An Unsupervised Neural Model to Analyse Thermal Properties of Construction Materials

Emilio Corchado[1,], Pedro Burgos[1], María del Mar Rodríguez[2], and Verónica Tricio[2]

[1] Department of Civil Engineering. University of Burgos,
09006 Burgos, Spain
`escorchado@ubu.es, pburgos@ubu.es`
[2] Department of Physics. University of Burgos,
09001 Burgos, Spain
`foulquie@arquinex.es, vtricio@ubu.es`

Abstract. This study proposes a new approach to feature selection and the identification of underlaying factors. The goal of this method is to visualize and extract information from complex and high dimensional data sets. The model proposed is an extension of Maximum Likelihood Hebbian Learning [14], [5], [15] based on a family of cost functions, which maximizes the likelihood of identifying a specific distribution in the data while minimizing the effect of outliers [7], [10]. We demonstrate a hierarchical extension method which provides an interactive method for identifying possibly hidden structure in the dataset. We have applied this method to study the thermal evolution of several construction materials under different thermal and humidity environmental conditions.

1 Introduction

We introduce a method which is closely related to exploratory projection pursuit. It is an extension of a neural model based on the Negative Feedback artificial neural network [3]. This method is called Maximum-Likelihood Hebbian learning (ML) [5], [14], [12].

In this paper we provide a hierarchical extension to the ML method. This extension allows the dynamic investigation of a data set in which each subsequent layer extracts structure from increasingly smaller subsets of the data. The method reduces the subspace spanned by the data as it passes through the layers of the network therefore identifying the lower dimensional manifold in which the data lies.

The Negative Feedback artificial neural network has been linked to the statistical techniques of Principal Component Analysis (PCA)[3], Factor Analysis [2] and Exploratory Projection Pursuit (EPP) [11]. The originality of this paper is the development and application of Hierarchical Maximum Likelihood Hebbian Learning (HML) to provide a novel approach different than those found by the contemporary statistical technique of PCA.

2   The Negative Feedback Neural Network

We introduce the Negative Feedback Network [3].

Consider an N-dimensional input vector, $\mathbf{x}$, and a M-dimensional output vector, $\mathbf{y}$, with W_{ij} being the weight linking input j to output i and let η be the learning rate.

The initial situation is that there is no activation at all in the network. The input data is fed forward via weights from the input neurons (the $\mathbf{x}$-values) to the output neurons (the $\mathbf{y}$-values) where a linear summation is performed to give the activation of the output neuron. We can express this as:

$$y_i = \sum_{j=1}^{N} W_{ij} x_j, \; \forall i \tag{1}$$

The activation is fed back through the same weights and subtracted from the inputs (where the inhibition takes place):

$$e_j = x_j - \sum_{i=1}^{M} W_{ij} y_i, \; \forall j \tag{2}$$

After that simple Hebbian learning is performed between input and outputs:

$$\Delta W_{ij} = \eta e_j y_i \tag{3}$$

The effect of the negative feedback is to stabilise the learning in the network. Because of that it is not necessary to normalise or clip the weights to get convergence to a stable solution. Note that this algorithm is clearly equivalent to Oja's Subspace Algorithm[6] since if we substitute equation (2) in equation (3) we get:

$$\Delta W_{ij} = \eta e_j y_i = \eta \left(x_j - \sum_k W_{kj} y_k \right) y_i \tag{4}$$

This network is capable of finding the principal components of the input data [3] in a manner that is equivalent to Oja's Subspace algorithm [6], and so the weights will not find the actual Principal Components but a basis of the Subspace spanned by these components.Writing the algorithm in this way gives us a model of the process which allows us to envisage different models which would otherwise be impossible [5,6].

3   A Family of Learning Rules

It has been shown [8] that the nonlinear PCA rule

$$\Delta W_{ij} = \eta \left(x_j f(y_i) - f(y_i) \sum_k W_{kj} f(y_k) \right) \tag{5}$$

can be derived as an approximation to the best non-linear compression of the data. Thus we may start with a cost function

$$J(W) = 1^T E\{(\mathbf{x} - Wf(W^T\mathbf{x}))^2\} \tag{6}$$

which we minimise to get the rule(5). [10] used the residual in the linear version of (6) to define a cost function of the residual

$$J = f_1(\mathbf{e}) = f_1(\mathbf{x} - W\mathbf{y}) \tag{7}$$

where $f_1 = \|\|^2$ is the (squared) Euclidean norm in the standard linear or nonlinear PCA rule. With this choice of $f_1(\ )$, the cost function is minimized with respect to any set of samples from the data set on the assumption that the residuals are chosen independently and identically distributed from a standard Gaussian distribution [13]. We may show that the minimization of J is equivalent to minimizing the negative log probability of the residual, $\mathbf{e}$, if e is Gaussian.

$$\text{Let } p(\mathbf{e}) = \left(\frac{1}{Z}\right)\exp(-\mathbf{e}^2) \tag{8}$$

The factor Z normalizes the integral of $p(\mathbf{y})$ to unity.

Then we [5] can denote a general cost function associated with this network as

$$J = -\log p(\mathbf{e}) = (\mathbf{e})^2 + K \tag{9}$$

where K is a constant. Therefore performing gradient descent on J we have

$$\Delta W \propto -\frac{\partial J}{\partial W} = -\frac{\partial J}{\partial \mathbf{e}}\frac{\partial \mathbf{e}}{\partial W} \approx \mathbf{y}(2\mathbf{e})^T \tag{10}$$

where we have discarded a less important term (see [9] for details).

   In general [7], the minimisation of such a cost function may be thought to make the probability of the residuals greater dependent on the probability density function (pdf) of the residuals. Thus if the probability density function of the residuals is known, this knowledge could be used to determine the optimal cost function.

[10] investigated this with the (one dimensional) function:

$$p(\mathbf{e}) = \left(\frac{1}{2+\varepsilon}\right)\exp(-|\mathbf{e}|_\varepsilon) \tag{11}$$

where

$$|\mathbf{e}|_\varepsilon = \begin{cases} 0 \_ \forall |\mathbf{e}| < \varepsilon \\ |\mathbf{e}| - \varepsilon \_ otherwise \end{cases} \tag{12}$$

with ε being a small scalar ≥ 0.

   [10] described this in terms of noise in the data set. However we feel [12] that it is more appropriate to state that, with this model of the pdf of the residual, the optimal $f_1(\ )$ function is the ε-insensitive cost function:

$$f_1(\mathbf{e}) = |\mathbf{e}|_\varepsilon \tag{13}$$

In the case of the Negative Feedback Network, the learning rule is

$$\Delta W \propto -\frac{\partial J}{\partial W} = -\frac{\partial f_1(\mathbf{e})}{\partial \mathbf{e}}\frac{\partial \mathbf{e}}{\partial W} \tag{14}$$

which gives:

$$\Delta W_{ij} = \begin{cases} 0 \text{ if } |\mathbf{e}_j| < \varepsilon \\ \eta y_i (sign(e_j)) \text{ otherwise} \end{cases} \tag{15}$$

In a more general situation let the residual after feedback has a probability density function (pdf)[5,12,14]:

$$p(\mathbf{e}) = \frac{1}{Z}\exp(-|\mathbf{e}|^p) \cdot \tag{16}$$

Then we [5] can denote a general cost function associated with this network as

$$J = -\log p(\mathbf{e}) = |\mathbf{e}|^p + K \tag{17}$$

where K is a constant. Therefore performing gradient descent on J we have

$$\Delta W \propto -\frac{\partial J}{\partial W} = -\frac{\partial J}{\partial \mathbf{e}}\frac{\partial \mathbf{e}}{\partial W} \approx y(p|\mathbf{e}|^{p-1} sign(\mathbf{e}))^T \tag{18}$$

where T denotes the transpose of a vector. We would expect that for leptokurtotic residuals (more kurtotic than a Gaussian distribution), values of p<2 would be appropriate, while for platykurtotic residuals (less kurtotic than a Gaussian), values of p>2 would be appropriate. Therefore the network operation is:

$$\text{Feedforward: } y_i = \sum_{j=1}^{N} W_{ij}x_j, \; \forall_i \tag{19}$$

$$\text{Feedback: } e_j = x_j - \sum_{i=1}^{M} W_{ij}y_i \tag{20}$$

$$\text{Weight change: } \Delta W_{ij} = \eta.y_i.sign(e_j)|e_j|^{p-1} \tag{21}$$

[10] described their rule as performing a type of PCA, but this is not strictly true since only the original (Oja) ordinary Hebbian rule actually performs PCA. It might be more appropriate to link this family of learning rules to Principal Factor Analysis since this method makes an assumption about the noise in a data set and then removes the assumed noise from the covariance structure of the data before performing a PCA. We are doing something similar here in that we are basing our PCA-type rule on the assumed distribution of the residual. By maximising the likelihood of the residual with respect to the actual distribution, we are matching the learning rule to the pdf of

the residual. Maximum Likelihood Hebbian Learning (ML) [5,12,14] has been linked to the standard statistical method of EPP [11,14].

4   Hierarchical Maximum Likelihood Hebbian Learning (HML)

There may be cases where the structure of the data may not be captured by a single linear projection. In such cases a hierarchical scheme may be beneficial.

This can be done in two ways, firstly by projecting the data using the ML [5], [12], [14] method, select the data points which are interesting and re-run the ML network on the selected data. Using this method only the projections are hierarchical.

A second more interesting adaptation is to use the resulting projected data of the previous ML network as the input to the next layer. Each subsequent layer of the network identifying structure among fewer data points in a lower dimensional subspace.

The HML method can reveal structure in the data which would not be identified by a single ML projection as each subsequent projection can analyse different sections of the subspace spanned by the data. This method has an additional advantage that with each sub-projection the number of samples and dimensionality of the data is reduced. As this is a linear method, by the linear combination of all previous layers we can identify the manifold in which the data points of interest lie.

5   Real Data Set and Experimental Results

The data used to illustrate our method is obtained from an experimental work which goal is the analysis of the thermal transmission of several construction materials [16]. We want to analysis their similarities and differences of behaviour. For that reason we have studied, in an independent way, the thermal transmission of these materials in a stationary state and in a transitory state. The data is collected from a "thermal house" with a heat source inside the house controlled by a thermostat outside of it. The house was inside of an inhabited building. It is a closed cube, in which all of the four vertical faces are made of different materials:

A piece of plywood (3 cm of thickness),	a simple glass (1,8 cm of thickness),
A piece of plaster (1,5 cm of thickness),	a piece of polystyrene (4 cm of thickness)

The experiment has been designed to register representative variables and them have been measured inside and outside the house. So we have measured the inside and outside environmental temperature, the inside and outside relative humidity, the inside and outside surface temperature of each vertical face and the thermal flux of each vertical face.

The initial situation is an equilibrium state between the inside and outside temperature for both cases. These are the surface temperature and the environmental one. Then the heat source is switched on until it reaches a stationary state. Once this state is reached, we kept the heating on for a while and then we switched it off. The data

recorded belong to a period which starts before the heating is switched on and finishes when the equilibrium state is reached after the heating is switched off.

We performs two series of measures with the same inside heat source. We do not provide humidity in the first serie and we do it during the second serie of measures. We have performs the same experiment several times.

In the following experiments we analyse the performance of the neural method (HML) and highlight the differences in the projections obtained by PCA and ML. We demonstrate also the HML method.

Comparison of PCA and ML on the Construction Materials Data Set
Figure 1 and Figure 2 show the behaviour of some thermal properties (the inside and outside environmental temperatures, the inside and outside relative humidity and the inside and outside surface temperature) of the four vertical faces during the ascending transitory state.
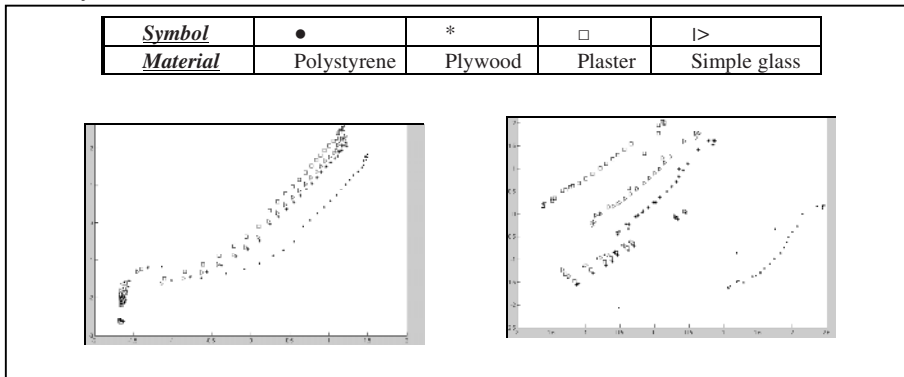
Symbol	●	*	□	|>
Material	Polystyrene	Plywood	Plaster	Simple glass

Fig. 1. PCA on construction materials data.   **Fig. 2.** ML on construction materials data

In Figure 1 and Figure 2 we show the comparison of PCA and ML projections of the construction material data. ML (Fig.2) clearly shows greater separation in the behaviour of different materials than is achieved with PCA (Fig.1). Figure 2 (ML) shows that Polystyrene behaves very different than the other three materials. These three materials (plywood, glass and plaster) have a similar behaviour. We could go further and say that plywood and glass have some times almost the same behaviour.

Behaviour of Two Vertical Faces, One Made of Glass and the Other One Made of Wood
In Figure 3 we show a representation of the behavior of some properties (inside and outside temperature, relative inside and outside humidity, the inside and outside surface temperature of each vertical face and their thermal flux) of two vertical faces, one made of glass and the other one made of wood.

If we analyse Fig.3 we could see that these two construction materials have quite the same behaviour in terms of temperature, but the thermal fluxes are different. These logic similarities and differences can not be observed by other classic models used by physicists [16].
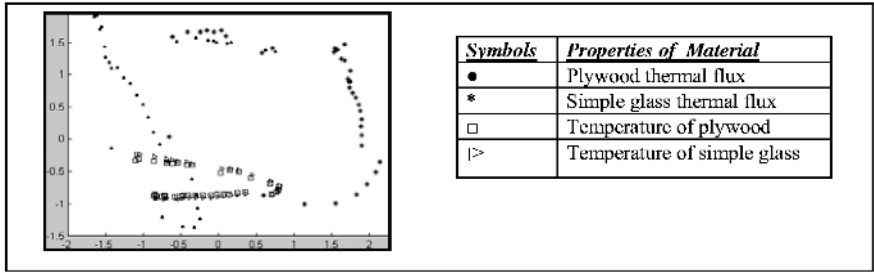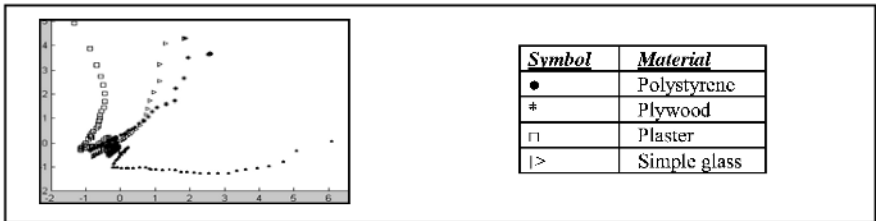
Fig. 3. ML on construction materials data.

Fig. 4. ML on construction materials data.

5.1 Results Using Hierarchical Maximum Likelihood Hebbian Learning

Fig.4. Representation of inside and outside temperature, relative inside and outside humidity and the inside and outside surface temperature of each vertical face versus time. The values correspond to the descendant transitory state. Fig.4 represents the results obtained using ML.

An important advantage of the hierarchical model is that provides more information in terms of inflexion points. This allows a better identification of different states in the studied transitory state and the identification of changes of state.as we could see in Fig.5.
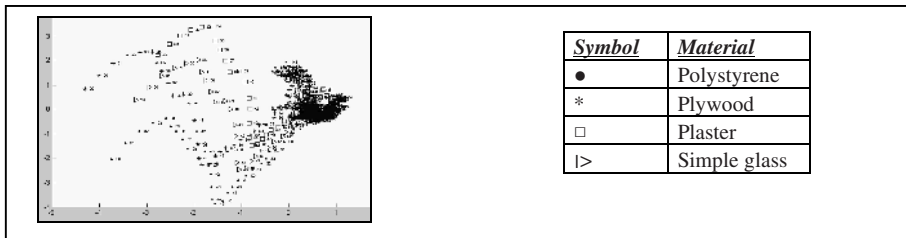

Fig. 5. Result of HML on left cluster in Fig.4

In Fig.5 we can see that HML identifies the inflexion points for the construction materials in a very clear way and gives us a new projection. It shows that the inflexion points of the four materials take places at the same time. If we compare Fig.5 and the temperature evolution plot (Fig.4) we may see some minimal temperature and relative humidity fluctuations in an inhabited place.

6   Conclusions

We have reviewed the recent method of ML, extend it to develop a hierarchical method and compare it with a typical statistical technique. The method proposed in this paper provides a novel way to extract more information from a particular good projection and it gives a very clear method to analyse the thermal behavior of construction materials. Future work will investigate further more complex data sets in order to study materials with an unpredictable behavior and in extreme conditions.

References

1. Bertsekas, D.P. Nonlinear Programming. Athena Scientific, Belmont, MA, 1995.
2. Charles, D. Unsupervised Artificial Neural Networks for the Identification of Multiple Causes in Data. PhD thesis, University of Paisley. 1999
3. Fyfe, C. Negative Feedback as an Organising Principle for Artificial Neural Networks, PhD Thesis, Strathclyde University. 1995
4. Fyfe, C. A. Neural Network for PCA and Beyond, Neural Processing Letters, 6:33-41. 1996.
5. Fyfe, C. and Corchado, E., Maximum Likelihood Hebbian Rules. European Symposium on Artificial Neural Networks. 2002.
6. Oja, E. Neural Networks, Principal Components and Subspaces, International Journal of Neural Systems, 1:61-68. 1989.
7. Smola, A.J. and Scholkopf, B. A Tutorial on Support Vector Regression. Technical Report NC2-TR-1998-030, NeuroCOLT2 Technical Report Series. 1998.
8. Xu, L. Least Mean Square Error Reconstruction for Self-Organizing Nets", Neural Networks, Vol. 6, pp. 627-648. 1993.
9. Karhunen, J. and Joutsensalo, J. Representation and Separation of Signals Using Nonlinear PCA Type Learning, Neural Networks, 7:113-127. 1994.
10. Fyfe, C. and MacDonald, D. ε-Insensitive Hebbian learning", Neuro Computing, 2002.
11. Friedman, J. and Tukey, J. A Projection Pursuit Algorithm for Exploratory Data Analysis. IEEE Transaction on Computers, (23): 881-890. 1974.
12. Corchado, E. and Fyfe, C. Orientation Selection Using Maximum Likelihood Hebbian Learning, International Journal of Knowledge-Based Intelligent Engineering Systems Volume 7 Number 2, April 2003. ISSN: 1327-2314. Brighton, United Kingdom.
13. Bishop, C.M., Neural Networks for Pattern Recognition, Oxford.,1995.
14. Corchado, E. MacDonald, D. and Fyfe, C. Maximum and Minimum Likelihood Hebbian Learning for Exploratory Projection Pursuit, Data Mining and Knowledge Discovery, Kluwer Academic Publishing, (In-Press).
15. Corchado, E. and Fyfe, C. Connectionist Techniques for the Identification and Suppression of Interference Underlying Factors. International Journal of Pattern Recognition and Artificial Intelligence. (IJPRAI). Vol. 17, No. 8 (Dec 2003).
16. Tricio, V.and Viloria, R. Microclimatic Study in a Historical Building: Protection and Conservation of the Cultural Heritage of the Mediterranean Cities. Topic: Environmental parameters of the Mediterranean Basin. 2002. Pg: 67-70.