Outlier Detection with Supervised Learning Method

By: Bawono, AH (Bawono, Aditya Hari)[1] ; Bachtiar, FA (Bachtiar, Fitra Abdurrachman)[1]

PROCEEDINGS OF 2019 4TH INTERNATIONAL CONFERENCE ON SUSTAINABLE INFORMATION ENGINEERING AND TECHNOLOGY (SIET 2019)
Book Group Author(s): IEEE
Pages: 306-309
Published: 2019
Document Type: Proceedings Paper

Conference

Conference: 4th International Conference on Sustainable Information Engineering and Technology (SIET)
Location: Senggigi, INDONESIA
Date: SEP 28-30, 2019
Sponsor(s): Brawijaya Univ, Fac Comp Sci; IEEE, Indonesia Sect; Univ Mataram

Abstract

Outliers are data points that can affect the quality of data and the results of analysis from data mining. Outlier detection can also be seen as a pre-processing step to find data points that do not properly placed in the data set. Previously outlier detection methods are unsupervised. Today, ODDS provide data set with outlier information as a ground truth for supervised learning. However, outliers are commonly minor in any data set. For supervised learning this will lead to an imbalanced data classification problem. This paper presents the result of popular classification method, k-Nearest neighbor, Centroid Classifier, and Naive Bayes to handle outlier detection task. Even the mentioned methods were not designed to detect outlier, they proved by achieving 81% average sensitivity which is good for further research.

Keywords

Author Keywords: outlier; outlier detection; supervised learning; classification; sensitivity
KeyWords Plus: ALGORITHMS; EFFICIENT

Author Information

Reprint Address:
Brawijaya University Univ Brawijaya, Fac Comp Sci, Malang, Indonesia.
Corresponding Address: Bawono, AH (corresponding author)

+ Univ Brawijaya, Fac Comp Sci, Malang, Indonesia.

Addresses:

+ [1] Univ Brawijaya, Fac Comp Sci, Malang, Indonesia

E-mail Addresses: ndaimen@ub.ac.id; fitra.bachtiar@ub.ac.id

Publisher

IEEE, 345 E 47TH ST, NEW YORK, NY 10017 USA

Categories / Classification

Research Areas: Computer Science; Science & Technology - Other Topics; Engineering
Web of Science Categories: Computer Science, Theory & Methods; Green & Sustainable Science & Technology; Engineering, Electrical & Electronic

See more data fields

1 of 1

Cited References: 24

Showing 24 of 24 View All in Cited References page

(from Web of Science Core Collection)

1. Distributed Strategies for Mining Outliers in Large Data Sets — Times Cited: 30
By: Angiulli, Fabrizio; Basta, Stefano; Lodi, Stefano; et al.
IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING Volume: 25 Issue: 7 Pages: 1520-1532 Published: JUL 2013

2. Title: [not available] — Times Cited: 1
By: Barnett; Lewis, T.
Int. J. Forecast. Volume: 12 Issue: 1 Pages: 175-176 Published: Mar. 1996
55.00 Outliers in statistical data: V
Publisher: John Wiley & Sons, Chichester

3. Nearest Neighbor Method Using Non-nested Generalized Exemplars in Breast Cancer Diagnosis — Times Cited: 1
By: Bohacik, Jan; Zabovsky, Michal
2017 IEEE 14TH INTERNATIONAL SCIENTIFIC CONFERENCE ON INFORMATICS Pages: 40-44 Published: 2017

4. LOF: Identifying density-based local outliers — Times Cited: 1,631
By: Breunig, MM; Kriegel, HP; Ng, RT; et al.
SIGMOD RECORD Volume: 29 Issue: 2 Pages: 93-104 Published: JUN 2000

5. Weighted nearest neighbors feature selection — Times Cited: 8
By: Bugata, Peter; Drotar, Peter
KNOWLEDGE-BASED SYSTEMS Volume: 163 Pages: 749-761 Published: JAN 1 2019

6. Least squares one-class support vector machine — Times Cited: 46
By: Choi, Young-Sik
PATTERN RECOGNITION LETTERS Volume: 30 Issue: 13 Pages: 1236-1240 Published: OCT 1 2009

7. EM Algorithms for Weighted-Data Clustering with Application to Audio-Visual Scene Analysis — Times Cited: 31
By: Gebru, Israel Dejene; Alameda-Pineda, Xavier; Forbes, Florence; et al.
IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE Volume: 38 Issue: 12 Pages: 2402-2415 Published: DEC 2016

8. Title: [not available] — Times Cited: 1,049
By: Hawkins, D.M.
Identification of Outliers Published: 1980
Publisher: Chapman and Hall, London

9. Multivariate functional outlier detection — Times Cited: 63
By: Hubert, Mia; Rousseeuw, Peter J.; Segaert, Pieter
STATISTICAL METHODS AND APPLICATIONS Volume: 24 Issue: 2 Pages: 177-202 Published: JUL 2015

10. Efficient biased sampling for approximate clustering and outlier detection in large data sets — Times Cited: 79
By: Kollios, G; Gunopulos, D; Koudas, N; et al.
IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING Volume: 15 Issue: 5 Pages: 1170-1187 Published: SEP-OCT 2003

11. Efficient and flexible algorithms for monitoring distance-based outliers over data streams — Times Cited: 30
By: Kontaki, Maria; Gounaris, Anastasios; Papadopoulos, Apostolos N.; et al.
INFORMATION SYSTEMS Volume: 55 Pages: 37-53 Published: JAN 2016

12. A new Centroid-Based Classification model for text categorization — Times Cited: 14
By: Liu, Chuan; Wang, Wenyong; Tu, Guanghui; et al.
KNOWLEDGE-BASED SYSTEMS Volume: 136 Pages: 15-26 Published: NOV 15 2017

13. Outlier Detection Using Improved Genetic K-Means — Times Cited: 1
By: Marghny, M.; Taloba, A.
Ssrn Volume: 28 Issue: 11 Pages: 33-36 Published: 2015

14. A Novel Method for Minimizing Loss of Accuracy in Naive Bayes Classifier — Times Cited: 1
By: Netti, K.
2015 IEEE INT C COMP Pages: 1-4 Published: 2015

15. An Unsupervised Approach for Combining Scores of Outlier Detection Techniques, Based on Similarity Measures — Times Cited: 6
By: Pasillas-Diaz, Jose Ramon; Ratte, Sylvie
ELECTRONIC NOTES IN THEORETICAL COMPUTER SCIENCE Volume: 329 Pages: 61-77 Published: DEC 9 2016

16. NEW ITERATIVE LEARNING STRATEGY TO IMPROVE CLASSIFICATION SYSTEMS BY USING OUTLIER DETECTION TECHNIQUES — Times Cited: 1
By: Pelletier, C.; Valero, S.; Inglada, J.; et al.
2017 IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM (IGARSS) Book Series: IEEE International Symposium on Geoscience and Remote Sensing IGARSS Pages: 3676-3679 Published: 2017

17. A Nearest Centroid Classifier based Clustering Algorithm for Solving Vehicle Routing Problem — Times Cited: 2
By: Praveen, V.; Kousalya, K.; Kumar, K. R. Prasanna
PROCEEDINGS OF THE 2016 IEEE 2ND INTERNATIONAL CONFERENCE ON ADVANCES IN ELECTRICAL & ELECTRONICS, INFORMATION, COMMUNICATION & BIO INFORMATICS (IEEE AEEICB-2016) Pages: 414-419 Published: 2016

18. Title: [not available] — Times Cited: 1

By: RAYANA S
OUTLIER DETECTION DA Published: 2016

19. Title: [not available] **Times Cited: 1**
By: Rousseeuw, P. J.; Leroy, A. M.
Robust Regression and Outlier Detection Volume: 39 Issue: 5 Published: 1987

20. **Lossless Pruned Naive Bayes for Big Data Classifications** **Times Cited: 3**
By: Sun, Nanfei; Sun, Bingjun; Lin, Jian(Denny) Jian(Denny); et al.
BIG DATA RESEARCH Volume: 14 Pages: 27-36 Published: DEC 2018

21. **Deleting or Keeping Outliers for Classifier Training?** **Times Cited: 7**
By: Tallon-Ballesteros, Antonio J.; Riquelme, Jose C.
2014 SIXTH WORLD CONGRESS ON NATURE AND BIOLOGICALLY INSPIRED COMPUTING (NABIC) Book Series: World Congress on Nature and Biologically Inspired Computing Pages: 281-286 Published: 2014

22. **Adapting centroid classifier for document categorization** **Times Cited: 15**
By: Tan, Songbo; Wang, Yuefen; Wu, Gaowei
EXPERT SYSTEMS WITH APPLICATIONS Volume: 38 Issue: 8 Pages: 10264-10273 Published: AUG 2011

23. **Road Targets Recognition Based on Deep Learning and Micro-Doppler Features** **Times Cited: 3**
By: Wang, Jie; Guo, Jianying; Shao, Xiang; et al.
2018 INTERNATIONAL CONFERENCE ON SENSOR NETWORKS AND SIGNAL PROCESSING (SNSP 2018) Pages: 271-276 Published: 2018

24. **A Centroid-Based Outlier Detection Method** **Times Cited: 1**
By: Wang, Xiaochun; Chen, Yiqin; Wang, Xia Li
PROCEEDINGS 2017 INTERNATIONAL CONFERENCE ON COMPUTATIONAL SCIENCE AND COMPUTATIONAL INTELLIGENCE (CSCI) Pages: 1411-1416 Published: 2017

Showing 24 of 24 **View All in Cited References page**