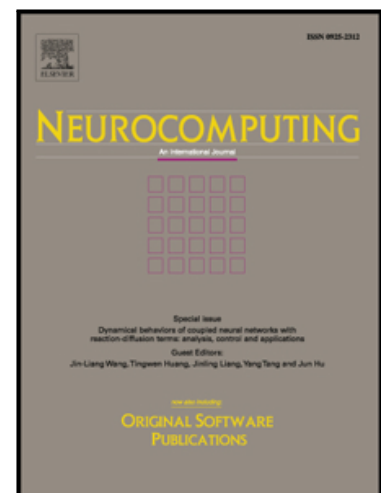


Accepted Manuscript

A Study on combining Dynamic Selection and Data Preprocessing for Imbalance Learning

Anandarup Roy, Rafael M.O. Cruz, Robert Sabourin,
George D.C. Cavalcanti

PII: S0925-2312(18)30093-6
DOI: [10.1016/j.neucom.2018.01.060](https://doi.org/10.1016/j.neucom.2018.01.060)
Reference: NEUCOM 19262



To appear in: *Neurocomputing*

Received date: 27 April 2017
Revised date: 8 October 2017
Accepted date: 24 January 2018

Please cite this article as: Anandarup Roy, Rafael M.O. Cruz, Robert Sabourin, George D.C. Cavalcanti, A Study on combining Dynamic Selection and Data Preprocessing for Imbalance Learning, *Neurocomputing* (2018), doi: [10.1016/j.neucom.2018.01.060](https://doi.org/10.1016/j.neucom.2018.01.060)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A Study on combining Dynamic Selection and Data Preprocessing for Imbalance Learning

Anandarup Roy^a, Rafael M.O. Cruz^{a,*}, Robert Sabourin^a, George D. C. Cavalcanti^b

^a*École de Technologie Supérieure, University of Quebec, Montreal, Quebec, Canada*

^b*Centro de Informática, Universidade Federal de Pernambuco, Recife, PE, Brazil*

Abstract

In real life, classifier learning may encounter a dataset in which the number of instances of a given class is much higher than for other classes. Such imbalanced datasets require special attention because traditional classifiers generally favor the majority class which has a large number of instances. Ensemble classifiers, in such cases, have been reported to yield promising results. Most often, ensembles are specially designed for data level preprocessing techniques that aim to balance class proportions by applying under-sampling and/or over-sampling. Most available studies concentrate on static ensembles designed for different preprocessing techniques. Contrary to static ensembles, dynamic ensembles became popular thanks to their performance in the context of ill defined problems (small size datasets). A dynamic ensemble includes a dynamic selection module for choosing the best ensemble given a test instance. This paper experimentally evaluates the argument that dynamic selection combined with a preprocessing technique can achieve higher performance than static ensemble for imbalanced classification problems. For this evaluation, we collect 84 two-class and 26 multi-class datasets of varying degrees of class-imbalance. In addition, we consider five variations of preprocessing methods and four dynamic selection methods. We further design a useful experimental framework to integrate preprocessing and dynamic selection. Our experiments show that the dynamic ensemble improves the F-measure and the G-mean as compared to the static ensemble. Moreover, considering different levels of imbalance, dynamic selection methods secure higher ranks than other alternatives.

Keywords: Imbalanced learning, ensemble of classifiers, multi-class imbalance, dynamic ensemble selection, preprocessing, SMOTE

Download English Version:

<https://daneshyari.com/en/article/6864418>

Download Persian Version:

<https://daneshyari.com/article/6864418>

[Daneshyari.com](https://daneshyari.com)