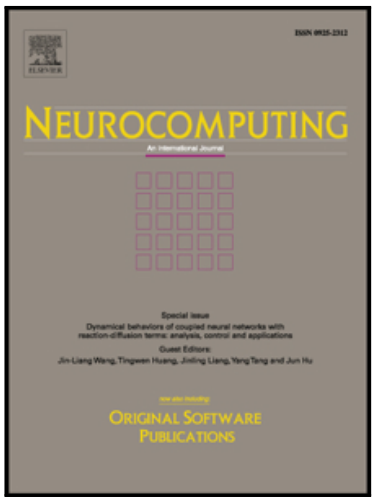

Multivariate Soft Repulsive System Identification for Constructing Rule-based Classification Systems: Application to Trauma Clinical Data

Abolfazl Doostparast Torshizi , Linda Petzold , Mitchel Cohen

PII: S0925-2312(17)30558-1
DOI: 10.1016/j.neucom.2017.03.043
Reference: NEUCOM 18271

To appear in: *Neurocomputing*

Received date: 12 December 2016
Revised date: 5 February 2017
Accepted date: 12 March 2017

Please cite this article as: Abolfazl Doostparast Torshizi , Linda Petzold , Mitchel Cohen , Multivariate Soft Repulsive System Identification for Constructing Rule-based Classification Systems: Application to Trauma Clinical Data, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.03.043

Multivariate Soft Repulsive System Identification for Constructing Rule-based Classification Systems: Application to Trauma Clinical Data

Abolfazl Doostparast Torshizi[1a], Linda Petzold[a], Mitchel Cohen[b]

[a]Department of Computer Science, University of California, Santa Barbara, USA.
[b]Department of Surgery, University of California, San Francisco, USA.

Abstract:

Rule-based classification systems constructed upon linguistic terms in the antecedent and consequent of the rules lack sufficient generalization capabilities. This paper proposes a new multivariate fuzzy system identification algorithm to design binary rule-based classification structures through making use of the repulsive forces between the cluster prototypes of different class labels. This approach is coupled with the potential discrimination power of each dimension in the feature space to increase the generalization potential. To address this issue, first the multivariate variant of a newly proposed soft clustering algorithm along with its mathematical foundations is proposed. Next, the discriminatory power of each individual feature is computed, using the multivariate membership values in the proposed clustering algorithm to achieve the most accurate firing degree in each rule. The main advantage of this method is to handle unbalanced datasets yielding superior true positive measure while keeping the false positive rate low enough to avoid the natural bias toward class labels containing larger number of training samples. To validate the proposed approaches, a series of numerical experiments on publicly available datasets and a real clinical dataset collected by our team were conducted. Simulation results demonstrated achievement of the primary goals of this research.

Keywords: Fuzzy Systems; System Identification; Multivariate Analysis; Clustering.

Introduction

Pattern recognition spans a wide range of applications from personalized medicine to engineering. In general, the primary objective of pattern recognition is to learn the dominating structure of the data and then to assign unseen samples into predetermined categories. In the realm of pattern recognition, a large body of classification methods can be found in the literature [1-3]. Some of these methods are quite well-known such as Support Vector Machines (SVM), Artificial Neural Networks (ANNs), AdaBoost [4], and Kernel Fisher Discriminant method (KFD) [5]. One of the major difficulties in using such state-of-the-art algorithms is that they are computationally complex while having poor interpretability. In recent years, many efforts have been made to overcome these issues through making radical changes in the existing algorithms. Some of these methods include: core vector machines [6-7], Lagrangian Support Vector

[1] Corresponding Author.
Email: ad3436 @cumc.columbia.edu.