Learning and Nonlinear Models - Revista da Sociedade Brasileira de Redes Neurais (SBRN), Vol. 5, No. 1, pp. 20-35, 2007

Morphological identification of inductive signatures using an Adaptive Resonance Theory based learning scheme

Glauston R. Teixeira de Lima[1], José Demísio S. Silva[2], Osamu Saotome[3]

[1]Programa de Pós-Graduação em Engenharia Eletrônica e Computação, Área de Dispositivos e Sistemas, ITA
[2]Laboratório Associado de Computação e Matemática Aplicada – LAC, INPE
[3]Divisão de Engenharia Eletrônica e Computação, Área de Dispositivos e Sistemas, ITA
E-mail: glau11@gmail.com, demisio@lac.inpe.br, osaotome@ita.br

Abstract -This paper presents an algorithm that uses a learning scheme based on the Adaptive Resonance Theory – ART2 to identify morphological patterns in two classes of vehicle inductive signatures and to construct inductive signatures subclasses based on morphological similarities among inductive signatures and previously identified patterns. The algorithm is initialized using all inductive signatures of a class as inputs to the learning scheme. During the iterations of the algorithm, the inputs whose morphological similarity exceeds a pre-established threshold value are merged into a single new input to the learning scheme. The merging process stops when the threshold value can not be overcome. The remaining inputs are considered to be the desired group of patterns representing the morphological diversity of the inductive signature class. Next, the morphological patterns identified by the learning scheme are submitted to a clustering algorithm to form the subclasses. In some classification tasks, the problem of classes overlap in the input space can be approached under the point of view of the morphological similarity among groups of patterns of each class. In these cases, pre-processing the original input space dividing it into morphological subclasses serves to identify the regions where the classes overlap is concentrated and to break down the initial classification task, whose solution is more difficult, in minor tasks, hypothetically, easier to be resolved. However, the supposition that the division in subclasses led to a better separation of the two classes of inductive signatures will be checked in subsequent stages of our research through tests with neural classifiers.

Keywords - Vehicle Inductive Signatures, Clustering, Adaptive Resonance Theory Neural Network

1 Introduction

The difficulty to identify the class of a pattern, which was randomly drawn from a set of patterns, is proportional to the overlap of the attribute spaces of the classes that form the set of patterns [1], [2]. In many cases, pre-processing techniques are used to map the patterns to a new domain in which the overlap of classes may decrease, or even (in the best cases) be eliminated. However, a common dilemma to all pattern classification tasks is related to the choice of the appropriate pre-processing techniques to achieve the reduction of the overlapping regions of the involved attribute spaces. An effective approach for many cases is to use the divide to conquer principle [3], that is, to start the pre-processing breaking the original problem into several minor ones.

Consider, for instance, a classification problem in which the input patterns are the inductive signatures of two classes of vehicles as shown in the Figure 1. These signals are collected by inductive loop traffic sensors and used in traffic surveillance and management systems to identify a vehicle class and estimate its speed among other expected results [4], [5], [6].

The morphology of the curves in Figure 1 is derived from the impedance alteration of the magnetic loop when the vehicle passes over [7]. It is hypothesized that the proximity of the metal parts of the axles, do alter the impedance of the loops and thus the presence of the axles is signalized. This way, the vehicle can be classified by the number of axles.

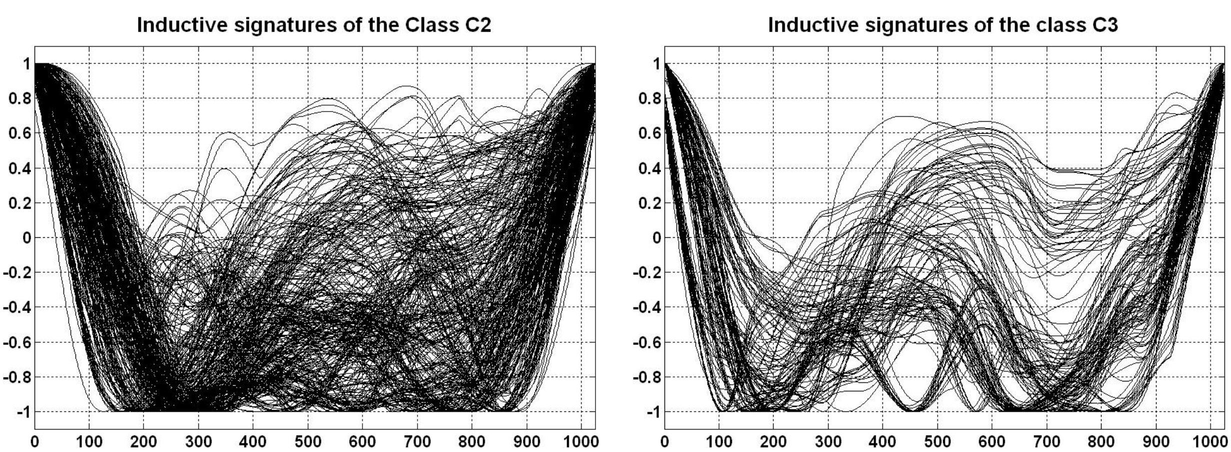


Figure 1 - Inductive signatures of two vehicles classes: on the left, signatures of trucks with one rear axle; on the right, signatures of trucks with double rear axles.

The left side of Figure 1 shows the inductive signatures of trucks with two axles (one rear axle) and the right side shows the inductive signatures of trucks with three axles (double rear axle). The inductive signatures of the trucks with two axles will be referred in this article as class C2 and the inductive signatures of the trucks with three axles will be referred as class C3. The data shown in Figure 1 was acquired in a real world setup assembled near a toll park on a road. There are 328 C2 signatures and 132 C3 signatures each one with 1024 samples. For classification purposes, in addition to the class overlapping, this database presents two more problems: class imbalances and reduced amount of data.

The signatures were filtered by a Savitzky-Golay' filter [8] and normalized to the range -1 to 1. The filtering and the amplitude normalization are justified as follows. Consider two moments of the passage of a truck over the inductive loop illustrated in Figure 2.

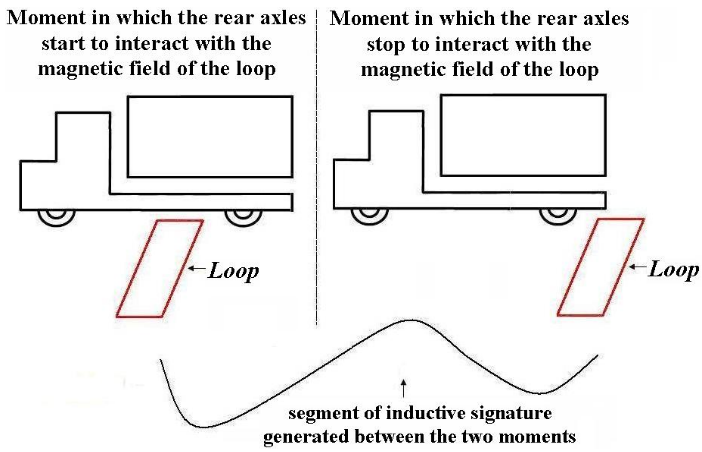


Figure 2 - Two moments of the passage of a truck over the loop: at left, the rear axles start to disturb the magnetic field; at right, the rear axles stop to disturb the magnetic field.

Figure 2 also shows a typical segment of inductive signature generated between the two moments. By considering that the situation sketched in Figure 2 is as a good approximation for real situations, we work with the following assumption to develop a method of classification for classes C2 and C3: the information that is useful to distinguish a class from the other, i.e., the information about the rear axle of the truck, is stored in the inductive signature in long duration segments or, in other words, in the low frequency components of the inductive signatures. So, to remove existing high frequency components from the raw data (which, according to our work hypothesis, don't carry useful information), the signatures have been smoothed with Savitzky-Golay' filter. This filter does a local polynomial fitting on a set of samples to calculate the smoothed value for each sample. The main advantage of this method is that the smoothed samples tend to preserve features of the distribution of the original samples.

The minimum value of an inductive signature is determined, primarily, by the proximity of the vehicle metal structure with the inductive loop. By considering that the distance from the rear axles to the road surface is standardized for C2 trucks and C3 trucks, any discrepancies in the value of this attribute (the minimum value of the inductive signature) are generated by other parts of the metal structure of the vehicle that are close to the rear axle. These "other parts of the metal structure of the vehicle" (unlike the rear axles) are not characteristics of the vehicle class and they can be present in trucks of both classes and its influence on the format of the inductive signatures should be suppressed. Hence, the inductive signature amplitudes have been normalized.

In order to better understand and formally characterize the overlap between the two classes, some preliminary tests were done with the inductive signatures by using 5 different neural networks to classify the data in the two classes [9]. The results showed that patterns in both classes share the same morphological patterns.

Suppose it is desired to classify the inductive signatures of classes C2 and C3 by using a supervised learning neural network. Bearing in mind the assumption that a number of C2 and C3 vectors share the same morphological patterns, the classification task may be implemented by a) using all the inductive signatures; or b) by dividing each class into "morphological subclasses", establishing a correspondence between the subclasses of C2 and the subclasses of C3 and accomplishing the classification task separately for each corresponding pair of subclasses.

The approach in b) has the disadvantage of requiring training and testing of more than one classifier. However, the amount of time to train each classifier may be smaller, due to the reduced number of signatures in each pair of subclass, and the complexity of each classification task may be reduced to the specific complexity for each pair of subclass. These assumptions may be verified by comparing the percentage of correct classification of the two classification approaches.

The remainder of this paper is organized in section 2, that presents a briefly revision of related work; section 3 that presents the algorithm to subdivide the two original classes, using an Adaptive Resonance Theory based learning scheme; and section 4 that presents the results of the classes subdivision and the results achieved with two benchmarks. Finally, section 5 brings some conclusions.

2 Related works

2.1 Classes subdivision in classification tasks

Some approaches to classification problems that make a previous division of the classes in subclasses, are proposed in [10], [11], [12], [13]. In [10], the regions of the subclasses in the input space are approximated by hyper-rectangles constructed through a process of random selection of samples. After definition, the hyper-rectangles are used to discriminate and to select attributes. In the work presented in [11] a template for each class is generated from the templates of the subclasses. Using correlation as a similarity measure, the aggregation of the templates is made to maximize the value of the lowest correlation coefficient between the subclasses templates and the single template that will substitute them. The method proposed in [12] builds the subclasses starting from the set of patterns and, at the same time, estimates the respective distributions of probability conditioned to the subclass. The subclasses are defined according to a criterion that minimizes the cross entropy of the conditional distribution functions. In [13], the input space is initially divided in areas according to the relative positions of the classes that are determined by their patterns. These regions overlap. A set of fuzzy rules, in the form of relational matrix, is elaborated based on this partition of the input space. A pattern is classified taking up the membership values of the pattern in each region and calculating the pattern possibility for each class by applying Zadeh' modified inference rule between the membership values and the relational matrix.

2.2 Time series clustering methods

Clustering methods can be divided into two main approaches: nonparametric methods or based on distance and parametric methods or based on models [14]. Two of these methods are related to the clustering method proposed in this article: agglomerative hierarchical distance-based methods and neural network models-based methods.

In the agglomerative methods, the number of clusters is not previously known and, in the beginning, each available pattern is considered a cluster. Next the algorithm creates a tree of clusters by merging patterns and finishes when the appropriate number of clusters is found. Agglomerative methods have two drawbacks: the process of merging patterns leaves no margin for further correction and the search for the correct number of clusters can be exhausting. Further, the computational complexity of agglomerative methods is generally proportional to the square of the number of patterns.

Neural network-based clustering methods explore the structures existent in the data to formulate models and to build data clusters according to these models. An advantage of this approach is the possibility to obtain good results even in databases with low representativeness. Neural networks such as the Self-Organizing Maps and Adaptive Resonance Theory Networks perform clustering by competitive learning and the cluster with the best model incorporates the pattern. This approach may not perform well if the data are statistically very similar.

Full reviews on time series clustering methods are provided in recent papers [15], [16].

2.3 The clustering method proposed

In the input space formed by the two classes of inductive signatures there is a large class overlap. As mentioned earlier, the inductive signatures share the same morphological prototype and, therefore, in an eventual classification task, it seems to be valid to approach the overlap problem under the point of view of the morphological similarity among groups of patterns in each class.

To implement this idea, we may perform a morphological characterization of the classes by dividing them into morphological subclasses. After the subdivision, the regions of the initial input space where the overlap is concentrated become explicit and each of these regions, for the purpose to classify the inductive signatures, can now be treated separately. Thus, we apply the "divide to conquer principle" aiming to make the classification task easier. This is our motivation for developing a clustering method.

A good strategy to accomplishing clustering tasks consists in integrating different approaches in order to compensate eventual deficiencies that each clustering approach has [15]. In this paper we proposed a clustering method using an Adaptive Resonance Theory based learning scheme to clustering inductive signatures into morphological subclasses and we believe that the proposed method is attractive for the following reasons:

First we must emphasize that, in the proposed method, most of the computational cost is destined to the choice of the centers of clusters. That is, we work with the idea that the representativeness of the selected centers determines the quality of the formed clusters.

The choice of the centers is an agglomerative process in which similar input patterns are merged into a single pattern. But, to merging two patterns, the similarity between them must be checked by the vigilance parameter. This is the main reason for using the ART2 network in our clustering method, namely, the possibility to exercise a finest control on the concept of similarity among patterns through the vigilance parameter.

The agglomerative process, from one phase to the next, reduces the number of patterns that remain in the centers selection process. The patterns go to the next phase and starting to compete again to remain in the selection process. So, if within a phase a merge decision is refined by using the vigilance parameter, in the passage from one phase to the next, possibilities to adjust earlier merge decisions are created. So, unlike the traditional agglomerative clustering methods, the agglomerative decisions, in our clustering method, are more flexible.

Therefore, we propose a hybrid clustering method and the main contribution of this approach is the ability to integrate characteristics of both an agglomerative clustering method and a model-based clustering method in order to improve the clustering quality.

3 The subdivision algorithm

3.1 The similarity criterion

Our aim is to build the subclasses clustering inductive signatures that are morphologically similar. For this purpose, take into account that correlation coefficient is related to the spatial distribution of the information in each vector, we think that the correlation coefficient is a similarity measure to be adopted in the subdivision algorithm.

However, before calculating the correlation coefficient the curves were pre-processed by applying a rescaling operation to each pattern during its presentation to the ART2 network. In order to better explain the idea of this pre-processing, consider the pairs of inductive signatures shown in Figure 3.

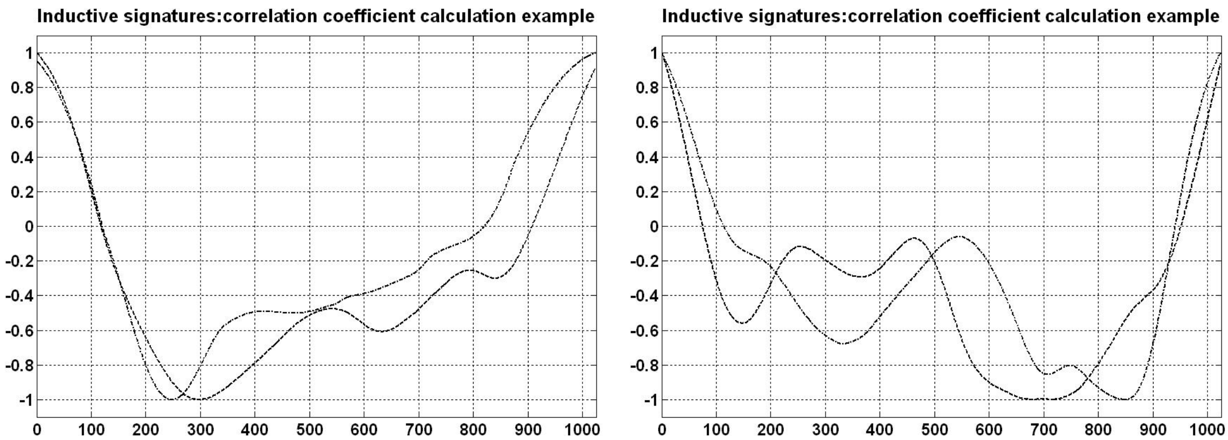


Figure 3 - Pairs of inductive signatures for correlation coefficient calculation

The correlation coefficient for the inductive signatures shown on the left of Figure 3 is 0.9401, and for the inductive signatures on the right of Figure 2 it is 0.7575.

The analysis of the curves in Figure 3 shows that there is a difference in the location of the first minimum points of the signatures. The location of the first minimum points may be approximately the same by linearly interpolating additional samples in the inductive signatures. Since the inductive signatures were smoothed by the low-pass filtering operation, increasing the number of samples may not affect the morphology of the curves.

3.1.1 Linear interpolation of the signatures on the left side of Figure 3

The first minimum points of the curves are located at positions 248 and 296, thus showing a difference of 48 samples. In the inductive signature, whose first minimum point is in the position 248, 48 samples are added to the initial segment (from first sample to sample 248). In the inductive signature, whose first minimum point is in the position 296, 48 samples are added to the final segment of the vector (from sample 297 to last sample), this way both curves will maintain the same length for the correlation operation.

3.1.2 Linear interpolation of the signatures on the right side of Figure 3

There is a difference of 186 sample points between the first minimum points of the curves on the right side of Figure 3 (the locations along the vectors are at positions 147 and 333). Thus 186 samples are added to the beginning of the curve whose first minimum is at position 147, and 186 samples are added to the end of the other curve. Both curves still present the same length for correlation calculation purposes.

After the modifications, the two pairs of inductive signatures were as shown in Figure 4.

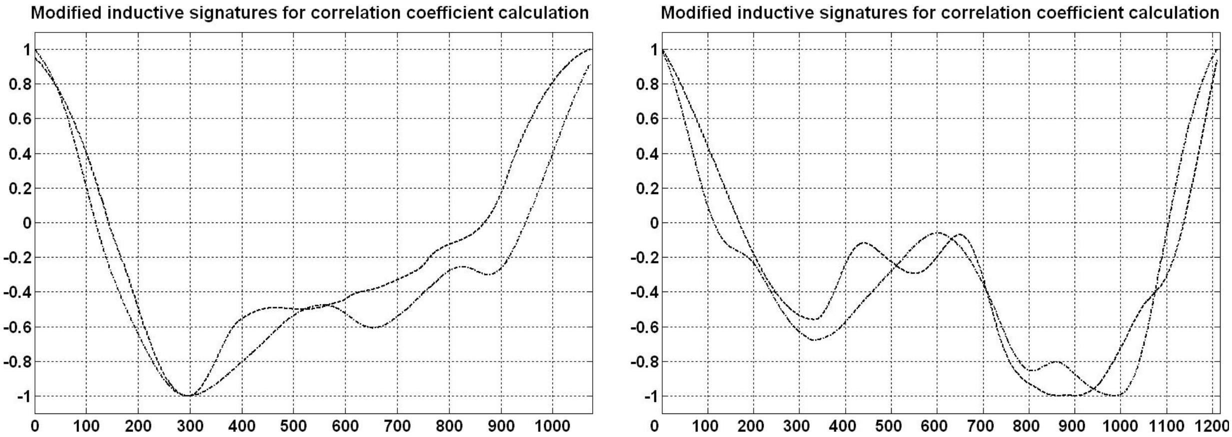


Figure 4 - Pairs of inductive signatures modified to calculate the correlation coefficient

After the described pre-processing phase the correlation coefficients of the two pairs of inductive signatures on the left and on the right sides of Figure 3 are 0.9686 and 0.9142, respectively.

The described pre-processing is possible because of the nature of the data in the curves. In the absence of a vehicle the loop impedance is kept constant. The identification of the vehicles is related only to the valleys of the curves that result from the passage of the axles over the loop. The initial segments of the inductive signatures are bounded by the first sampled and minimum points of the curves, and are quite similar for all the vectors in the two classes. In these regards, such segments are less relevant for discrimination purposes. The most relevant morphological information is located after the first minimum and before the last minimum of the curves. Thus, the morphological discrimination between two inductive signatures, expressed by the correlation coefficient, should be mostly related to the inner segments of the curves. Therefore, the removal of this difference contributes to make the correlation value be predominantly determined by the inner parts of the inductive signature segments with the most relevant morphological features.

3.2 The algorithm

The proposed subdivision algorithm is based on the Adaptive Resonance Theory architecture illustrated in Figure 5 and described in Carpenter and Grossberg [17]. In the proposed algorithm, the input layer I is composed of two types of units: U units that receive the input patterns, and P units that receive the signal from U. In layer I, normalization and noise filtering are not implemented because the inductive signatures have already been submitted to other forms of normalization and filtering. The output layer has the same structure proposed in the model of Carpenter and Grossberg [17], [18].

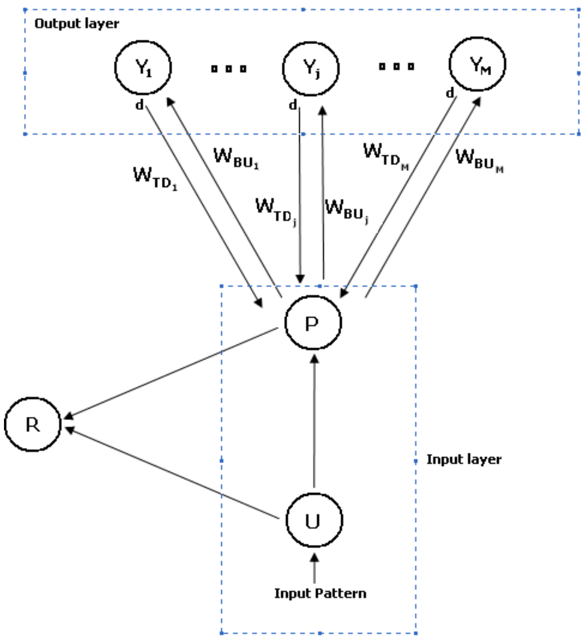


Figure 5 – An Adaptive Resonance Theory based learning scheme to extract morphological patterns

The subdivision of the classes is achieved in two stages. In the first stage the prototypes that represent the morphological diversity of the classes are found. In the second stage, each class is divided into subclasses based on the group of prototypes by using a clustering algorithm.

The first stage of the algorithm is divided into phases that are divided into epochs. An epoch corresponds to the presentation of all the input patterns. At each phase F, the set of input patterns to the neural network is different. In phase 1, for instance, the input patterns are all the inductive signatures of the class. The first step in each phase is the calculation of the "phase correlation matrix", CM_F. The CM_F is a square matrix in which the i-th column contains the values of the correlation coefficient between the i-th input pattern and all the other input patterns. While the CM_F is calculated the output layer of the neural network contains no output unity.

The initial value of the vigilance parameter is set based on the values of the first phase CM_F matrix according to the following calculation steps: a vector MV is formed with the maximum value of each column of CM_1; the average value of MV is calculated and a vector MV' is formed by taking all values in MV greater than the average value; the initial vigilance parameter is chosen in the range from the average value of MV' to the maximum value of CM_1. For C2 and C3, for instance, the maximum values of CM_1 and the average values of MV' are, respectively, (0.9999, 0.9954) and (0.9995, 0.9938). Thus, in the first phase of the C2 and C3 subdivision process, the vigilance parameter value is initialized with a high value equal to 0.999. In each new phase, the vigilance parameter is decreased by 0.001 until it reaches the lower limit, set at a medium vigilance value equal to 0.95. Taking into account that de initial value of the vigilance parameter is very high, we consider that when the vigilance parameter reaches a medium value (for instance, 0.95), the input patterns that remain not yet merged with other patterns (because the correlation between them do not overcome 0.95) are sufficiently dissimilar to represent the entire morphological diversity of the class. This way to vary the vigilance parameter, from high to medium values in small decreases, the vigilance parameter is allowed to be used more efficiently to control the similarity between patterns and, hence, increasing the possibility that the similar patterns merging process produces the expected number of patterns reduction while the few remaining patterns absorb the entire morphological diversity of the class.

The input patterns are randomly presented to the learning scheme. For each input pattern presented to the network, units U receive the input pattern, and units P receive the information from U (P=U). The CM_F column, corresponding to the input pattern, is verified. If the maximum correlation value in the column of CM_F overcomes the vigilance value of the phase, a new unit Y is created in the output layer. The current input pattern becomes the corresponding top-down weight vector, W_{TD}, connecting the new inserted output unit Y to unit P and the input vector which is the most similar to the current input pattern becomes the bottom-up weight vector, W_{BU}, connecting unit P to the new inserted output unit Y. The insertion of a new unit Y only happens in this way and, when that happens, the input pair involved, that is, the current input pattern and its most similar corresponding input, is registered in a vector V1. The vector V1 will always be checked to avoid the insertion of a new unit in the output layer for the same input pair. This control prevents the indiscriminate insertion of units Y and thus it guarantees a reduction of the number of input patterns from one phase to the next.

After the insertion of an output unit Y and its corresponding weight vectors [W_{TD}, W_{BU}], it starts competing to learn the pattern presented at the input. Thus, in the subsequent input patterns presentations, the correlation values between the input pattern and all the weight vectors, W_{BU}, are calculated.

$$\mathbf{C\_V(j) = correlation\ (input\ pattern\ ,\ W_{BUj})} \tag{1}$$

The output unit Y_J with the largest C_V value is the winner. But, to learn the input pattern, Y_J needs go through the vigilance test described as follows.

The output of the unit P is now given by

$$\mathbf{P = U + d^* W_{TDJ}} \tag{2}$$

where d is the activation of the Y units in output layer and its value is set to 0.99.

The reset unit, R, receives the outputs of units U and P and calculates the correlation coefficient. The CM_F matrix is visited again. Let M be the maximum correlation coefficient value occurred in the column of CM_F corresponding to the current input pattern. The following situations should be analyzed:

1. R and M overcome the vigilance value
   - If R > M, the weights vectors [W_{TDJ}, W_{BUJ}] are adjusted by

     $$\mathbf{W_{TDJ} = \alpha^*d^*U + [\ 1 + \alpha^*d^*(1\text{-}d)]\ ^*\ W_{TDJ}} \tag{3}$$

     $$\mathbf{W_{BUJ} = \alpha^*d^*U + [\ 1 + \alpha^*d^*(1\text{-}d)]\ ^*\ W_{BUJ}} \tag{4}$$

     where, α is the learning rate set as 0.001.
   - If M > R
     - If the current input pattern does not appear in V1, a new unit Y will be added to the output layer (following the steps explained before).
     - If the current input pattern appears in V1, the unit Y_J is inhibited by making C_V(J) equal to -1. The output unit Y with the next largest C_V value is chosen and the vigilance test is repeated.

2. R and M are lower than vigilance value
   - The unit Y_J is inhibited by making C_V(J) equal to -1. The output unit Y with the next largest C_V value is chosen and the vigilance test is repeated.

The search among the units Y is limited to the units with the three largest C_V values. If none of them can learn the input pattern, the structure of the neural network is not modified by the presentation of the input pattern. That is, the existing weight vectors [W_{TDJ}, W_{BUJ}] are not adjusted and none new unit Y is added to the output layer.

Each phase F may finish for one of the three reasons:

1. If during an epoch, each input pattern is learned by an existing output unit Y. That is, for all the input patterns presented to the network, R > vigilance parameter and R > M.
2. From the tenth epoch, the following control is started.
   - In each epoch, the input patterns whose presentation did not change the structure of the network are recorded in a vector, V2.
   - At the end of each epoch, the vector V2 is compared with the vector V2 of the previous epoch. If the two vectors are equal, the value of a control variable is increased by 1. When the control variable value reaches 10 (which means that during ten successive epochs the input patterns that do not alter the structure of the network are the same), the phase F is finished.
3. The maximum number of epochs, allowed for the algorithm to run, is achieved in phase F.

When a phase F is finished, the following actions are performed:

1. The weight vectors W_{TD} and the input patterns, that appear in vector V2, are gathered to form the new set of input patterns for the next phase of the algorithm. This is the mechanism that reduces the number of input patterns from one phase to the next. It is important to recall that the weight vectors [W_{TD}, W_{BU}] are initialized by pairs of input patterns in phase F. However, when the phase F finishes, only the weight vectors W_{TD} are passed to the next phase as input patterns.
2. The current set of input patterns and the weight vectors [W_{TD}, W_{BU}], are discarded. That is, the output layer of the network is emptied.
3. A variable that controls the total number of epochs of the algorithm is increased by the number of epochs of the phase F.

There are two stopping criteria for the execution algorithm

1. The maximum number of epochs allowed to be exceeded (set to 5000 epochs in the experiments).
2. The vigilance parameter reaches its minimum value (0.95)

4 Results

4.1 Results for inductive signatures

The algorithm was executed 10 times for each class of inductive signatures. The number of morphological prototypes identified by the proposed learning scheme, ranged between 19 and 25 for class C2, and between 13 and 17 for class C3. These differences were already expected because, according to [19], if the input patterns are randomly presented to the ART2 network , the number of formed clusters can vary. However, regarding the morphologic representativeness, the 10 groups of prototypes are quite similar and this enables us to consider each of them, for practical purposes, as the same result. It may be said, therefore, that the algorithm has converged.

Figure 6 shows two sets of prototypes obtained for class C3 and Figure 7 shows two sets of prototypes obtained for class C2. In each case the two most similar results were chosen to be shown. Note in Figures 6 and 7, that the similar prototypes have been placed in the same column. Also at the bottom of both figures are the corresponding linear models of the general geometric shape shared by the curves in the same column. These general geometric shapes may be viewed as "piecewise linear models".

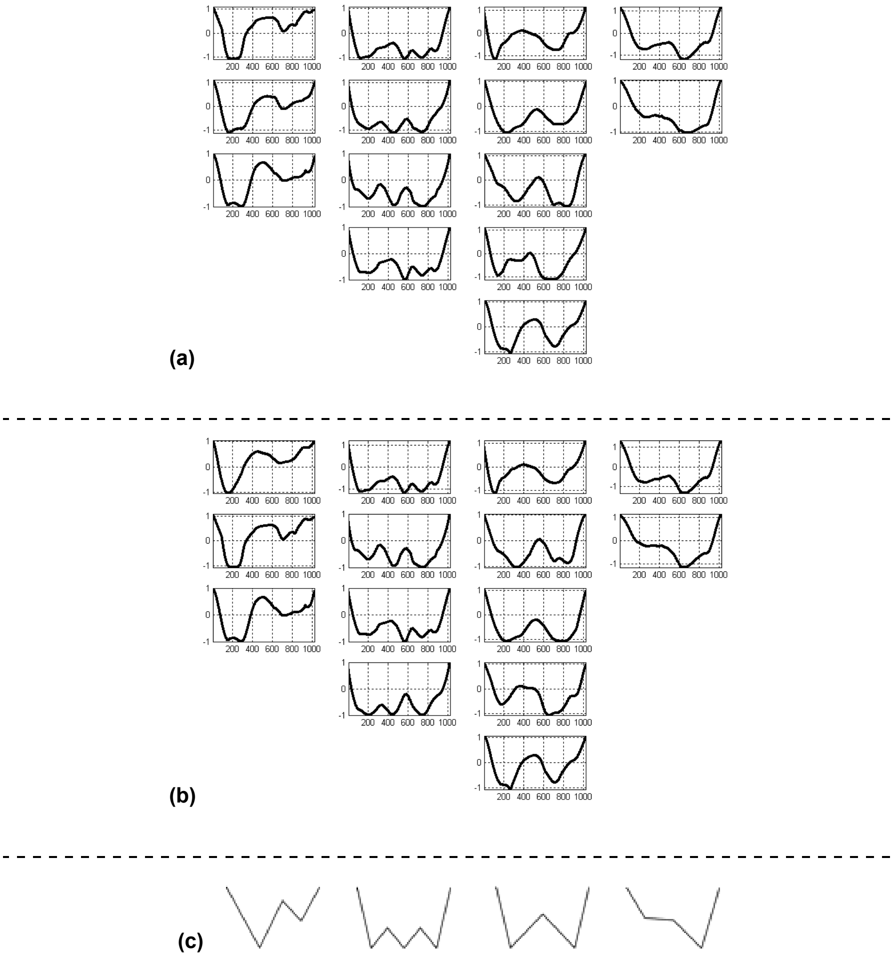


Figure 6 – a) and b) groups of morphological prototypes for class C3 obtained with the proposed ART – based learning scheme; and c) simplified geometric models for the prototypes allocated in columns.

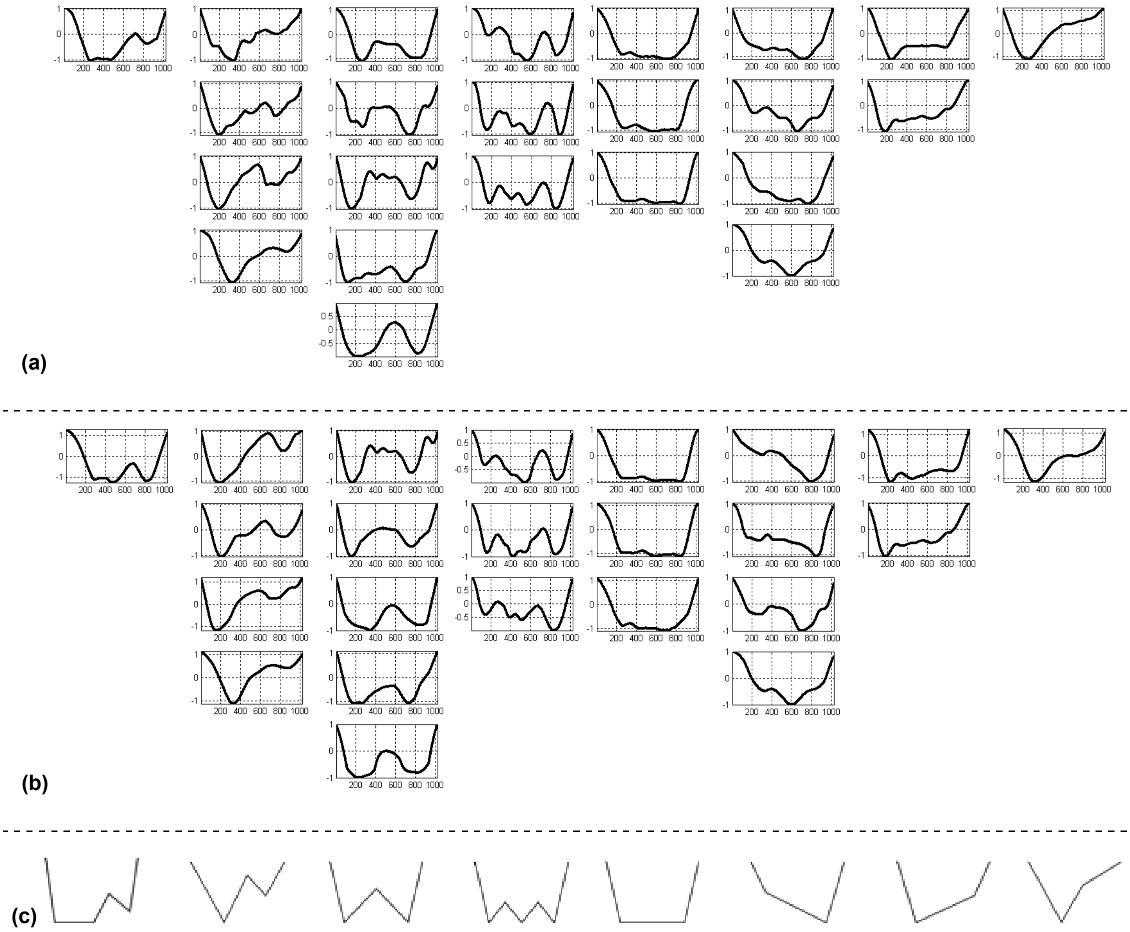


Figure 7 – a) and b) groups of morphological prototypes for class C2 obtained with the proposed ART – based learning scheme; and c) simplified geometric models for the prototypes allocated in columns.

The subclasses were formed by a clustering algorithm using the prototypes previously obtained by the Adaptive Resonance Theory learning scheme. This clustering process only involves competition in which the correlation coefficients between each inductive signature and the prototypes were calculated and the largest correlation value found defined the subclass to which the inductive signature should be assigned. However, as shown in columns of Figures 6 and 7, there are similar prototypes within the group of prototypes. These are the prototypes that have the same simplified geometric model (also shown in Figures 6 and 7). Thus, to make the result more practical, the subclasses formed around the prototypes with the same simplified geometric models were assembled in a single subclass. This subdivision is shown in Figures 8 and 9 for classes C2 and C3, respectively.

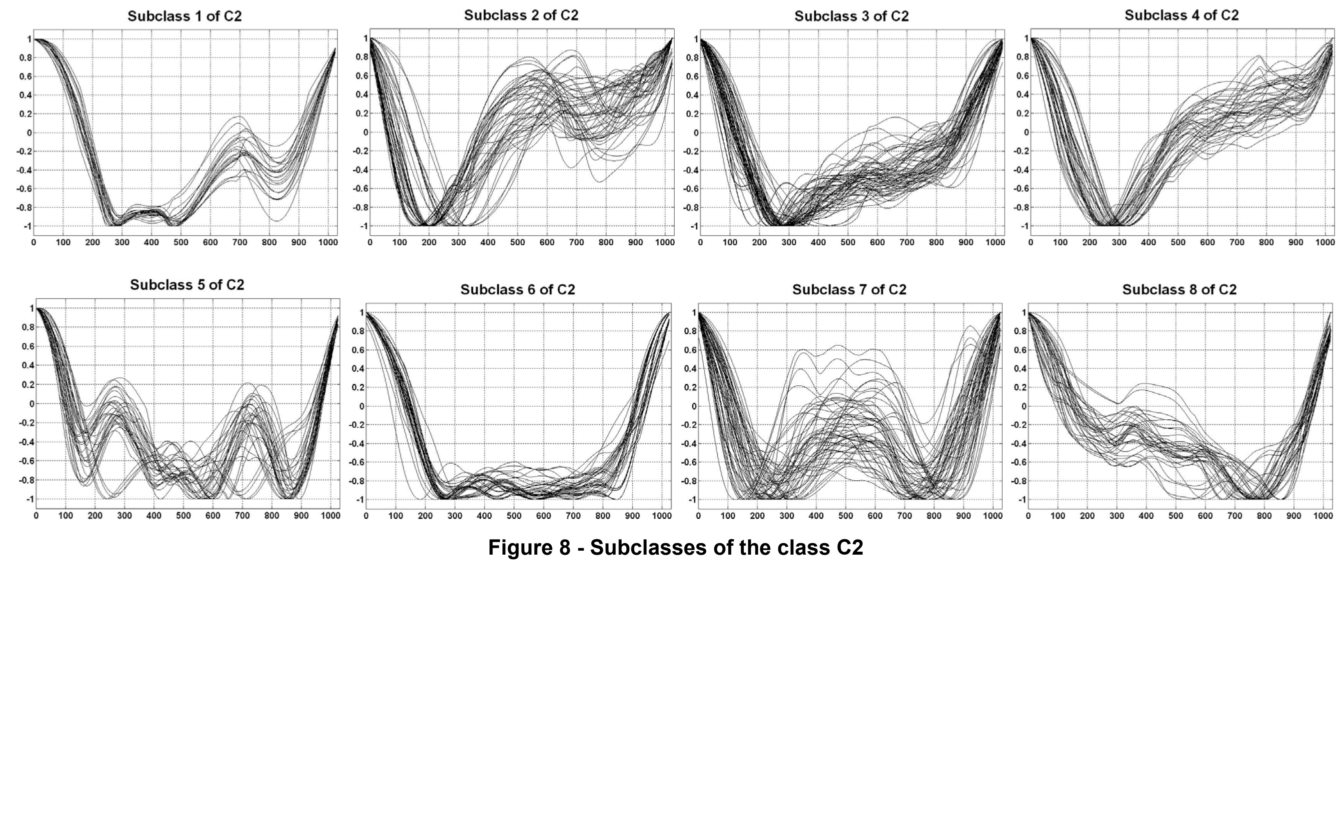


Figure 8 - Subclasses of the class C2

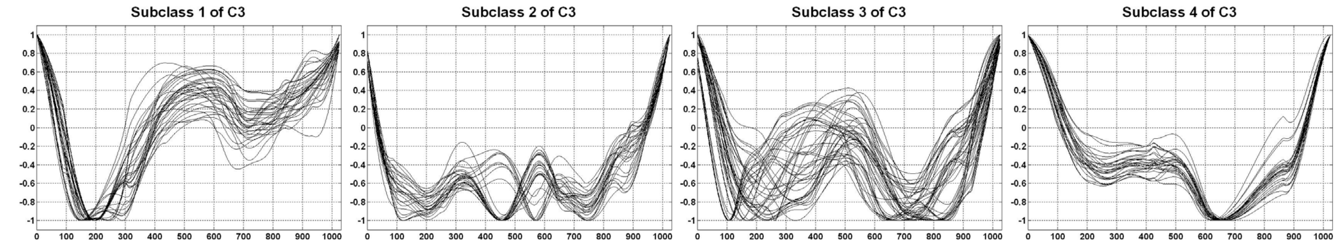


Figure 9 - Subclasses of the class C3

Although the subdivision shown in Figures 8 and 9 is satisfactory, because all the relevant morphological patterns of the two classes were well identified and separated, it is necessary to remind that the motivation to perform this subdivision in the input spaces of the inductive signatures, is the belief that it may be possible to make an eventual classification task easier. In some classification tasks, depending on the established goal, it is suitable to divide the classes into the same number of subclasses.

In these regards, a correspondence one-to-one among subclasses may be established and the classifiers may be trained with pairs of corresponding subclasses. This further subdivision may be implemented by transforming the subclasses (1, 2, 3, 4) and (5, 6) of class C2 in only two subclasses. After this gathering process the number of subclasses in C2 matches the number of subclasses in C3, which stayed unchanged. The result is shown in Figure 10 in which the pairs of corresponding subclasses are in the same column.

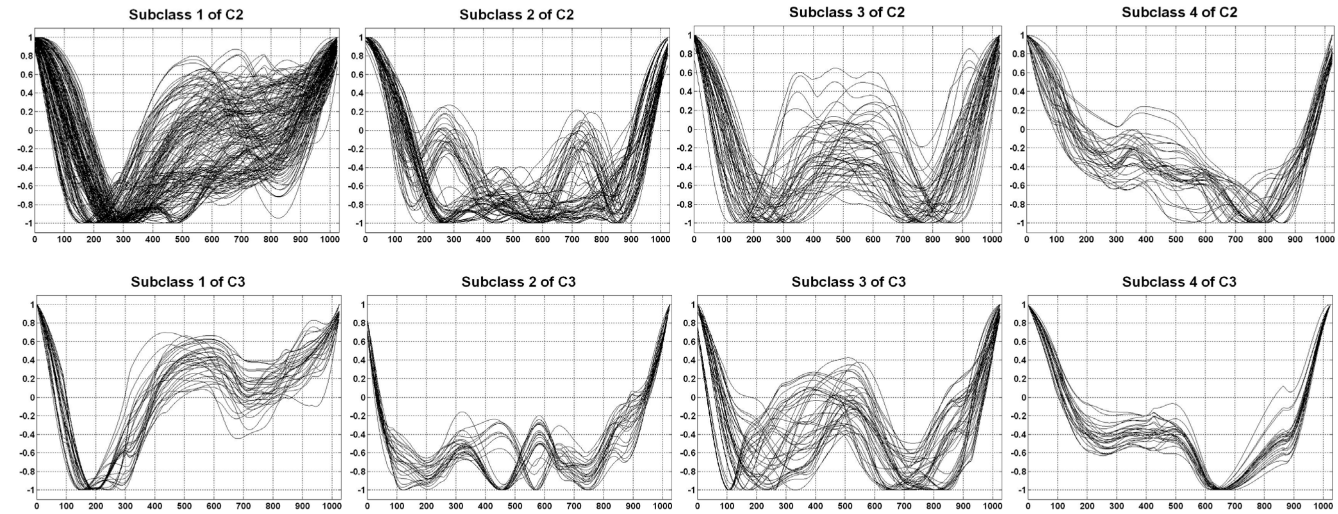


Figure 10 - Final subdivision of the classes C2 (top) and C3 (bottom)

4.2 Computational complexity

In order to obtain a reference to the computational complexity of our clustering algorithm, we identified morphological subclasses in the inductive signatures by using the traditional k-means algorithm. The number of clusters for class C2 was set as 8 and for Class C3 was set as 4 (according to the results obtained with our clustering algorithm). The running time measured for k-means was greater than the running time measure for the Adaptive Resonance Theory based learning scheme. Then, the computational complexity of our clustering algorithm is lower than computational complexity of the algorithm k-means that according to [20], is O(ktn), where n is the total number of patterns, k is the number of clusters, and t is the number of iterations

4.3 Benchmarks

With the purpose of benchmarking our subdivision algorithm, we used two databases: Iris database and a temperature database.

Database Iris was published by Fisher in 1936 [21] and it has been used broadly for analysis of clustering algorithms. This database contains measures taken on three subclasses of the flower Iris: Iris setosa, Iris versicolor, Iris virginica. For each subclass there are 50 vectors with measures of the following attributes: the sepal length, sepal width, petal length and petal width.

Data about temperature was taken from the National Climatic Data Center site [22] and comprises 44 time series of daily temperatures measured between April 2007 and March 2008 in the areas north and south of the state of Florida. There are 22 temperature records for the north (latitudes from 30°04'N to 30°57'N; longitudes from 81°24'W to 87°08'W) and 22 temperature records for the south (latitudes from 24°33'N to 25°57'N; longitudes from 80°09'W to 82°52'W), each one with 366 samples.

Just like in the inductive signatures database, in the two databases above, the amount of information is reduced and the degree of class overlap is considerable. Moreover, temperature records database is composed by not smooth patterns and the abrupt changes in these patterns are like noise added to a smooth pattern. Therefore, the temperature records test the robustness of the clustering method proposed in the presence of noises.

It is necessary take into consideration that there is a difference between the way the algorithm subdivision was applied to the databases: in the case of inductive signatures the morphological prototypes were found separately for each class and the division into subclasses was made also separately for each class; in the case of Iris database, the search for the morphological prototypes was performed among all the patterns in the database and, next , all the prototypes and all the patterns in the database were used in clustering stage; in the case of temperature database, the morphological prototypes were found separately for each class and, next , all the prototypes and all the patterns in the database were used in clustering stage .That is, for the two benchmarks, in the clustering stage a classification was performed in which the class models were the morphological prototypes.

4.3.1 Results for Iris database

The Adaptive Resonance Theory learning scheme was tested with the Iris database with the following execution parameters:

- Vigilance parameter range: from 0.9999 to 0,999, in step of -0.0001.
- learning rate set as 0.001.
- output units activations set as 0.9.

The correlation coefficient was adopted as the similarity metric, but no data pre-processing operation was necessary before its calculation.

The algorithm identified six prototypes in the data set, as shown in Table 1.

Prototypes	Attributes (in mm)			
	Sepal length	Sepal width	Petal length	Petal width
Prototype 1	82.5	35.6	58.5	19.1
Prototype 2	83.7	37.1	67.9	20.0
Prototype 3	81.1	39.9	61.0	18.6
Prototype 4	67.3	44.5	18.9	2.8
Prototype 5	60.7	27.1	42.3	11.9
Prototype 6	57.6	27.2	46.1	12.6

Table 1: Prototypes found by the Adaptive Resonance Theory learning scheme for Iris database

In Table 1, the prototypes 1, 2 and 3 as well as prototypes 4 and 5 are very similar. The average standard deviation (taken on all the attributes) for the prototypes (1, 2, 3) is ≈2.3 and for the prototypes (5, 6) is ≈1.4. These values indicate that in the Iris database or there is subclass overlap or some subclass need more than one prototype to be properly represented or both.

Figure 11 illustrates the obtained prototypes. Below each prototype in Figure 11, the numbers that correspond to the patterns grouped around the prototype are shown. The numbers from 1 to 50 correspond to the patterns of the Iris setosa subclass, the numbers from 51 to 100 correspond to the patterns of the Iris versicolor subclass and numbers from 101 to 150 correspond to the patterns of the Iris virginica subclass.

By considering that one cluster may be assigned to the subclass with the largest number of patterns, the six output clusters are labelled as fallows:

Cluster 1: Iris versicolor	Cluster 2: Iris virginica	Cluster 3: Iris versicolor
Cluster 4: Iris setosa	Cluster 5: Iris versicolor	Cluster 6: Iris versicolor

Six patterns (indicated by arrows in Figure 11) were incorrectly clustered. In the first cluster, patterns 104 and 114 (Iris virginica) were allocated together with the Iris versicolor patterns. In the second cluster, patterns 55 and 88 (Iris versicolor) were allocated together with Iris virginica Patterns. In the sixth cluster, patterns 118 and 133 (Iris virginica) were allocated together with Iris versicolor patterns. The percentages of correct clustering, calculated by subclasses, were thus:

Iris setosa: 100 % ; Iris versicolor : 92 % ; Iris virginica : 98%

Among the 6 patterns incorrectly clustered, 4 belong to clusters 1 and 2 whose prototypes, as mentioned above, are very similar. Therefore, Iris versicolor and Iris virginica subclass overlap. The other 2 patterns incorrectly clustered, confirm the overlap between these subclasses. It should also be noted that 4 clusters are assigned to the subclass versicolor what indicate a high within variance in this subclass. This characteristic increases the possibility that the subclass versicolor overlap other subclasses of the Iris database.

As overall result, a percentage of (144/150) * 100 = 96% of correct clustering was obtained. This percentage is consistent with the results obtained by other clustering approaches applied to the same data base [23], [24], [25], [26], [27].

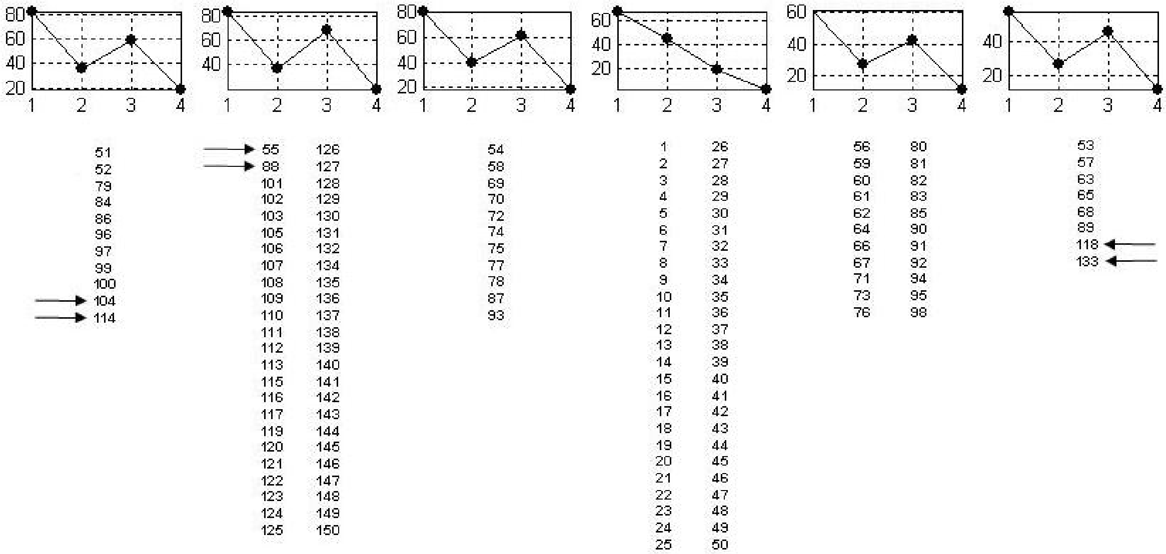


Figure 11 - Prototypes found for Iris database. Below each prototype, are shown the numbers of the patterns allocated in the cluster generated around the prototype; 1:50 - Iris setosa; 51:100 - Iris versicolor; 101:150 - Iris virginica.

4.3.2 Results for temperature database

Figure 12 shows the patterns of the two sets of daily temperatures for north Florida and south Florida,

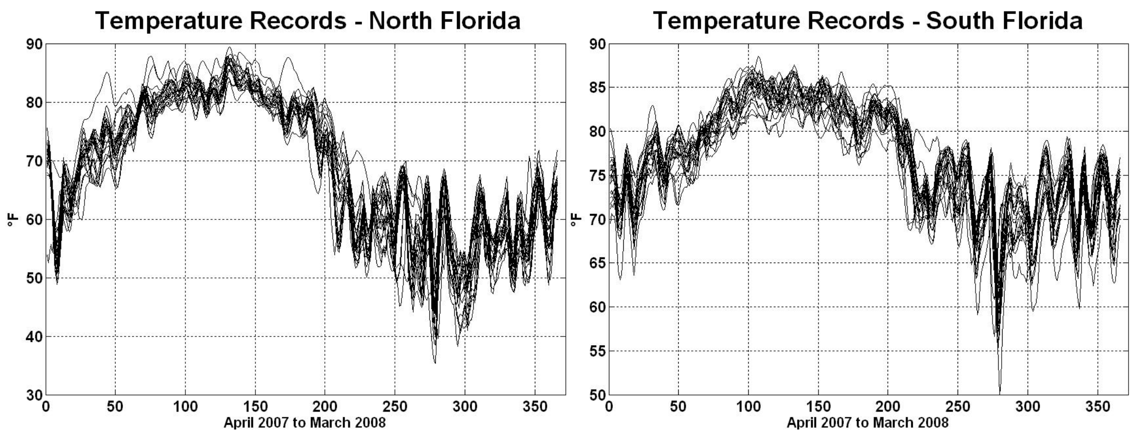


Figure 12 - Patterns of daily temperature recorded between 04/2007 and 03/2008 for two regions of the state of Florida

Temperature records in Figure 12 were first smoothed by moving average filter on segments of 5 samples.

The Adaptive Resonance Theory learning scheme was tested with this database with the following execution parameters:

- vigilance parameter range: from 0.980 to 0.850, in step of -0.001.
- learning rate set as 0.0001.
- output units activations set as 0.9.

The measure of similarity was the correlation coefficient

For this database it is interesting to evaluate the clustering results of our clustering method. We take for reference the results presented in [28]. In [28], a database, also composed of temperature records is used to check the quality of a clustering method where the similarity among the time series is given by the Euclidian distance of cepstral coefficients of the time series' ARIMA models.

Then, in each simulation, 11 temperature records were randomly drawn from each group and each subgroup with 11 records was used to identify the morphological prototypes of the respective group. Next, the group formed by the two remaining subgroups of 11 temperature records was divided into clusters based on the identified prototypes. So, the two remaining subgroups of 11 temperature records worked as a ground truth in our tests.

The algorithm was executed 100 times for each set of temperature records. The number of prototypes identified by the algorithm, ranged between 2 and 5 for both sets, but, in most of the results, 3 or 4 prototypes were identified for each set. In the clustering stage, the percentage of correct clustering was 100% in most simulations. In the worst cases, two temperature records were incorrectly clustered. A worst result example is presented below. The quality of the clustering results obtained with the ART2 learning scheme is evaluated by this example.

Figures 13 and 14 shows the two subsets with 11 temperature records used to identify the morphological prototypes and the respective identified prototypes.

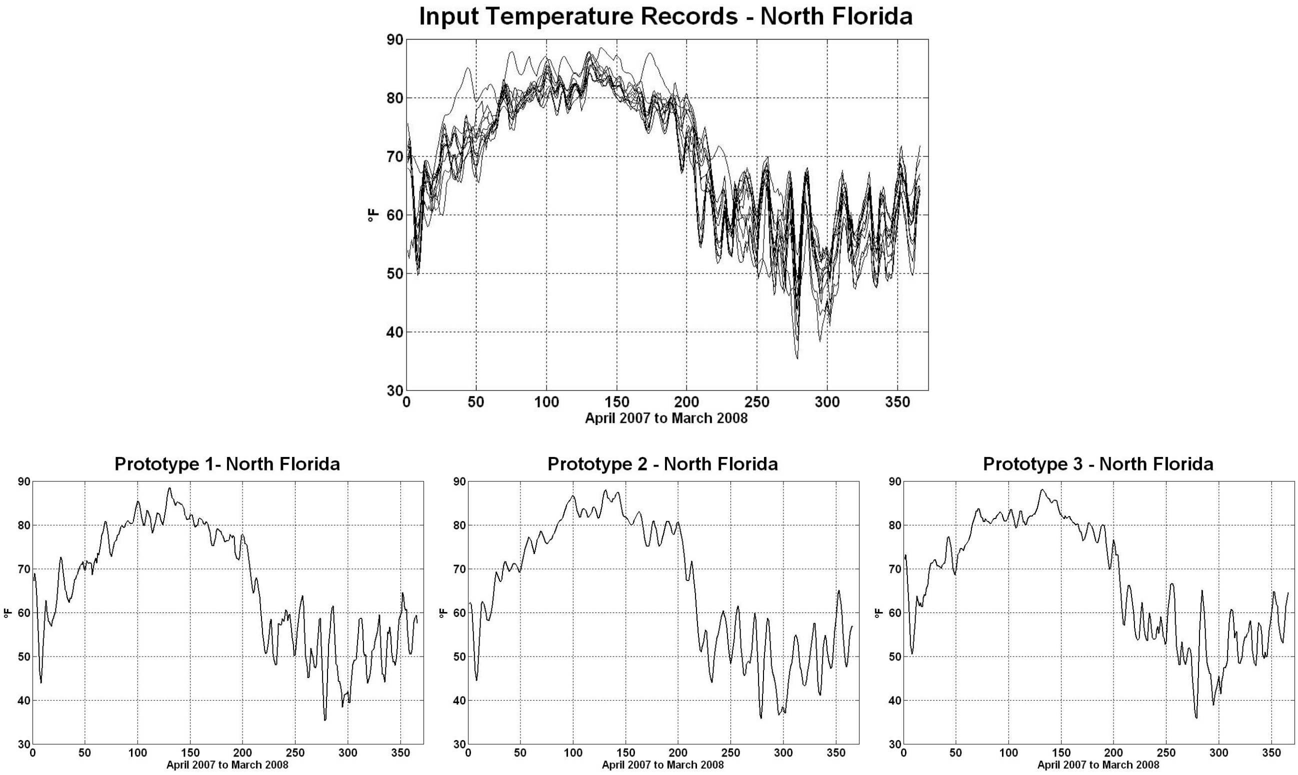
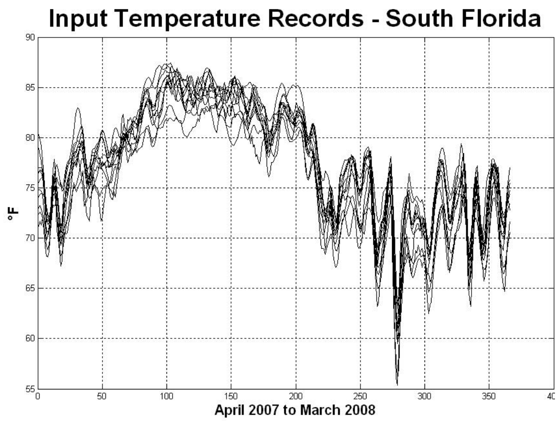


Figure 13 – North Florida input temperature records and prototypes identified by the Adaptive Resonance Theory learning scheme



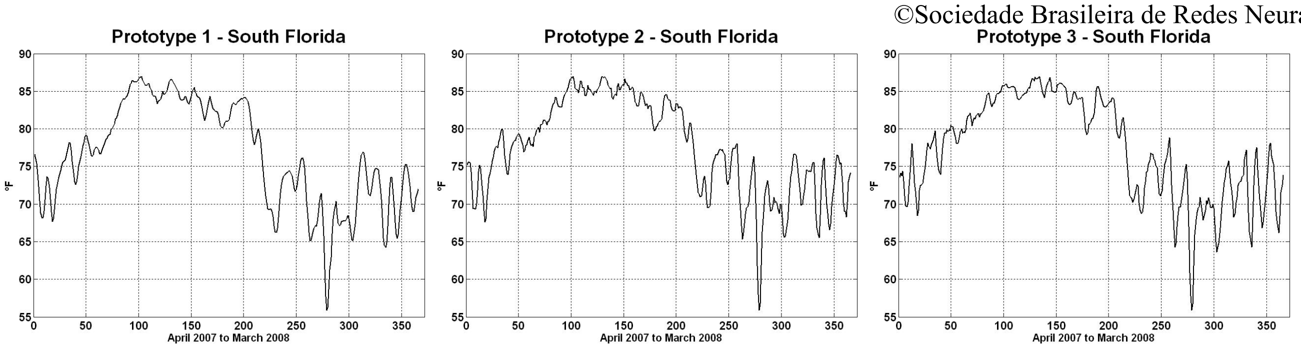


Figure 14 – South Florida input temperature records and prototypes identified by the Adaptive Resonance Theory learning scheme

Figure 15 shown the sets of temperature records used as ground truth clusters

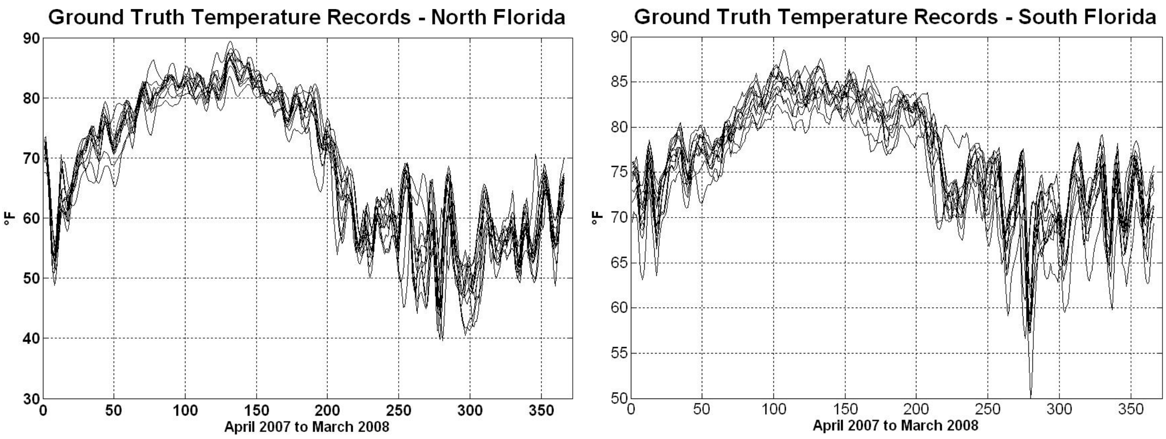


Figure 15 – Two ground truth clusters for clustering operation based on prototypes shown in the Figures 13 and 14; at left, ground truth cluster for north Florida clusters and at right, ground truth cluster for south Florida clusters

Figure 16 presents clustering results.

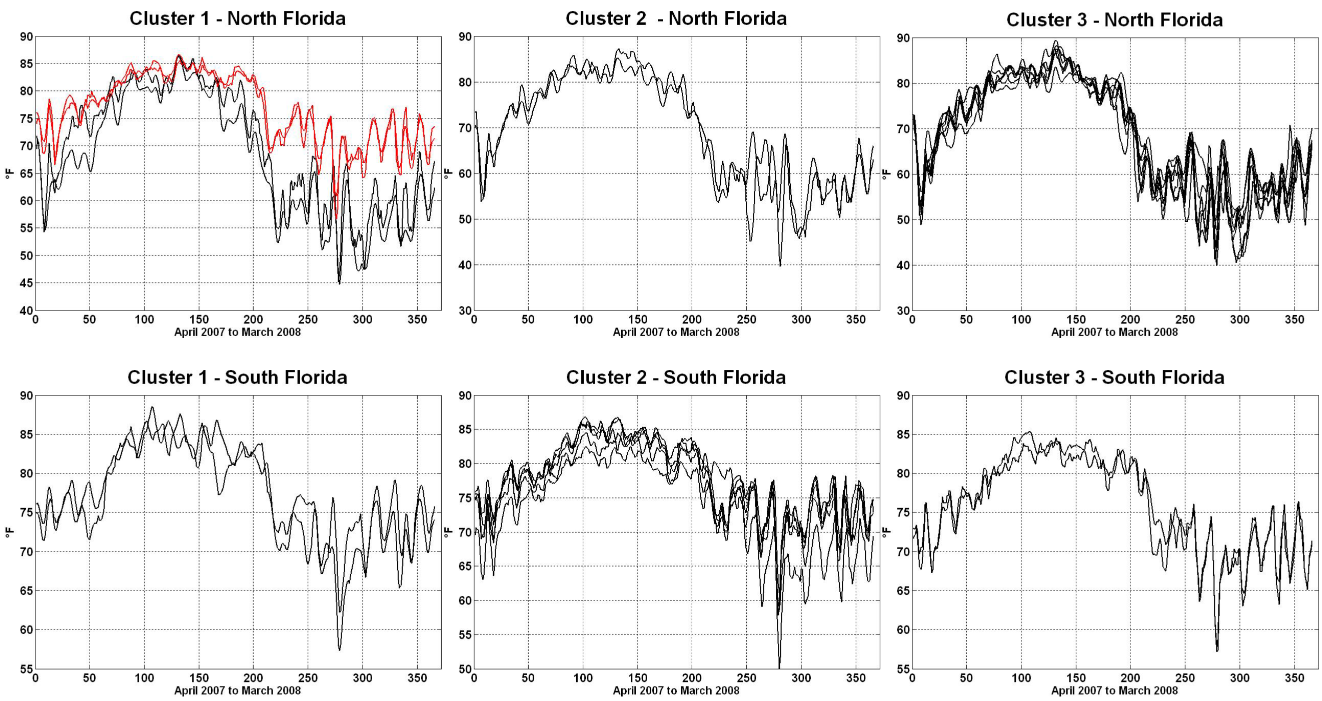


Figure 16 – Clusters of the temperature records in Figure 15 built around the morphological prototypes shown in the Figures 13 and 14

To evaluate the clustering results in Figure 16 we adopted the cluster similarity criterion:

$$\mathbf{Sim(G,C)} = \frac{1}{k}\sum_{i=1}^{k} \max_j \mathbf{Sim(G_i, C_j)} \qquad \mathbf{(5)}$$

where G e C are, respectively, ground truth clusters and under evaluation clusters, and

$$\mathbf{Sim}(\mathbf{G_i}, \mathbf{C_j}) = \frac{2\left|\mathbf{G_i} \cap \mathbf{C_j}\right|}{\left|\mathbf{G_i}\right| + \left|\mathbf{C_j}\right|} \qquad \textbf{(6)}$$

In (6) | . | symbolize the cardinality of the set.

In the calculation of Sim(G, C) we consider the clusters on the top of the Figure 16 based on prototypes for north Florida as a single cluster C_1 and the clusters in the bottom of the Figure 16 based on prototypes for south Florida as a single cluster C_2. The respective G_1 e G_2 are as presented in the Figure 15. Thus, in (5) i = j = k = 2.

In this example two temperature records of south Florida set (the patterns in red in the Figure 16) were incorrectly allocated in the cluster 1 based on morphological prototypes for north Florida set. Thus, for this example, the cardinalities in (6) are:

$|G_1| = |G_2| = 11$; $|C_1| = 13$; $|C_2| = 9$; $|G_1 \cap C_1| = 11$; $|G_1 \cap C_2| = 2$; $|G_2 \cap C_1| = 0$; $|G_2 \cap C_2| = 9$.

For the above cardinalities values, Sim(G,C) ≈ 0,908. This cluster similarity index is close to index obtained in [28] for a similar database. This result indicates that our clustering method also works well for noisy patterns.

4 Conclusions

This paper presents an algorithm to reduce a class of patterns into a small set of prototypes able to represent the morphological diversity of the class and, next, divide the class into subclasses that are formed according to a criterion of morphological similarity with the prototypes identified. In the search for the set of prototypes, the algorithm uses a learning scheme based on the Adaptive Resonance Theory with a decreasing strategy for the vigilance parameter. This process reduces the number of input patterns from one execution phase to the next.

The algorithm was executed several times and the sets of prototypes obtained are equivalent, from the point of view of morphological representativeness. The resulting subclasses based on the set of prototypes, are coherent with the aim of discriminating the predominant morphological groups within the classes. This division of the classes into subclasses was performed under the hypothesis that this pre-processing may contribute to make easier an eventual classification task, providing highest rates of correct classifications. This hypothesis should be confirmed in the next steps of this research by comparing the performance of the classification system in the following configurations: with just one classifier trained with all the inductive signatures; with multiple classifiers, each of them trained with a pair of correspondent subclasses of inductive signatures.

Two different databases were used to benchmarking the proposed clustering algorithm and the performances reached in both cases are consistent with the results of other approaches in literature. These results make the algorithm proposed an interesting alternative to clustering tasks, considering the simplicity of the foundations in which its conception is based.

5 References

[1] Duda, R. O.; Hart, P. E.; Stork, D. G. *Pattern Classification*. 2. ed. New York: John Wiley & Sons, Inc., 2000.

[2] Bishop, C. M. *Neural Networks for Pattern Recognition*. Oxford: Clarendon Press, 1995.

[3] Haykin S. Neural Networks – A Comprehensive Foundation, Prentice Hall, second edition, 1999.

[4] Sun, C.; Ritchie, S.G. Individual Vehicle Speed Estimation Using Single Loop Inductive Waveforms. Journal of Transportation Engineering, v. 125, n. 6, p.531-538, Nov-Dec 1999.

[5] Gajda, J.; Sroka, R.; Stencel, M.; Wajda, A.; Zeglen, T. A Vehicle Classification Based on Inductive Loop Detectors. Proceedings of the 18th IEEE Instrumentation and Measurement Technology Conference, Budapest, p. 460-464, May 21-23, 2001.

[6] Oh, S.; Ritchie, S. G.; Oh, C. Real Time Traffic Measurement from Single Loop Inductive Signatures. 81st Annual meeting of the Transportation Research Board, Washington D.C, January 13-17, 2002.

[7] Mimbela, L. E. Y.; Klein, L. A. A Summary of Vehicle Detection and Surveillance Technologies used in Intelligent Transportation Systems. Vehicle Detector Clearinghouse Project - Federal Highway Administration's (FHWA) Intelligent Transportation Systems Program Office. April 30 2003.

[8] Savitzky, A.; Golay, M. J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36, pp 1627-1639, July 1964.

[9] Lima, G.R.T. "Initial analysis of an overlap case between two attribute spaces by using artificial neural networks" - Technical-Scientific Report – 11/2007 (This document is under examination to be published by the ITA library)

[10] Kudo, M.; Yanagi, S.; Shimbo M. Construction of class regions by a randomized algorithm: A randomized subclass method, Pattern Recognition, 29, pp 581–588,1996.

[11] Avi-Itzhak, H. I.; Van Mieghem, J. A.; Rub, L. Multiple subclass pattern recognition: A maximin correlation approach, IEEE Transactions on Pattern Analysis and Machine Intelligence, 17, 4, pp 418–431.

[12] Zhu, Q.; Cai, Y. A subclass model for non-linear pattern classification, Pattern Recognition Letters, 19, pp. 19-29,1998.

[13] Mandal, D. P. Partitioning of feature space for pattern classification, Pattern Recognition, v. 30, n. 12, pp. 1971-1990, 1997

[14] Xiong, Y.; Yeung, D-Y. Time Series Clustering with ARMA Mixtures, Pattern Recognition, v. 37, pp. 1675-1689, 2004.

[15] Liao, T.W. Clustering of time series data—a survey, Pattern Recognition, v. 38, pp. 1857 – 1874, 2005.

[16] Pharn, D. T.; Afify, A. A. Clustering techniques and their applications in engineering, Proceedings of the Institution of Mechanical Engineers Part C-Journal of Mechanical Engineering Science, v. 221, issue 11 pp. 1445-1459, Nov 2007.

[17] Carpenter, G. A.; Grossberg, S. “ART 2: Self-organization of stable category recognition codes for analog input patterns” Applied Optics, v. 26, n. 26, pp. 4919-4930, 1987.

[18] Fausett, L. Fundamentals of Neural Networks:architectures, algorithms and applications, Prentice-Hall, Inc., pp.218-283, 1994.

[19] Kung, S.Y. Digital Neural Networks, Prentice-Hall, Englewood Cliffs, 1993.

[20] Khaled A.; Sanjay R.; Vineet S. An efficient k-means clustering algorithm, Proceedings of the First Workshop on High-Performance Data Mining (Second Edition), IEEE, Orlando, FL ,1998.

[21] Fisher, R.A. The Use of Multiple Measurements in Taxonomic Problems, Annals of Eugenics, 7, pp.179 -188, 1936.

[22] http://www.ncdc.noaa.gov/oa/ncdc.html

[23] Hruschka, E. R.; Ebecken, N. F. F. Using a Clustering Genetic Algorithm for Rule Extraction from Artificial Neural Networks, Proceedings of the First IEEE Symposium on Combinations of Evolutionary Computation and Neural Networks, USA, pp 199-206, 2000.

[24] Salvini, R. L.; Carvalho, L. A. V. Elastic Neural Net Algorithm for Cluster Analysis, Proceedings Sixth Brazilian Symposium on Neural Networks, Rio de Janeiro, pp 191-195 2000.

[25] Jiang, M. F.; Tseng, S. S.; Su, C. M. Two-phase clustering process for outliers detection, Pattern Recognition Letters, v. 22, pp. 691-700, 2001.

[26] Alves, C. A. M. Uma Ferramenta de Extração de Regras de Redes Neurais, Dissertação de Mestrado, COPPE/UFRJ, 2001.

[27] Das, S.; Abraham, A.; Konar, A. Automatic Clustering Using an Improved Differential Evolution Algorithm, IEEE Transactions On Systems, Man, And Cybernetics — Part A: Systems And Humans, v.. 38, n. 1, January 2008.

[28] Kalpakis, K.; Gada, D.; Puttagunta, V. Distance Measures For Effective Clustering of ARIMA time-series, Proceedings IEEE International Conference on Data Mining-ICDM, pp. 273-280, Nov. /Dec.-2001.