Abstract

Nowadays learning from imbalanced data sets are a relatively a very critical task for many data mining applications such as fraud detection, anomaly detection, medical diagnosis, information retrieval systems. The imbalanced learning problem is nothing but unequal distribution of data between the classes where one class contains more and more samples while another contains very little. Because of imbalance learning problems, it becomes hard for the classifier to learn the minority class samples. The Aim of this paper is to review on various techniques which are used for resolving imbalanced learning problem. This paper proposes a taxonomy for various methods used for handling the class imbalance problem where each method can be categorized depending on the techniques it uses. To handle imbalanced learning problem significant work
has been done, which can be categorized into four categories: sampling-based methods, cost-based methods, kernel-based methods, and active learning-based methods. All these methods resolve the imbalanced learning problem efficiently.

References

- R. Ade and P. R. Deshmukh, "Incremental learning in students classification system with efficient knowledge transformation," Int'l Conf. on PDGC, Dec 2014.
- Reshma C. Bhagat and Sachin S. Patil, "Enhanced SMOTE Algorithm for Classification of Imbalanced Big-Data using Random Forest," IEEE International Advance
A Review on Imbalanced Learning Methods

Computing Conference (IACC), 2015

- Xingyi LIU, "Cost-sensitive Decision Tree with Missing Values and Multiple Cost Scales," Int'l Conf. on Artificial Intelligence, 2009
- ZhiQiang ZENG and ShunZhi ZHU, "A Kernel-based Sampling to Train SVM with Imbalanced Data Set," Conference Anthology, IEEE, January 2013
- Bo ZHOU, Cheng YANG, Haixiang GUO and Jinglu HU, "A Quasi-linear SVM Combined with Assembled SMOTE for Imbalanced Data Classification," Int'l Conf. on Neural Networks, August 2013

Index Terms
Computer Science
Information Sciences
Keywords
Imbalanced Learning  Active Learning  Cost-sensitive Learning