Supervised Classification with Associative SOM

Rafael del-Hoyo[1], David Buldain[2], Alvaro Marco

[1] Instituto Tecnológico de Aragón, C/ María Luna nº6,
50005 Zaragoza, Spain
rdelhoyo@ita.es
http://www.ita.es
[2] Departamento de Electrónica y Comunicaciones, University of Zaragossa,
C/ María Luna 1, 50005 Zaragoza, Spain
buldain@posta.unizar.es

Abstract. This paper presents an extension of the Self Organizing Map model called Associative SOM that is able to process different types of input data in separated data-paths. The ASOM model can easily deal with situations of incomplete data-patterns and incorporate class labels for supervisory purposes. The ASOM is successfully compared with Multilayer Perceptrons in the incremental classification of six erythemato–squamous diseases, where only partial data is available in successive steps.

1 Introduction

Supervised neural models as Multilayer Perceptrons (MLP) and Radial Basis Functions (RBF) have demonstrated in many applications their capability for resolving classification problems. However in a real situation, the necessary information for a classification task often is not completely obtained or is not acquired at the same time. The incomplete data and the great heterogeneity of information are usual situations in real data, a very common problem found in medical diagnosis. The correct diagnosis is a classification problem done in two steps of incremental data acquisition: a compilation of symptoms and, if it is necessary, several laboratory analyses. The first step is critical in many diseases, and physicians usually can decide the medication using only the clinical inspection. However, the second step sometimes is necessary to obtain a correct diagnosis. This second data acquisition is modified when new laboratory tests are discovered and old ones are abandoned. This is a good example of what we call an incremental classification problem.
This variability and heterogeneity in the data sources is not well handled by the most known neural models as MLP and RBF networks supervised with the mean squared erro (MSE). An automated classification system must present high flexibility in its structural implementation to deal with the continued apparition of new data sources or the deletion of obsolete ones. A neural classification system should be able to deal with the different data sources in an associative process, where each data source could be easily included or excluded during the classification process. The solution consists

in including separated data-paths for the different data sources with certain modulator mechanism to select the information incoming to the neural system.
A well-known neural model with associative behavior is the SOM network [3]. In this paper the multi-path data structure is implemented in the SOM model by extending the description of the SOMPACK available in [7]. We call it the Associative SOM (ASOM). In Sect. 2, we describe the problem selected to represent an incremental classification task. It is a real dermatology database of the differential diagnosis of six erythemato-squamous diseases. Sect. 3 contains an outlook of the ASOM and its learning schemes. Finally, the learning experiments with the ASOM and the MLP are presented in Sect. 4.

2 Differential diagnosis of erythemato-squamous disease

The differential diagnosis of erythemato-squamous diseases is a difficult problem in dermatology. It has been studied using several IA algorithms [2]. Usually a biopsy is necessary for the correct and definitive diagnosis. Patients are evaluated by the physician in two steps: first the clinical inspection of the degree of erythema, scaling and the compilation of historical data that configure 12 features. In the second step, skin samples have to be taken for the evaluation of 22 histopathological features determined by an analysis under a microscope. This is an expensive process, not always necessary to have a correct diagnosis, and the patient is not medicated until the results are obtained.
The database available in [6] contains 34 features: the ‘family history’ has the value 1 if any of these diseases has been observed in the family, and 0 otherwise. The ‘age’ represents the age of the patient. The rest of the features are separated in two groups: 10 clinical features and 22 histopathological features, all of them with numeric values in the range of 0 to 3. Here, 0 indicates that the feature was not present, 3 indicates the largest amount possible, and 1, 2 indicate the relative intermediate values.
We separated the features in three groups: 12 clinical features, 22 histopathological features and 6 features with the class labeling. The class labels codification contains 6 binary input components assigning one-class-to-one-component. The rest of the features were normalized to mean zero and variance scaled to value one. This preprocess ensures that all input components present a similar importance for the neural classifiers. Also, in the case of learning simulations with the MLP, this preprocess allows to manage the situations where some input data is not presented, by assigning null values to the missing input components (their mean values).
The results found by Güvenir et al. [2] were obtained using 10-fold cross-validation evaluation. They claim that the VF15 algorithm (Voting Features Intervals) achieves 96.2 % accuracy on the dermatology dataset and the 99.2% when VF15 weights are selected with a genetic algorithm.

3 Associative Self-Organized Map (ASOM)

The ASOM model is a SOM extension that allows the incremental classification and introduces a certain associative supervision during the learning process by means of the inputs, instead of using the class labels for supervising the outputs. The map receives different paths of information, that we call data-paths, each one of them processing features from similar sources. Another data-paths can introduce information about the correct classification of the pattern. With such an input structure, the neurons in the map generate prototypes associating all the incoming information from different data sources and are able of managing situations with missing data by a simple modulator method of the gains assigned to the input-paths. This idea was originally proposed in [1] using Radial Basis Units with a competitive algorithm.

3.1 The ASOM description

Let's denote the neuron label by 'i' and the weights or prototypes associated to it by '$\mathbf{w}_i$'. Consider the input data divided in a number of sub-patterns D, in such a form that the different data sub-patterns 'd' would be processed by the corresponding data-path 'd', denoted as super indexes. The symbol $N^{(d}$ denotes the number of components or data inputs of the data-path 'd'. The sub-weights $\mathbf{w}_i^{(d}$, represent the fraction of weights of the neuron 'i' assigned to the data-path 'd'. Each data-path processes similar information and calculates the Euclidean distance between sub-patterns and sub-weights like the SOM model does. So we calculate the distance in the data-path 'd' by:

$$dist(\mathbf{w}_i^{(d}, \mathbf{x}^{(d}) = \sum_{j=1}^{N^d} (w_{ij}^{(d} - x_j^{(d})^2 \qquad (\mathbf{1})$$

The path-excitation is calculated by multiplying the path-distance by a particular path-gain coefficient, $g^{(d}$ (eq. 2). The sum of the path-excitations gives the whole excitation of the neuron 'Exc_i' (eq. 3). Notice that if all the path-gains present value one, the whole excitation is the Euclidean distance between the complete data pattern and the weights of the neuron, as in the SOM model.

$$exc_i^{(d} = g^{(d} dist(\mathbf{w}_i^{(d}, \mathbf{x}^{(d}) \qquad (\mathbf{2})$$

$$Exc_i = \sum_{d=1}^{D} exc_i^{(d} \qquad (\mathbf{3})$$

If a certain path-gain is set to zero, the data introduced by the corresponding path provide a low contribution in the whole excitation of the neurons, and does not decide which neuron is the winner. However, if the path-gain has a high value, its data-path predominates in the selection of the winner neuron and, therefore leads the selection of the winner neuron in both the recall and the learning processes. During the recall phase, the class-gain must be annulled and the rest of path-gains can be modulated to process situations with incomplete data patterns, assigning null values to the gains in paths with missing data.

Another difference with the SOM model is the output response. The output of the SOM is the index of the winner neuron, but the ASOM outputs the winner's sub-weights in the class-path. If the winner prototype corresponds to a data region with a clear classification result, the output of the network should only present a value one in the corresponding class-component and the rest with value zero. When the neuron prototype is positioned in a data region between two classes, the output will present two high responses (near value 0.5) for the corresponding classes, as the result of the interpolation between two class labels. These special neurons must be recognized as neurons associated to map regions where there exist a high indetermination about the class. The output response with the sub-weights of the class-path also permits the calculation of error measures like the MSE, so the comparison with other neural models based in the MSE is straightforward.

The most common features of a disease can also be identified by recognizing the winner neuron among the neurons of the same class. This best neuron of the class corresponds to the best matching neuron when we exclusively evaluate the map with the class labels. The ASOM also presents all the interesting properties of the SOM model like the possibility of the visual inspection of the topological map representation.

However we must realize that the path-gains only influence in the choice of the winner neuron, but the training algorithm is the original SOM algorithm, that remains essentially like an unsupervised learning process. The ASOM is like an associative memory [4] providing association of data patterns and labels.

We used, for training the maps in batch mode and for the weight initialization, the functions available in the SOMPACK [7]. The map size chosen along all the experiments was 6x12 neurons, because this map size was suited to the two first PCA projections.

4. Experiments

The original dataset contains 366 patterns and 6 classes of diseases. We applied a 10-fold cross validation method for evaluating the performance of the maps, because the number of samples is quite low in certain classes. Ten data groups where generated randomly by separating the dataset in a training set with the 95% of the samples and the evaluation set with the 5%, avoiding duplication of the samples in the datasets.

For the evaluation of the neural classifiers, we formulate three possibilities: use exclusively clinical data, use only histopathological data, or use both of them. This scheme is suited for real diagnosis, where the physician obtains first the clinical information and later can access to the histopathological information. At the first step only the clinical information is processed, so if the classification result is clear, the extraction of the histopathological data can be avoided. However if the clinical data classification were unclear between two classes, the physician would decide if either is possible to find medication focused in both diseases or to obtain extra information from a histopathological analysis that would refine the response of the classifier. This is a clear example of that we call the incremental classification problem.

4.1 Experiments with ASOM

The maps must face with three types of data: clinical, histopathological and class labels, therefore their input structure presents three data paths. The maps were trained in two fixed number of cycles (600 and 1200 cycles), although better methods based in the evaluation of the MSE will be considered in future studies. We defined three learning schemes for the path-gains. The first scheme, called the unbalanced-gain training (UG), assigns value 1 for the three gains along the training process. The second scheme, called balanced-gain training (BG), tries to balance the excitations of the data-paths assigning values to the gains proportional to the inverse of the number of input components in the path. The values selected were: value 1 for the class-path, value 0.25 for the histopatological path and value 0.5 for the clinical path.
The third scheme, called balanced increasing training (BIG), presents path-gains varying during the training process in order to lead the associative learning of the map with the classification labels as the predominant influence. The previous results obtained in the UG and BG schemes were better for the second one, so we decided to balance the evolving gains. A possible training strategy gives initially a high value to the class-gain (value 1) and low values to the path-gains of the rest of the data-paths (for example value 0.1). As training progresses, the value of the class-gain is maintained constant, but the rest of the path-gains are increased exponentially to promote an increasing importance of the data pattern in the development of the map. With this scheme, in the beginning of the learning process, the map neurons are mainly clustered by the class labels, and the topological ordering of the map is established by the data-pattern that have a less influence in the choice of the winners. As the training progresses, the gains of the data are higher and the pattern gets more influence in the result of the competition process.

The BIG learning scheme needs a step-wise increase in the gains. If gains are continuously increased, the subspaces of data are expanded quicker than the development of the map, and the resulting map remains contracted after many cycles with the training sets. This behavior reveals that the map needs several cycles with constant gains to get expanded over the data subspaces. The solution for increasing the gains, is to maintain them stable during a number of cycles that we call step. Several values of steps were tested taking the values: 3, 6, 10, 60 and 120. The best MSE was obtained with the map trained 100 steps of 6 cycles (600 cycles). The main results are resumed in table 2. The ASOM in the UG and BG schemes presented the same errors whether trained 600 or 1200 cycles, while the BIG scheme showed different errors depending of the training cycles. The results of the ASOM for two BIG schemes with step value 6 during 600 cycles (case 100x6) and 1200 (case 200x6) cycles are also presented in table 2.

4.2 Experiments with Multilayer Perceptrons (MLP)

The MLP model was simulated with the same data groups in a 10-fold cross validation, separating in each data group a 5% of the training data for early stopping evaluation. The chosen learning algorithm was the Levenberg-Marquardt algorithm

implemented in the Neural-Toolbox of Matlab. Two different types of architectures where simulated: the first network architecture with one hidden layer and the output layer with six units associated to the six class labels (represented by 34-X-6). The second network architecture consisted in six separated networks including one hidden layer and only one output-class unit (represented by 34-X-1). Each one of the six networks was assigned to recognize only one disease-class and reject the rest of the classes. Both architectures received the 34 input components. In the evaluations with partial data, the unavailable components were assigned value zero.

As the numbers of samples in the classes were so uneven, in the first architecture the training sets were augmented by replicating the samples of classes with less representation to equilibrate the number of samples in all classes. In the second architecture, the number of samples of the class recognized by the network was also augmented to equilibrate the sum of samples in the rest of the classes.

The number of hidden units in the MLP networks were estimated with an previous learning tests with the whole dataset during a fixed number of cycles (50 was enough in all cases). The first network was estimated using only even numbers of units from 6 to 30 (13 numbers), and the second networks were estimated using even numbers of units from 4 to 24 (11 numbers). Each network was simulated five times (in total 120 simulations) with a target error of value 10^{-12} (with this target all networks were trained till the cycle 50). The MSE evaluated in the 50^{th} cycle were averaged in the five exemplars of the networks. The numbers of hidden units of the networks with the lowest MSE were selected for the training in the cross validation method (see table 1).

Table 1. Number of hidden units selected in the MLP networks

MLP Networks	Number of hidden units
MLP with 6 outputs	24
MLP-Psoriasis	22
MLP-seboreic dermatitis	18
MLP-lichen planus	12
MLP-pityriasis rosea	10
MLP-cronic dermatitis	12
MLP-pityriasis rubra pilaris	12

The selected architectures were trained during 50 cycles with the training data and each cycle the resulting network was stored. The exemplar with the minimum MSE in the 5% of the training data (separated for this early stopping method) was selected for the evaluation with the test dataset. The same combinations of incomplete data that were evaluated with the maps were performed during the evaluation of the MLP classifiers, and the resulting classification errors are presented in table 2.

The graphs in the figure 1 represent a detailed comparison between the performance of the networks in the best MLP classifier (figure 1b) and the ASOM networks trained with the BIG scheme (figure 1a).

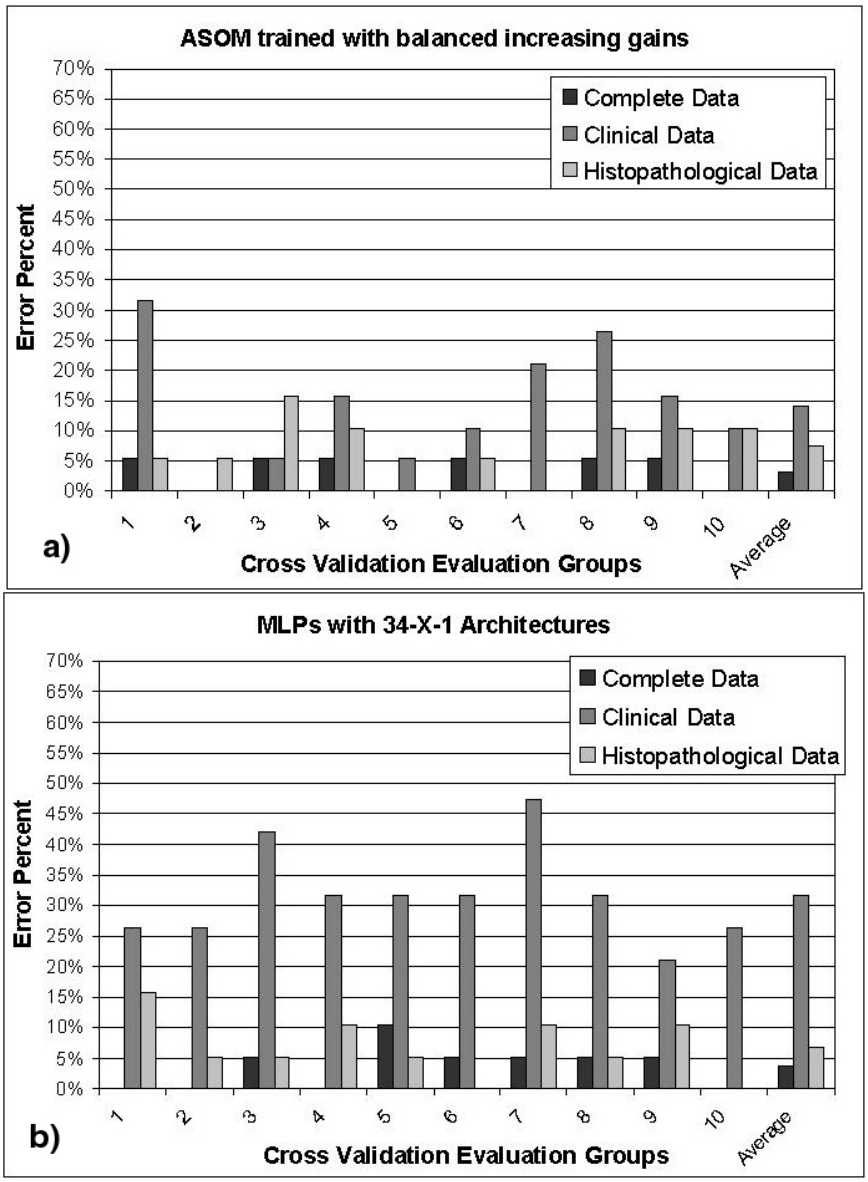

Fig. 1. Both graphs resume the classification percent errors measured in the 10 validation groups and the resulting average. The bars represent the resulting errors with only the clinical data, only the histopathological data and the complete data patterns. Graph a) presents the results obtained in the ASOM network trained with balanced increasing gains with step value 6 (ASOM-BIG-100x6 in table 2). Graph b) presents the results in the six MLP networks with one class output (MLP-34-X-1 architecture in table 2)

Table 2. Average classification errors from 10-fold cross validation in the ASOM and MLP networks, for the three types of evaluation with the two data groups.

Neural Network Model	Only Clinical Data	Only Histopathological Data	Complete Data
ASOM – UG(600)	13,68%	6,31%	4,21%
ASOM – BG(600)	12,10%	7,36%	3,68%
ASOM-BIG(200x6)	15,78%	6,31%	3,15%
ASOM-BIG(100x6)	14,21%	5,78%	3,15%
MLP-34-X-6	45%	8%	6%
6 MLP-34-X-1	32%	7%	4%

5. Conclusions

The paper presents a SOM extension that allows introducing different information paths to the maps. These data-paths can be processed separately by modulating their corresponding gains. This extension turns the non-supervised SOM model into a supervised one by means of the associative supervision with the class-labels used as inputs in a certain class-path. It was tested with the differential diagnosis of six erythemato–squamous diseases. This classification problem is suited for evaluations of incomplete data information as it deals with the combination of two kinds of information attributes: the clinical and the histopathological information. The MLP classifiers and the Güvenir results with the VF15 algorithm have been compared with those of ASOM and all of them obtained for the complete data classification a similar classification error around 4%. However if we compare those situations where incomplete data is presented to the classifier, only clinical or histopathological data, the ASOM performs quite better than the MLP networks.

References

1. J.D.Buldain, A.Roy: Association with Multi-Dendritic Radial Basis Units. IWANN 99, Lecture Notes in Computer Science 1606, Foundations and Tools for Neural Modeling (1999) 573-581
2. H.A.Güvenir, G. Demiröz and N. Ilter,: Learning Differential Diagnosis of Erythemato-Squamous Diseases using Voting Feature Intervals. Artificial Intelligence in Medicine, Vol. 13, No. 3 (1998) 147-165
3. T.Kohonen. The self-organizing map. In Proc. IEEE, volumen 78, pages 1464-1480,1990.
4. T.Kohonen. Self-Organization and Associative Memory, Springer Series In Information Sciences 8. Springer, Heidelberg, 1984.
5. H.Ritter. Parametrized selft-organizing maps. In S. Gielen and B. Kappen, editors, ICANN93-Proceddings, Amsterdam, pages 568-575. Springer Verlag, Berlin, 1993.
6. ftp://ftp.ics.uci.edu/pub/machine-learning-databases/dermatology/
7. http://www.cis.hut.fi/projects/somtoolbox/