

AN EFFECTIVE SEMANTIC ENCRYPTED RELATIONAL DATA USING K-NN MODEL

Subhashini Padma.M and Mrs.Vanitha.P

M.E.Scholar, Department of Computer Science & Engineering Nandha Engineering
College, Erode, Tamil Nadu, India

Assistant Professor, Department of Computer Science & Engineering, Nandha
Engineering College, Erode, Tamil Nadu, India

ABSTRACT

Data exchange and data publishing are becoming an important part of business and academic practices. Data owners need to maintain the rights over the datasets they share. A right-protection mechanism can be provided for the ownership of shared data, without revealing its usage under a wide range of machine learning and mining. In the approach provide two algorithms: the Nearest-Neighbors (NN) and determiner preserves the Minimum Spanning Tree (MST). The K-NN protocol guarantees that relations between object remain unaltered. The algorithms preserve the both right protection and utility preservation. The right-protection scheme is based on watermarking. Watermarking methodology preserves the distance relationships.

KEYWORDS

k-NN classifier, Watermarking, Privacy preserving

1. INTRODUCTION

Data mining referred to as Knowledge Discovery from Data (KDD), in which extraction of patterns representing knowledge is stored in large databases. Data mining is a multidisciplinary field, draws work from areas including database technology and machine learning. Privacy preserving and security are becoming a major issue in data mining. When data are highly sensitive, the data need to be encrypted before obtaining. When data are encrypted by performing any data mining tasks it becomes very challenging without ever decrypting the data. The users now have the opportunity to obtain the data form as well as the data mining tasks to the cloud.

2. LITERATURE SURVEY

The literature review includes Privacy preserving methodology. By using different methodologies we can eliminate the data privacy and right protection.

2.1 Optimal- Fourier Descriptors Domain

V. Solachidis and I. Pitas [1] describes Polygonal lines are key graphical primitives in vector graphics. In addition, plane lines can be used to define the boundary of Video Objects. The ability to apply a digital watermark in which the benefits of copyright are achieved to a wide range such as Geographical Information Systems data or MPEG-4 video. It builds on and extends the contour watermarking algorithm proposed. The detection of such watermarks is been carried out through a correlator detector. In this paper, the statistics of the Fourier Descriptors are appraised, and it is to analysis and utilized to devise an optimal detector, designed according to the Bayesian

decision theory. A watermark is hidden information within a digital signal, used predominantly used for the right protection. The main method are the so slight of the imposed modifications and its persistence against processing that may result in its removal, is done intentionally or by unintentionally. A general framework for digital watermarking has been presented, in which it is provided with an excellent overview of the watermarking principles and its techniques.

2.2 Segmentation with Geodesic Distance

George Economies, Vassilios Pathos and Apostolos Ifantis [2] propose a new approach for the integration of spatial approximation information in which graph based segmentation algorithms. This is done by means of the geodesic distance. The calculation and implementation is done using the Minimal Spanning Tree, constructed on a watershed image partition. Distance is given with MST edges it presents a measure of both spatial and feature coherence. It is incorporated in MST based on color image segmentation applications, by means of a new density feature, which is computed with spatial location in which it is about to come to an end. The image analysis is done in which has two features that are usual embeddings used for representation of visual data.

2.3 Spanning Tree with Multiple Accessing

Michail Vlachos¹, Bahar Taneri, Eamonn Keogh, and Philip S. Yu [3], present methods for DNA sequences in which it has two forms. The DNA method calls out for a result that gives a transformation of gene method can be changed into numerical form. The course captures the given content that allows for fast moving and easy viewing visualization of long DNA sequences. The project relatively gives a route on the 2D plane using a spanning-tree in which they are arranged that allows the efficient comparison of multiple sequences. They demonstrate with various examples. The applicability of the technique in evolutionary biology is specified in which expressing along with envisage the molecular phylogeny between species. Identification of evolutionary far with the species that has always been a interest for the researchers. There are different methods in which the evolutionary relationships of the species can be understood and through the DNA approach is can be studied.

2.4 Watermarking Method with Data Privacy

Rakesh Agrawal Jerry Kiernan [4], discuss the need for watermarking database relations is to give the privacy, identify the unique characteristics of relational data which possess challenges for watermarking and gives some desirable system for relational data. A watermark is implemented or can be possessed to any database relation having attributes which may change with values that do not affect the applications that are used in watermarking. They present an effective watermarking technique geared for relational data. The technique ensures that a positions of the attributes of some of the tuples that contain some values. The tuples within a tuple a small positions in an attribute, and specific bit values are all algorithmically defined with the control of a private key which can be used and accessed only by the owners of the data. This pattern gives the watermark. If one needs to access the privately used key detection of watermark is easy. Detecting the watermark need to access the original data or the watermark. Even a small subset can be detected using watermark with some samples.

3. WATERMARK DETECTION

The detection process in which goal is to the presence of a watermark in particular in a watermarked dataset. It gives to plan the correlation between a tested watermark and the watermarked dataset. The higher the correlation between the two, the higher the probability that the embedded watermark was the one tested. Because the watermark is embedded that objects of a dataset that one option is to quantify the correlation between the watermark and the average

value of the descriptors across all objects of the dataset is given. However directly measuring of the correlation is not so effective under multiplicative embedding. The reason is that since we want to minimize distortion, a small embedding power is preferred, when the magnitudes of the Fourier descriptors are controlled by the original level of the average.

3.1 TYPES OF WATERMARKING DETECTION

This section gives out the types of watermarking and their detection

3.1.1. Audio Watermarking Detection

3.1.2. Digital Watermarking Detection

3.1.3. Stamp Watermarking Detection

3.1.4. Secure Watermarking Detection

3.1.1. Audio Watermarking Detection

This watermarking is electronic in which it is implanted with the audio signal which is used to get the copyright which is same as photograph on watermarking.

3.1.2. Digital Watermarking Detection

This kind of watermarking is a marker in which implanted with noise tolerant for example audio or video.

3.1.3. Stamp Watermarking Detection

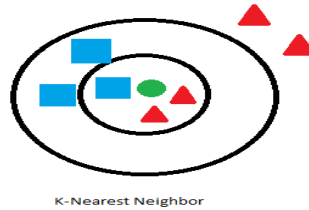
The watermark in which a stamp watermarking, that are used to find the contract between a common and a rare stamp.

3.1.4. Secure Watermarking Detection

This technique is used to provide the confidentiality and privacy protection of once secure data.

4. K-NEAREST NEIGHBORS ALGORITHMS

The **K-Nearest Neighbors algorithm** (or **k-NN** for short) it is a non-parametric that is used especially with the classification and regression. In regression and classification, the input has k closest training feature space. The output belongs on whether k -NN which is used for classification or regression: In k -NN classification, class membership is a object, and the object is classified by majority neighbors vote, with the object being assigned to the class most common among its k nearest neighbors (k is + integer, if $k = 1$, then the object is simply assigned to the class of that single nearest neighbor. In k -NN regression, the object is the property value of the output. The value is the average of its nearest neighbors. k -NN is a instance-based learning in which function is localized and computation gets different until classification gets completed. This algorithm that is the simple algorithm of all machine learning. For classification and regression, the algorithm is useful to weight the vote of the neighbors so that the nearer neighbors gives much to the average than the information which are far. Consider an example, weighting method contains each neighbor a weight of $1/d$, where d is the distance of the neighbor. From the set of objects the neighbors are taken and the value is known . This can be done of the data set which are of training set in which though no explicit training step is needed.



4.1. ALGORITHM 1

4.4.1 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.

4.2 ALGORITHM 2

4.4.1 NN-Preservation Algorithm for Numeric Data Set

Embed 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 $I \geq L$ 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. $W(I) = \text{Observation Data's third value} - 301$
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.

5. RELATED WORK

The work utilise the spread-spectrum method. This gives that the watermarking is used across multiple frequencies of object and across objects of the dataset. It involves the taking out of the watermark is difficult without compromising the data utility. The watermark is embedded in the coefficients that exhibit, over the dataset, the largest Fourier magnitudes.

In right-protect a dataset, at the same time guarantee preservation of the outcome of important distance-based mining operations. In the approach provide two variants: one that preserves K-Nearest Neighbors (NN) and next that the Minimum Spanning Tree (MST). The output of this algorithm is to provide the right protection and then to preserve the privacy. For that privacy and right protection we have to have a look at the critical watermark intensity of both protection of the dataset and to guarantee that some parts of the distance graph are not wrapped.

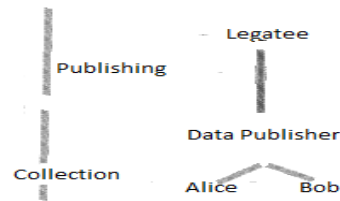


Fig.1.1 Data Collection and Publishing

6. CONCLUSIONS

The watermarking is used without altering the KNN property. Numeric data set is chosen in which the K-NN property remains the same when watermarking is applied, if data is corrupted, it can be easily got. Watermarking is given out in image data set and numeric data set.

REFERENCES

- [1] JemmaWua”A Framework For Learning Comprehensible Theories In Xml Document Classification” IEEE Transactions On Knowledge And Data Engineering VOL. 24, NO. 1, JANUARY 201
- [2] Robert Varga, S. Marie Matheson, and Andrew Hamilton-Wright” Aggregate Features in Multi sample Classification Problems” IEEE Journal of Biomedical and Health Informatics VOL. 19, NO. 2, MARCH 2015
- [3] Kartick Subramanian, Sundaram Suresh and NarasimhanSundrarajan”A Metacognitive Neuro-Fuzzy Inference System (McFIS) for Sequential Classification Problems”IEEE Transactions on Fuzzy Systems, VOL. 21, NO. 6, DECEMBER 2013
- [4] AnirbanMukhopadhyay, UjjwalMaulik and SanghamitraBandyopadhyay” A Survey of Multiobjective Evolutionary Algorithms for Data Mining: Part I” IEEE Transactions On Evolutionary Computation, VOL. 18, NO. 1, FEBRUARY 2014
- [5] JayasimhaRao, NooraPartamies, Olga Amariutei, MikkoSyrjäsuo, and Koen E. A. van de Sande” Automatic Auroral Detection in Color All-Sky Camera Images” IEEE Journal Of Selected Topics In Applied Earth Observations And Remote Sensing, VOL. 7, NO. 12, DECEMBER 2014
- [6] Paolo Melillo, Nicola De Luca, Marcello Bracale, and Leandro Pecchia” Classification Tree for Risk Assessment in Patients Suffering From Congestive Heart Failure via Long-Term Heart Rate Variability” IEEE Journal Of Biomedical And Health Informatics, VOL. 17, NO. 3, MAY 2013
- [7] Tomohiro Shiraishi, Takeshi Motohka,Rajesh BahadurThapa, Manabu Watanabe”Comparative Assessment of Supervised Classifiers for Land Use–Land Cover Classification in a Tropical Region Using Time-Series PALSAR Mosaic Data” IEEE Journal Of Selected Topics In Applied EarthObservations And Remote Sensing, VOL. 7, NO. 4, APRIL 2014
- [8] Souneil Park, Jungil Kim, Kyung Soon Lee and Junehwa Song”Disputant Relation- Based Classification for Contrasting Opposing Views of Contentious News Issues” IEEE Transactions On Knowledge And Data Engineering, VOL. 25, NO. 12, DECEMBER 2013
- [9] IreneuszCzarnowski,PiotrJe,drzejowicz” Ensemble classifier for mining data streams” International Conference on Knowledge- Based and Intelligent Information & Engineering Systems ,Vol. 10, , No 1,2014
- [10] Rodrigo C. Barros, Márcio P. Basgalupp, Alex A. Freitas, and André C. P. L. F. De Carvalho”Evolutionary Design of Decision- Tree Algorithms Tailored to Microarray Gene Expression Data Sets” IEEE Transactions On Evolutionary Computation, VOL. 18, NO. 6, DECEMBER 2014
- [11] Ben D.FulcerAnd Nick S.Jones “Highly Comparative Feature-Based Time-Series Classification” IEEE Transcations On Knowledge And Data Engineering VOL. 26, NO.12, DECEMBER 2014

- [12] Bharath K. Samanthula, YousefElmehdwi, and Wei Jiang” k-Nearest Neighbor Classification over Semantically Secure Encrypted Relational Data”IEEE Transactions On Knowledge And Data Engineering,2014
- [13] Aaron M. Swan and David G. Long”Multiyear Arctic Sea Ice Classification Using QuikSCAT”IEEE Transactions on Geoscience and Remote Sensing, VOL. 50, NO. 9, SEPTEMBER 2012
- [14] Xiaoshuang Ma,”Polarimetric-Spatial Classification of SAR Images Based on the Fusion of Multiple Classifiers”IEEE Journal Of Selected Topics In Applied Earth Observations And Remote Sensing, VOL. 7, NO. 3, MARCH 2014

AUTHORS

M.Subhashini Padma received the B.E. degree in Computer science and Engineering from Velalar College of Engineering in 2014.She is currently doing her M.E Computer science and Engineering in Nandha engineering college, Erode, India



P.Vanitha, (AP/CSE) working as Assistant Professor in Nandha Engineering College. She has published many national and international research papers. She has depth of knowledge in her research areas.

