Abstract

As searching methods have advanced the increased risk of privacy disclosure makes it important to protect privacy of user during data publishing. Many of the algorithms used for the data de-identification are not efficient because resulted dataset can easily linked with the public database and it reveals the users identity. One of the method uses for protecting the privacy of user is to apply anonymization algorithms. TDS and TDR using generalization of method to anonymized the dataset. Major drawback as these algorithm is they requires a manually generated domain hierarchy taxonomy for every quasi-identifier in the data set on which k-anonymity has to be performed. Therefore, in this paper we propose new approach which will makes use of suppression based k-anonymization method to allow data publisher to de-identify datasets and in this method only certain attributes from record are suppressed based on values other attributes. As suppression method is used in algorithm, it does not required manually created taxonomy tree of quasi-identifiers. We applied this algorithm on 3 different data sets to evaluate its accuracy as compared to other k-anonymity generalization algorithms. It is found that predicative performance of this algorithm is better than existing generalization methods. This method is expected to provide privacy and accuracy measures to data publishers.
K-Anonymization using Multidimensional Suppression for Data De-identification


- L. Sweeney, "Datafly: A System for Providing Anonymity in Medical Data", Proc. IFIP TC11 WG11. 3 11th Int&apos;l Conf. Database Security XI: Status and Prospects,
K-Anonymization using Multidimensional Suppression for Data De-identification

- S. Kisilevich, L. Rokach, Y. Elovici, B. Shapira, Efficient multidimensional suppression for k-anonymity, IEEE Transaction on Knowledge and Data Engineering 22 (3) (2010) 334–347

Index Terms

Computer Science
Algorithms

Keywords

Privacy Preservation Data Mining Data De-identification PPDM k-Anonymization
Suppression