A Connectionist Knowledge Management Model

Bruno Baruque [1], Álvaro Herrero [1], Emilio Corchado[1],
Lourdes Sáiz [1], Ana Lara[1] , Colin Fyfe[2]
[1] Department of Civil Engineering, University of Burgos, Spain.
[2]Applied Computational Intelligence Research Unit, University of Paisley, Scotland.
escorchardo@ubu.es, lsaiz@ubu.es, amlara@ubu.es, Colin.Fyfe@paisley.ac.uk

Abstract. In this paper, we illustrate the particular use of a novel connectionist architecture, characterised for its ability to capture some type of topological ordering. It has been used to build a part of a Global and Integral Model of Business Management, which provides a global improvement in the firm, in terms of flexibility, value and competitiveness. In this paper we approach knowledge management from a theoretical and practical point of view, showing the repercussions that their transformations from the inferior states (data and information) to the highest ones (knowledge and their management) have in the individual and organizational responsibilities.

1 Background

Knowledge management has become one of the hottest subjects of the day. The rising tide of data can be viewed as a vital, big and necessary source of information. This information may become knowledge [12]. Knowledge management increases speed to market through the reuse of proven resources and methods; reduces costly mistakes; and ensures consistent excellent service. It enables rapid absorption and diffusion of new ideas, allowing organizations to sustain a competitive advantage by improving some aspects as operational efficiency, rate of innovation, organizational agility.

We apply a novel method which is closely related to factor analysis (FA) and exploratory projection pursuit (EPP). It is a neural model based on the Negative Feedback artificial neural network, which has been extended by the combination of two different techniques. Initially by the selection of a proper cost function from a family of them, to identify the right distribution related to the data problem. This method is called Maximum-Likelihood Hebbian learning (MLHL) [5]. Then, lateral connections derived from the Rectified Gaussian Distribution [10] [11] were added to the MLHL architecture [5]. These enforced a greater sparsity in the weight vectors. In this study we have focus our attention, specifically, on the problem of knowledge management from a pragmatic and managerial point of view that contemplates the possibility that knowledge can be classified and organised in order to achieve a better understanding. This issue is based, above all, on understanding the distinctions between transformations in forms of knowledge, starting from an inferior level -data and information- and advancing towards other higher levels, such as knowledge itself and its management, individual, and even organizational responsibilities.

2 A Family of Learning Rules

The model used in this study is based on the Negative Feedback Network [7]. 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.
We can express this as:

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

The activation is fed back through the same weights and subtracted from the inputs

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

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

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

This network is capable of finding the principal components of the input data [7] in a manner that is equivalent to Oja's Subspace algorithm [9], and so the weights will not find the actual Principal Components (PCA) but a basis of the subspace spanned by these components. Factor Analysis is a technique similar to PCA in that it attempts to explain the data set in terms of a smaller number of underlying factors. However FA begins with a specific model and then attempts to explain the data by finding parameters which best fit this model to the data.

Let the residual [8] after feedback have probability density function

$$p(\mathbf{e}) = \frac{1}{Z}\exp(-|\mathbf{e}|^p) \tag{4}$$

A general cost function associated with this network is

$$J = -\log p(\mathbf{e}) = |\mathbf{e}|^p + K \tag{5}$$

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

$$\Delta W \propto -\frac{\partial J}{\partial W} = -\frac{\partial J}{\partial \mathbf{e}}\frac{\partial \mathbf{e}}{\partial W} \approx y(p|\mathbf{e}|^{p-1}\, sign(\mathbf{e}))^T \tag{6}$$

where T denotes the transpose of a vector.

Therefore the network operation is as before (feedforward (Eq.1), feedback (Eq.2)), but now the weight change is as follows:

weight change: $$\Delta W_{ij} = \eta . y_i . sign\left(e_j\right) | e_j |^{p-1} \qquad (7)$$

This method has been linked to the standard statistical method of EPP [6, 5].

3 Lateral Connections

The Rectified Gaussian Distribution (RGD) [11] is a modification of the standard Gaussian distribution in which the variables are constrained to be non-negative, enabling the use of non-convex energy functions.

The multivariate normal distribution can be defined in terms of an energy or cost function in that, if realised samples are taken far from the distribution's mean, they will be deemed to have high energy and this will be equated to low probability. More formally, the standard Gaussian distribution may be defined by:

$$p(\mathbf{y}) = Z^{-1} e^{-\beta E(\mathbf{y})}, \qquad (8)$$

$$E(\mathbf{y}) = \frac{1}{2}\mathbf{y}^T \mathbf{A}\mathbf{y} - \mathbf{b}^T \mathbf{y} \qquad (9)$$

The quadratic energy function $E(\mathbf{y})$ is defined by the vector **b** and the symmetric matrix **A**. The parameter $\beta = 1/T$ is an inverse temperature. Lowering the temperature concentrates the distribution at the minimum of the energy function.

An example of the RGD is the cooperative distribution. The modes of the cooperative distribution are closely spaced along a non-linear continuous manifold.

Neither distribution can be accurately approximated by a single standard Gaussian. Using the RGD, it is possible to represent both discrete and continuous variability in a way that a standard Gaussian cannot.

The sorts of energy function that can be used are only those where the matrix A has the property:

$$\mathbf{y}^{\mathbf{T}}\mathbf{A}\mathbf{y} > \mathbf{0} \text{ for all } \mathbf{y} : y_i > 0, i = 1 \ldots N \qquad (10)$$

where N is the dimensionality of **y**. This property blocks the directions in which the energy diverges to negative infinity.

The cooperative distribution in the case of N variables is defined by:

$$A_{ij} = \delta_{ij} + \frac{1}{N} - \frac{4}{N}\cos\left(\frac{2\pi}{N}(i-j)\right) \text{ and} \qquad (11)$$

$$\mathrm{b_i} = 1 \qquad (12)$$

where δ_{ij} is the Kronecker delta and *i* and *j* represent the identifiers of output neuron.

To speed learning up, the matrix **A** can be simplified [2] to:

$$A_{ij} = \left(\delta_{ij} - \cos\left(2\pi(i-j)/N\right)\right) \qquad (13)$$

The matrix **A** is used to modify the response to the data based on the relation between the distances between the outputs.
We use the projected gradient method, consisting of a gradient step followed by a rectification:

$$y_i(t+1) = [y_i(t) + \tau(b - \mathrm{A}y)]^+ \tag{14}$$

where the rectification $[\ ]^+$ is necessary to ensure that the y-values keep to the positive quadrant. If the step size τ is chosen correctly, this algorithm can provably be shown to converge to a stationary point of the energy function [1]. In practice, this stationary point is generally a local minimum.

The mode of the distribution can be approached by gradient descent on the derivative of the energy function with respect to **y**. This is:

$$\Delta \mathbf{y} \propto -\frac{\partial E}{\partial \mathbf{y}} = -(\mathrm{A}\mathbf{y} - \mathbf{b}) = \mathbf{b} - A\mathbf{y} \tag{15}$$

which is used as in Eq.14.

Now the rectification in Eq.14 is identical to the rectification which Corchado [3] used in the Maximum-Likelihood Network. Thus we will use this movement towards the mode in the FA version [2] of the Maximum-Likelihood Network before training the weights as previously. The net result will be shown to be a network which can find the independent factors of a data set but do so in a way which captures some type of global ordering in the data set.

We use the standard Maximum-Likelihood Network but now with a lateral connection (which acts after the feed forward but before the feedback). Thus we have:

Feedforward:
$$y_i = \sum_{j=1}^{\mathbf{N}} W_{ij} x_j, \forall i \tag{16}$$

Lateral Activation Passing:
$$y_i(t+1) = [y_i(t) + \tau(b - Ay)]^+ \tag{17}$$

Feedback:
$$e_j = x_j - \sum_{i=1}^{M} W_{ij} y_i, \tag{18}$$

Weight change:
$$\Delta W_{ij} = \eta . y_i . sign(e_j) | e_j |^{p-1} \tag{19}$$

Where the parameter τ represents the strength of the lateral connections. This method has been called Cooperative Maximum-Likelihood Hebbian learning (CMLHL) [10].

4 Experiments.

In this study we have analysed a multinational group, leader in the design and production of a great variety of components for the automotive industry. The justification of this choice lies in the fact that the characteristics of its management represent a favourable environment and opportune moment for the introduction of Knowledge Management. It is an undergoing organizational change and faces great growth and expansion, which requires a rapid adaptation to the demands of the sector, with greater

resources, imminent transfers and accurate forecasting of knowledge, together with the immediate demand to capitalise on them, to share and use them within the firm.
The design of the preliminary theoretical model of Knowledge Management is based on three components: the Organisation -Strategy and People-, Processes -Acquisition, Transfer and Updating of Knowledge- and Technology –Technological Aids-, from which the propositions of the model are defined. The population sample used came to 277 registries (individuals) that correspond with the "necessities of knowledge" showed by the head of eleven departments of the company studied. This knowledge gathers different stages (knowledge levels) that depict the current situation of each department for the tasks or activities assigned to each department to be successfully accomplished. Also, it has been possible to obtain valuable data on the degree of importance for the company of the gathered knowledge. This way, it is possible to identify the lack of the knowledge that it is necessary to perform the activity, so as to make the right decision on its acquisition in terms of how it is acquired, or what is the cost or time needed. In the same way, it is possible to specify the knowledge possessedwhich is not comprehensively employed, either because the person does not use it in its entirety, or because it has additional value and potential use, also, for other departments. Furthermore, it is possible to include the analysis corresponding to the necessary evolution of the present knowledge to detect new knowledge, to eliminate the obsolete knowledge and to validate new needs, among others.

The results obtained from the application to this date set is shown in Fig. 1a. Fig. 1b explains the result shows by Fig.1. The following explanation is based on Fig. 1.

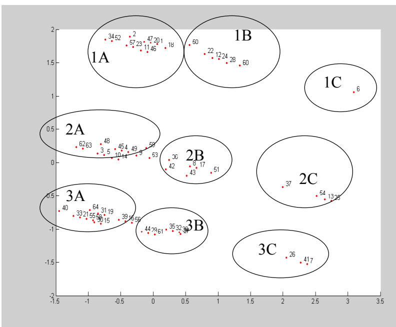

***Figure 1a**: CMLHL on the real data.*

CLOUD 1A **CRITICAL SITUATION** *a lot of urgency basic level*	*CLOUD 1B* **ALMOST GOOD** *during this year basic level*	*CLOUD 1C* **GOOD** *later basic level*
CLOUD 2A **CRITIC. SITUATION** *a lot of urgency half level*	*CLOUD 2B* **ALARM** *during this year half level*	*CLOUD 2C* **IMPROVE STRATEGY** *later half level*
CLOUD 3A **CHAOS** *a lot of urgency wide level*	*CLOUD 3B* **ALARM** *during this year wide level*	*CLOUD 3C* **GROWTH STRATEGY** *later wide level*

***Figure 1b**: Results Representation.*

Fig. 1. Fig.1a shows the result of CMLHL on the data. The projection identifies separated clusters (clouds), each of then has been labeled. We have identified mainly 9 cluster or clouds. Fig. 1.b is a graphical explanation of Fig.1a.

In terms of firm type, the positions of cloud 1C are related with a GOOD SITUATION. The firm is located in this place because the level of knowledge required is low and therefore the acquisition of knowledge is not a priority and also it would be at a very basic level. Also the fact that only one point (point 6 in Fig. 1a.)

appears underlines the fact that the company only has to acquire knowledge only in one specific area.

In the contrasting case, in the area occupied by the clouds labelled as 3A, there is a lot of urgency to acquire knowledge at a wide level. This area is called “CHAOS”. In a similar way, in the area occupied by the clouds 1A and 2A there is a need to acquire knowledge urgently at a half and a basic level. It could be that in these cases there is a holding of knowledge that can put the company in a CRITICAL SITUATION, since it may depend on the concession of new projects, the incorporation of new clients and all those parameters that somehow help to generates activity within the firm.

The area occupied by the points of the cloud 2C outlines the possibility to acquire knowledge at a later stage but at a half level. This could mean an IMPROVE STRATEGY in the firm, where it needs to improve in what it already possesses.

However, cloud 3C, represents the situation on that the firm has to acquire the knowledge later but at a wide level. This means that the company should think about the idea of enlarging and growing, both in terms of new processes and in new products. This is: GROWTH ESTRATEGY.

The points corresponding to area 1B, are related to an ALMOST GOOD area, because the knowledge is needed urgently and at a basic level. Cloud 2B and 3B identify an ALARM area, because there is no urgency and the level needed is half.

Conclusions and Future Work

We have presented and applied a recent model called CMLHL as a novel and robust tool to identify critical situations that allow firms to take decisions about acquisition, transfer and updating processes about knowledge management.

We have applied some other methods such as PCA or MLHL. CMLHL provides more sparse projections than the others [10] and captures some type of global ordering in the data set.

Future work will be based on the study of different distributions and learning rules to improve the whole architecture.

References

1. D.P. Bertsekas, Nonlinear Programming. Athena Scientific, Belmont, MA, (1995)
2. D. Charles, Unsupervised Artificial Neural Networks for the Identification of Multiple Causes in Data. PhD thesis, University of Paisley. (1999)
3. E. Corchado and C. Fyfe, Orientation Selection Using Maximum Likelihood Hebbian Learning, International Journal of Knowledge-Based Intelligent Engineering Systems Volume 7 Number 2, ISSN: 1327-2314. Brighton, United Kingdom. April (2003).
4. E. Corchado, Y. Han, C. Fyfe. Structuring global responses of local filters using lateral connections. J. Exp. Theor. Artif. Intell. 15(4): 473-487 (2003)
5 E. Corchado, D. MacDonald, and C. Fyfe, Maximum and Minimum Likelihood Hebbian Learning for Exploratory Projection Pursuit, Data mining and Knowledge Discovery, Kluwer Academic Publishing. May 2004, vol. 8, no. 3, pp. 203-225. (2004)
6. J. Friedman and J. Tukey. A Projection Pursuit Algorithm for Exploratory Data Analysis. IEEE Transaction on Computers, (23): 881-890. (1974)

7. C. Fyfe A Neural Network for PCA and Beyond, Neural Processing Letters, 6:33-41. 1996.
8. C. Fyfe, and E. Corchado, Maximum Likelihood Hebbian Rules. European Symposium on Artificial Neural Networks. (2002)
9. E. Oja, Neural Networks, Principal Components and Subspaces, International Journal of Neural Systems, 1:61-68. (1989)
10. E. Corchado, C. Fyfe: Connectionist Techniques for the Identification and Suppression of Interfering Underlying Factors. International Journal of Pattern Recognition and Artificial Intelligence. Vol. 17, No 8, 1447-1466. (2003)
11. H.S. Seung, N.D. Socci, and D. Lee, The Rectified Gaussian Distribution, Advances in Neural Information Processing Systems, 10. 350 (1998)
12. Alberthal, Les. Remarks to the Financial Executives Institute, October 23, 1995, Dallas, TX.