ABSTRACT

Data mining techniques, in spite of their benefit in a wide range of applications have also raised threat to privacy and data security. Privacy is becoming an increasingly important issue in many data mining applications. This has triggered the development of many privacy preserving data mining techniques. A frequently used disclosure protection method is data perturbation. When used for data mining, it is desirable that perturbation preserves statistical relationships between attributes, while providing adequate protection for individual confidential data.

Data owners must not only meet privacy requirements but also guarantee valid clustering results. In this project, we show that this dual goal can be achieved by transforming a database using a histogram based equalisation technique. This will maintain a balance between privacy and accuracy. We focus primarily on privacy preserving data clustering. Our proposed method distorts confidential attributes to meet privacy requirements, while preserving general features for clustering analysis. Our experiments demonstrate that our methods are effective and provide acceptable values in practice for balancing privacy and accuracy.

Keywords: privacy, data mining, perturbation, accuracy, histogram equalisation, data clustering.