Preservation of Privacy using Multidimensional K-Anonymity Method for Non-Relational Data

Abhijit J. Patankar, Kotrappa Sirbi, Kshama V. Kulhalli

II. REVIEW OF LITERATURE

2.1 PRIVACY DEFINITION:-
Privacy related to how data is collected, used and shared by the users. So privacy definition varies from one environment to the other, Privacy is the generalized term which is used for securing confidential or personal information associated with individual or any group or organization.

2.2 REVIEW OF LITERATURE
To achieve K-Anonymity we have studied 16 different papers in these papers different methods and techniques were discussed for obtaining animalization on dataset in Survey paper [1] advantages and drawbacks of PPDM discussed but here all probable techniques were not discussed. In Protecting Privacy by Multi-dimensional K-anonymity [8] basic method for Anonymity was discussed and also advantages and drawbacks were discussed Detail comparative Review of Literature as follows including Gaps

2.3 RESEARCH OBJECTIVES:-
The research objectives of this research work are as follows:-
- At the time of displaying of real critical and minute level-data need to avoid losses of data.
- To protect individuals private data and improve confidentiality.
- For Preservation of Privacy of persons or groups identified in released data
- To avoid from Linking, Background Knowledge and Homogeneity Attacks.
- To prevent from various information disclosure.
- To improve anonymity and data security on web 3.0.
- To use Nearest Neighborhood strategy to improve k-Anonymity on multidimensional data.

2.4 PROBLEM DEFINITION:
To obtain K-anonymity, quasi-identifier (Key attributes) attributes which we found from dataset are required to be processed, for reducing the chances that intruder will obtain the sensitive data by linking datasets on Key attributes attributes and prevent from homogeneous type of attack and background knowledge type of attack. to increase the anonymization and information loss should be less, and less time complexity for web 3.0 we can build K-Anonymization as a web based service which will convert normal data to protected data before release.
III. METHODOLOGY

3.1 PROPOSED SYSTEM ARCHITECTURE

In above Architecture as shown in Fig 3.1 after data is preprocessed if divisible dimensions are obtained then we will map Multi to single dimension and form the dimension selection Array and later we apply nearest neighborhood strategy and add obtained resultsets or records to improve Anonymization and later we release the Results this is a continuous process till we release all records

3.2 MODULES Implemented

3.2.1. Preprocessing and cleaning of Data
In data preprocessing all basic data mining processes were carried out like data cleaning, data Transformation to make uniform data.

3.2.2. Multidimensional to single dimensional Mapping.
In this module mapping of different multidimensional datasets with single dimensional dataset takes place this mapping will convert all multi dimensionally sets to single dimension

3.2.3. To array for dimension selection
If information loss need to be reduced, by not compromising availability, while selecting the dimensions closer value tuples are grouped together in same partition also there is a need to check interdependency among the relations which will separate selection of dimensions.

3.2.4. Nearest Neighborhood Strategy applied on datasets to calculate distance
As per requirement of Normalized Euclidean distance measures are used for calculating the distance.

\[ d_i = \sqrt{\sum_{i=1}^{v} \left( \frac{P_{1i} - P_{2i}}{v} \right)^2} \]

\( p \) = indicates how many attributes used per person,
\( v \) = indicates difference at maximum scale
\( d \) = indicates total distance between person/attribute.

IV. SCOPE OF THE WORK

4.1 Scope
The proposed system will help to protect privacy and minimize information loss on web 3.0
Following are the identified problems to achieve privacy
- How to protect individuals data on web 3.0?
- How to quantify privacy protection?
- How to maximize usefulness of published data?
- How to implement privacy protection in Relational data?

Here our system implement multidimensional K-Anonymity using nearest neighborhood strategy To achieve security and confidentiality on micro data release. Also in future work we can do data protection for relational data using Access Control Mechanism and we can achieve K-Anonymity with relational data. This Technique will provide wide scope to maintain Data Integrity and Confidentiality.

4.2. Facilities Available at research Centre
- The adequate library facility at the research centre.
- The computing facility in the computer laboratory of RRC and in the personal computer at home.
- Full access to Digital library including IEEE and ACM
- Good Research laboratory with 24 X 7 Internet Facility
- Availability of Tools like Weka
- Availability of Training Dataset like LIC, Election Database, Medical Database.
- State of Art Systems with Printing Facility

V. OUTCOME OF THE WORK

The outcome of this work is justified by creating different records of Non-Relational dataset such as Adult dataset and LIC dataset After applying preprocessing activities on the nonrelational dataset we apply K-Anonymity and also prevent from background knowledge attack. Linking attack so that integrity of data is maintained this is called as protecting data privacy when doing micro-release Above fig shows step by step execution and outcome

5.1. Common Records with zipode and Date of Birth this record is having values without anonymity
5.2 Data with zipcode and Diseases Anatomized
Here we applied anonymization on datasets such as zipcode and disease to obtain K-anonymity and to protect from attacks.

5.3 Common Records with Zipcode and Date of Birth Anonymized
Here if we try to implement any attack then linking and background knowledge attacks are not possible on datasets.

5.4 Graph showing result of Normal distribution of values (Cavg) with the value of k and we obtain better results using Nearest Neighbour than Monderian and close k methods.

5.5 Graph shows change of values of (Cavg) when value of k varies it proves that for normal distribution CDM values of Nearest Neighbour are better than Median and close-k method.

VI. CONCLUSION
With the obtained results and best on comparison of different methods for two quality valued variables Cavg and Cdm based on variation of value of k it proves that nearest neighbor with multidimensional k-anonymity is the better method for Anonymity improvement the and reduce the loss of data and different types of attacks.

REFERENCES


Preservation of Privacy using Multidimensional K-Anonymity Method for Non-Relational Data


