Chapter 25
Hybrid Privacy Preservation Technique Using Neural Networks

R. VidyaBanu
Sri Krishna College of Engineering and Technology, India

N. Nagaveni
Coimbatore Institute of Technology, India

ABSTRACT

A novel Artificial Neural Network (ANN) dimension expansion-based framework that addresses the demand for privacy preservation of low dimensional data in clustering analysis is discussed. A hybrid approach that combines ANN with Linear Discriminant Analysis (LDA) is proposed to preserve the privacy of data in mining. This chapter describes a feasible technique for privacy preserving clustering with the objective of providing superior level of privacy protection without compromising the data utility and mining outcome. The suitability of these techniques for mining has been evaluated by performing clustering on transformed data and the performance of the proposed method is measured in terms of misclassification and privacy level percentage. The methods are further validated by comparing the results with traditional Geometrical Data Transformation Methods (GDTMs). The results arrived at are significant and promising.

1. INTRODUCTION

Data mining is one of the main steps in Knowledge discovery in databases. Data mining is a technique of extrapolating useful information and valid knowledge from a collection of data that can be used to predict future behavior. There are several data mining techniques that have been developed for fulfilling these objectives. Some of the common techniques include associations, classifications, sequential patterns and clustering. Data mining brings a lot of advantages when used in a specific industry. Data mining can aid direct marketers by providing them with useful and accurate trends about the purchasing behavior of their customers, assist financial institutions in areas such as credit reporting and loan

DOI: 10.4018/978-1-5225-7113-1.ch025

information, aid law enforcers in identifying criminal suspects as well as apprehending these criminals by examining trends in location, crime type, habit, and other patterns of behaviors and assist researchers by speeding up their data analyzing process.

Data mining has been used extensively in the banking, health care, business and financial sectors to model and predict credit fraud, evaluate risk, perform trend analysis, profitability analysis, in stock-price forecasting, disease prediction, option trading, bond rating, portfolio management, commodity price prediction, key phrase extraction from digital libraries (Qi et al., 2011), intelligent information retrieval (Veeramalai & Kannan, 2011), software effort estimation (Deng, 2011), comparison opinion, stock prediction in mergers and acquisitions, forecasting financial disasters etc.

Data mining, with its promise to efficiently discover valuable, non obvious information from large datasets, is particularly vulnerable to misuse (Agarwal and Srikant 2000). Although mining is expected to produce remarkable knowledge and uncover interesting patterns, there are growing concerns about the privacy of personal and sensitive information. This creates a great threat to privacy. The sensitive data used in the process of data mining often get exposed to several parties including data collectors, owners, users and miners. Trends obtained through data mining intended to be used for marketing purpose or for some other ethical purposes, may be misused. People hesitate to share their personal data. This can result in skewing the outcome of the data mining because the data collected may then contain incorrect or incomplete information.

Privacy refers to the ability of an individual or group to protect themselves or information about themselves from unwanted exposure. Flaherty (1989) forwards an idea of privacy as "information control", where the individuals want to be left alone and to exercise some control over how information about them is used. The right to privacy is our right to keep a domain around us, which includes all those things that are part of us, such as our body, home, thoughts, feelings, secrets and identity. Onn (2005) define the right to privacy as, "the ability to choose which parts in this domain can be accessed by others and to control the extent, manner and timing of the use of those parts we choose to disclose".

1.1 Privacy Issues in Data Mining

Due to explosion of data, demands for privacy issue are increasing at an alarming rate. Privacy of individuals can be violated in different ways and with different intentions. Privacy may be voluntarily sacrificed, in exchange for perceived benefits and very often with specific dangers and losses. We are witnessing many threats to data through our day to day activities such as, using credit cards, swapping security cards, using emails etc. Ideally, the data should be collected with the consent of the individual or the organization, with some assurance that the individual privacy will be protected. With the advent of technology and internet, people voluntarily provide personal information to banks, hospitals, surveys, super markets, government, commercial organizations and social networking web sites for different purposes without realizing that this information may cause serious threats to their privacy.

There are multiple factors that contribute to a violation of privacy in data mining, and data can be misused in a number of ways. One source of data privacy violation is the use of "data magnets" which are tools used for collecting private data. They include techniques such as collecting information through on-line registration, identifying users through IP addresses and indirectly collecting information for secondary usage. In most of the cases, the users will be totally or partially unaware of the fact that their