples $v_1, v_2, \ldots, v_n$, are drawn from a distribution V. The owner of the data provides the perturbed values $u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n$ and the cumulative distribution function $F_V(r)$ of V. The reconstruction problem is to estimate the distribution $F_U(x)$ of the original data, from the perturbed data.

4.2. Estimation of Distribution Function from the Perturbed Dataset

The authors [2] suggest the following method to estimate the distribution $F_U(u)$ of U, given n independent samples $w_i = u_i + v_i$, $i = 1, 2, \ldots, n$ and $F_V(v)$. Using Bayes' rule, the posterior distribution function $F'_U(u)$ of U, given that $U + V = w$, can be written as

$$F'_U(u) = \frac{\int_{-\infty}^{u} f_V(w - z) f_U(z) dz}{\int_{-\infty}^{\infty} f_V(w - z) f_U(z) dz},$$

which upon differentiation with respect to u yields the density function

$$f'_U(u) = \frac{f_V(w-u) f_U(u)}{\int_{-\infty}^{\infty} f_V(w-z) f_U(z) dz},$$

where $f_U(.)$, $f_V(.)$ denote the probability density function of U and V respectively. If we have n independent samples $u_i + v_i = w_i$, $i = 1, 2, \ldots, n$, the corresponding posterior distribution can be obtained by averaging:

$$f'_U(u) = \frac{1}{n} \sum_{i=1}^{n} \frac{f_V(w_i-u) f_U(u)}{\int_{-\infty}^{\infty} f_V(w_i-z) f_U(z) dz}. \quad (1)$$

For sufficiently large number of samples n, we expect the above density function to be close to the real density function $f_U(u)$. In practice, since the true density $f_U(u)$ is unknown, we need to modify the right-hand side of equation 1. The authors suggest an iterative procedure where at each step $j = 1, 2, \ldots$, the posterior density $f_U^{j-1}(u)$ estimated at step $j-1$ is used in the right-hand side of equation 1. The uniform density is used to initialize the iterations. The iterations are carried out until the difference between successive estimates becomes small. In order to speed up computations, the authors also discuss approximations to the above procedure using partitioning of the domain of data values.

5. Randomness and Patterns

The random perturbation technique "apparently" distorts the sensitive attribute values and still allows estimation of the underlying distribution information. However, does this apparent distortion fundamentally prohibit us from extracting the hidden information? This section presents a discussion on the properties of random matrices and presents some results that will be used later in this paper.

Random matrices [13] exhibit many interesting properties that are often exploited in high energy physics [13], signal processing [16], and even data mining [10]. The random noise added to the data can be viewed as a random matrix and therefore its properties can be understood by studying the properties of random matrices. In this paper we shall develop a spectral filter designed based on random matrix theory for extracting the hidden data from the data perturbed by random noise.

For our approach, we are mainly concerned about distribution of eigenvalues of the sample covariance matrix obtained from a random matrix. Let V be a random $m \times n$ matrix whose entries are V_{ij}, $i = 1, \ldots, m$, $j = 1, \ldots, n$, are i.i.d. random variables with zero mean and variance σ^2. The covariance matrix of X is given by $Y = \frac{1}{m} V'V$. Clearly, Y is an $n \times n$ matrix. Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the eigenvalues of Y. Let

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} U(x - \lambda_i),$$

be the empirical cumulative distribution function (c.d.f.) of the eigenvalues λ_i, $(1 \leq i \leq n)$, where

$$U(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$$

is the unit step function. In order to consider the asymptotic properties of the c.d.f. $F_n(x)$, we will consider the dimensions $m = m(N)$ and $n = n(N)$ of matrix X to be functions of a variable N. We will consider asymptotics such that in the limit as $N \to \infty$, we have $m(N) \to \infty$, $n(N) \to \infty$, and $\frac{m(N)}{n(N)} \to Q$, where $Q \geq 1$. Under these assumptions, it can be shown that [8] the empirical c.d.f. $F_n(x)$ converges in probability to a continuous distribution function $F_Q(x)$ for every x, whose probability density function (p.d.f.) is given by

$$f_Q(x) = \begin{cases} \frac{Q\sqrt{(x-\lambda_{\min})(\lambda_{\max}-x)}}{2\pi\sigma^2 x} & \lambda_{\min} < x < \lambda_{\max} \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are as follows:

$$\lambda_{\min} = \sigma^2(1 - 1/\sqrt{Q})^2.$$
$$\lambda_{\max} = \sigma^2(1 + 1/\sqrt{Q})^2. \quad (3)$$

Further refinements of this result and other discussions can be found in [16].

6. Separating the Data from the Noise

Consider an $m \times n$ data matrix U and a noise matrix V with same dimensions. The random value perturbation technique generates a modified data matrix $U_p = U + V$. Our objective is to extract U from U_p. Although the noise matrix V may introduce seemingly significant difference between U and U_p, it may not be successful in hiding the data.

Consider the covariance matrix of U_p:

$$\begin{aligned} U_p^T U_p &= (U+V)^T (U+V) \\ &= U^T U + V^T U + U^T V + V^T V. \end{aligned} \quad (4)$$

Now note that when the signal random vector (rows of U) and noise random vector (rows of V) are uncorrelated, we have $E[U^T V] = E[V^T U] = 0$. The uncorrelated assumption is valid in practice since the noise V that is added to the data U is generated by a statistically independent process. Recall that the random value perturbation technique discussed in the previous section introduces uncorrelated

noise to hide the signal or the data. If the number of observations is sufficiently large, we have that $U^TV \sim 0$ and $V^TU \sim 0$. Equation 4 can now be simplified as follows:

$$U_p^TU_p = U^TU + V^TV \tag{5}$$

Since the correlation matrices U^TU, $U_p^TU_p$, and V^TV are symmetric and positive semi-definite, let

$$\begin{aligned} U^TU &= Q_u\Lambda_uQ_u^T, \\ U_p^TU_p &= Q_p\Lambda_pQ_p^T, \text{ and} \\ V^TV &= Q_v\Lambda_vQ_v^T, \end{aligned} \tag{6}$$

where Q_u, Q_p, Q_v are orthogonal matrices whose column vectors are eigenvectors of U^TU, $U_p^TU_p$, V^TV, respectively, and Λ_u, Λ_p, Λ_v are diagonal matrices with the corresponding eigenvalues on their diagonals.

The following result from matrix perturbation theory [20] gives a relationship between Λ_u, Λ_v, and Λ_v.

Theorem 1 *[20] Suppose $\lambda_{1,(a)} \geq \lambda_{2,(a)} \geq \ldots \lambda_{n,(a)} \geq 0$, $a \in \{u, p, v\}$ are the eigenvalues of U^TU, $U_p^TU_p$, and V^TV, respectively. Then, for $i = 1, \ldots, n$,*

$$\lambda_{i,(p)} \in [\lambda_{i,(u)} + \lambda_{n,(v)}, \lambda_{i,(u)} + \lambda_{1,(v)}].$$

This theorem provides us a bound on the change in the eigenvalues of the data correlation matrix U^TU in terms of the minimum and maximum eigenvalues of the noise correlation matrix V^TV. Now let us take a step further and explore the properties of the eigenvalues of the perturbed data matrix U_p for large values of m.

Lemma 1 *Let data matrix U and noise matrix V be of size $m \times n$ and $U_p = U + V$. Let Q_u, Q_p, Q_v be orthogonal matrices and Λ_u, Λ_p, Λ_v be diagonal matrices as defined in 6. If $m/n \to \infty$ then $\Lambda_p = \Delta\Lambda_u\Delta^T + \Lambda_v$ where $\Delta = Q_p^TQ_u$.*

Proof:
Using Equations 5 and 6 we can write,

$$\begin{aligned} Q_p\Lambda_pQ_p^T &= Q_u\Lambda_uQ_u^T + Q_v\Lambda_vQ_v^T \\ \Rightarrow \Lambda_p &= Q_p^TQ_u\Lambda_uQ_u^TQ_p + Q_p^TQ_v\Lambda_vQ_v^TQ_p \\ &= \Delta\Lambda_u\Delta^T + Q_p^TQ_v\Lambda_vQ_v^TQ_p \end{aligned} \tag{7}$$

Let the minimum and maximum eigenvalues of V be $\lambda_{\min,(v)}$ and $\lambda_{\max,(v)}$ respectively. It follows from equation 2 that $m/n \to \infty$ all the eigenvalues in Λ_v become identical since $\lim_{m/n=Q\to\infty} \lambda_{\max,(v)} = \lim_{m/n=Q\to\infty} \lambda_{\min,(v)} = \sigma^2$ (say). This implies that, as $m/n \to \infty$, $\Lambda_v \to \sigma^2 I$, where I is the $n \times n$ identity matrix. Therefore, if the number of observations m is large enough (note that, in practice, number of features n is fixed), $V^TV = Q_v\Lambda_vQ_v^T = \sigma^2Q_vQ_V^T = \sigma^2 I$. Therefore Equation 7 becomes

$$\begin{aligned} \Lambda_p &= \Delta\Lambda_u\Delta^T + Q_p^TQ_p\Lambda_vQ_p^TQ_p \\ \Lambda_p &= \Delta\Lambda_u\Delta^T + \Lambda_v. \end{aligned} \tag{8}$$

■

If the norm of the perturbation matrix V is small, the eigenvectors Q_p of $U_p^TU_p$ would be close to the eigenvectors $Q_u^TQ_u$ of U^TU. Indeed, matrix perturbation theory provides precise bounds on the angle between eigenvectors (and invariant subspaces) of a matrix U and that of its perturbation $U_p = U + V$, in terms of the norms of the perturbation matrix V. For example, let (x_u, λ_u) be an eigenvector-eigenvalue pair for matrix U^TU and $\epsilon = \|V^TV\|_2 = \sigma_{\max}(V^TV)$ be the two-norm of the perturbation, where $\sigma_{\max}(V^TV)$ is the largest singular value of V^TV. Then there exists an eigenvalue-eigenvector pair (x_p, λ_p) of $U_p^TU_p$ satisfying [20, 17]

$$\tan(\angle(x_u, x_p) < 2\frac{\epsilon}{\delta - \epsilon},$$

where δ is the distance between λ_u and the closest eigenvalue of U^TU, provided $\epsilon < \delta$. This shows that the eigenvalues of U^TU and $U_P^TU_P$ are in general close, for small perturbations. Moreover,

$$|\lambda_u - x^*U_px_u| < 2\frac{\epsilon^2}{\delta - \epsilon},$$

where x^* is the conjugate-transpose of x. Consequently, the product $\Delta = Q_p^TQ_u$, which is the matrix of inner products between the eigenvectors of U^TU and $U_p^TU_p$ would be close to an identity matrix; i.e., $\Delta = Q_p^TQ_u \approx I$. Thus equation 8 becomes

$$\Lambda_p \approx \Lambda_u + \Lambda_v. \tag{9}$$

Suppose the signal covariance matrix has only a few dominant eigenvalues, say $\lambda_{1,(u)} \geq \cdots \geq \lambda_{k,(u)}$, with $\lambda_{i,(u)} \leq \epsilon$ for some small value ϵ and $i = k+1, \ldots, n$. This condition is true for many real-world signals. Suppose $\lambda_{k,(u)} > \lambda_{1,(v)}$, the largest eigenvalue of the noise covariance matrix. It is then clear that we can separate the signal and noise eigenvalues Λ_u, Λ_v from the eigenvalues Λ_p of the observed data by a simple thresholding at $\lambda_{1,(v)}$.

Note that equation 9 is only an approximation. However, in practice, one can design a filter based on this approximation to filter out the perturbation from the data. Experimental results presented in the following sections indicate that this provides a good recovery of the data.

7. Random Matrix-Based Data Filtering

This section describes the proposed filter for extracting the original data from the noisy perturbed data. Suppose actual data U is perturbed by a randomly generated noise matrix V in order to produce $U_p = U + V$. Let $u_{p,i} = \mathbf{u}_i + \mathbf{v}_i$,

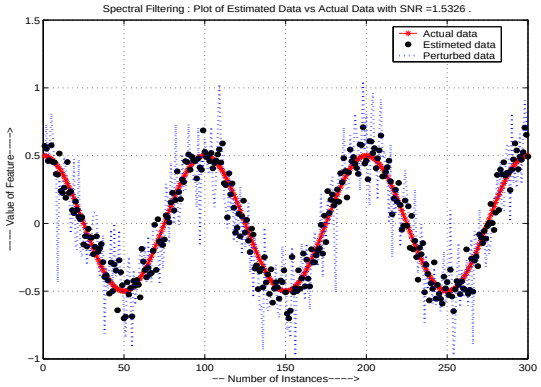

Figure 1. Estimation of original sinusoidal data with known random noise variance.

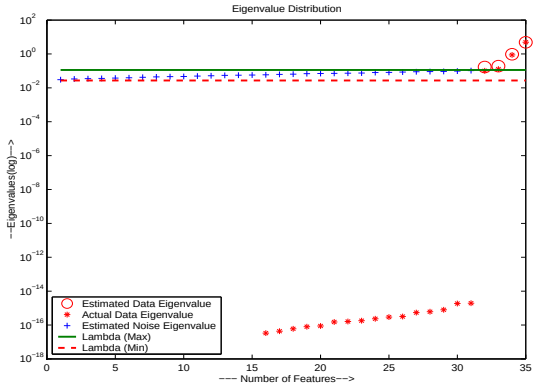

Figure 2. Distribution of eigenvalues of actual data, and estimated eigenvalues of random noise and actual data.

$i = 1, 2, \ldots, m$, be m (perturbed) data points, each being a vector of n features.

When the noise distribution $F_V(v)$ of V is completely known (as required by the random value perturbation technique [2]), the noise variance σ^2 is first calculated from the given distribution. Equation 2 is then used to calculate λ_{max} and λ_{min} which provide the theoretical bounds of the eigenvalues corresponding to noise matrix V. From the perturbed data, we compute the eigenvalues of its covariance matrix Y, say $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Then we identify the noisy eigenstates $\lambda_i \leq \lambda_{i+1} \leq \cdots \leq \lambda_j$ such that $\lambda_i \geq \lambda_{min}$ and $\lambda_j \leq \lambda_{max}$. The remaining eigenstates are the eigenstates corresponding to actual data. Let, Λ_v = diag $(\lambda_i, \lambda_{i+1}, \ldots, \lambda_j)$ be the diagonal matrix with all noise-related eigenvalues, and A_v be the matrix whose columns are the corresponding eigenvectors. Similarly, let Λ_u be the eigenvalue matrix for the actual data part and A_u be the corresponding eigenvector matrix which is an $n \times k$ matrix ($k \leq n$). Based on these matrices, we decompose the covariance matrix Y into two parts, Y_s and Y_r with $Y = Y_s + Y_r$, where $Y_r = A_v \Lambda_r A_v^T$, is the covariance matrix corresponding to random noise part, and $Y_s = A_u \Lambda_u A_u^T$, is the covariance matrix corresponding to actual data part. An estimate $\hat{U}$ of the actual data U is obtained by projecting the data U_p on to the subspace spanned by the columns of A_u. In other words, $\hat{U} = U_p A_u A_u^T$.

8. Experimental Results

In this section, we present results of our experiments with the proposed spectral filtering technique. This section also includes discussion on the effect of noise variance on the performance of the spectral filtering method.

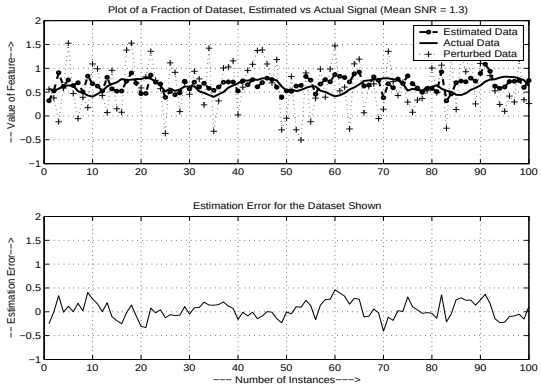

Figure 3. Spectral filtering used to estimate real world audio data. Waveform of a audio signal is closely estimated from its perturbed version.

8.1. Estimation with Known Perturbing Distribution

We tested our privacy breaching technique using several datasets of different sizes. We considered both artificially generated and real data sets. Towards that end, we generated a dataset with 35 features and 300 instances. Each feature has a specific trend like sinusoidal, square, and triangular shape, however there is no dependency between any two features. The actual dataset is perturbed by adding Gaussian noise (with zero mean and known variance), and our proposed technique is applied to recover the actual data from the perturbed data. Figure 1 shows the result of our spectral filtering for one such feature where the actual data has a sinusoidal trend. The filtering technique appears to pro-

vide an accurate estimate of the individual values of the actual data. Figure 2 shows the distribution of eigenvalues of the actual and perturbed data. It also identifies the estimated noise eigenvalues and the theoretical bounds $\lambda_{\max}$ and $\lambda_{\min}$. As we see, the filtering method accurately distinguishes between noisy eigenvalues and eigenvalues corresponding to actual data. Note that the estimated eigenvalues of actual data is very close to eigenvalues of actual data and almost overlap with them above $\lambda_{\max}$. The eigenvalues of actual data below $\lambda_{\min}$ are practically negligible. Thus, the estimated eigenvalues of the actual data capture most of the information and discard the additive noise.

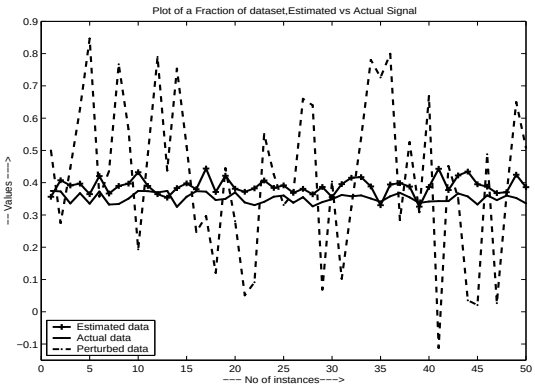

Figure 4. Plot of the individual values of a fraction of the dataset with 'Triangular' distribution. Spectral filtering gives close estimation of individual values.

The random matrix-based filtering technique can also be extended to datasets with a single feature, i.e when the dataset is a single column vector. The data vector is perturbed with a noise vector with the same dimension. The perturbed data vector is then split into a fixed number of vectors with equal length and all of these vectors are appended to form a matrix. The spectral filtering technique is then applied to this matrix to estimate the original data. After the data matrix is estimated, its columns are concatenated to form a single vector.

We used a real world single feature data set to verify the performance of the spectral filtering. The dataset used is the scaled amplitude of the waveform of an audio tune recorded using a fixed sampling frequency. The tune recorded is fairly noise free with 10000 sample points. We perturbed this data with additive Gaussian noise.

We define the term *Signal-to-Noise Ratio* (SNR) to quantify the relative amount of noise added to actual data to perturb it:

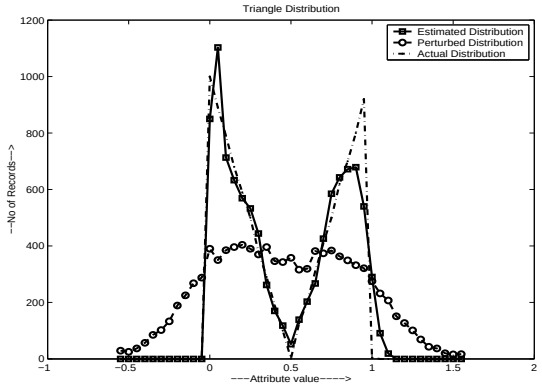

Figure 5. Reconstruction of the 'Triangular' distribution. Perturbed data distribution does not look like a triangular distribution, but reconstructed distribution using spectral filtering resembles the original distribution closely.

$$\text{SNR} = \frac{\text{Variance of Actual Data}}{\text{Noise Variance}}. \qquad (10)$$

In this experiment, the noise variance was chosen to yield a signal-to-noise ratio of 1.3. We split this vector of perturbed data into 40 columns, each containing 250 points and applied the spectral filtering technique to recover the actual data. The result is shown in Figure 3. For the sake of clarity, only a fraction of dataset is shown, and estimation error is plotted for that fraction. As shown in Figure 3, the perturbed data is very different from the actual data, whereas the estimated data is a close approximation of the actual data. The estimation performance is similar to that for a multi-featured data (see Figure 1).

8.2. Comparison With Results in [2]

The proposed spectral filtering technique can estimate values of individual data-points from the perturbed dataset. This point-wise estimation can then be used to reconstruct the distribution of actual data as well. The methods suggested by [2, 1] can only reconstruct the distribution of the original data from the data perturbed by random value distortion; but it does not consider estimation of the individual values of the data-points. The spectral filtering technique, on the other hand, is explicitly designed to reconstruct the individual data-points and hence, also the distribution of the actual dataset.

We tried to replicate the experiment reported in [2] using our method to recover the triangular distribution. We used a vector data of 10000 values having a triangular distribution as shown in Figure 2 in [2]. The individual values of actual

data are within 0 and 1 and are independent of each other. We added Gaussian random noise with mean 0 and standard deviation $\sigma = 0.25$ to this data and split the data vector into 50 columns, each having 200 values. We then applied our spectral filter to recover the actual data from the perturbed data. Figure 4 shows a portion of the actual data, their values after distortion, and their estimated values. Note that the estimated values are very close to the actual values, compared to the perturbed values. Using the estimate of individual data-points, we reconstruct the distribution of the actual data. Figure 5 shows estimation of the distribution from the estimated value of individual data-points. The distribution of the perturbed data is very different than the actual triangular distribution, but the estimated distribution looks very similar to the original distribution. This shows that our method recovers the original distribution along with individual data-points, similar to the result reported in [2]. The estimation accuracy is greater than 80% for all datapoints. Since spectral filtering can filter out the individual values of actual data and its distribution from a perturbed representation, it breaches the privacy preserving protection of the randomized data perturbation technique [2].

8.3. Effect of Perturbation Variance and the Inherent Random Component of the Actual Data

Quality of the data recovery depends upon the relative noise content in the perturbed data. We use the SNR (see equation (10)) to quantify the relative amount of noise added to actual data to perturb it. As the noise added to the actual value increases, the SNR decreases. Our experiments show that the proposed filtering method predicts the actual data reasonably well up to a SNR value of 1.0 (i.e. 100% noise). The results shown in Figure 1 corresponds to an SNR value nearly 2, i.e. noise content is about 50%. Figure 4 shows a data-block where the SNR is 1.9. As the SNR goes below 1, the estimation becomes too erroneous. Figure 6 shows the difference in estimation accuracy as the SNR increases from 1. The dataset used here has a sinusoidal trend in its values. The top graph corresponds to 23% noise (SNR = 4.3), whereas the bottom graph corresponds to 100% noise (SNR = 1.0).

Another important factor that affects the quality of recovery of the actual data is the inherent noise in the actual dataset (apart from the perturbation noise added intentionally). If the actual dataset has a random component in it, and random noise is added to perturb it, spectral filtering method does not filter the actual data accurately. Our experiments with some inherently noisy real life dataset show that the eigenvalues of signal and noise no longer remains clearly separable since the their eigenvalues may not be distributed over two non-overlapping regimes any longer.

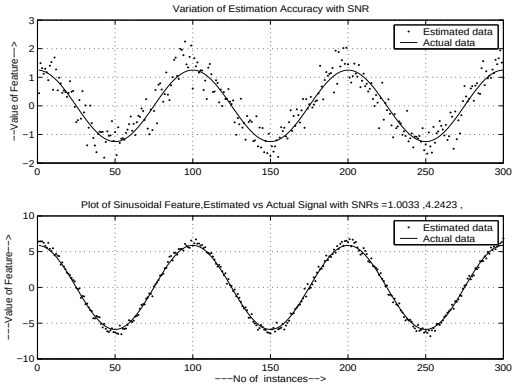

Figure 6. A higher noise content (low SNR)leads to less accurate estimation. SNR in the upper figure is 1, while that for the lower figure is 4.3.

We have performed experiments with artificial dataset with specific trend in its value as well as real world dataset containing a random component. Figure 1 in fact shows that our method gives a close estimation of actual data when the dataset has some specific trend (sinusoid). We also applied our method to "Ionosphere data" available from [14], which is inherently noisy. We perturbed the original data with random noise such that mean SNR is same as the artificial dataset, i.e. 1.1. Figure 7 shows that recovery quality is poor compared to datasets having definite trend.

However, this opens up a different question: Is the random component of the original data set really important as far as data mining is concerned? One may argue that most data mining techniques exploit only the non-random structured patterns of the data. Therefore, losing the inherent random component of the original data may not be important in a privacy preserving data mining application.

9. Conclusion and Future Work

Preserving privacy in data mining activities is a very important issue in many applications. Randomization-based techniques are likely to play an important role in this domain. However, this paper illustrates some of the challenges that these techniques face in preserving the data privacy. It showed that under certain conditions it is relatively easy to breach the privacy protection offered by the random perturbation based techniques. It provided extensive experimental results with different types of data and showed that this is really a concern that we must address. In addition to raising this concern the paper offers a random-matrix based data filtering technique that may find wider application in developing a new perspective toward developing better privacy-preserving data mining algorithms.

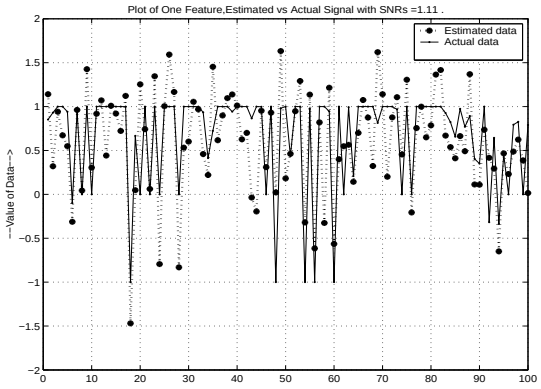

Figure 7. Spectral filtering performs poorly on a dataset with a random component in its actual value. However, it is not clear if loosing the random component of the data is a concern for data mining applications.

Since the problem mainly originates from the usage of additive, independent "white" noise for privacy preservation, we should explore "colored" noise for this application. We have already started exploring multiplicative noise matrices in this context. If U be the data matrix and V be an appropriately sized random noise matrix then we are interested in the properties of the perturbed data $U_p = UV$ for privacy-preserving data mining applications. If V is a square matrix then we may be able to extract signal using techniques like independent component analysis. However, projection matrices that satisfy certain conditions may be more appealing for such applications. More details about this possibility can be found elsewhere [11].

Acknowledgments

The authors acknowledge supports from the United States National Science Foundation CAREER award IIS-0093353, NASA (NRA) NAS2-37143, and TEDCO, Maryland Technology Development Center.

References

[1] D. Agrawal and C. C. Aggawal. On the design and quantifi cation of privacy preserving data mining algorothms. In *Proceedings of the 20th ACM SIMOD Symposium on Principles of Database Systems*, pages 247–255, Santa Barbara, May 2001.

[2] R. Agrawal and R. Srikant. Privacy-preserving data mining. In *Proceeding of the ACM SIGMOD Conference on Management of Data*, pages 439–450, Dallas, Texas, May 2000. ACM Press.

[3] V. Estivill-Castro and L. Brankovic. Data swaping: Balancing privacy against precision in mining for logic rules. In *Proceedings of the fi rst Conference on Data Warehousing and Knowledge Discovery (DaWaK-99)*, pages 389 – 398, Florence, Italy, 1999. Springer Verlag.

[4] A. Evfi mevski, J. Gehrke, and R. Srikant. Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the ACM SIMOD/PODS Conference*, San Diego, CA, June 2003.

[5] A. Evfi mevski, R. Srikant, R. Agrawal, and J. Gehrke. Privacy preserving mining of association rules. In *Proceedings of the ACM SIKDD Conference*, Edmonton, Canada, 2002.

[6] S. Evfi mievski. Randomization techniques for privacy preserving association rule mining. In *SIGKDD Explorations*, volume 4(2), Dec 2002.

[7] S. Janson, T. L. , and A. Rucinski. *Random Graphs*. Wiley Publishers, 1 edition, 2000.

[8] D. Jonsson. Some limit theorems for the eigenvalues of a sample covariance matrix. *Journal of Multivariate Analysis*, 12:1–38, 1982.

[9] M. Kantarcioglu and C. Clifton. Privacy-preserving distributed mining of association rules on horizontally partitioned data. In *SIGMOD Workshop on DMKD*, Madison, WI, June 2002.

[10] H. Kargupta, K. Sivakumar, and S. Ghosh. Dependency detection in mobimine and random matrices. In *Proceedings of the 6th European Conference on Principles and Practice of Knowledge Discovery in Databases*, pages 250–262. Springer, 2002.

[11] K. Liu, H. Kargupta, and J. Ryan. Random projection and privacy preserving correlation computation from distributed data. Technical report, University of Maryland Baltimore County, Computer Science and Electrical Engineering Department, Technical Report TR-CS-03-24, 2003.

[12] D. G. Manolakis, V. K. Ingle, and S. M. Kogon. *Statistical and Adaptive Signal Processing*. McGraw Hill, 2000.

[13] M. L. Mehta. *Random Matrices*. Academic Press, London, 2 edition, 1991.

[14] U. M. L. Repository. http://www.ics.uci.edu/ mlearn/mlsummary.html.

[15] S. J. Rizvi and J. R. Haritsa. Maintaining data privacy in association rule mining. In *Proceedings of the 28th VLDB Conference*, Hong Kong, China, 2002.

[16] J. W. Silverstein and P. L. Combettes. Signal detection via spectral theory of large dimensional random matrices. *IEEE Transactions on Signal Processing*, 40(8):2100–2105, 1992.

[17] G. W. Stewart. Error and perturbation bounds for subspaces associated with certain eigenvalue problems. *SIAM Review*, 15(4):727–764, October 1973.

[18] J. F. Traub, Y. Yemini, and H. Woz'niakowski. The statistical security of a statistical database. *ACM Transactions on Database Systems (TODS)*, 9(4):672–679, 1984.

[19] J. Vaidya and C. Clifton. Privacy preserving association rule mining in vertically partitioned data. In *The Eighth ACM SIGKDD International conference on Knowledge Discovery and Data Mining*, Edmonton, Alberta, CA, July 2002.

[20] H. Weyl. Inequalities between the two kinds of eigenvalues of a linear transformation. In *Proceedings of the National Academy of Sciences*, volume 35, pages 408–411, 1949.