Abstract

Real world application often found the problem of unbalanced dataset. This then create the problem in machine learning methods. In this paper we have surveyed the imbalance dataset problem at the algorithmic level. By over sampling and under sampling some researchers artificially prove that updated svm, cost sensitive classifier, class orientation methods can be good on imbalanced dataset. This imbalance problem is also switching towards hybrid algorithm.

References

1. Szil’ard Vajda, Gernot A. Fink Strategies for Training Robust Neural Network Based Digit Recognizers on Unbalanced Data Set 2010 12th International Conference on Frontiers in Handwriting Recognition
2. C.V. KrishnaVeni, T. Sobha Rani On the Classification of Imbalanced Datasets IJCST Vol. 2, SP 1, December 2015
3. Nitesh V. Chawla, Nathalie Japkowicz, 
4. Special Issue on Learning from Imbalanced Data Sets— Sigkdd Explorations. Volume 6, Issue 1 
11. Gary M. Weiss, Kate McCarthy, and Bibi Zabar Cost-Sensitive Learning vs. Sampling: Which is Best for Handling Unbalanced Classes with Unequal Error costs? 
13. Charles X. Ling, Qiang Yang, Jianning Decisions tree with minimum costs. 
15. Wei Liu, Sanjay Chawla, David A. Nitesh v Chawala for imbalanced datasets. 2013 
18. NGUYEN HA VO, YONGGWAN WON Classification of Unbalanced Medical Data with Weighted Convergence of bio science technology 2015. 
20. Yanmin Sun, Mohamed S. Kamel, Andrew K cost sensitive boosting on imbalanced dataset 2013. 
22. TAO Xiao-yan, JI Hong-bing AModifiedPSVM and its Application to Unbalanced Data Classification. Thir International Conference on NaturalICNS 2017

Index Terms

Computer Science

Information Sciences

Keywords
cost-sensitive learning, imbalanced data set, modified SVM, oversampling, undersampling