World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:14, No:12, 2020

Machine Learning Facing Behavioral Noise Problem in an Imbalanced Data Using One Side Behavioral Noise Reduction: Application to a Fraud Detection

Authors : Salma El Hajjami, Jamal Malki, Alain Bouju, Mohammed Berrada

Abstract : With the expansion of machine learning and data mining in the context of Big Data analytics, the common problem that affects data is class imbalance. It refers to an imbalanced distribution of instances belonging to each class. This problem is present in many real world applications such as fraud detection, network intrusion detection, medical diagnostics, etc. In these cases, data instances labeled negatively are significantly more numerous than the instances labeled positively. When this difference is too large, the learning system may face difficulty when tackling this problem, since it is initially designed to work in relatively balanced class distribution scenarios. Another important problem, which usually accompanies these imbalanced data, is the overlapping instances between the two classes. It is commonly referred to as noise or overlapping data. In this article, we propose an approach called: One Side Behavioral Noise Reduction (OSBNR). This approach presents a way to deal with the problem of class imbalance in the presence of a high noise level. OSBNR is based on two steps. Firstly, a cluster analysis is applied to groups similar instances from the minority class into several behavior clusters. Secondly, we select and eliminate the instances of the majority class, considered as behavioral noise, which overlap with behavior clusters of the minority class. The results of experiments carried out on a representative public dataset confirm that the proposed approach is efficient for the treatment of class imbalances in the presence of noise.

Keywords : machine learning, imbalanced data, data mining, big data
Conference Title : ICSRD 2020 : International Conference on Scientific Research and Development
Conference Location : Chicago, United States
Conference Dates : December 12-13, 2020