Abstract

Data Mining plays a vital role in today's information-oriented world where it has been widely applied in various organizations. The current trend is that organizations need to share data for mutual benefit. This has led to a lot of concern over privacy in the recent years. It has also raised a potential threat of revealing sensitive data of an individual when the data is released publicly. Various methods have been proposed to tackle the privacy preservation problem. But the recurring problem is information loss. The loss of sensitive information about certain individuals may affect the data quality and in extreme cases the data may become completely useless. In recent years Privacy preserving data mining has emerged as a key domain of research. One of the methods used for preserving privacy is k-anonymization. k-anonymity demands that every tuple in the dataset released be indistinguishably related to no fewer than k respondents. But the distribution preservation is not guaranteed. In this work a modified k-anonymity model is introduced where the privacy in a dataset is preserved while preserving the distribution also.
References

- A. C. Chaaru and P. S. Yu, "On Static and Dynamic Methods for Condensation-Based Privacy-Protection Data Mining", ACM transaction on Database Systems, Volume 33, no 1 article 22, Publication Date: March 2008.
- V. Jaideep, C. Clifton, M. Kantarcioglu, P. A. Scott, "Privacy-preserving decision trees over vertically partitioned data", ACM transaction on Knowledge Discovery from data, vol 2 No 3, article 14, Publication date: October 2008.

Index Terms

Computer Science
Artificial Intelligence

Keywords
Data Mining  Privacy preserving  k- anonymity  Sensitive attributes