International Journal On Engineering Technology and Sciences – IJETS™
ISSN(P): 2349-3968, ISSN (O): 2349-3976
Volume III, Issue V, May- 2016

RIGHT-PROTECTED DATA PUBLISHING WITH PROVABLE DISTANCE-BASED MINING

Ms.k.Manikam
Department of Computer Science and Engineering
SSM College of Engineering, Komarapalayam,
Tamil Nadu, India
manikkamssm@gmail.com@gmail.com

Mr.A.T.Ravi
Department of Computer Science and Engineering
SSM College of Engineering, Komarapalayam,
Tamil Nadu, India
csehod@ssmce.ac.in

ABSTRACT

Data exchange and data publishing are becoming an inherent part of business and academic practices. Data owners, nonetheless, also need to maintain the principal rights over the datasets that they share, which in many cases have been obtained after expensive and laborious procedures. This thesis work presents a right-protection mechanism that can provide detectable evidence for the legal ownership of a shared dataset, without compromising its usability under a wide range of machine learning, mining, and search operations. It is accomplished by guaranteeing that order relations between object distances remain unaltered.

Protection of one's intellectual property is a topic with important technological and legal facets. This project provides mechanisms for establishing the ownership of a dataset consisting of multiple objects. The algorithms also preserve important properties of the dataset, which are important for mining operations, and so guarantee both right protection and utility preservation. The project considers a right-protection scheme based on watermarking. Watermarking may distort the original distance graph. The proposed watermarking methodology preserves important distance relationships, such as: the Nearest Neighbors (NN) of each object of the original dataset. This leads to preservation of any mining operation that depends on the ordering of distances between objects, such as NN-search and classification, as well as many visualization techniques. It proves fundamental lower and upper bounds on the distance between objects post-watermarking.

In particular, it establishes a restricted isometric property, i.e., tight bounds on the contraction/expansion of the original distances. This analysis used to design fast algorithms for NN-preserving watermarking that drastically prunes the vast search space. The application is designed using Microsoft Visual Studio .Net 2005 as front end. The coding language used is Visual C# .Net. MS-SQL Server 2000 is used as back end database

OBJECTIVE OF PROJECT

A multipurpose watermarking scheme which can be applied to achieve both authentication and protection of data set has been presented in this proposed system. Watermarks are embedded once in the hiding process and can be blindly extracted for different applications in the detection process. The proposed scheme has three special features:

- The approximation information of a host image is kept in the hiding process by utilizing masking thresholds.

- Oblivious and robust watermarking is achieved for copy- right protection.
- Fragile watermarking is achieved for detection of malicious modifications and tolerance of incidental manipulations.
- In addition to images (gray-scale and color), this method has been extended to audio watermarking. To the best of our knowledge, this is the first method that combines both robust

LITERATURE REVIEW

Rakesh Agrawal Jerry Kiernan et al [2004] [1] describe the piracy of digital assets such as software, images, video, audio and text has long been a concern for owners of these assets. Protection of these assets is usually based upon the insertion of digital watermarks into the data The watermarking software introduces small errors into the object being watermarked. These intentional errors are called marks and all the marks together constitute the watermark. The marks must not have a significant impact on the usefulness of the data and they should be placed in such a way that a malicious user cannot destroy them without making the data less useful. Thus, watermarking does not prevent copying, but it deters illegal copying by providing a means for establishing the original ownership of a redistributed copy.

These differences include:

- A multimedia object consists of a large number of bits, with considerable redundancy. Thus, the watermark has a large cover in which to hide. A database relation consists of tuples, each of which represents a separate object. The watermark needs to be spread over these separate objects.
- The relative spatial/temporal positioning of various pieces of a multimedia object typically does not change. Tuples of a relation on the other hand constitute a set and there is no implied ordering between them.
- Portions of a multimedia object cannot be dropped or replaced arbitrarily without causing perceptual changes in the object. However, the pirate of a relation can simply drop some tuples or substitute them with tuples from other relations.

Claudio Lucchese et al [2001] [2] describe the sharing is an important aspect of scientific or business collaboration. However, data owners are also concern with the protection of their rights on the datasets, which is many cases have been obtained after expensive and laborious procedures. The ease of data exchange through the Interne has compounded the need to assemble technological mechanisms for effectively protecting one's intellectual or pragmatic property. Trajectories abound in applications such as GPS tracking experiments, video and motion capture data, and even image shapes can be considered as 2-dimensional

trajectories. We provide ownership assurances on such datasets using watermarking principles. While there is a rich literature on watermarking for multimedia datasets, previous work is primarily concerned with watermarking a single object and not a collection of objects. Here, we consider the watermarking problem from a new perspective, by focusing on the additional maintenance of the inter-relationship between objects.

The technique embeds a secret key in each of the dataset objects, distorting them imperceptibly, while taking special consideration in retaining the original neighboring object. We call this operation Neighbor Preserving (NP) watermarking. Guaranteeing preservation of the nearest objects is very important for an array of search and mining operations, such as similarity search or Nearest-Neighbor(NN)-classification. Contrary to privacy-preserving approaches for data-mining that first add noise and then reconstruct the original data distributions based on the known noise model, our approach learns/calculates the largest amount of noise that can be added, so that nearest neighbors are not distorted. While the naive algorithms for determining the said watermarking power are costly, we show efficient ways of speeding up the process making it more than 2 orders of magnitude faster, thus allowing the technique to be applicable to large datasets.

Vıctor R. Doncel, Nikos et al [2000] [3] describe a watermark is a hidden information within a digital signal, used primarily for copyright protection of multimedia data. Its main features are the imperceptibility of the imposed medications and its persistence against processing (attacks) that may result in its removal, either intentionally or unintentionally. A general framework for digital water marking has been presented, whereas provides an excellent overview of the watermarking principles and techniques. Digital watermarking has been mainly applied to still image, audio and video data. However, little work has been done in watermarking vector graphics data, that are typically used in Geographic Information Systems (GIS) or in Computer Aided Design (CAD).

Digital watermarking is a polygonal line, which are a key graphics primitive in vector graphics data and thus can be used for the copyright protection of such data. Furthermore, the method can be used for the watermarking of MPEG-4 natural video data by watermarking the outline of the Video Objects in MPEG-4 stream. In that case, the method should be accompanied by a way of extrapolating existing textures in case the watermarked boundary defines a bigger area than the original one. This paper extends the work presented. The same embedding method is adopted here, and efforts focus the design of a new, enhanced

performance detector. Theoretical and experimental analysis show that a substantial improvement in detection performance can be achieved if the statistics of the watermarked polygon are considered.

Rakesh Agrawal and Jerry Kiernan et al [2006] describe watermarking database relations to deter their piracy, identify the unique characteristics of relational data which pose new challenges for watermarking, and provide desirable properties of a watermarking system for relational data. A watermark can be applied to any database relation having attributes which are such that changes in a few of their values do not affect the applications. An effective watermarking technique geared for relational data. This technique ensures that some bit positions of some of the attributes of some of the tuples contain specific values. The tuples, attributes within a tuples, bit positions in an attribute, and specific bit values are all algorithmically determined under the control of a private key known only to the owner of the data.

N. F. Johnson, Z. Duric [2000] describe the piracy of digital assets such as software, images, video, audio and text has long been a concern for owners of these assets. Protection of these assets is usually based upon the insertion of digital watermarks into the data. The watermarking software introduces small errors into the object being watermarked. These intentional errors are called marks and all the marks together constitute the watermark. The marks must not have a significant impact on the usefulness of the data and they should be placed in such a way that a malicious user cannot destroy them without making the data less useful. Thus, watermarking does not prevent copying, but it deters illegal copying by providing a means for establishing the original ownership of a redistributed copy. The increasing use of databases in applications beyond "behind-the-firewalls data processing" is creating a similar need for watermarking databases.

SYSTEM ANALYSIS

EXISTING SYSTEM AND ITS DRAWBACKS

The existing system uses a spread-spectrum approach. This embeds the watermark across multiple frequencies of each object and across multiple objects of the dataset. As such, it renders the removal of the watermark particularly difficult without substantially compromising the data utility. The data locations are altered before applying the watermarking.

The robustness of the watermark embedding depends on the choice of coefficients. The watermark is embedded in the coefficients that exhibit, on average over the dataset, the largest Fourier magnitudes. This makes the removal of the watermark difficult. However, during the data distribution, all kind of receiver users receive the same data.

DRAWBACKS

- Data about the receiving user is not embedded in watermarking information.
- Watermark data applied on image data only.
- Simulation is applied on image data only.

PROPOSED SYSTEM AND ITS ADVANTAGES

Like the existing system, proposed system also uses watermarking without altering the KNN property. In addition, numeric data set is choosen for applying the watermark without altering the KNN property. In addition, if watermark data is corrupted, it can be found out.

ADVANTAGES

- Watermarking is applied in both image and numeric data set.
- Data about the receiving user is also embedded in watermarking information.
- Watermark corrupted information can be found out.

PROBLEM DEFINITION

This thesis presents a new spectral hiding data clustering method called novel KNN properties, which is performed in the data publishing with distances. In this framework, the data publishing are projected into a high dimensional level in which the distances between the patients records are maximized data sets. The problem of literature review is a spread-spectrum approach and embeds the watermark across multiple frequencies of each object and across multiple objects of the dataset. As such, it renders the removal of the watermark particularly difficult without substantially compromising the data utility.

The data locations are altered before applying the watermarking. The robustness of the watermark embedding depends on the choice of coefficients. The watermark is embedded in the coefficients that exhibit, on average over the dataset, the largest Fourier magnitudes. This makes the removal of the watermark difficult. However, during the data distribution, all kind of receiver users receive the same data. The propose system is problems solve in watermarking without altering the KNN property. In addition, numeric data set is choosing for applying the watermark without altering the KNN property. In addition, if watermark data is corrupted, it can be found out.

OVERVIEW OF PROJECT

Discover to right-protect a dataset, but at the same time guarantee preservation of the outcome of important distance-based mining operations. In the approach provide two variants: one that preserves Nearest-Neighbors (NN) and another that preserves the Minimum Spanning Tree (MST). Therefore, the output of any algorithm based on these two properties will be preserved after right protection. To guarantee this, we study the critical watermark intensity to both protect the dataset, as well as ensure that important parts of the object distance graph are not distorted.

It is essential to discover the maximum watermark intensity for right protection. This provides assurances of better detectability and hence better security for the right protection scheme. We first study how (Euclidean) distances between the objects are distorted as a function of the watermark embedding strength. This gives us insight on how to design fast variants of our algorithms that still guarantee

preservation of the NN and the MST, but operate significantly faster than the exhaustive algorithms

MODULE DESCRIPTION

The following modules are present in the project.

1. PATIENT PROFILE ADDITION

2. PATIENT OBSERVATION ENTRY

3. WATERMARK CONTENT ADDITION

4. EMBED WATERMARK DATA IN PATIENT OBSERVATION NUMERIC DATA SET

5. EXTRACT WATERMARK DATA IN PATIENT OBSERVATION NUMERIC DATA SET AFTER KNN CHECK

6. IMAGE ADDITION

7. EMBED WATERMARK DATA IN PATIENT OBSERVATION IMAGE

8. EXTRACT WATERMARK DATA IN PATIENT OBSERVATION IMAGE AFTER KNN CHECK

1. PATIENT PROFILE

In this module, the patient details such as patient id, name, gender, date of birth, entry date, address, mobile, occupation, annual income, father name, guardian name, symptom and previous treatment taken details are keyed in and saved in to 'Patients' table.

2. PATIENT OBSERVATION DETAILS

In this module, the patient id is selected, entry date becomes today date, test observation data 1, test observation data 2 and test observation 3 are keyed in and saved in to 'Observations' table.

3. WATERMARK CONTENT ADDITION

In this module, the watermark content details are added. The details are saved in 'Watermark' Table.

4. EMBED WATERMARK DATA IN PATIENT OBSERVATION NUMERIC DATA SET

In this module, the watermark content details are converted into bytes and stored in patient observations third column along with a numeric value 301 is added. The first observation values are taken as X position and second observation values are taken as Y position and are pointed initially. Then the first watermark byte value is added with X and then the X Position is modified. This repeats for all watermark bytes (each one watermark byte is stored in one patient's all observation data). The process is listed in embed steps of algorithm 1 of algorithms section.

5. EXTRACT WATERMARK DATA IN PATIENT OBSERVATION NUMERIC DATA SET AFTER KNN CHECK

In this module, the modified patient data set is taken and record values for third observation column values greater than 301 is filtered out. Then for each patient, the third observation column value is subtracted with 301 and the numeric value's character is found out (ascii value). This repeats for all the patients and watermark content is appended. The result watermark content is displayed. The KNN value is found out before and after watermarking and checks for same result display. The process is listed in extract steps of algorithm 1 of algorithms section.

6. IMAGE ADDITION

In this module, the image is browsed and saved in 'Images' table.

7. EMBED WATERMARK DATA IN PATIENT OBSERVATION IMAGE

In this module, the watermark content details are converted into bytes and stored in last image browsed and saved in project folder. The process is listed in embed steps of algorithm 2 of algorithms section.

8. EXTRACT WATERMARK DATA IN PATIENT OBSERVATION IMAGE AFTER KNN CHECK

In this module, the watermark content details are found out from the embedded image. The process is listed in extract steps of algorithm 2 of algorithms section.

ALGORITHMS USED

ALGORITHM 1

NN-Preservation Algorithm For Image

Embed Steps:

Input: Image, Watermark Data

Output: Embedded Image, Extracted Watermarked Data

1. Select the Image (I).
2. Enter watermark content (W).
3. Find the Pixels (PG-Pixel Group) where Red component values falls from 137 to 147, 157 to 167 and 187 to 207.
4. Convert the watermark data to bytes and find the length of watermark data (L).
5. In the first pixel of the PG, store the 'L' value in Blue component.
6. Then store all the 'W' bytes in Blue component one by one starting from second pixel to all other successive pixels in PG group until the 'W' bytes are completed.

Extract Steps:

Input: Embedded Image

Output: Extracted Image, Extracted Watermarked Data

1. Select the Image (Watermark Embedded) (I).
2. Find the Pixels (PG-Pixel Group) where Red component values falls from 137 to 147, 157 to 167 and 187 to 207.
3. From the first pixel of the PG, get the 'L' value from the Blue component.
4. Fetch all the 'W' bytes from Blue components one by one starting from second pixel to all other successive pixels in PG group until the 'W' bytes are completed.
5. Convert the watermark bytes to data.
6. Check the KNN Property.
7. Output Watermark Data.

ALGORITHM 2

NN-Preservation Algorithm for Numeric Data Set

mbed Steps:

Input: Patient Observations, Watermark Data

Output: Modified Patient Observation Data

1. Add the Patient Profiles (P).
2. Add the Patient Observation Data (O).
3. Enter watermark content (W).

4. Convert the watermark data to bytes and find the length of watermark data (L).
5. Sort the Patient Observation Data (O) Patient wise.
6. I=0
7. For Each Patient's Observation Set in (O)
8. Alter the Observation Data's third value such that OD(3) = 301 + W(I)
9. Change the OD(1) position = OD(1) position + W(I)
10. I=I+1
11. If I>=L Then
12. Break
13. End If
14. Next
15. Output the New Patient Data Set.

Extract Steps:

Input: Modified Patient Observation Data

Output: Patient Observation Data, Extracted Watermarked Data

1. Select the Patient Data Set (where Watermark Data Embedded) (P).
2. I=0;
3. For Each Patient's Observation Set in (O)
4. W(I) = Observation Data's third value - 301
5. Change the OD(3) value= OD(3) value -301
6. If I=0 Then
7. L= W(I)
8. End If
9. I=I+1
10. If I >L Then
11. Break
12. End If
13. Next
14. Convert the watermark bytes to data.
15. Check the KNN Property.
16. Output Watermark Data.

CONCLUSION AND FUTURE WORKS

CONCLUSION

The project uses watermarking without altering the KNN property. Numeric data set is chosen for applying the watermark without altering the KNN property. In addition, if watermark data is corrupted, it can be found out. Watermarking is applied in both image and numeric data set. Data about the receiving user is also embedded in watermarking information. Watermark corrupted information can be found out. The proposed watermarking methodology preserves the Nearest Neighbors (NN) property of each object of the original dataset. This leads to preservation of any mining operation that depends on the ordering of distances between objects, such as NN-search and classification, as well as many visualization techniques. It proves fundamental lower and upper bounds on the distance between objects post-watermarking.

SCOPE FOR FUTURE ENHANCEMENTS

The future work of watermarking methodology will be studying and preserving the Minimum Spanning Tree (MST) property of each object of the original dataset. The analysis will be focused on the above area to provide fast versions of the exhaustive algorithms. The future work will demonstrate impressive speed-up in all experiments. This analysis will be used to design fast algorithms for MST-preserving watermarking that drastically prunes the vast search space

REFERENCE

1. P. Das, N. R. Chakraborti, and P. K. Chaudhuri, "Spherical mini-max location problem ," Comput. Optim. Appl., vol. 18, no. 3, pp. 311–326, 2001

2. G. Economou, V. Pothos, and A. Ifantis, "Geodesic distance and MST-based image segmentation," in Proc. 12th EUSIPCO, Vienna, Austria, 2004, pp. 941–944.

3. I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," IEEE Trans.Image Process., vol. 6, no. 12, pp. 1673–1687, Dec. 2000.

4. C. G. Atkeson, A. W. Moore, and S. Schaal, "Locally weighted learning," Artif. Intell. Rev., vol. 11, pp. 11–73, Feb. 2006.

5. N. Paivinen, "Clustering with a minimum spanning tree of scale-free-like structure," Pattern Recognit. Lett., vol. 26, no. 7, pp. 921–930, 2006.

6. Y. Xu, V. Olman, and D. Xu, "Minimum spanning trees for gene expression data clustering," Genome Inform., vol. 12, pp. 24–33, 2001.

7. J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geo-metric framework for nonlinear dimensionality reduction," Sci., vol. 290, no. 5500, pp. 2319–2323, 2000.

8. M. Vlachos, C. Lucchese, D. Rajan, and P. S. Yu, "Ownership protection of shape datasets with geodesic distance preservation," in Proc. 11th Int. Conf. EDBT, Nantes, France, 2008, pp. 276–286.

9. J.-P. M. G. Linnartz and M. van Dijk, "Analysis of the sensitivity attack against electronic watermarks in images," in Proc. 2nd Int. Workshop IH, Portland, OR, USA, 1998, pp. 258–272.

10. F. Hartung, J. Su, and B. Girod., "Spread spectrum watermarking: Malicious attacks and counterattacks," in Proc. SPIE Security Watermarking Multimedia Contents, vol. 3657, San Jose, CA, USA,1999.

11. V. Solachidis and I. Pitas, "Watermarking polygonal lines usingFourier descriptors," IEEE Comput. Graph. Appl., vol. 24, no. 3, pp. 44–51, May/Jun. 2004.

12. M. Vlachos, B. Taneri, E. J. Keogh, and P. S. Yu, "Visual exploration of genomic data," in Proc. 11th Eur. Conf. PKDD, vol. 4702, Warsaw, Poland, 2007, pp. 613–620.

13. S. J. Shyu, Y. T. Tsai, and R. C. T. Lee, "The minimal spanning tree preservation approaches for DNA multiple sequence alignment and evolutionary tree construction," J. Combinat. Optim., vol. 8, no. 4, pp. 453–468, 2004.