International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622
NATIONAL CONFERENCE on Developments, Advances & Trends in Engineering Sciences
(NCDATES- 09th & 10th January 2015)

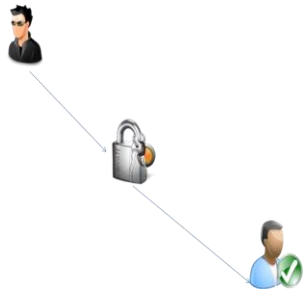
RESEARCH ARTICLE OPEN ACCESS

Theoretic Information on Utility-Privacy Exchange in Databases

A.Ravi Kishore, T.Sree Latha, K.Niveditha
Student,M.Tech Bapatla Engg College Bapatla
Student,Mtech Bapatla Engg College Bapatla
Student,Mtech Bapatla Engg College apatla

Abstract:
Ensuring the usefulness of electronic data sources while providing necessary privacy guarantees is an important unsolved problem. This problem drives the need for an analytical framework that can quantify the privacy of personally identifiable information while still providing a quantifable benefit (utility) to multiple legitimate information consumers. This paper presents an information-theoretic framework that promises an analytical model guaranteeing tight bounds of how much utility is possible for a given level of privacy and vice-versa. Specific contributions include: i) stochastic data models for both categorical and numerical data; ii) utility-privacy tradeoff regions and the encoding (sanization) schemes achieving them for both classes and their practical relevance; and iii) modeling of prior knowledge at the user and/or data source and optimal encoding schemes for both cases.

Architecture:



EXISTING SYSTEM:

We divide the existing work into two categories, heuristic and theoretical techniques, and outline the major milestones from these categories for comparison. The earliest attempts at systematic privacy were in the area of census data publication where data was required to be made public but without leaking individuals' information. A number of ad hoc techniques such as sub-sampling, aggregation, and suppression were explored. The first formal definition of privacy was k-anonymity by Sweeney. However k-anonymity was found to be inadequate as it only protects from identity disclosure but not attribute-based disclosure and was extended with t-closeness and l-diversity . All these techniques have proved to be non-universal as they were only robust against limited adversaries. Heuristic techniques for privacy in data mining have focused on using a mutual information-based privacy metrics.

PROPOSED SYSTEM:

Our work is based on the observation that large datasets (including databases) have a distributional basis; i.e., there exists an underlying (sometimes implicit) statistical model for the data. Even in the case Of data mining where only one or a few instances of the dataset are ever available, the use of correlations between attributes used an implicit distributional assumption about the dataset. We explicitly model the data as being generated by a source with a finite or infinite alphabet and a known distribution. Each row of the database is a collection of correlated attributes (of an individual) that belongs to the alphabet of the source and is generated according to the probability of occurrence of that letter (of the alphabet). Our statistical model for databases is also motivated by the fact that while the attributes of an individual may be correlated, the records of a large number of individuals are generally independent or weakly correlated with each other. We thus model the database as a collection of n observations generated by a memory less source whose outputs are independent and identically distributed.

Modules :

1. **Registration**
2. **Login**
3. **Admin**
4. **Encryption and Decryption**
5. **Chart_view**

Modules

Description

Registration:

In this module Sender/User have to register first, then only he/she has to access the data base.

Login:

In this module, any of the above mentioned person have to login, they should login by giving their email id and password .Admin login by giving username and password

Admin:

Admin can see the details of the people who are published their personal data. Data are in encrypted form. He then decrypt it by using decryption and then only he will be able to see the original data

Encryption and Decryption Code:

```
public class EBCDIC
{
public static void main(String arg[])
      {
EBCDIC a=new EBCDIC();
      System.out.println("EBCDIC:"+
a.decrypt(a.encrypt("abcdhello")));
      }
public static String encrypt(String str)
      {
      byte b[] = new byte[str.length()];
      byte result[] = new byte[str.length()];
//    byte mod[] = new byte[str.length()];
      b=str.getBytes();
      for(int i=0;i<str.length();i++)
            {
            result[i] = (byte) ((byte) b[i] -(byte) 4);
//mod[i]=(byte) ((byte) b[i] % (byte) 4);

            System.out.println(b[i]+"-"+result[i]);
            }
      return ( new String(result) );
      }
public static String decrypt(String str)
      {
      byte b[] = new byte[str.length()];
      byte result[] = new byte[str.length()];
      b=str.getBytes();
      for(int i=0;i<str.length();i++)
            {
            result[i] = (byte) ((byte) b[i]+(byte) 4);
            System.out.println(b[i]+"-"+result[i]);
            }
      return ( new String(result) );
      }
}
```

Chart_View:

The Receiver can only view the senders personal data by pictorial representation i.e chart.Chart will be prepared by applying the senders input.Also he can see the personal data in encrypted form.Registered users only can decrypt the data.We hide the correct income of the senders who pass the data to receivers.Receivers will be able to see the actual income of senders by applying some side informations.

CONCLUDING REMARKS:

The ability to achieve the desired level of privacy while guaranteeing a minimal level of utility and vice-versa for a general data source is paramount. Our work defines privacy and utility as fundamental characteristics of data sources that may be in conflict and can be traded off. This is one of the earliest attempts at systematically applying information theoretic techniques to this problem. Using rate-distortion theory, we have developed a U-P tradeoff region for i.i.d. data sources with known distribution. We have presented a theoretical treatment of a universal (i.e. not dependent on specific data features or adversarial assumptions) theory for privacy and utility that addresses both numeric and categorical (non-numeric) data. We have proposed a novel notion of privacy based on guarding existing uncertaintyabout hidden data that is intuitive but also supported by rigorous theory. Prior to our work there was no comparable model that applied to both data types, so no side-by-side comparisons can be made across the board between different approaches. The examples developed here are the first step towards understanding

Result:

The ability to achieve the desired level of privacy while guaranteeing a minimal level of utility and vice-versa for a general data source is paramount. Our work defines privacy and utility as fundamental characteristics of data sources that may be in conflict and can be traded off. This is one of the earliest attempts at systematically applying information theoretic techniques to this problem. Using rate-distortion theory, we have developed a U-P tradeoff region for i.i.d. data sources with known distribution.

REFERENCES

[1] H. Yamamoto, "A source coding problem for sources with additional outputs to keep secret from the receiver or wiretappers," IEEE Trans. Inform. Theory, vol. 29, no. 6, pp. 918–923, Nov. 1983.

[2] L. Sankar, S. R. Rajagopalan, S. Mohajer, and H. V. Poor, "Smart meter privacy: A theoretical framework," IEEE Trans. Smart Grid, no. 99, pp. 1 –10, 2012, early access article.

[3] L. Sweeney, "k-anonymity: A model for protecting privacy," Intl. J. Uncertainty,

Fuzziness, and Knowledge-based Systems, vol. 10, no. 5, pp. 557–570, 2002.

[4] C. Dwork, "Differential privacy," in Proc. 33rd Intl. Colloq. Automata, Lang., Prog., Venice, Italy, Jul. 2006.

[5] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," in Proc. IEEE Intl. Symp. Security and Privacy, Oakland, CA, May 2008, pp. 111–125.

[6] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "An information-theoretic approach to privacy," in Proc. 48th Annual Allerton Conf. on Commun., Control, and Computing, Monticello, IL, Sep. 2010, pp. 1220–1227.

[7] T. E. Raghunathan, J. P. Reiter, and D. B. Rubin, "Multiple imputation for statistical diclosure limitation," IEEE Trans. Inform. Theory, vol. 43, no. 6, pp. 1877–1894, Nov. 1997.

[8] A. Dobra, S. Fienberg, and M. Trottini, Assessing the Risk of Disclosure of Confidential Categorical Data. Oxford University Press, 2000, vol. 7, pp. 125-144.

[9] S. Chawla, C. Dwork, F. McSherry, and K. Talwar, "On privacy-preserving histograms," in Proc. 21st Conf. Uncert. Art. Intell., Edinburgh, Scotland, Jul. 2005.

[10] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "L-diversity: Privacy beyond k-anonymity," ACM Trans. Knowledge Discov. Data, vol. 1, no. 1, 2007.

[11] D. Agrawal and C. Aggarwal, "On the design and quantification of privacy preserving data mining algorithms," in Proc. 20th Symp. Principles of Database Systems, Santa Barbara, CA, May 2001.

[12] C. Dwork, "A firm foundation for private data analysis," Jan. 2011, http://research.microsoft.com/apps/pubs/?id=116123.

[13] G. Jagannathan, K. Pillaipakkamnatt, and R. N. Wright, "A practical differentially private random decision tree classifier," in Proc. ICDM Intl. Workshop Privacy Aspects of Data Mining, Miami, FL, Dec. 2009.

[14] T. Li and N. Li, "On the tradeoff between privacy and utility in data publishing," in Proc. 15th ACM SIGKDD Intl. Conf. Knowledge discovery and data mining, Paris, France, 2009, pp. 517–526.

[15] M. Alvim and M. Andr´es, "On the relation between differential privacy and quantitative information flow," in Proc. 38th Intl. Conf. Automata, Languages and Programming - Volume Part II, Zurich, Switzerland, 2011, pp. 60–76.

[16] F. du Pin Calmon and N. Fawaz, "Privacy against statistical inference," in Proc. 50th Annual Allerton Conf. on Commun., Control, and Computing, Monticello, IL, Sep. 2012.

[17] T. M. Cover and J. A. Thomas, Elements of Information Theory. New York: Wiley, 1991.

[18] H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar, "Random data perturbation techniques and privacy preserving data mining," J. Know. Inform. Sys., vol. 7, no. 4, pp. 387–414, May 2005.

[19] D. Rebollo-Monedero, J. Forne, and J. Domingo-Ferrer, "From t-closeness-like privacy to postrandomization via informationtheory," IEEE Trans. Knowledge Data Engg., vol. 22, no. 11, pp. 1623–1636, Nov. 2010.

[20] R. Tandon, L. Sankar, and H. V. Poor, "Lossy discriminatory source coding: Side-information privacy," May 2011, under revision; arXiv:1106.2057.

[21] J. T. Pinkston, "An application of rate-distortion theory to a converse to the coding theorem," IEEE Trans. Inform. Theory, vol. 15, no. 1, pp. 66–71, Jan. 1969