

Entropy Analysis Applied to NFIR Models

Oswaldo Ludwig Júnior, A. C. de Castro Lima, Leizer Schnitman, J. A. M. Felipe de Souza

Abstract— This present work has been developed in view of a research project that aims a mapping process between infrared images and thermo pair sensors readings in Diesel motors by the use of Artificial Neural Network. The consistence analysis of a set of examples used in the supervised training of Artificial Neural Networks is presented. The proposed approach is based on the entropy analysis of both input and output data. The Principal Component Analysis method is applied to avoid redundant information that can lead to an entropy level overestimation.

I. INTRODUCTION

ARTIFICIAL Neural Networks (ANN) applications have been attracted several research interests. In spite of countless types of applications for this tool, a supervised ANN is actually a function that relates an input space with an output space [1]. Equations 1.1 and 1.2 illustrate a function that exemplifies a multi-layer perceptron with one hidden layer and a linear output layer respectively:

$$y_h = \varphi(xW_1 + b_1) \quad (1.1)$$

$$y = y_h W_2 + b_2 \quad (1.2)$$

where $\varphi(\cdot)$ is a non-linear function, W_1 , W_2 are the synaptic weights matrix and y_h is the output vector of the hidden layer.

Functions cannot present entropy level [2] bigger than its arguments. If the training set has a target data with more information than the input data, then the ANN cannot learn the relationships between these data pairs. In other words, ANNs is not able to create information.

II. PRINCIPAL COMPONENTS ANALYSIS

In applications of ANN to dynamic system identification, a supervised ANN receives a set of examples

Oswaldo Ludwig Júnior is with the Universidade Federal da Bahia, Salvador-BA-Brazil (e-mail: oludwig@terra.com.br).

A. C. de Castro Lima is with the Universidade Federal da Bahia, Salvador-BA-Brazil (e-mail: aclima@terra.com.br).

Leizer Schnitman, is with **AREA I Faculdade de Ciência e Tecnologia**, Av. Santiago de Compostela, 216, Campus Sede, 40.279-150, Salvador, BA, Brazil (e-mail: leizer@areal.br).

J. A. M. Felipe de Souza is with the Electrical Engineering Department, Universidade da Beira Interior, Covilhã - Portugal, (e-mail: felippe@dem.ubi.pt).

composed by the matching of vectors regressed in n interactions in time. Due to this fact, the input data has high covariance level. This fact implies in redundant information, which can lead to an overestimation of the entropy level. Therefore, Principal Components Analysis (PCA) is indispensable to reach the real entropy evaluation.

In this work the PCA method is applied to both input data and target in order to eliminate the covariance between its coordinates. In a second stage, the entropy analysis is applied.

The PCA method [1] allows the coordinates covariance elimination by a basis change.

When applying PCA method, the first step is the calculation of the covariance matrix C . In a discrete approach, each element of this matrix represents the covariance $s^2(x_i, x_j)$ between two coordinates of a vector x :

$$s^2(x_i, x_j) = \sum_{n=1}^N x_i[n]x_j[n] - \frac{\sum_{n=1}^N x_i[n] \sum_{n=1}^N x_j[n]}{N} \quad (2.1)$$

where x_i and x_j are coordinates of vector x and N is the number of examples.

The second step is the determination of the eigenvectors v_n of C . The eigenvectors form the rows of an eigenvectors matrix P :

$$P = \{v_1, v_2, \dots, v_n\} \quad (2.2)$$

which is employed in the basis change of C in order to obtain a diagonal matrix D of eigenvalues of C :

$$CP = PD \therefore P^{-1}CP = D \quad (2.3)$$

Clearly the matrix D (i.e., C in the new basis) does not present covariance. This diagonal matrix has only variance $s^2(x_i, x_i)$. The D matrix is the covariance matrix of vector x in the basis P :

$$C = f(x) \therefore D = P^{-1}CP = f(P^{-1}x) \quad (2.4)$$

and therefore the vector x in the basis P has no covariance among its coordinates. In other words, this vector in basis P has no redundant information.

III. THE ENTROPY EVALUATION

It is well known that the entropy is the expected value of the information that a random variable may contain. Considering that the coordinate is represented by a

constraint set of discrete values, it is possible to associate each coordinate to a discrete random variable.

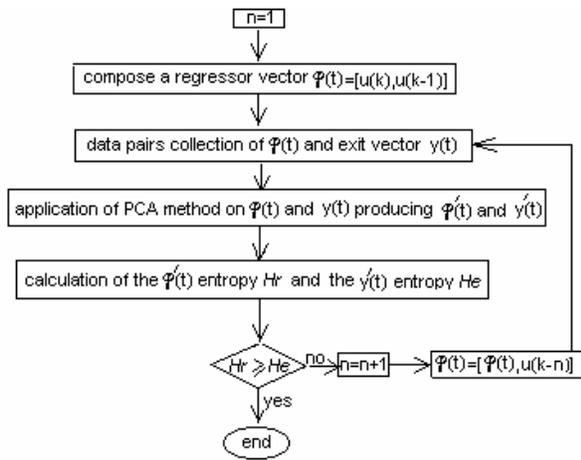
Computation of the entropy of a discrete random variable X requires the computation of the amount of information I revealed after occurrence of the event $X = x_i$. Considering p_i as being the probability of the event $X = x_i$ occurrence, then the amount of information I is defined as:

$$I(x_i) = \log\left(\frac{1}{p_i}\right) = -\log(p_i) \quad (3.1)$$

Considering N possible values of x_i that X can assume, the entropy is computed as

$$H(X) = E[I(x_i)] = \sum_{i=1}^N -\log(p_i) \cdot p_i \quad (3.2)$$

where $E[\cdot]$ is the expectation statistical operator.



IV. APPLICABILITY TO NFIR MODEL

ANNs can be applied to dynamic system identification [5, 6]. In this case, the regressor vector must contain all the necessary information to describe the system. However, each coordinate of this vector corresponds to an adjustable parameter of the ANN. A biggest coordinate number implicates in more computational effort, a biggest sample data set and increase of the sensor noise.

The method proposed here, illustrated in Figure 1 can help in the estimation process of the regressor vector length. This method can be applied to NFIR models (Non-linear Finite Impulse Response). The regressor vector of the NFIR models is just formed by the input data regressed in the time.

Note that this method (illustrated in figure 1) is a necessary but not a sufficient condition to a successful identification process of the system.

V. CONCLUSIONS

The proposed method can explain some problems that frequently occur in the training process convergence of ANN's feedforward multilayer.

This macro analysis can be applied to other mapping tolls as Fuzzy inference systems [8, 9].

This present work has been developed in view of a research project that aims a mapping process between infrared images and thermo pair sensors readings in Diesel motors by the use of ANN [7]. The regressor vector is composed by images matrix regressed in time and the output vector is composed by thermo pair sensor readings. The NFIR model is adopted here.

REFERENCES

- [1] Haykin, S.: *Redes Neurais, principios e práticas*. Bookman.
- [2] Papoulis, A., "Probability, Random Variables and Stochastic Processes", McGraw-Hill Book Co.(1991).
- [3] Lippmann, R.P., "Pattern classification using neural networks", *IEEE Communications Magazine*, vol.27, 1989.
- [4] Ludwig, O., Schmitman L., de Souza J.A.M.F., Lepikson H., "A Comparative analysis between conventional approaches and connectionist methods in pattern recognition tasks", *Proceedings of the IASTED International Conference on Artificial Intelligence and Applications (AIA 2004)*, Innsbruck, Austria, pp. 639-644.
- [5] S. Chen, S. Billings, "Neural networks for nonlinear dynamic system modeling and identification", *International Journal of Control*, v.56, n.2, p. 319-346, 1992.
- [6] K.S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks", v.1, n.1, p.4-27, 1990.
- [7] Merryman, S.A.; Nelms, R.M.; "Diagnostic technique for power systems utilizing infrared thermal imaging", *Industrial Electronics, IEEE Transactions*, Volume: 42 Issue: 6, Dec. 1995, page(s): 615 - 628.
- [8] So, A.T.P.; Chan, W.L.; Tse, C.T.; Lee, K.K.; "Fuzzy logic based automatic diagnosis of power apparatus by infrared imaging", *Neural Networks to Power Systems, 1993. ANNPS '93*.
- [9] Ludwig, O., Schmitman L., Lepikson H., "Uma arquitetura neural híbrida para extração de regras nebulosas", *Magazine IEEE Latin America*, Vol 3, jul. 2004.