Automatic Prognostic Determination and Evolution of Cognitive Decline using Artificial Neural Networks

Patricio García Báez[1], Carmen Paz Suárez Araujo[2], Carlos Fernández Viadero[3], and José Regidor García[2]

[1] Departament of Statistics, Operating Reseach and Computation, University of La Laguna, 38071 La Laguna, Canary Islands, Spain
pgarcia@ull.es
[2] Institute of Cybernetic Sciences and Technology, University of Las Palmas de Gran Canaria, 35017 Las Palmas de Gran Canaria, Canary Islands, Spain
cpsuarez@dis.ulpgc.es
[3] Unidad de Atención a la Dependencia de Santander, Gobierno de Cantabria, 39012 Santander, Cantabria, Spain

Abstract. This work tries to go a step further in the development of methods based on automatic learning techniques to parse and interpret data relating to cognitive decline (CD). There have been studied the neuropsychological tests of 267 consultations made over 30 patients by the Alzheimer's Patient Association of Gran Canaria in 2005. The Sanger neural network adaptation for missing values treatment has allowed making a Principal Components Analysis (PCA) on the successfully obtained data. The results show that the first three obtained principal components are able to extract information relating to functional, cognitive and instrumental sintomatology, respectively, from the test. By means of these techniques, it is possible to develop tools that allow physicians to quantify, view and make a better pursuit of the sintomatology associated to the cognitive decline processes, contributing to a better knowledge of these ones.

1 Introduction

The progressive aging of the world population has supposed an increase in the incidence of diseases associated to the age. Among them, the neurodegenerative processes and dementia in particular have acquired a special relevance. Its high prevalence, 8-10% in older than 65 years and more than 25% in the very old ones [1], originates very important repercussions, not only in the scope of the patient, affecting to its autonomy and quality of life, but also in the familiar, social and, of course, sanitary spheres [2][3].

These consequences urge us for the accomplishment of possible precocious diagnosis, and the most accurate as possible, and trying to resist to the maximum the devastating effects of this pathology by means of the accomplishment of a therapeutic plan and a suitable pursuit [4]. For this reason, the early detection of the cognitive decline is an advisable practice in all the scopes of attention. Nevertheless, the high uncertainty diagnosis [5], the degree of infradiagnosis, that can be so important that reaches 95% in some scopes, [6], advise to develop detection programs, new instruments of diagnose and guiding for the sanitary professionals. These actions have demonstrated their effectiveness as much in the attitude of the doctor on these pathologies like in evident improvements in the process of detection, diagnosis and treatment of the old ones affected by the dementia [7].

The evaluation of cognitive decline can be made by opened, semi structured form, or by means of the application of a series of cognitive evaluation scales. We must remember, nevertheless, that a test or a cognitive-based scale does not constitute a deterioration or normality diagnosis, but it only indicates the possible existence of a cognitive deficit. Throughout the years, there have been developed and have been used different neuropsychological tests [8] that evaluate different cognitive sections of a patient, but does not turn out simple to relate these tests to concrete sintomatologies or different levels of cognitive decline. It is necessary that they have been widely validated and contrasted and, of course, to know which is its level of sensitivity, specificity and predictive value. Other non-despicable problems in the use of these scales are the absence of universal cut points, the trans-cultural problematic and the precision that seems similar as much with the use of short and/or long scales [7][9].

It is an usual practice to put patients under different tests. Their scores are strongly collinear, because the different tests are composed of common cognitive components. In order to approach this problem and to advance towards the diagnosis we present a proposal based on neural computation. Thus, using an artificial neural network to carry out a PCA, we can approach the inherent multicolineality between the different tests. This way, we obtain more accurate information about the patient state, which allows us to make prognoses and studies of the evolution of the cognitive decline.

2 Data Environment

The dataset is made of the results of the 267 clinical consultations made over 30 patients during 2005 at Alzheimer's Patient Association of Gran Canaria. Its structure includes a patient identifier, resulting of 5 neuropsychological tests and diagnose of cognitive decline (DCD) as well as diagnose of differential dementia. These data has as advantage that they are very homogeneous; each patient has scores of his/her monthly-made tests, excepting the Mini Mental test which is made twice a year. But, although the majority of the patients have been tested 12 times, there are some patients with fewer consultations made and some of data missing in the consultations. Thus, we work with a dataset where the missing data feature is present, neither in all the consultations nor all the patients are applied the complete test battery.

The 5 different used data tests are the following ones:

1. Mini Mental Status Examination (MMSE) [10] is the most spread, employed and quoted standardized instrument to value the cognoscitive function. It consists of a set of short and simple questions that allow a quick evaluation of several cognitive areas: orientation, fixation, calculation and attention, memory, language, reading, writing and visoconstructive abilities. Its score ranges from 0 to 30 points. It also constitutes the most used pruning test on the international epidemiological investigation, as well as in the clinical tests that require evaluating the patient's intellective functions.
2. FAST scale (Functional Assessment Staging) [11] is used to evaluate the possible relation between functional stage and survival. It consists of 7 stages (1 to 7) being the very last two ones, severe dementias and advanced ones, subdivided (6A to 6E and 7A to 7F). Derivated of our experience, it seems that the evaluation by means of a specific scale as FAST, in which the function is included, clearly orientates on the different stages towards a less accurate diagnose while the functional affectation rises [12]. A nondespicable disadvantage is derived from the generalization of the FAST scale, initially designed for patients of Alzheimer's, and therefore it does not have a sufficiently contrasted use in other dementias.
3. Katz's index [13] evaluates the pure function in the basic activities of the daily life: bath, dress, cleanliness, mobilization, continence and food. It is defined as an observation instrument and an objective guide of the course of the chronic disease, as an aid to study the process of aging and as an aid in the rehabilitation. Its score ranges from A to H.
4. Barthel's index [14] is similar to the one of Katz's with the difference of a numerical result, which is adapted for a continuous gradual evaluation, since the Katz's index is a ordinal scale with items of dichotomizing character. Like Katz's index, it evaluates the same functions this one does, although in a wider way. We use a summary score that consists of 5 stages: Independent, Slight, Moderate, Severe and Totally dependent.
5. Lawton-Brody's index [15] evaluates the behavior aspects of instrumental character, for which it is necessary to be able to make the basic functions of a suitable form. It implies therefore a much greater and more complex functional integrity. The use of the telephone, the capacity to make purchases, being able to cook, the house care, washing the clothes, the use of transport means, being able to handle the proper medication and the handling of the financial aspects are evaluated. Its score ranges from 0 to 8 points for women, and 0 through 5 for men.

The missing data corresponding to the MMSE test have been completed, in agreement with the clinical experts, by means of interpolation from the annual results of both annual tests that almost all the patients have been put under. Even so, other values have left empty, 71 of the total of 1335, whose distribution by test and number of patients is indicated in Table 1.

In order to facilitate the convergence, as previous step to their use, the different fields that constitute the successfully obtained information were preprocessed. The neuropsycological tests were standardized between 0 and 10 from

Table 1. Statistics of missing data in data set.

$Testtype$	$Number of missings data$	$Number of patients$
$MMSE$	36	4
$FAST$	0	-
$Katz$	33	12
$Barthel$	1	1
$Lawton - Brody$	1	1
$Total$	71	16

the minimum and maximum values that can be reached in these tests. Those fields not being filled up labeled as lost values or missing ones, had a later special treatment.

At a control stage, the values of diagnosis are used. The diagnosis of the CD can contain four different values: Without CD, Slight CD, Moderate Cd and Severe CD. From the entire set of consultations, 15% were diagnosed as Slight CD, 34% as Moderate CD and 51% as Severe CD, due to the data source does not include patients without cognitive decline. In the diagnosis of dementias, 73% of consultations correspond to Alzheimer-type dementias, being the other distributed 27% in other types of dementias.

3 Neural Approximation

The dimensionality and the data type to be treated make it difficult to observe the existing differences between different patients, or the differences that can be between diverse consultations of a patient. This is, in part, due to the co-linearity of the tests: some of the tests in fact measure diverse aspects or several tests measure similar components. In order to be able to approach this difficulty, we transformed the representative data of the different consultations into a more reduced dimensionality data space so that a minimum loss of information exist, and where each dimension represents features that have greater independence from the other ones and which can be assigned a concrete meaning to.

Projective methods for feature extraction generate a transformation of the input space into a new feature space with a lower dimensionality. Usually, this transformation is created on the basis of linear combinations that try to maximize some interesting measurement, like preserving the max possible amount of information, reducing remarkably the data dimensionality. Its utility is focused into the data compression or optimal codification, and the display of high dimensionality data. Some of the processes that make this reduction are the related to PCA. Among the neural approaches whose processes converge into PCA we have the Sanger network, also called Generalized Hebbian Algorithm (GHA) [16], the Adaptive Principal-component Extractor (APEX) [17], the Foldiak's model [18] and the Rubner's model [19]. A detailed comparative of PCA neural networks can be found in [20]. These processes allow generating a hierarchical structure

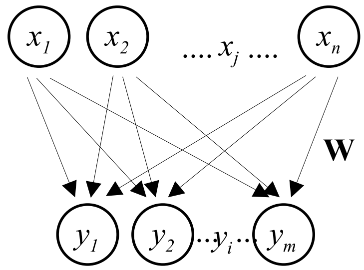

Fig. 1. Topology of Sanger neural network.

of each of the new coordinates so that these are ordered based on the amount of information that represent on the total of data. Analyzing the existing relationship between the original coordinates space and the new coordinates space we can give meaning to each of the main components, this way is possible to represent each consultation by means of a reduced number of features that qualify the state of the patient at the moment of accomplishment of the consultation. The comparison between these features for different types of patients or consultations at different dates for a same patient can bring to light important information related to the prognosis and temporary evolution of the cognitive decline from the examined patients.

Using the Sanger network, there have been extracted and analyzed the first three main components of our data. As the shown in Figure 1, Sanger network is totally interconnected with the inputs; it is a linear processing network where the i-th unit has the y_i output that is given by the following expression:

$$y_i = \sum_{j=1}^{n} w_{ij} x_j \tag{1}$$

where n is the dimensionality of the input vector, x_j is the value of j-th input and w_{ij} is the weight between the j-th input and the i-th unit. The learning rule that progressively fits this matrix of weights is given by the following expression:

$$\Delta w_{ij} = \eta(t) y_i \left(x_j - \sum_{k=1}^{i} y_k w_{kj} \right) . \tag{2}$$

In the expression given by Equation (2), the learning ratio $\eta(t)$ is a linearly decreasing function in time t. This formula causes the Sanger weights to converge to the main training data set components.

An important aspect for our developments is that the compression process carried out by this network is reversible, the original input vector ($\mathbf{x}$) can be reconstructed ($\mathbf{Rx}$) with a minimum loss of information from the m-outputs vector ($\mathbf{y}$) and the Sanger weights ($\mathbf{W}$) (decompression procedure). The equation that carries out the decompression is the following one:

$$Rx_j = \sum_{i=1}^{m} w_{ij} y_i. \quad (3)$$

3.1 Sanger Network Extension for Missing Data Treatment

The previous expressions of the Sanger network must be adapted to be able to face the processing of missing data, to accomplish it we have followed a scheme similar to the one found in [21], causing these missing values not to contribute the output nor modify the weights of the network. This way, Equation (1) will be as it follows:

$$y_i = \sum_{j \in P_t}^{n} w_{ij} x_j \quad (4)$$

where P_t is the set of input units, j, whose values x_j are available at time t. Also, Equation (2) will be modified the following way:

$$\Delta w_{ij} = \begin{cases} \eta(t) y_i \left(x_j - \sum_{k=1}^{i} y_k w_{kj} \right) & \text{if } j \in P_t; \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

This way, once the learning has ended, and having used the decompression procedure, it is possible to make an estimation of the value of the missing data.

4 Results and Discussion

4.1 Patients Prognostic

Once the Sanger network has been trained with our data set, we obtain the values for each one of the first three principal components in relation to the five tests, in Figure 2 are shown the obtained values for each one of the first three main components in relation to the five tests. We see that the first component depends on a great extent on the values of the Katz's and Barthel's tests, which takes us, according to the meaning of these tests, to the conclusion that this component represents the functional sintomatology. The second component is inversely proportional to the value of MMSE test, and directly proportional, but to a lesser extent, with the FAST test. It indicates to us that this second component is strongly related to the cognitive sintomatology. Finally, the third component depends mainly on the value of the Lawton-Brody test, and to a lesser extent of the value of Barthel test and inversely to the value of the MMSE. The representation of this component is then bound to the instrumental sintomatology that features the Lawton-Brody test. It is possible to observe therefore that the extraction of main components can isolate the main characteristics that are mixed in the different tests.

In Figure 3a, it can be seen data of all the consultations facing the first against the second main components. In addition, each consultation has been

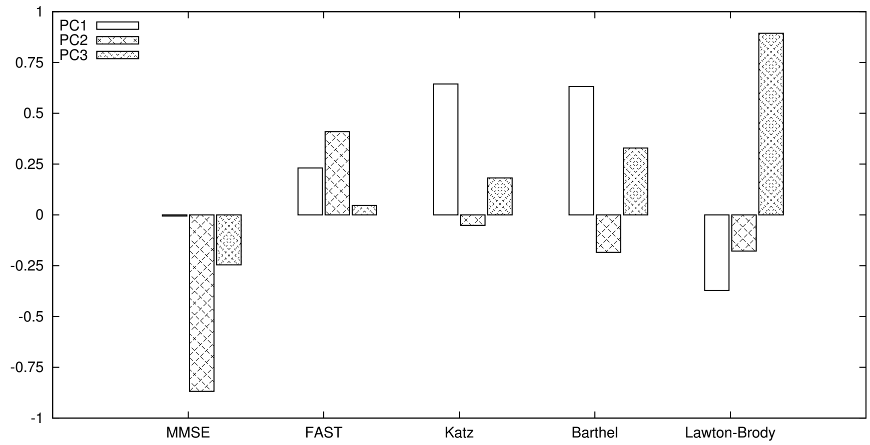

Fig. 2. Values of first three principal components.

labeled with the patient's DCD at that moment. The CD is clearly represented within the second component, grouping the most of the slight CD below value -3, moderate CD between -3 and 1 and severe CD over 0. On the other hand, the functional skills of the patients are represented by the first component, having those patients with greater functional problems higher values. There exists a relationship between this component and DCD, since habitually patients with greater deterioration have more diminished their functional skills. Thus, we see in the graph that most of patients with slight CD or moderate CD have negative values of this component, having positive values of the most of patients with severe CD.

Figure 3b represents the second against the third components. The analysis carried out by the clinical experts on these data verifies to us the relationship of the third main component with the instrumental skills of patients. In the case of this component, a direct relationship with DCD is not appraised, since the different groups distribute themselves the same in their rank.

4.2 Temporary Evolution of Patients

Once our data has been processed by means of the Sanger network, we obtain for each of the consultations a point of the new coordinates space given by the values of each of the principal components that conform it. If we unit the points of a same patient from the first to the last consultation we will have a path in this space, this path will indicate us the temporary evolution that the patient has had, that is to say, the trajectory that follows between each point indicates for each coordinate if there have been improvements or worsenings in the sintomatology associated to this coordinate.

In Figure 4, we can observe the traces corresponding to the consultations associated to three different patients. Two of them change of state of moder-

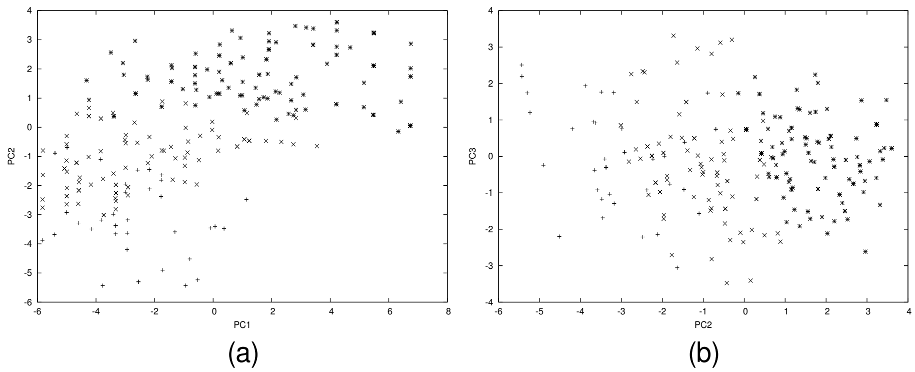

Fig. 3. Graphical projection of (a) scores of first and second factors from the PCA and (b) scores of second and third factors from the PCA. + show patients with slight CD, x with moderate CD and $*$ with servere CD.

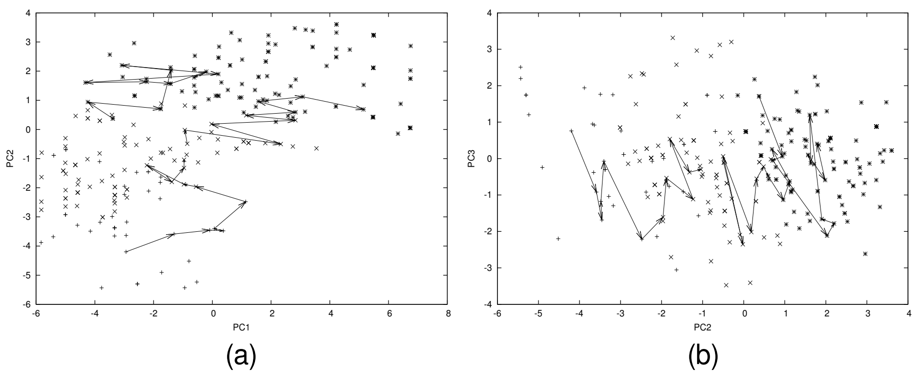

Fig. 4. The arrows show the temporal evolution of (a) scores of first and second factors from the PCA and (b) scores of second and third factors from the PCA of three different patients.

ate CD to severe CD, and the third one starts from slight CD and reaches the moderate CD. It is possible to observe that throughout the trajectories the appearance of oscillations is frequent, according to our clinical team these changes are in agreement with the real evolutionary situation of the patients. The decline processes are usually not stable, but patient can undergo small improvements or worsenings throughout which can be caused by diverse external factors.

It is possible that the maximum passed time between the first and last consultation, a year, turns out short to be able to see the global evolution of the patient with clarity over the commented oscillations in the data we have. Despite, indeed the utilities of this analysis relative to the pursuit of the evolution of the disease seem clear, comparison between evolutions of different patients and even prognosis from paths that could follow in the future.

5 Conclusions

The present work supposes an important advance in the area of the neuroinformatics and medical computer science, since it proposes new techniques of automatic learning to analyze and process data relative to neurodegenerative processes habitually related to the aging. On the basis of its results the creation of tools that allow the doctors to quantify, visualize and make a better pursuit of the associated sintomatology to these processes will be possible, contributing to a better knowledge of these ones.

In clinical practice, it is usual not to have the required clinical tests to evaluate a patient. In order to be able to face this difficulty in the PCA accomplishment we have extended the Sanger network equipping it with capacities for processing missing data.

PCA on subjects that present CD has allowed us to dissect the inherent multicolinearity from the different results of neuropsychological tests. We have been able to quantify in an independent way for each subject the functional, cognitive and instrumental sintomatologies that this one presents, and that is expressed by the different associated values from the first three principal components.

The analysis of the temporary evolution of the main components values has been like a powerful tool for the pursuit of patients. Allowing to detect the different sintomatologic changes and to compare these changes with the produced ones in the rest of patients. this analysis could allow us in a future to categorize the different evolutionary pictures, as well as to orient the doctor onto the possible prognosis that can experience his patients sintomatology. Our proposal will also make it easier to prognose the degeneration in a dementia as well as to make studies on pharmacological treatments, all which will result in the possibility of reaching one better quality of life of the patient.

Acknowledgement

We wold like to thank to Gobierno de Canarias for their support in de undertaking of this research under Research Project PI2003/178A and to Alzheimer's Patient Association of Gran Canaria for their generous input as data provider.

References

[1] López-Pousa, S.: La demencia: concepto y epidemiología. Enfermedad de Alzheimer y otras demencias, Mdica Panamericana, Madrid (2006) 33–42

[2] Andersen, C.K., Wittrup-Jensen, K.U., Lolk, A., Andersen, K., Kragh-Sorensen, P.: Ability to perform activities of daily living is the main factor affecting quality of life in patients with dementia. Health Qual Life Outcomes **2** (2004)

[3] DeKosky, S.: Early intervention is key to successful management of Alzheimer disease. Alzheimer Disease and Associated Disorders **17** (2003) S99–S104

[4] Doody, R.S.: Refining treatment guidelines in Alzheimer's disease. Geriatrics Suppl(2005) 14–20

[5] Neuropathology Group of the Medical Research Council Cognitive Function and Ageing Study: Pathological correlates of late-onset dementia in a multicentre, community-based population in England and Wales. The Lancet **357** (2001) 169–175

[6] Solomon, P., Murphy, C.: Should we screen for Alzheimer's disease? Geriatrics **60** (2005) 26–31

[7] Borson, S., Scanlan, J.M., Watanabe, J., Tu, S.P., Lessing, M.: Simplifying detection of cognitive impairment: comparison of the Mini-Cog and Mini-Mental State Examination in a multiethnic sample. Journal of the American Geriatrics Society **53** (2005) 871–874

[8] Nestor, P.J., Scheltens, P., Hodges, J.R.: Advances in the early detection of Alzheimer's desease. Nature Reviews Neuroscience **5** (2004) S34–S31

[9] Damian, M., Kreis, M., Krumm, B., Hentschel, F.: Optimized neuropsychological procedures at different stages of dementia diagnostics. Journal of the Neurological Sciences **229** (2005) 95–101

[10] Folstein, M.F., Folstein, S.E., McHugh, P.R.: Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. Journal of psychiatric research **12** (1975) 189–198

[11] Reisberg, B.: Functional Assessment Staging (FAST). Psychopharmacology Bulletin **24** (1988) 653–659

[12] Fernandez-Viadero, C., Verduga, R., Crespo, D.: Biomarcadores biológicos del envejecimiento. Biogerontología, Universidad de Cantabria, Cantabira (2006)

[13] Katz, S.C., Ford, A.B., Moskowitz, R.W.: Studies of illness in the aged. The index of ADL: a standardized measure of biological and psychosocial function. JAMA **185** (1963) 914–919

[14] Mahoney, F.I., Barthel, D.: Functional evaluation: The Barthel Index. Maryland State Medical Journal **14** (1965) 61–65

[15] Lawton, M.P., Brody, E.M.: Assessment of older people: self-mantaining and instrumental activities of daily living. Gerontologist **9** (1969) 179–186

[16] Sanger, T.D.: Optimal Unsupervised Learning in a Single-Layer Linear Feedforward Neural N. Neural Networks **2** (1989) 459–473

[17] Dimantaras, K., Kung, S.Y.: Principal Component Neural Networks: Theory and Applications. John Wiley and Sons, New York, NY (1996)

[18] Foldiak, P.: Adaptive network for optimal linear feature extraction. Proceedings of the International Joint Conference on Neural Networks I (1989) 401–405

[19] Rubner, J., Tavan, P.: A self-organizing network for principal component analysis. Europhysics Letters **10** (1989) 693–698

[20] Fiori, S. An Experimental Comparison of Three PCA Neural Networks. Neural Processing Letters. **11** (2000) 209–218

[21] Samad, T., Harp, S.A.: Self-organization with partial data. Network **3** (1992) 205–212