World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:10, No:02, 2016

Adaptive Swarm Balancing Algorithms for Rare-Event Prediction in Imbalanced Healthcare Data

Authors : Jinyan Li, Simon Fong, Raymond Wong, Mohammed Sabah, Fiaidhi Jinan

Abstract : Clinical data analysis and forecasting have make great contributions to disease control, prevention and detection. However, such data usually suffer from highly unbalanced samples in class distributions. In this paper, we target at the binary imbalanced dataset, where the positive samples take up only the minority. We investigate two different meta-heuristic algorithms, particle swarm optimization and bat-inspired algorithm, and combine both of them with the synthetic minority over-sampling technique (SMOTE) for processing the datasets. One approach is to process the full dataset as a whole. The other is to split up the dataset and adaptively process it one segment at a time. The experimental results reveal that while the performance improvements obtained by the former methods are not scalable to larger data scales, the later one, which we call Adaptive Swarm Balancing Algorithms, leads to significant efficiency and effectiveness improvements on large datasets. We also find it more consistent with the practice of the typical large imbalanced medical datasets. We further use the meta-heuristic algorithms to optimize two key parameters of SMOTE. Leading to more credible performances of the classifier, and shortening the running time compared with the brute-force method.

Keywords : Imbalanced dataset, meta-heuristic algorithm, SMOTE, big data
Conference Title : ICDMA 2016 : International Conference on Data Mining and Applications
Conference Location : Melbourne, Australia
Conference Dates : February 04-05, 2016