Abstract

Privacy has become an important issue in the progress of data mining techniques. Many laws are being enacted in various countries to protect the privacy of data. This privacy concern has been addressed by developing data mining techniques under a framework called privacy preserving data mining. Presently there are two main approaches popularly used - data perturbation and secure multiparty computation. In this paper we propose a technique for privacy preserving clustering using Principal component Analysis (PCA) based transformation approach. This method is suitable for clustering horizontally partitioned or centralized data sets. The framework was implemented on synthetic datasets and clustering was done using Self organizing Map (SOM). The accuracy of clustering before and after privacy preserving transformation was estimated.

Reference

[5]. Lorrie Cranor, Marc Langheinrich, Massimo Marchiori, Martin Presler-Marshall, and Joseph
Reagle. The Platform for Privacy Preferences 1.0 (P3P1.0) Specification. W3C Recommendation, 16 April 2002.


Index Terms

Computer Science

Security

Key words

PCA

SOM

Rand

Transformation matrix