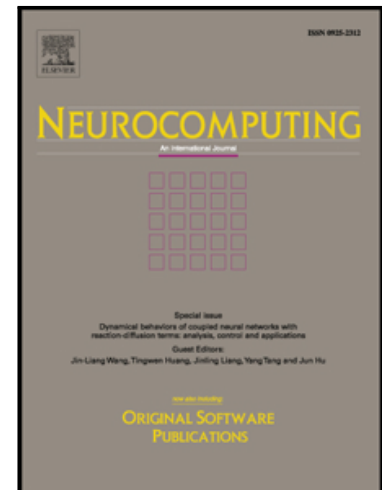


B²FSE Framework for High Dimensional Imbalanced Data: A Case Study for Drug Toxicity Prediction

Nishtha Hooda, Seema Bawa, Prashant Singh Rana

PII: S0925-2312(17)31545-X
DOI: [10.1016/j.neucom.2017.04.081](https://doi.org/10.1016/j.neucom.2017.04.081)
Reference: NEUCOM 18908



To appear in: *Neurocomputing*

Received date: 1 February 2016
Revised date: 26 April 2017
Accepted date: 29 April 2017

Please cite this article as: Nishtha Hooda, Seema Bawa, Prashant Singh Rana, B²FSE Framework for High Dimensional Imbalanced Data: A Case Study for Drug Toxicity Prediction, *Neurocomputing* (2017), doi: [10.1016/j.neucom.2017.04.081](https://doi.org/10.1016/j.neucom.2017.04.081)

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.

B²FSE Framework for High Dimensional Imbalanced Data: A Case Study for Drug Toxicity Prediction

Nishtha Hooda¹, Seema Bawa, Prashant Singh Rana

Computer Science and Engineering Department, Thapar University, Patiala 147004, India

Abstract

The life of people is imperiled by umpteen chemicals unwittingly through the diverse sources like food, cleaning products, medicines, etc. At times, these chemicals can be toxic. Assessing and analyzing the toxicity of these chemical compounds can lead us to prospects to improve the environmental chemicals and invent new medicines. Tox21 crowdsourcing program initiative brings an evolutionary breakthrough for the researchers to develop better toxicity assessment techniques. Machine learning has received much attention in the domain of predictive analytics as it applies computational statistics and offers automation environment to expedite the data modeling process. The goal is to develop an efficient prediction model, combined with the machine learning algorithmic characteristics, which can predict whether a chemical compound is toxic and can affect the health adversely or not. Hence, an efficient pre-processing method should be adopted to achieve the best performance of the machine learning classifier. This work is a specific case study which proposes a Better Balanced Feature Selection Ensemble(B²FSE) framework for the classification of drug toxicity molecules, carried out on imbalanced and high dimensional complex drug data. We show that, an ensemble feature selection and an ensemble classifier, integrated with random subset selection, and a class balancer have the potential to generate more accurate, lower cost, and balanced classification framework. The performance of the proposed framework, when evaluated with different evaluation parameters and compared with the state-of-the art methods like SVM, random forest, bagging, etc., is found to be superior than the available methods.

Keywords: Imbalanced Learning, Toxicity Prediction, Machine Learning, Ensemble, Drug Designing

1. Introduction

Drugs are small organic molecules administered to instigate any biological problem. Biologically, it is a ligand designing which involves a technique of designing a molecule that can bind tightly to the target. Identifying targets, stratifying patients, monitoring drug toxicity, and safety are the key challenges in the area of drug designing and development. Determining the toxicity of a chemical compound is of crucial importance in order to minimize our exposure to many harmful substances in everyday products. For instance, ten thousand additional chemicals like preservatives, flavors, etc. are allowed to put into the food products in the United States (US) [1]. Toxicity is also a central issue in the development of the new drugs, with more than 30 percent of the drug candidates failing in clinical trials because of undetected toxic effects [2]. Therapeutic prediction and decision making regarding the drug toxicity is an eminent research problem to effectuate clinical outcomes [3]. Government agencies NIH, EPA etc. instigated the Tox21 data challenge to encourage the ingenious computational techniques for the drug toxicity prediction [4]. The main intent of this research is to classify a new drug into a class (toxic or non-toxic), based on the different chemical descriptors of a drug molecule. Ongoing techniques of toxicity assessment for testing a vast number of chemicals are

Email address: 27nishtha@gmail.com (Nishtha Hooda)

Download English Version:

<https://daneshyari.com/en/article/6864769>

Download Persian Version:

<https://daneshyari.com/article/6864769>

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