

# Selecting and constructing features using grammatical evolution

Dimitris Gavrilis<sup>b</sup>, Ioannis G. Tsoulos<sup>a,\*</sup>, Evangelos Dermatas<sup>b</sup>

<sup>a</sup> *Department of Computer Science, University of Ioannina, Greece*

<sup>b</sup> *Department of Electrical and Computer Engineering, University of Patras, Greece*

Received 5 November 2007; received in revised form 5 February 2008

Available online 4 March 2008

Communicated by T. Vasilakos

## Abstract

A novel method for feature selection and construction is introduced. The method improves the classification accuracy, utilizing the well-established technique of grammatical evolution by creating non-linear mappings of the original features to artificial ones in order to improve the effectiveness of artificial intelligence tools such as multi-layer perceptron (MLP), Radial-basis-function (RBF) neural networks and nearest neighbor (KNN) classifier. The proposed method has been applied on a series of classification and regression problems and an experimental comparison is carried out against the accuracy obtained on the original features as well as on features created by the PCA method.

© 2008 Elsevier B.V. All rights reserved.

*Keywords:* Artificial neural networks; Feature selection; Feature construction; Genetic programming; Grammatical evolution

## 1. Introduction

In classification theory, a set of features composes the pattern, typically classified in two categories; the primitive features, and a set of non-linear projections of the primitive features. The number and type of features are critical to the classification accuracy and computational complexity. As the number of features increases, additional examples are required to complete a reliable training process, allowing for more robust generalization capabilities without over-fitting effects. In the case of limited number of available data, the “curse of dimensionality” introduces two major problems: the selection of a subset from the original set of features that preserves the classification scheme and the construction of a robust set of new features from non-linear-transformations of the original set.

According to the Covers’ theorem (Cover, 1965), at least one non-linear extension of the features vector exists which

defines a features space where the classes of an arbitrary set of examples are linearly separated. In this direction, many methods have already been proposed for automatic definition of a non-linear extension of the features vector such as, multi-layer perceptron (MLP), polynomial, Radial-basis-function (RBF) neural networks etc. The proposed method utilizes the grammatical evolution technique to select and construct a subset of the original features set that satisfies or approaches a priori defined classification accuracy. In contrast to the traditional features selection approach, where experts or semi-automatic methods derive or transform the original set of features, the proposed method (Fig. 1) is fully automatic. The proposed method is quite different from others because:

- (1) It utilizes a BNF grammar to represent the search space. This allows the easy manipulation of the search space through the grammar.
- (2) The method can be easily scaled to use complex functions (like statistics, neural networks, etc.). In order to achieve that, one can define a new function and then simply use it in the grammar.

\* Corresponding author. Tel.: +30 2651098871; fax: +30 2651098890.  
E-mail address: [itsoulos@cs.uoi.gr](mailto:itsoulos@cs.uoi.gr) (I.G. Tsoulos).

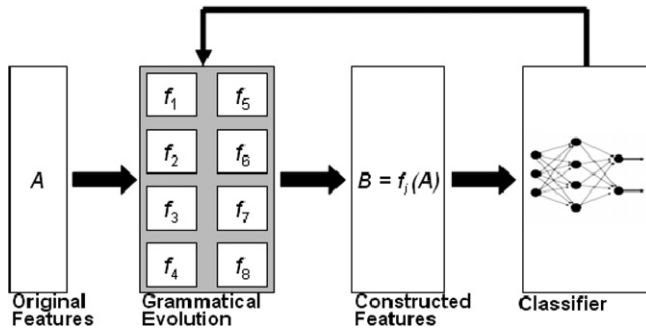


Fig. 1. Features selection and construction using grammatical evolution and neural networks.

- (3) The results produced by the proposed method are in a human – readable form. In some problems this can lead scientists to a better evaluation and better understanding of the resulting features.

In Georgoulas et al. (2007) the proposed method was compared against three known classifiers (Ldc, Qdc, 1-nn) which reduced the original number of features using PCA. The proposed method achieved an overall accuracy of 88.13% which was the best against the other three (Ldc: 68.75%, Qdc: 72.50%, 1-nn: 80.63%). In Gavrilis et al. (2006), a variation of the proposed method was used to compress a large set of original features (15,000) into 20 features that were used for classification. After the compression, an average accuracy of 96–97% was achieved.

## 2. Related work

Representation transformations in classification and regression problems can be grouped into three different categories: feature selection, construction, and reduction through linear transformation. In the first case, a subset of the original features is defined, eliminating the irrelevant and less significant features. In the second case, a number of efficient solutions have already been proposed using dimensionality reduction techniques including genetic programming (GP) (Raymer et al., 2000), singular value decomposition, and principal component analysis. In a more sophisticated approach, the linear weights are estimated by a fully automatic algorithm (Jarmulak, 1999). The feature construction method is more complex, since it usually involves features selection, followed by linear and non-linear-transformations of the primitive set. The evolutionary approach is among the most recent methods of constructing new features from the original set of primitive features (Krawiec, 2002).

Related work in the area of feature selection and construction, mostly involves feature selection rather than feature construction. In Jarmulak (1999), a genetic algorithm is used to reduce the initial set of 26 features, eliminating irrelevant features and improve the classification accuracy using a C4.5 decision tree and the nearest neighbor rule. The authors proposed an evolutionary approach to select

the relevant features, and real valued genes by assigning weights to the original features. The experimental results show that the selection method improves the classification rate.

Raymer et al. (2000) a genetic algorithm (GA) in combination with a K-nearest neighbor classifier to select a subset of the original feature set in the bio-chemistry domain. A binary representation in the chromosome is used to select or discard a feature. The number of centers of the KNN classifier is also encoded in the chromosome and weights are assigned to each feature. Experimental results in a bio-chemistry domain dataset show that the GA/KNN performed well along eight the following classifiers:

1. The linear discriminant
2. The quadratic discriminant
3. Nearest neighbor
4. Bayes (independent)
5. Bayes (2nd order)
6. Neural network (back prop)
7. Predictive value max.
8. CART tree

The genetic programming framework and a tree-like representation have already been presented in discovering new features for classification problems by Krawiec (2002).

In this paper, a fully automatic features selection and construction (FSC) method for robust pattern recognition is presented, introducing a novel feature construction method with certain advantages. Taking into account the close relation between the features vector and the classification system, three problems are solved simultaneously in a single optimization process: the features selection, construction, and the training process of the classification system. The proposed FSC method is based on the original genetic programming method proposed by Krawiec (2002), enriched by several innovations and extensions: The use of a Backus Naur form (BNF) description is less restrictive, as the features construction form is concerned. Thus, the resulting function can become very complicated, if necessary, using the appropriate number of the original features. New basic functions can be easily introduced, and complex restrictions can be encoded in the BNF description. The genetic search prefers simpler expressions and the new features are normalized automatically. The proposed data-driven FSC method detects an effective set of features dedicated to the pattern classification module involved in the GP optimization function including also the training process of the pattern classification system.

The proposed FSC method is evaluated using twenty artificial and real well-known datasets and the experimental results are compared with a genetic feature selection method (Raymer et al., 2000), and the popular features reduction method known as principal component analysis. In all experiments, the multi-fold cross-validation methodology is used in order to make the datasets less prone to both training and classification error estimation due to the limited number of examples.

Download English Version:

<https://daneshyari.com/en/article/535402>

Download Persian Version:

<https://daneshyari.com/article/535402>

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