

{tag}

{/tag}

International Journal of Computer Applications
© 2015 by IJCA Journal

Volume 122 - Number 8

Year of Publication: 2015

Authors:

A. Rama Satish

P. Bala Krishna Prasad

10.5120/21717-4858

{bibtex}pxc3904858.bib{/bibtex}

Abstract

Outliers are abnormal instances or observations. Detecting data outliers is a very important concept in Knowledge data discovery. Outlier detection has been studied in the context of a large number of research areas like large distributed systems, data mining, wireless sensor networks (WSN), health monitoring, environmental science, statistics, etc. , Density based (DB) outlier detection techniques are robust in detecting outliers. In many applications, too much voluminous distributed data is generating every day. Finding deviating observations in the large distributed database rather than in any individual database is not a simple task. Integrating distributed database cause two major problems. First, render massive data from different databases. In addition, data integration may cause violation of data security and leakage of sensitive information. In this work we propose cell density based mechanism for outlier detection (CDOD) in large distributed databases. A centralized detection paradigm is used; it allows overcoming the expensive data integration and information leakage. The experimental results show robustness for finding outliers in large number of databases, instances and attributes.

ences

Refer

- Rakesh Agrawal, Johannes Gehrke, Dimitrios Gunopulos, and Prabhakar Raghavan. Automatic subspace clustering of high dimensional data for data mining applications, volume 27. ACM, 1998.
- Fabrizio Angiulli, Stefano Basta, Stefano Lodi, and Claudio Sartori. A distributed approach to detect outliers in very large data sets. In Euro-Par 2010-Parallel Processing, pages 329–340. Springer, 2010.
- Vic Barnett and Toby Lewis. Outliers in statistical data, volume 3. Wiley New York, 1994.
- Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. Lof: identifying density-based local outliers. In ACM sigmod record, volume 29, pages 93–104. ACM, 2000.
- Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In Kdd, volume 96, pages 226–231, 1996.
- Douglas M Hawkins. Identification of outliers, volume 11. Springer, 1980.
- Alexander Hinneburg and Daniel A Keim. An efficient approach to clustering in large multimedia databases with noise. In KDD, volume 98, pages 58–65, 1998.
- Wen Jin, Anthony KH Tung, and Jiawei Han. Mining top-n local outliers in large databases. In Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, pages 293–298. ACM, 2001.
- Edwin M Knorr and Raymond T Ng. Finding intensional knowledge of distance-based outliers. In VLDB, volume 99, pages 211–222, 1999.
- Edwin M Knox and Raymond T Ng. Algorithms for mining distancebased outliers in large datasets. In Proceedings of the International Conference on Very Large Data Bases, pages 392–403. Citeseer, 1998.
- Ankita Dubey Muruganantham B. Outlier detection using distributed mining technology in large database. International Journal of Computer Science and Engineering, 2(2):6–11, 2015.
- Raymond T Ng and Jiawei Han. Efficient and effective clustering methods for spatial data mining. In Proc. of, pages 144–155, 1994.
- Yaling Pei, Osmar R Zaiane, and Yong Gao. An efficient reference-based approach to outlier detection in large datasets. In Data Mining, 2006. ICDM'06. Sixth International Conference on, pages 478–487. IEEE, 2006.
- Sridhar Ramaswamy, Rajeev Rastogi, and Kyuseok Shim. Efficient algorithms for mining outliers from large data sets. In ACM SIGMOD Record, volume 29, pages 427–438. ACM, 2000.
- Jian Tang, Zhixiang Chen, Ada Wai-Chee Fu, and David W Cheung. Enhancing effectiveness of outlier detections for low density patterns. In Advances in Knowledge Discovery and Data Mining, pages 535–548. Springer, 2002.
- Ji Zhang, Wynne Hsu, and Mong Li Lee. Clustering in dynamic spatial databases. Journal of intelligent information systems, 24(1):5–27, 2005.
- Ji Zhang, Meng Lou, Tok Wang Ling, and Hai Wang. Hos-miner: a system for detecting outlying subspaces of high-dimensional data. In Proceedings of the Thirtieth international conference on Very large data bases-Volume 30, pages 1265–1268. VLDB Endowment, 2004.

- Ji Zhang, Xiaohui Tao, and HuaWang. Outlier detection from large distributed databases. *World Wide Web*, 17(4):539–568, 2014.
- Ji Zhang and Hai Wang. Detecting outlying subspaces for high-dimensional data: the new task, algorithms, and performance. *Knowledge and information systems*, 10(3):333–355, 2006.
- Tian Zhang, Raghu Ramakrishnan, and Miron Livny. Birch: an efficient data clustering method for very large databases. In *ACM SIGMOD Record*, volume 25, pages 103–114. ACM, 1996.

Computer Science

Index Terms

Databases

Keywords

Data Mining KDD Large distributed databases Density based outlier detection.