Abstract

Privacy preserving data mining deals with the effectiveness of preserving privacy and utility of the data. Privacy becomes a key concern when the medical data is published for research purposes. Anonymization techniques can be used to transform the dataset into less specific values before publishing to overcome the security breaches. Privacy preservation may reduce the utility value of data. Classification helps to improve the utility of the anonymized data. We propose a model in which a multi-decision tree classifier is built on the anonymized dataset to improve the utility. Multi-decision tree classifier is constituted by Improved ID3 based ADABOOST classifier. The proposed approach is different as the decision tree built is multi-decision tree and as it is constructed on the anonymized dataset. It is proved to be better than the pure decision tree classifier as the multi-decision tree classifier has accuracy better than and training duration shorter than the normal ID3 based ADABOOST classifier.

References

- Duncan, G. T., and Lambert, D. 1986. Disclosure-limited data dissemination. In:
Post Anonymization Techniques in Privacy Preserved Data Mining

- Yan ZHU., and Lin PENG. 2007. Study on K-anonymity Models of Sharing Medical Information. Beijing, China, IEEE.
- Duan Xiaochen., and Hong Xue. 2011. Multi-Decision-Tree Classifier in Master Data Management System. IEEE.

Index Terms

Computer Science Information Sciences

Keywords

Data Privacy Anonymization Classification