

# Algorithm Optimization of Anomaly Detection Based on Data Mining

**MICANS INFOTECH**

# ABSTRACT

- ▶ In this paper, firstly two improved algorithm methods are introduced, namely INFLOF and COF, which are based on LOF, then the motivation of each algorithm, the definition of the algorithm and the specific steps of the algorithm are described respectively.
- ▶ Then through summarizing LOF, INFLOF and COF it can find out the intrinsic link between them:
- ▶ INFLOF can solve the problem of edge misjudgment caused by different density cluster's closing to each other in data set, while COF can solve the problem of outliers, but these kinds of two algorithms are from different steps to solve the outlier factor.
- ▶ Finally, the advantages of the these two algorithms are presented, thus the algorithm of this paper is introduced.
- ▶ Moreover, the definition of the algorithm, as well as the specific steps of the algorithm is respectively introduced, besides it also analyzed the time complexity of algorithm.

# EXISTING SYSTEM

- ▶ Tang et. al. proposed COF (Connectivity based Outlier Factor) based on link outlier factor chain distance can be divided by the average value of chain distance of its all nearest neighbor point distance, so as to define the outlier factor of data.
- ▶ Thus, when the data distribution is sparse and some patterns are distributed, there will be a good effect of outlier detection. Subsequently, Hui Cao and other people proposed DSNOF (Density Similarity Neighbor based Outlier Factor), which can further strengthen the effect of outlier detection when COF presents the case of deviation.

# PROPOSED SYSTEM

- ▶ In this paper, it will firstly introduce the two main algorithm methods based on LOF, namely, INLOF and COF, then putting focus on the proposed improved algorithms according to the shortcomings of these two kinds of algorithms, moreover it analyzes the time complexity of the algorithm, in the next chapter it will analyze the effectiveness of the proposed algorithm through the experiment.

# HARDWARE REQUIREMENTS

- ▶ Processor
  - Pentium -III
- ▶ Speed
  - 1.1 Ghz
- ▶ RAM
  - 256 MB(min)
- ▶ Hard Disk
  - 20 GB
- ▶ Floppy Drive
  - 1.44 MB
- ▶ Key Board
  - Standard Windows Keyboard
- ▶ Mouse
  - Two or Three Button Mouse
- ▶ Monitor
  - SVGA

**MICANS INFOTECH**

# SOFTWARE REQUIREMENTS

- ▶ Operating System : Windows 8
- ▶ Front End : Java /DOTNET
- ▶ Database : Mysql /HEIDISQL

**MICANS INFOTECH**

# CONCLUSION

- ▶ In this paper, an improved local outliers detection algorithm based on density is proposed.
- ▶ Through having in-depth analysis on two improved algorithms of outliers detection algorithm based on density namely, INFLOF and COF, we can find out their shortcomings, through integrating the advantages of two algorithms, an improved algorithm is proposed in this paper,
- ▶ thus the algorithm and specific steps are given, moreover it also analyzes the time complexity of the algorithm in this paper.

# REFERENCE

- [1] Sun Huanliang, Bao Yubin, Yu Ge, et al. An algorithm based on partition for outlier detection [J]. Journal of software, 2006, 17(5): 1009–1016.
- [2] Breunig M, Kriegel H, Ng R, et al. lof: Identifying density-based local outliers [C] // Proc. SIGMOD Conf. IEEE, 2000: 93–104.
- [3] Knorr M E, Ng R T, Tucakov V. Distance-based outliers: Algorithms and applications [J]. The VLDB Journal, 2000, 8 (3–4) : 237–253.
- [4] Ester M, Kriegel H, Sander J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise [C] // Proceeding of the 2nd International Conference on Knowledge Discovery and Data Mining. 1996: 226–231.