A Beta-Cooperative CBR System for Constructing a Business Management Model

Emilio S. Corchado[1], Juan M. Corchado[2], Lourdes Sáiz[1], and Ana Lara[1]

[1]Department of Civil Engineering,
University of Burgos, Spain
{escorchardo, lsaiz, amlara}@ubu.es
[2]Departamento Informática y Automática,
Universidad de Salamanca, Plaza de la Merced s/n,
37008, Salamanca, Spain
corchado@usal.es

Abstract. Knowledge has become the most strategic resource in the new business environment. A case-based reasoning system has been developed for identifying critical situations in business processes. The CBR system can be used to categorize the necessities for the Acquisition, Transfer and Updating of Knowledge of the different departments of a firm. This technique is used as a tool to develop a part of a Global and Integral Model of Business Management, which brings about a global improvement in the firm, adding value, flexibility and competitiveness. From this perspective, the data mining model tries to generalize the hypothesis of organizational survival and competitiveness, so that the organization that is able to identify, strengthen, and use key knowledge will reach a pole position. This case-based reasoning system incorporates a novel artificial neural architecture called Beta-Cooperative Learning in order to categorize the necessities for the Acquisition, Transfer and Updating of Knowledge of the different departments of a firm. This architecture is used to retrieve the most similar cases to a given subject.

1 Introduction

In this paper, we have centre our attention specifically on the problem of knowledge management from a pragmatic and managerial point of view that contemplates, first and foremost, 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 higher levels, such as knowledge itself and its management, individual, and even organizational responsibilities. This paper outlines the results obtained with a case-based reasoning system (CBR) developed to identify critical situations that allow firms to take decisions about acquisition, transfer and updating processes in knowledge management. Case-based reasoning (CBR) systems have been successfully used in several domains such as diagnosis, monitoring, prediction, control and planning [1] [2]. CBR systems require adequate

P. Perner (Ed.): ICDM 2004, LNAI 3275, pp. 42–49, 2004.

retrieval and reuse mechanisms to provide successful results. Such mechanisms need to be consistent with the problem that has to be solved and with the data used to represent the problem domain. A CBR system is a methodology used to construct software tools to assist experts in the resolution of problems. The CBR system presented in this paper incorporates mechanisms that facilitate the data clustering and indexation and automates the retrieval and adaptation stages of the CBR system.

A novel method which is closely related to exploratory projection pursuit has been used for the clustering and retrieval stages of the developed CBR system. 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 data, to identify the right distribution related to the data problem. This method is called Beta learning (BL). Then, lateral connections derived from the Rectified Gaussian Distribution are added to the Beta architecture [3]. These enforce a greater sparcity in the weight vectors. After presenting the Beta-cooperative network, the case-based reasoning system is employed and finally the results obtained with it are outlined.

2 Beta-Cooperative Learning

The model used in this study is based, as mentioned above, on the Negative Feedback Network [4]. 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:

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

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:

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

This network is capable of finding the principal components of the input data in a manner that is equivalent to Oja's Subspace algorithm [5], and so the weights will not find the actual Principal Components but a basis of the Subspace spanned by these components.

We can start with the probability density function of the Beta distribution and derive rules which will optimally find the distribution. If the residual is drawn from the Beta distribution, $B(v,\omega)$, with the following probability density function:

$$p(e) = \mathbf{e}^{v-1}(1-\mathbf{e})^{\omega-1} = (\mathbf{x} - W\mathbf{y})^{v-1}(1-\mathbf{x}+W\mathbf{y})^{\omega-1} \tag{4}$$

then if we wish to maximise the likelihood of the data with respect to the weights, we will perform gradient ascent using:

$$\frac{\partial p}{\partial W} = \mathbf{e}^{v-2}(1-\mathbf{e})^{\omega-2}\mathbf{y}(-(v-1)(1-\mathbf{e})+\mathbf{e}(\omega-1)) \tag{5}$$

which is a rather cumbersome rule.

So the learning rule is:

$$\Delta W = \eta \mathbf{e}^{v-2}(1-\mathbf{e})^{\omega-2}\mathbf{y}(-(v-1)(1-\mathbf{e})+\mathbf{e}(\omega-1)) \tag{6}$$

However for the case in which $v = \omega = 2$:

$$\frac{\partial p}{\partial W} = \mathbf{y}(-(1-\mathbf{e})+\mathbf{e}) \tag{7}$$

So the learning rule is simplified to:

$$\Delta W = \eta \mathbf{y}(2\mathbf{e}-1) \tag{8}$$

This method has been linked to the standard statistical method of Exploratory Projection Pursuit (EPP) [6].

The Rectified Gaussian Distribution [3] 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, we may define the standard Gaussian distribution by:

$$p(\mathbf{y}) = Z^{-1}e^{-\beta E(\mathbf{y})}, \tag{9}$$

$$E(\mathbf{y}) = \frac{1}{2}\mathbf{y}^T\mathbf{A}\mathbf{y} - \mathbf{b}^T\mathbf{y} \tag{10}$$

The quadratic energy function $E(\mathbf{y})$ is defined by the vector $\mathbf{b}$ and the symmetric matrix $\mathbf{A}$. 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 Rectified Gaussian distribution is the cooperative distribution. The modes of the cooperative distribution are closely spaced along a non-linear continuous manifold. Our experiments focus on a network based on the use of the cooperative distribution.

Neither distribution can be accurately approximated by a single standard Gaussian. Using the Rectified Gaussian, 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:

$$\mathbf{y}^{\mathbf{T}}\mathbf{A}\mathbf{y} > \mathbf{0} \text{ for all } \mathbf{y}: y_i > 0, i = 1 \ldots N \tag{11}$$

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:

$$A_{ij} = \delta_{ij} + \frac{1}{N} - \frac{4}{N}\cos\left(\frac{2\pi}{N}(i-j)\right) \text{ and} \tag{12}$$

$$b_i = 1 \tag{13}$$

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

To speed up learning, the matrix **A** can be simplified [7] [8] to:

$$A_{ij} = \left(\delta_{ij} - \cos\left(2\pi(i-j)/N\right)\right) \tag{14}$$

The matrix **A** is used to modify the response to the data based on the relation between the distances among the outputs. Note that the modes of the Rectified Gaussian are the minima of the energy function, subject to non-negativity constraints. We use the projected gradient method, consisting of a gradient step followed by a rectification:

$$y_i(t+1) = \left[y_i(t) + \tau(b - \mathrm{Ay})\right]^{+} \tag{15}$$

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 [9]. 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:

$$\Delta\mathbf{y} \propto -\frac{\partial E}{\partial \mathbf{y}} = -(\mathrm{A}\mathbf{y} - \mathbf{b}) = \mathbf{b} - A\mathbf{y} \tag{16}$$

which is used as in Eq. 15.

Thus we will use this movement towards the mode in the Beta Learning 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 Beta Learning Network but now with a lateral connection (which acts after the feed forward but before the feedback). Thus we have:

Feed forward:
$$y_i = \sum_{j=1}^{\mathbf{N}} W_{ij} x_j, \; \forall i \tag{17}$$

Lateral Activation Passing:
$$y_i(t+1) = \left[y_i(t) + \tau(b - Ay) \right]^{+} \tag{18}$$

Feedback:
$$e_j = x_j - \sum_{i=1}^{M} W_{ij} y_i, \tag{19}$$

Weight change:
$$\Delta W = \eta \mathbf{e}^{v-2} (1-\mathbf{e})^{\omega-2} \mathbf{y}(-(v-1)(1-\mathbf{e}) + \mathbf{e}(\omega-1)) \tag{20}$$

Where the parameter τ represents the strength of the lateral connections.

3 Problem Case

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 undergoing organisational 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 possessed which is not comprehensively employed, either because the person does not use it in its entirety or because it has additional value and potential use 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.

4 The Beta Cooperative Case-Based Reasoning System

The case-based reasoning developed assists experts in the decision processes taken, and helps to improve business processes. Case-based reasoning systems are used to solve new problems by adapting solutions that were used to solve previous similar problems [10]. The operation of a CBR system involves the adaptation of old solutions to match new experiences, using past cases to explain new situations, using previous experience to formulate new solutions, or reasoning from precedents to interpret a similar situation.

The reasoning cycle of a typical CBR system includes four steps that are cyclically carried out and in a sequenced way: retrieve, reuse, revise, and retain [1][2]. During the retrieve phase those cases that are most similar to the problem case are recovered from the case-base. The recovered cases are adapted to generate a possible solution during the reuse stage. Such a solution is reviewed and if it is appropriate, a new case is created and stored, during the retain stage, in the memory. Therefore CBR systems update (with every retain step) their case-bases and evolve with their environment.

CBR systems represent a methodology that requires the use of different techniques in each of the four stages. In this particular problem the case memory is indexed and the cases are classified with the help of the Beta-cooperative learning technique previously presented. The knowledge necessities are classified and once a new one is identified or reviewed the Beta-cooperative learning technique is recalled to identify similar knowledge necessities. During the reuse stage a radial basis function is used to identify the risk level associated with the new knowledge necessity [10]. The revision is manually carried out and the knowledge needs evaluated are added to or updated in the CBR system case base.

5 Results and Conclusions

The system presented above has been applied to the business management problem. The Beta-cooperative learning technique has been applied to classify the knowledge needs and the result obtained is presented in Figure 1. The knowledge needs with the same risk level associated are grouped together. Figure 1 presents the labelled groups.

The Beta-cooperative architecture can be used as a knowledge generation tool and can be used to analyse the business requirements. In terms of firm type, the points of the top right cloud are related to 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.

In a contrasting case, in the area occupied by the bottom left clouds, there is a lot of urgency to acquire knowledge at a wide level. This area is called “CHAOS”. In a similar way, in the top left area there is a need to acquire knowledge urgently at a half

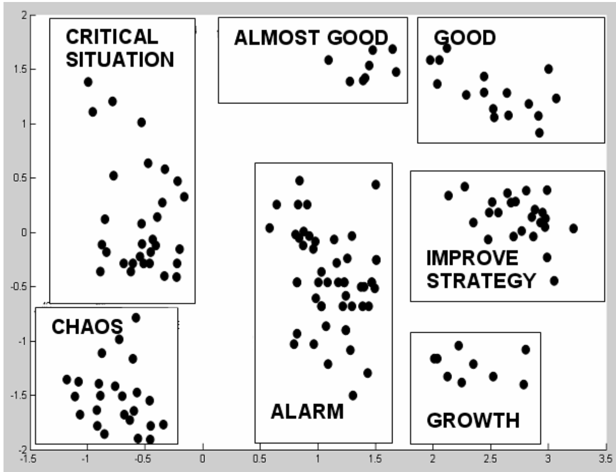

Fig. 1. The result of the application of the Beta-learning algorithm to the data. The projection identifies separate clusters (clouds), each of the groups has been labelled

or 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 generate activity within the firm.

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

However, bottom right cloud represents the situation 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 new products. This is: GROWTH STRATEGY.

The points corresponding to top area are related to an ALMOST GOOD area, because the knowledge is needed urgently at a basic level. The centre and bottom cloud identify an ALARM area, because there is no urgency and the level needed is half.

As a final conclusion we could say that we have presented and applied a model called Beta-Cooperative Learning 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 [5] [6] or MLHL [7], but Beta-Cooperative Learning provides more sparse projections [8] and captures a form of global ordering in the data set.

The initial results show that the system presented is a reliable tool for classifying the different situations (clusters) which the firm may face and to identify whether the firm is at a situation that requires it to take decisions about acquisition, transfer and updating processes about knowledge management. A second prototype of this system is under construction.

References

1. Watson I. and Marir F. (1994) *Case-Based Reasoning: A Review*. Cambridge University Press, 1994. The knowledge Engineering Review. Vol. 9. No. 3.
2. Pal S. K., Dillon T. S. and Yeung D. S. (2000) Soft Computing in Case-based Reasoning. (eds.). Springer Verlag, London, U.K.
3. Seung H. S., Socci N. D., and Lee D. (1998) The Rectified Gaussian Distribution, Advances in Neural Information Processing Systems, 10. 350 (1998).
4. Fyfe C. (1996) A Neural Network for PCA and Beyond, Neural Processing Letters, 6: 33-41.
5. Oja E. (1989) Neural Networks, Principal Components and Subspaces, International Journal of Neural Systems, 1:61-68.
6. Friedman J. and Tukey J. (1974) A Projection Pursuit Algorithm for Exploratory Data Analysis. IEEE Transaction on Computers, (23): 881-890.
7. Charles D. (1999) Unsupervised Artificial Neural Networks for the Identification of Multiple Causes in Data. PhD thesis, University of Paisley.
8. Corchado E., Han Y. and Fyfe C. (2003) Structuring global responses of local filters using lateral connections. J. Exp. Theor. Artif. Intell. 15(4): 473-487.
9. Bertsekas D. P. (1995) Nonlinear Programming. Athena Scientific, Belmont, MA.
10. Corchado J. M. and Aiken J. (2002) Hybrid Artificial Intelligence Methods in Oceanographic Forecasting Models. IEEE SMC Transactions Part C. Vol. 32, No.4. pp. 307-313.