t-Closeness through Microaggregation: Strict Privacy with Enhanced Utility Preservation

Abstract

Microaggregation is a technique for disclosure limitation aimed at protecting the privacy of data subjects in microdata releases. It has been used as an alternative to generalization and suppression to generate k-anonymous data sets, where the identity of each subject is hidden within a group of k subjects. Unlike generalization, microaggregation perturbs the data and this additional masking freedom allows improving data utility in several ways, such as increasing data granularity, reducing the impact of outliers and avoiding discretization of numerical data. k-Anonymity, on the other side, does not protect against attribute disclosure, which occurs if the variability of the confidential values in a group of k subjects is too small. To address this issue, several refinements of k-anonymity have been proposed, among which t-closeness stands out as providing one of the strictest privacy guarantees. Existing algorithms to generate t-close data sets are based on generalization and suppression (they are extensions of k-anonymization...
algorithms based on the same principles). This paper proposes and shows how to use microaggregation to generate k-anonymous t-close data sets. The advantages of microaggregation are analyzed, and then several microaggregation algorithms for k-anonymous t-closeness are presented and empirically evaluated.

EXISTING SYSTEM

- generalization-based approaches is to find the minimal generalization that satisfies the requirements of the underlying privacy model.
- Target privacy model when checking whether a specific generalization is viable.

DRAWBACK OF EXISTING SYSTEM

Generalization-based approaches suffer from some Drawbacks

- Data generalization usually results in a significant
- Loss of granularity, because input values can only be replaced by a reduced set of generalizations
PROPOSED SYSTEM

Microaggregation is a technique for disclosure limitation aimed at protecting the privacy of data subjects in microdata releases. It generates k-anonymous data sets, where the identity of each subject is hidden within a group of k subjects. Several refinements of k-anonymity have been proposed.

ADVANTAGE OF PROPOSED SYSTEM

- Microaggregation allows improving data utility in several ways, such as increasing data granularity, reducing the impact of outliers and avoiding discretization of numerical data.
- The advantages of microaggregation are analyzed, and then several microaggregation algorithms for k-anonymous t-closeness are presented and empirically evaluated.
SYSTEM SPECIFICATION

Hardware Requirements

- System : Pentium IV 2.4 GHz
- Hard Disk : 40 GB
- Floppy Drive : 1.44 Mb
- Monitor : 15 VGA Colour
- Mouse : Logitech
- Ram : 512 Mb

Software Requirements

- Operating system : Windows Family
- Tools : eclipse/Netbeans
- Technology Used : Java
- Backend Used : SQL Server