Abstract

To preserve the electronic fitness records of data mining applications, Data perturbation is used. It is a form of privacy-preserving data mining. An improved amount and range of information stored in databases has direct to an enhancement in the desire for ranked and "best match" queries. Such queries are mainly applicable when dealing with privacy-sensitive information. To facilitate privacy preservation in data mining or machine learning algorithms over horizontally partitioned or vertically partitioned data, many protocols have been proposed using SMC and various secure building blocks. In this work, we plan to present an effective and efficient cluster based privacy preserving data perturbation technique to mine Multi-partitioned data sets that comprises of both vertical and horizontal data sets which is current demand of e-business and e-commerce data mining environment. To evaluate a trade of between data privacy and transparency of individual's data, data perturbation technique is presented with validation and authentication. In multi-partitioned data distribution, data perturbation raised ambiguity between vertical and horizontal partitions of the data. To overcome the ambiguity, we plan to introduce divisive k-neighbor clusters for multi-partitioned data sets. The performance is evaluated with bench data sets obtained from popular e-business / e-commerce sites.
(Amazon, e-bay etc.).

References

- Jaideep Vaidya,, Christopher W. Clifton ,, Privacy-Preserving Kth Element Score over Vertically Partitioned Data and Knowledge and Data Engineering, Vol. 21, No. 2, February 2009
- Shuguo Han, Wee Keong Ng, Li Wan, and Vincent C. S. Lee, Privacy-Preserving Gradient-Descent Methods and Knowledge and Data Engineering, Vol. 22, No. 6, June 2010
- W. Jiang and C. Clifton, Ac-Framework for Privacy-Preserving Collaboration and Knowledge and Data Engineering (SDM), 2007.
- M. Shaneck, Y. Kim, and V. Kumar, Privacy Preserving Nearest Neighbor Search and Knowledge and Data Mining (ICDM Workshops), pp. 541-545, 2006.
- H. Yu, J. Vaidya, and X. Jiang, Privacy-Preserving SVM Classification on Vertically Partitioned Data and Knowledge and Data Mining (PAKDD), pp. 647-656, Apr. 2006.
- S. Han and W. K. Ng, Privacy-Preserving Linear Fisher Discriminant Analysis and Knowledge and Data Mining (PAKDD), May 2008
- S. Han and W. K. Ng, Privacy-Preserving Genetic Algorithms for Rule Discovery and Data Warehousing and Knowledge Discovery (DaWak), pp. 407-417, Sept. 2007.
- Rozita A. Dara, Masoud Makrehchi, Mohamed S. Kamel, Filter-Based Data Partitioning for Training Multiple Classifier Systems and Knowledge and Data Engineering, Vol. 22, No. 4, April 2010.
- Keng-Pei Lin and Ming-Syan Chen, On the Design and Analysis of the Privacy-Preserving SVM Classifier and Knowledge and Data Engineering, Volume 23 Issue 11, November 2011

Index Terms

Computer Science Information Sciences
Keywords
Multi-partitioned Data Sets  Privacy Preservation  Data Perturbation Technique