

{tag}

{/tag}

International Journal of Computer Applications

© 2014 by IJCA Journal

Volume 96 - Number 6

Year of Publication: 2014

Authors:

Syed Md. Tarique Ahmad

Shameemul Haque

Prince Shoeb Khan

10.5120/16797-6509

{bibtex}pxc3896509.bib{/bibtex}

Abstract

Extracting previously unknown patterns from massive volume of data is the main objective of any data mining algorithm. In current days there is a tremendous expansion in data collection due to the development in the field of information technology. The patterns revealed by data mining algorithm can be used in various domains like Image Analysis, Marketing and weather forecasting. As a side effect of the mining algorithm some sensitive information is also revealed. There is a need to preserve the privacy of individuals which can be achieved by using privacy preserving data mining. In this paper we use min- max normalization approach for preserving privacy during the mining process. We clean the original data using min- max normalization approach before publishing. For experimental purpose we have used k- means algorithm and from our results it is obvious that our approach preserves both privacy and accuracy.

Refer

ences

- Cohen, E. and Strauss, M. , "Maintaining Time Decaying Stream Aggregates," Proceedings of the 22th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, San Diego, California, U. S. A. , pp. 223-233 (2003).
- Chang, J. H. and Lee, W. S. , "Finding Recent Frequent Itemsets Adaptively over Online Data Stream," Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, D. C. , U. S. A. , pp. 487-492 (2003).
- Agrawal, R. and Srikant, R. , "Privacy-Preserving Data Mining,", Proceeding of the ACM SIGMOD Conference on Management of Data, Dallas, Texas, U. S. A. , pp. 439-450 (2000).
- Kumari, P. K. , Raju, K. and Rao, S. S. , "Privacy Preserving in Clustering Using Fuzzy Sets," Proceedings of the 2006 International Conference on Data Mining, Las Vegas, Nevada, U. S. A. , pp. 26-29 (2006).
- Clifton, C. , Kantarcioglu, M. , Vaidya, J. , Lin, X. and Zhu, M. Y. , "Tools for Privacy Preserving Distributed Data Mining," ACM SIGKDD Explorations Newsletter, Vol. 4, pp. 28-34 (2002).
- Meregu, S. and Ghosh, J. , "Privacy-Preserving Distributed Clustering Using Generative Models," Proceedings of the 3th IEEE International Conference on Data Mining, Melbourne, Florida, U. S. A. , pp. 211-218 (2003).
- Oliveira, S. R. M. and Zaiane, O. R. , "Privacy Preserving Clustering by Data Transformation," Proceedings of the 18th Brazilian Symposium on Databases, Manaus, Brazil, pp. 304-318 (2003).
- Liu, L. and Thuraisingham, B. , "The Applicability of the Perturbation Model-Based Privacy Preserving Data Mining for Real-World Data," Proceedings of the 6th IEEE International Conference on Data Mining, Hong Kong, China, pp. 507-512 (2006).
- Vaidya, J. and Clifton, C. , "Privacy-Preserving KMeans Clustering over Vertically Partitioned Data," Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, D. C. , U. S. A. , pp. 206-215 (2003).
- Chen, T. S. , Lin, C. C. , Chiu, Y. H. and Chen, R. C. , "Combined Density-Based and Constraint-Based Algorithm for Clustering," Journal of Donghua University, Vol. 23, pp. 36-38 (2006).
- Hulten, G. , Spencer, L. and Domingos, P. , "Mining Time-Changing Data Streams," Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, U. S. A. , pp. 97-106 (2001).
- Domingos, P. and Hulten, G. , "Mining High-Speed Data Streams," Proceedings of the Association for Computing Machinery 6th International Conference on Knowledge Discovery and Data Mining, Boston, U. S. A. , pp. 71-80 (2000).
- Ordonez, C. , "Clustering Binary Data Streams with K-means," Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, San Diego, California, U. S. A. , pp. 12-19 (2003).
- Aggarwal, C. , Han, J. , Wang, J. and Yu, P. S. , "A Framework for Clustering Evolving Data Streams," Proceedings of the 29th International Conference on Very Large Data Bases, Berlin, Germany, pp. 81-92 (2003).
- Gaber, M. M. , Krishnaswamy, S. and Zaslavsky, A. , "On-Board Mining of Data Streams in Sensor Networks," Springer, Berlin Heidelberg, Germany, pp. 307-335 (2005).

- Yang, C. and Zhou, J. , "HClustream: A Novel Approach for Cluster-ing Evolving Heterogeneous Data Stream," Proceedings of the 6th IEEE International Conference on Data Mining, Hong Kong, China, pp. 682-688 (2006).
- Manikandan G, Sairam N et al. "Privacy preserving clustering by shearing based data transformation", Proceedings of International Conference on Computing and Control Engineering. (2012).
- Han J, and Kamber M, "Data mining-concepts and techniques", 2nd Edn. San Francisco: Morgan Kaufmann Publishers. (2006).
- <http://archive.ics.uci.edu/ml/datasets.html>.

Computer Science

Index Terms

Data Mining

Keywords

Clustering K- Means Accuracy Privacy Min-Max Normalization Normalization.