

A Hybrid Ensemble and Swarm Based Approach for Imbalanced Classification

¹K. Lokanayaki and ²M. Sailaja

Abstract—This paper proposes an ensemble algorithm named of SSO-Adaboost-KNN for multi-class imbalanced data classification. The main goal of this paper is to integrate feature selection, boosting into ensemble and improve accuracy rate for minority class than existing algorithm in imbalanced data classification. First, In this model Simplified Swarm Optimization (SSO) is proposed to adjust the weight as the feature selection algorithm with fitness. Second, Adaboost uses KNN algorithm to sever down majority class weights which is near to minority class, so that the classify can pay more attention to minority class. The results show that the proposed algorithm improves both the stability and the accuracy of boosting after carrying out feature selection, and the performance of proposed algorithm.

Key Index Terms—Imbalanced data, SSO, ensemble learning, machine learning.

I. INTRODUCTION

Class imbalance problem become greatest issue in data mining. Imbalanced data become an obstacle in data mining nowadays, minority class sometimes are more important than majority class, just like in medical diagnosis, credit card, fraud and etc. Classification of data becomes difficult because of unbounded size and imbalance nature of data. Imbalance problem occur where one of the two classes having more sample than other classes. The most of algorithm are more focusing on classification of major sample while ignoring or misclassifying minority sample. The minority samples are those that rarely occur but very important. There are different methods available for classification of imbalance data set[1]. The challenge of learning from imbalanced data is that the relatively or

absolutely Under represented class cannot draw equal attention to the learning algorithm Compared to the majority class, which often leads to very special classification rules or missing rules for the minority class without much generalization ability for future prediction[10]. How to better recognize data from the minority class is a major research question in class imbalance learning. Its learning objective can be generally described as "obtaining a classifier that will provide high accuracy for the minority class without severely jeopardizing the accuracy of the majority class"[2]. Data sampling has received much attention in data mining related to class imbalance problem. Data sampling tries to overcome imbalanced class distributions problem by adding samples to or removing sampling from the data set [3].

II. LITERATURE REVIEW

This method improves the classification accuracy of minority class but, because of infinite data streams and continuous concept drifting, this method cannot suitable for skewed data stream classification. Most existing imbalance learning techniques are only designed for two class problem. Multiclass imbalance problem mostly solve by using class decomposition. AdaBoost.NC [3-6] is an ensemble learning algorithm that combines the strength of negative correlation learning and boosting method. This algorithm mainly used in multiclass imbalance dataset. The results suggest that AdaBoost.NC combined with random oversampling can improve the prediction accuracy on the minority class without losing the overall performance compared to other existing class imbalance learning methods. Wang et al. proposed the classification algorithm