Chapter 35
Optimal Privacy Preserving Scheme Based on Modified ANN and PSO in Cloud

N.G. Nageswari Amma
Vins Christian College of Engineering, Nagercoil, India

F. Ramesh Dhanaseelan
Computer Application Department, St Xavier's Catholic, College of Engineering, Nagercoil, India

ABSTRACT

In the cloud, various privacy-preserving and security threats on data retrieval processes exist. In this article, the authors propose an efficient method for secure privacy preserving in cloud. Initially, the shared file is encrypted using a Vigenere encryption algorithm before uploading. For creating the privacy map, the efficient classification algorithm is recommended. Here, a Modified Artificial Neural Network (MANN) is used to generate the privacy map. The weight value of the neural network is optimized using a Particle Swarm Optimization (PSO) algorithm. While retrieving files initially, the authorization of the person is verified by providing basic information, then the OTP of the respective files is verified. Since the user can retrieve the files only after authorization, verification and decryption of the files is highly secured and privacy is preserved. The performance of the proposed method is evaluated in terms of time and accuracy.

1. INTRODUCTION

With the explosive growth of data and their shared and distributed sources, the need for cooperative and profitable analysis of the data has become increasingly high across organizations. Associated to this, however, are the concerns of privacy breaches of the shared data which might have important legal and strategic consequences for organizations (Mukherjee et al., 2008). Such privacy concerns often limit trajectory data holders' enthusiasm in providing data for further research and applications (Chen et al., 2013). In these days data mining techniques have been viewed as a threat to the sensitive content of

DOI: 10.4018/978-1-7998-8954-0.ch035

personal information. This kind of privacy issue has led to research for privacy preserving data mining techniques (Lin & Chen, 2011). When personal information about people is used in the linking of databases across organizations, then the privacy of this information needs to be carefully protected (Vatsalan et al., 2013). So it is more appropriate to protect every party's data privacy in a distributed way. Hence privacy preserving machine learning models have been introduced. Privacy Preserved Data Mining (PPDM) is a new type which has entered the market and which claims to take care of this particular issue (Banu & Nagaveni, 2013). The goal of privacy preserving data mining is to develop data mining methods without increasing the risk of misuse of the data used to generate those methods (Shi et al., 2014).

Most of the traditional PPDM algorithms preserve the privacy of data by transforming the original data in such a way that the utility of the data is not lost. The ability to analyze private data without violating the privacy of the individuals has contributed to the popularity of PPDM. Redaction is a privacy-preserving method that aims to avoid (or at least mitigate) the disclosure of raw confidential data, such as textual documents (in contrast with specific privacy protection methods focusing only on relational databases (Sánchez et al., 2014). They are utilized in many software applications such as defect prediction, defect classification and clustering models. For example, a group of privacy preserving techniques produces synthetic data from an original data set, and instead of the original data set it releases the synthetic data set that maintains some characteristics of the original data set (Islam & Brankovic, 2011). Recently, many privacy preserving methods based on machine learning techniques have been proposed to assist network experts to analyze the security risks and detect attacks against their systems (Fahad et al., 2014). New privacy models and data anonymization methods have been iteratively proposed, broken, and patched with the discovery of new types of privacy attacks (Khokhar et al., 2014).

The main goal of the all these privacy preserving machine learning models is to hide sensitive defect rules in inter and intra network communication from unauthorized users (Moparthi & Geethanjali, 2016). Even though number of privacy-preserving data mining protocols has been proposed such as those for association rule mining, clustering, naive Bayes classifiers and etc they suffer from limitations. Researchers cites a large number of methods, most of which use some form of transformation on the original data to ensure privacy preservation, called key interchange mapping methods, but these methods are quite complex and compute and memory intensive, thus leading to limited usage of these methods (Bhat et al., 2015). Designing privacy-friendly measurement collection architecture and an associated set of procedures involves several layers: the secure transport of the data over the communication network, the secure storage of collected measurements and suitable procedures for accessing the data (Rottondi et al., 2013). Hence so far, there have been two main approaches for privacy-preserving data mining which are as follows. One is the randomization approach. Another is the cryptographic approach (Yi & Zhang, 2007). On the basis of this different fuzzy methods have been used for classification, regression, feature selection and data mining model which are applied on several databases by different researchers. But there is very few awareness about privacy preserving sub-feature selection using fuzzy model (Bhuyan & Kamila, 2015).

2. LITERATURE SURVEY

Bilge, *et al.* (Bilge & Polat, 2013) proposed a novel privacy-preserving collaborative filtering plan in view of bisecting k-means clustering in which they applied two preprocessing strategies. The primary preprocessing plan managed with reliability issue by developing a binary decision tree via a bisecting