Title: Feature Selection for the Classification of Traced Neurons

Authors: José D. López-Cabrera, Juan V. Lorenzo-Ginori

PII: S0165-0270(18)30105-5
DOI: https://doi.org/10.1016/j.jneumeth.2018.04.002
Reference: NSM 7977

To appear in: *Journal of Neuroscience Methods*

Received date: 28-9-2017
Revised date: 19-3-2018
Accepted date: 4-4-2018

Please cite this article as: López-Cabrera José D, Lorenzo-Ginori Juan V.Feature Selection for the Classification of Traced Neurons.*Journal of Neuroscience Methods* https://doi.org/10.1016/j.jneumeth.2018.04.002

Feature Selection for the Classification of Traced Neurons

José D. López-Cabrera[a,*], Juan V. Lorenzo-Ginori[a]

[a] Centro de Investigaciones en Informática, Universidad Central "Marta Abreu" de Las Villas, Santa Clara, Cuba, CP 54830

Corresponding Author:

Tel: 00 53 281157

Highlights

- Several feature selection methods where used to improve the classification performance of traced neurons which belong to the neocortex of mice, divided into pyramidal cells and interneurons.
- Random Forest algorithm was the best one, reaching an AUC value close to 0.9
- L-measure features EucDistanceSD, PathDistanceSD, Branch_pathlengthAve, Branch_pathlengthSD and EucDistanceAve were present in more than 60% of the selected subsets

e-mail: josedaniellc@uclv.cu

Abstract

Background

The great availability of computational tools to calculate the properties of traced neurons leads to the existence of many descriptors which allow the automated classification of neurons from these reconstructions. This situation determines the necessity to eliminate irrelevant features as well as making a selection of the most appropriate among them, in order to improve the quality of the classification obtained.

Methods

The dataset used contains a total of 318 traced neurons, classified by human experts in 192 GABAergic interneurons and 126 pyramidal cells. The features were extracted by means of the L-measure software, which is one of the most used computational tools in neuroinformatics to quantify traced neurons. We review some current feature selection techniques as filter, wrapper, embedded and ensemble methods. The stability of the feature selection methods was measured. For the ensemble methods, several aggregation methods based on different metrics were applied to combine the subsets obtained during the feature selection process.

Results

The subsets obtained applying feature selection methods were evaluated using supervised classifiers, among which Random Forest, C4.5, SVM, Naïve Bayes, Knn, Decision Table and the Logistic classifier were used as classification algorithms.

Comparison with Existing Methods

Feature selection methods of types filter, embedded, wrappers and ensembles were compared and the subsets returned were tested in classification tasks for different classification algorithms.

Conclusions