Computers and Electronics in Agriculture 134 (2017) 27–32

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Ensembles of wrappers for automated feature selection in fish age classification

Sergio Bermejo

Departament d'Enginyeria Electrònica, Universitat Politècnica de Catalunya (UPC), Jordi Girona 1-3 (C4 building), 08034 Barcelona, Spain

ARTICLE INFO

Article history:
Received 21 April 2016
Received in revised form 6 January 2017
Accepted 10 January 2017
Available online 20 January 2017

Keywords:
Automated fish age classification
Atlantic cod otoliths
Feature selection
Nearest neighbor classifiers
Statistical pattern recognition
Support vector machines

ABSTRACT

In feature selection, the most important features must be chosen so as to decrease the number thereof while retaining their discriminatory information. Within this context, a novel feature selection method based on an ensemble of wrappers is proposed and applied for automatically select features in fish age classification. The effectiveness of this procedure using an Atlantic cod database has been tested for different powerful statistical learning classifiers. The subsets based on few features selected, e.g. otolith weight and fish weight, are particularly noticeable given current biological findings and practices in fishery research and the classification results obtained with them outperforms those of previous studies in which a manual feature selection was performed.

1. Introduction

One of the most challenging problems in the field of pattern recognition (PR) is feature extraction (Guyon et al., 2006), which aims finding the most compact and discriminative set of properties or "features" presented in data. Although many research in feature extraction has been addressed to automate such a process, it has traditionally been considered a task much more problem- or domain-dependent than others in PR (Duda et al., 2001) since a good knowledge of the domain could be used to obtain such features, at least tentatively.

Fish age classification, a PR task of vital relevance among others for stock assessment and management (Girdler et al., 2010), usually relies on such manual procedures for feature extraction. In this direction, several fish features have been proposed for use in statistical fish age prediction and classification, with special emphasis in recent years to fish otolith features based on Fourier descriptors (Fablet and Le Josse, 2005; Galley et al., 2006) and different morphological parameters (Burke et al., 2008; Bermejo et al., 2007; Robotham et al., 2010; Hua et al., 2012).

However, the generalization error of statistical classifiers –i.e. their ability to mistake new examples taken on the same problem– tends to increase as of the number of features (Raudys and Jain, 1991) and, accordingly, the use of an arbitrary number of them leads to poor performance. One example of such behavior was demonstrated in (Bermejo, 2014) using multi-class support vector machines for fish age classification of an Atlantic cod database. Hence, if automatic feature extraction methods were additionally employed for reducing the complexity of the feature space a better performance could presumably be obtained. Other important benefits of such strategy includes speeding up computation (e.g. decreasing training times) and data understanding or reverse engineering (i.e. to increase knowledge about the problem, which can be of vital significance in natural sciences like fisheries science).

While some authors (e.g. Webb, 2002) consider feature extraction a process only concerning transformation of the original variables, it is generally agreed that feature extraction comprises the following steps: feature construction or generation that performs some kind of preprocessing –e.g. a linear or non-linear transformation– of the original raw variables (Theodoridis and Koutroumbas, 2008) and feature selection (Guyon and Elisseeff, 2003) that chooses a subset of the original or transformed variables.

There are three main approaches to feature selection (Blum and Langley, 1997; Guyon and Elisseeff, 2003, 2006): filter methods, wrappers and embedded methods. While filters can be viewed as a preprocessing step since they select a subset of variables independently of the chosen predictor (e.g. a classifier), wrappers use it as a black box or subroutine to score subsets of variables and

E-mail address: sergio.bermejo@upc.edu

http://dx.doi.org/10.1016/j.compag.2017.01.007
0168-1699/

embedded methods perform variable selection in its training phase. In this way, wrappers are based on an arguably better estimate of accuracy obtained with the predictor that will employ the feature subset than a separate measure that may have a completely unrelated inductive bias, but, at the expense of a higher computational cost (Blum and Langley, 1997). However, the inherent variance (or instability) of feature subset selection methods (Guyon and Elisseeff, 2006) produces a plethora of very different subsets attained for different conditions, i.e. different parameter tuning, small perturbations of the dataset or presence of redundant features.

In this paper, a novel wrapper that use a form of ensemble learning (Dietterich, 2003), which are based on a strategic combination of several predictors, have been proposed to attain a greater stabilization and thus a better generalization of the feature selection process. Feature subsets obtained with the ensemble of wrappers which employ as base classifiers support vector machines and nearest neighbor classifiers allow achieving a classification performance that outperforms a previous study (Bermejo, 2014). Moreover, these subsets that have very few features, e.g. only otolith weight and fish weight, are of relevance in accordance with recent findings in fisheries research.

2. Materials and methods

2.1. Atlantic cod database

This dataset contains morphological and biological features for codfish age classification. Traditional methods for determining the age of fish usually focus on analyzing hard parts of the body, such as otoliths, which are small particles in the inner ear composed of a gelatinous matrix and calcium carbonate, since the macroscopic growth patterns of otholiths are correlated with the fish' age.

The fish database consists of one hundred forty-five Atlantic cod of known age (varying from two to six years) from the Plateau stock that were hatched the same year and later kept and reared in pen cages. This dataset was created from originally fish of known-age sampled at different years in captivity since a number of samples were recaptured once a year. Otoliths were taken from this stock and weighed and also four morphological features were recorded following an image analysis method defined in (Bermejo et al., 2007). Additionally, fish length, weight and sex were available for each sample.

The leave-one-out (LOO) error using a 1-NN rule (Devroye et al., 1996; pp. 407-421) were computed for this set (19.31%) as a way to estimate the Bayes error, i.e. the minimum amount of classification error achievable. In a previous study with this database using SVMs (Bermejo, 2014), the minimum obtained error was 21.79% for otolith weight, fish length, weight and sex acting as features, which is lower than an error rate of 22% obtained for a related dataset, combining five experts' readings, who were given low and intermediate levels of information about fishes and the conditions that they were obtained (Doering-Arjes et al., 2008). According to the above considerations, some improvement in accuracy is still possible with SVMs taking the value of the LOO estimate as an approximate lower bound to the attainable misclassification rate. Table 1 displays the results of the LOO estimate and also includes other relevant information of this dataset. A more comprehensive description of the cod database is presented in (Bermejo, 2014).

2.2. Ensemble of wrappers

Ensemble learning methods, such as bagging, boosting and variants (Bauer and Kohavi, 1999) are based on the formation of a set of predictors $\{\phi(\mathbf{x};\mathbf{D}_k)\}$ trained on a sequence of learning sets $\{\boldsymbol{D}_k\}$, which are typically generated from a single dataset $\boldsymbol{D}$ using a resampling technique such as bootstrapping (Efron and Tibshirani, 1994). The second core element of any ensemble method is a combination strategy: the most obvious and effective procedure for combining a sequence of K predictors $\{\phi_k\}$ whose outputs are continuous is averaging (Breiman, 1996a), i.e. $\bar{\phi} = \sum_k \phi_k / K$. Ensembles have been built specifically to select features; for example, variants of AdaBoost for feature selection have been proposed using decision stumps (Long and Vega, 2003) and a mutual information measure (Liu et al., 2008), random subspace methods have also been employed in feature ranking for removal of irrelevant variables (e.g. Tuv et al., 2009), and ensembles based on bootstrapping have been combined with recursive feature elimination and feature ranking (Windeatt et al., 2007). Furthermore, several studies have analyzed the use of averaging and voting for the combination of multiple feature selection criteria with the hope that several criteria would reflect different properties in feature subsets (e.g. Somol et al., 2009), although none of them has analyzed the effect of these procedures using a sole criterion to obtain a single feature subset. Our proposal addresses this problem in the context of wrappers.

Wrappers (Kohavi, 1995) select features from a pool of feature sets based on a decision rule of the form $\phi_W = \arg\min_j L_{CV}(C^j_{\mathbf{D}};\mathbf{D})$, that is, they select the jth feature set for which $L_{CV}(C^j_{\mathbf{D}};\mathbf{D})$ is the minimum, where L_{CV} is the cross-validation error based on the dataset $\boldsymbol{D}$ computed in the base classifier $C^j_{\mathbf{D}} = C(\mathbf{x}^j;\mathbf{D})$, whose inputs belong to the jth feature set space. If the database is divided into a learning set $\boldsymbol{D}$ for performing cross-validation and a test set $\boldsymbol{T}$ for final assessment of the classifier after feature selection, a sequence of learning sets $\{\boldsymbol{D}_k\}$ and test sets $\{\boldsymbol{T}_k\}$ can be generated for different random splits of the database. Then, and in accordance to the theoretical analysis given in (Breiman, 1996a, 1996b), we propose in this paper a stabilized feature selection rule that can be obtained through averaging over L_{CV} in order to stabilize the metric used in wrappers directly, so the feature selection rule based on an ensemble of wrappers (EW) can be computed as $\bar{\phi}_{EW} = \arg\min_j(\sum_k L_{CV}(C^j_{\mathbf{D}_k};\mathbf{D}_k)/K)$. The proposed stabilization of the assessment criterion can be simply seen as an averaging of several k-fold cross-validation estimates (based on the output of the wrapper's base classifier) similarly to the way in which the outputs of several classifiers are stabilized through averaging. The reader is referred to Breiman, 1996a,b for further discussion, and definition, of stability.

A baseline algorithm for feature selection with wrappers using internal cross-validation (Flach, 2012) is suggested in Algorithm no. 1. The ensemble approach using rule $\bar{\phi}_{EW}$ is detailed in Algorithm no. 2 as a straightforward variation of

Table 1
Codfish dataset summary.

Size	No. of Features	Features/feature vector	No. of classes	Minimum leave-one-out error
145	8	Fish sex (S), fish length (L), fish weigh (W), otolith weight (OW), otolith contour length (C), otolith area (A), otolith maximum internal distance (I), otolith maximum perpendicular distance (P)/(P I A C OW W L S)	5 [fish age: 2–6]	0.1931 [for feature set 12 = $(00001100)_2$]