CONCLUSION AND FUTURE WORK

With rapid advance of the network and data mining techniques, the protection of the confidentiality of sensitive information in a database becomes a critical issue to be resolved. Privacy preserving data mining is a new body of research focusing on the implications originating from the application of data mining algorithms to large public databases. Data privacy and security concerns are very important for performing the data mining operations in a safe manner. Privacy preserving data mining is an important issue which can be applied to various domains, such as web commerce, crime reconnoitering, health care, and customer's consumption analysis. This research work delved into the deep waters of knowledge hiding, which is primarily concerned with the privacy of knowledge that is hidden in large databases. Many privacy preserving techniques and algorithms are available to perform the data mining tasks in a privacy preserving way. This research work has concentrated on two important and challenging research problems such as microdata protection and association rule hiding.

Statistical disclosure control is one of the privacy preserving data mining techniques which mainly deals with the protection of confidential information available in the statistical databases. Microdata protection is the youngest sub-discipline of the statistical disclosure control technique and the various microdata protection techniques are used for protecting the confidential attributes of the micro data. In this research thesis, perturbative masking techniques are used and proposed for modifying the sensitive numerical attributes of the microdata. The most important, popular and simple existing techniques used in this research work are additive noise, rounding and micro aggregation. The newly proposed perturbative masking techniques are data transformation, bit transformation, bit++ and bit--. The efficiency of the existing and the proposed perturbative masking techniques are analyzed by using the various performance factors such as statistical accuracy, privacy protection, k-means clustering accuracy, information loss and disclosure risk. Based on the experimental results, it is observed and found that the “DATA TRANSFORMATION PERTURBATIVE MASKING TECHNIQUE” is the
best technique for modifying the data items of the sensitive numeric attribute in the micro
data.

In microdata protection, this research work has given importance to the numerical
attributes only. The sensitive numerical attribute(s) are selected and the data items of
these attributes are modified by the existing and the proposed masking techniques. In
future, the techniques are to be proposed for protecting the data items of the categorical
attributes also. There are many different microdata protection techniques available. This
research work used and proposed perturbative masking techniques only. In future, the
other microdata protection techniques, such as non-perturbative masking techniques and
synthetic data generation techniques can be used and new techniques based on these
concepts can also to be proposed for protecting both the numerical and the categorical
confidential attributes.

Association analysis is a powerful and popular tool for discovering relationships
hidden in large data sets. The relationships can be represented in a form of frequent item
sets or association rules. One rule is categorized as sensitive if its disclosure risk is above
some given threshold. Many approaches are used for protecting the sensitive association
rules. In this research thesis, heuristics and non-heuristics distortion based approaches are
applied for protecting the sensitive association rules. The existing hiding techniques used
in this research work is ISL, DSR and genetic algorithms. The heuristics based hiding
techniques proposed for modifying the sensitive items in order to protect the sensitive
association rules are modified genetic algorithm, tabu search based hiding and ant colony
optimization. The non-heuristic proposed hiding technique is the dummy items creation.
The existing and the proposed hiding techniques performance efficiency are analyzed
using various factors, namely hiding failure, misses cost, artifactual errors and efficiency.
From the experimental results analysis it is observed and found that the proposed “TABU
SEARCH BASED HIDING” technique is the most optimal hiding technique for
protecting the sensitive association rules when compared with the other existing and the
proposed techniques.

In association rule hiding, this research work has concentrated mainly on
developing heuristics algorithms based on distortion concepts. In future, the other concept
blocking can also be used for sensitive association rule protection. This research mainly
focuses on heuristic hiding approaches. In future, the other classes of hiding approaches,
such as border based hiding approaches and the exact hiding approaches can also be considered and new techniques are to be proposed for hiding the sensitive association rules in privacy preserving data mining.