Studies on Protecting Privacy of Anonymized Medical Data

T.Kowshiga¹, T.Saranya², T.Jayasudha³, Prof.M.Sowmiya⁴ and Prof.S.Balamurugan⁵

Department of IT, Kalaignar Karunanidhi Institute of Technology, Coimbatore, TamilNadu, India¹,²,³,⁴,⁵

ABSTRACT: This paper details about various methods prevailing in literature for protecting privacy of anonymized medical data. Ontology Based measure to compute semantic similarity in Biomedicine is studied. Ordinal, continuous and heterogeneous K-Anonymity through Microaggregation are dealt in detail. Protecting patient privacy by quantifiable control of disclosure in disseminated databases and achieving k-Anonymity privacy protection using generalization and suppression are discussed in detail. Efficient Multivariate data-Oriented Micro aggregation of Categorical data for confidential documents is examined. Differential Privacy for Automatic De-Identification of textual documents in the electronic health records and Statistical Disclosure control for patient records in biomedical information System is considered. Density-based microaggregation for statistical disclosure control and anonymization of Set-Valued Data via Top-Down, Local Generalization are also aggregated in brief. This paper would promote a lot of research in the area of protecting privacy of anonymized medical data.

KEYWORDS: Ontology, Microaggregation, Differential Privacy, De-Identification, Biomedical Information System, Anonymous Authentication.

I. INTRODUCTION

Cloud computing offers unique opportunities for supporting long-term record preservation. MyPHRMachines [1], a patient owned health record system prototype based on remote virtual machines hosted in the cloud. MyPHRMachines is particularly promising for countries with a very heterogeneous architecture of systems across hospitals and other care institutions. In the view of developer PHRs should be portable. PHR systems typically offer functionality to share, visualize and analyze PHR data. Secure lifelong management of patient medical records since data are stored in the cloud and do not have to be carried around by patients.

II. ONTOLOGY BASED MEASURE TO COMPUTE SEMANTIC SIMILARITY IN BIOMEDICINE

In this paper the author describe understanding the textual data requires exploitation and integration of clinical resources. In past several approaches for assessing word similarity by exploiting different knowledge source have been proposed. These measures have been adapted to the biomedical field by incorporating domain information extracted from clinical data.

III. ORDINAL, CONTINUOUS AND HETEROGENEOUS K-ANONYMITY THROUGH MICROAGGREGATION

In this paper, k-Anonymity concept used for solve the tension between data utility and respondent privacy for individual data protection. The generalization and suppression approaches proposed in literature to achieve K-Anonymity is not equally suited for all types of attributes:
- Generalization/suppression is one of the few possibilities for nominal categorical attributes.
- It is one possibility for ordinal categorical attributes.
- It is completely unsuitable for continuous attributes, as it causes then to lose their numerical meaning.

IV. PROTECTING PATIENT PRIVACY BY QUANTIFIABLE CONTROL OF DISCLOSURE IN DISSEMINATED DATABASES

In this paper, one of the fundamental rights of patient to have their privacy protected by health care organizations. This information used to identify particular individual is not used to reveal sensitive patient such...
diagnoses, etc. If the degree of anonymity of a disseminated data set could be measured. Privacy protection in disseminated databases can be facilitated by use of special ambiguities algorithm.

V. ACHIEVING K-ANONYMITY PRIVACY PROTECTION USING GENERALIZATION AND SUPPRESSION

In this paper the author have released records adhere to K-anonymity which means each records in the release whose value are indistinct over fields that appear in external data. Generalization involves replacing a value with less specific but semantically consistent value. Suppression involves not replacing a value at all. In this paper authors plead on the need of knowledge intensive tools is data privacy more especially we discuss the role of knowledge related tools in data protection and in disclosure risk assessment.

VI. EFFICIENT MULTIVARIATE DATA-ORIENTED MICRO AGGREGATION

Statistical Disclosure Control (SDC) family of method for micro data masking the micro data so that they can be released while preserving the privacy. Aggregate original database records into small groups prior to publication. Each group should contain k-records, k->constant value. Recently micro aggregation achieve k-Anonymity in addition. Optimal micro aggregation computed in polynomial time for univariate data. This paper present new data oriented heuristics which improve on the trade off between computational complexity and information loss and are thus usable for large data sets.

VII. TOWARDS SEMANTIC MICRO AGGREGATION OF CATEGORICAL DATA FOR CONFIDENTIAL DOCUMENTS

In data privacy context, specifically in SDC. Microaggregation is the well known Microdata protection method, ensuring confidentiality. Authors propose and use for new approach like text documents. This method relies on word net framework that provide full semantic relationship taxonomy between words. Authors aim to ensure confidentiality of text document, at the same time preserve general meaning by applying some measures to evaluate the quality of the protection method relying on information loss.

VIII. PRIVACY PRESERVING DATA MINING

Inference control in data base also known as SDC. This is an important application in several areas such as official statistics, health statistics, e-commerce, etc. Hence it refers to data modification, challenge for SDC is to achieve protection with minimum loss of accuracy database we discuss several information loss and disclosure risk measures and analyse several ways of combining them to assess the performance of the various method.

IX. DIFFERENTIAL PRIVACY

Nothing about an individual should be learnable from the data base that cannot be learned without access to the data base. Generally we say Dalenius formalization is impossible. This state of affairs suggests a new measure, differential privacy intuitively, captures the increased by participants in a database. We achieve any desired level of privacy under this measure. In many cases, extremely accurate information about the database can be provided while simultaneously ensuring very high level of privacy.

X. AUTOMATIC DE-IDENTIFICATION OF TEXTUAL DOCUMENTS IN THE ELECTRONIC HEALTH RECORDS:A REVIEW OF RECENT RESEARCH

In US, the Health Insurance Portability and Accountability Act (HIPAA) protects the confidentiality of patient data and approval of internal review Board to use data for research but these requirements can be waived if data is de-identified. The De-identification of narrative text documents often realized and require significant resources. In this method based performed better with PHI is rarely mentioned in clinical text but are more difficult to generalize.
XI. STATISTICAL DISCLOSURE CONTROL FOR PATIENT RECORDS IN BIOMEDICAL INFORMATION SYSTEM

Patient record data are highly sensitive so their secondary use raises both ethical and data protection issues. Disclosure of patient data could cause serious difficulties so individual damaging for patient and clinicians. In this paper grid based medical data repository accessing risk and suggest a new model for Statiscal Disclosure Control (SDC) of patient data. It provides enormous opportunities for data mining tasks. K-Anonymity which are vulnerable to privacy attacks based on background knowledge. Set-value data could be efficiently released under differential privacy with guaranteed utility help of taxonomy trees. Top down partitioning algorithm to generate a differentially privacy release scale with input data size.

XII. DENSITY-BASED MICROAGGREGATION FOR STATISTICAL DISCLOSURE CONTROL

Protect of personal data in statistical data base has become major concern before they released to public use they applied statistical data bases. Microaggregation for SDC is to protect micro data that is record on individual compares. Micro data into groups at least K-records replace the record in each group. DBA (Density Based Algorithm) it form descending order of their densities in reverse order and compare with latest microaggregation methods.

XIII. ANONYMIZATION OF SET-VALUED DATA VIA TOP-DOWN, LOCAL GENERALIZATION

➢ Thus individual records in common databases unfortunately the bulk of existing anonymization technique each individual is associated with only one sensitive value are well suited for set valued data.
➢ Proposed top-down partitioning based approach linear with input size and reduced information loss by query logs to using this approach.

XIV. CONCLUSION AND FUTURE WORK

This paper detailed about various methods prevailing in literature for protecting privacy of anonymized medical data. Ontology Based measure to compute semantic similarity in Biomedicine is studied. Ordinal, continuous and heterogeneous K-Anonymity through Microaggregation are dealt in detail. Protecting patient privacy by quantifiable control of disclosure in disseminated databases and achieving k-Anonymity privacy protection using generalization and suppression are discussed in detail. Efficient Multivariate data-Oriented Micro aggregation of Categorical data for confidential documents is examined. Differential Privacy for Automatic De-Identification of textual documents in the electronic health records and Statistical Disclosure control for patient records in biomedical information System is considered. Density-based microaggregation for statistical disclosure control and anonymization of Set-Valued Data via Top-Down, Local Generalization are also aggregated in brief. This paper would promote a lot of research in the area of protecting privacy of anonymized medical data.

REFERENCES

11. 11:195–212.
33. Rogers J. Publically reported breaches in EPR confidentiality; 2005.
40. Torra V. Microaggregation for categorical variables: a median based approach.
43. B.Powmeya , Nikita Mary Alett, V,Mohanapriya,S.Balmarugan,"An Object Oriented approach to Model the secure Health care Database systems,"In proceedings of International conference on computer , communication & signal processing(ICSP)in association with IEEE students forum and the society of digital information and wireless communication,SDIWC,2011,pp.2-3
54. V.M.Prabhakaran, Prof.S.Balamurugan, S.Charanyaa, “Privacy Preserving Personal Health Care Data in Cloud”, International Advanced Research Journal in Science, Engineering and Technology Vol 1, Issue 2, October 2014
57. P.Andrew, J.Anish Kumar, R.Sanithya, Prof.S.Balamurugan, S.Charanyaa, “Survey on Approaches Developed for Preserving Privacy of Data Objects” International Advanced Research Journal in Science, Engineering and Technology Vol 1, Issue 2, October 2014