DIPKIP: A Connectionist Knowledge Management System to Identify Knowledge Deficits in Practical Cases

Álvaro Herrero[1], Emilio Corchado[1], Lourdes Sáiz[1], and Ajith Abraham[2]

[1] Department of Civil Engineering, University of Burgos

C/ Francisco de Vitoria s/n, 09006 Burgos, Spain

Tel/Fax: +34 947 25 93 95

{ahcosio, escorchado, lsaiz}@ubu.es

[2] Centre for Quantifiable Quality of Service in Communication Systems

Norwegian University of Science and Technology

Trondheim, Norway

ajith.abraham@ieee.org

Abstract - This study presents a novel, multidisciplinary research project entitled DIPKIP (Data Acquisition, Intelligent Processing, Knowledge Identification and Proposal), which is a Knowledge Management (KM) system that profiles the KM status of a company. Qualitative data is fed into the system that allows it not only to assess the KM situation in the company in a straightforward and intuitive manner, but also to propose corrective actions to improve that situation.

DIPKIP is based on four separate steps. An initial 'Data Acquisition' step, in which key data is captured, is followed by an 'Intelligent Processing' step, using neural projection architectures. Subsequently, the 'Knowledge Identification' step catalogues the company into three categories, which define a set of possible theoretical strategic knowledge situations: knowledge deficit, partial knowledge deficit and no knowledge deficit. Finally, a 'Proposal' step is performed, in which the 'knowledge processes' - creation/acquisition, transference/distribution and putting into practice/updating– are appraised to arrive at a coherent recommendation. The knowledge updating process (increasing the knowledge held and removing obsolete knowledge) is in itself a novel contribution. DIPKIP may be applied as a decision support system, which, under the supervision of a KM expert, can provide useful and practical proposals to senior management for the improvement of KM, leading to flexibility, cost savings and greater competitiveness.

The research also analyses the future for powerful neural projection models in the emerging field of KM by reviewing a variety of robust unsupervised projection architectures, all of which are used to visualize the intrinsic structure of high-dimensional data sets. The main projection architecture in this research, known as Cooperative Maximum-Likelihood Hebbian Learning (CMLHL), manages to capture a degree of KM topological ordering based on the application of cooperative lateral connections.

The results of two real-life case studies in very different industrial sectors corroborated the relevance and viability of the DIPKIP system and the concepts upon which it is founded.

Keywords- Data and knowledge visualization, Connectionism and neural nets, Knowledge-based systems, Knowledge management applications, Discovery-based science.

I. INTRODUCTION

Knowledge Management (KM), a relatively new and fast-growing discipline, enables organizations to capture, share, and apply the collective experience and know-how (knowledge) of their staff, which is fundamental to competing in the knowledge economy. Exponential increases in data volumes are increasingly viewed as important and essential sources of information that may eventually be turned into knowledge.

KM can be successfully applied in organizations by developing and implementing knowledge infrastructures (Sivan 2000). These knowledge infrastructures consist of three main dimensions: people, organizational and technological systems. However, from a KM point of view, knowledge is defined as information that is relevant for business activities (Strohmaier 2003). This research is mainly based on an understanding of the distinctions between transformations in states of data and knowledge: from the lowest level (raw data and information) up to higher levels, such as knowledge itself and its management, and individual or organizational responsibilities.

Nowadays, a heterogeneous set of KM technologies (Maier and Remus 2002), (Maurer and Tochtermann 2002), (Rollett 2003), (Nielsen and Michailova 2007) is available from industrial vendors (Hyperwave), (OPENTEXT), (IBM) as well as from academia (Woitsch and Karagiannis 2002), (Dustdar 2002), (Wang 2004), (Chen et al. 2005).

In recent years, the deployment of information technology has become a crucial tool for enterprises to achieve a competitive advantage and organizational innovation (Shu-Mei 2008). In keeping with this idea, Artificial Intelligence (AI) can be applied in KM systems in order to speed up processes, classify unstructured data formats that KM is unable to organize, visualize the intrinsic structure of data sets, and select employee-related knowledge from large amounts of data, among other processes.

AI methods include those that analyze massive data sets. For instance, clustering reduces the quantity of data items by grouping them together. Methods also exist that can be used to reduce the dimensionality of the data sets and provide an interesting visualization of their internal structures. Among these are the so-called "projection" methods that are based on the identification of "interesting" directions. The interestingness of a direction is measured in terms of any one specific index or projection (see section II for further details). In this research, unsupervised learning models are applied. Their main advantage is that the neural network undergoes reorganization with respect to its internal parameters without external prompting. This is mainly caused by reactions with certain aspects of the input data. Typically, these aspects will be either redundancy in the input data or clusters in the data. In other words, an inner structure must exist in the data to which the network can respond.

This research approaches KM from both a theoretical and a practical point of view. It describes the impact that transformations can have on individual and organizational responsibilities, from the lowest states (data and information) to the highest (knowledge and its management). The research describes the development and testing of DIPKIP (Data Acquisition, Intelligent Processing, Knowledge Identification and Proposal), a novel KM system that identifies the KM status of an entire company or of one or more individual company unit (staff members, departments, divisions, etc.). It is based on 4 steps: Data Acquisition, Intelligent Processing, Knowledge Identification and a final Proposal. Both the Intelligent Processing step (based on the application of neural projection methods) and the Proposal step are worth emphasizing.

The purpose of DIPKIP is to support decision making that relates to knowledge acquisition, sharing and updating processes that are key to KM processes in the company. Given their complexity, the proposed KM system approaches the task in stages. It advances solutions so that the firm may access the knowledge it requires (at the right time) to develop its activities in a satisfactory way. In short, to manage and to build on what it knows how to do and to detect what it does not know about its business. As demonstrated by many studies since those of (Prescott and Visscher 1980) or (Edvinsson and Malone 1997), and as well as other more recent ones (Davenport and Hall 2002), (Senge et al. 2004), (Agndal and Nilsson 2006), (Grevesen and Damanpour 2007), the identification, acquisition, sharing and updating of knowledge increases the ability of the company to select and implement its key processes more efficiently. This leads to increased business productivity and profitability, which adds greater value to the company.

Empirical tests on two real-life case studies confirm the validity of this projection-based approach. Selected because of their diversity, they represent two different perspectives: from a departmental level within a multinational company (from the automotive industry sector) and from a worker level within a sector (group of companies that undertake the same economic activity) in the autonomous region of Castilla y León - Spain.

Contributions made by authors such as (Wiig 1994), (Marguardt 1996), (Ruggles 1997), (Beckman 1997), (Holsapple and Joshi 2002), (Taylor 2007), (Weissor, Sheng-Tun, and Kuan-Ju 2008), (Wei-Wen 2008), (Wu, Ong, and Hsu 2008) were given preferential attention in the formulation of the proposed model, to which the “updating of knowledge” process has been added. This process is one of the main strengths of DIPKIP because it helps to complement the design of KM models that are currently of such great interest to companies.

The paper is structured in the following way. Section 2 introduces some unsupervised neural projection methods. Section 3 describes DIPKIP, the proposed KM system, while Section 4 presents the main connectionist model applied in this research along with earlier ones that have supported its development. Section 5 describes the empirical validation of this research, starting with the description of the two real-life case studies and the high-dimensional data sets they generated. This section also presents results, discussions and a comparative study (several neural models are applied to the same case studies to compare their projection performance). Finally, Section 6 presents the conclusions and a number of proposals for future work in the same field.

II. UNSUPERVISED NEURAL MODELS AS A VISUALIZATION TOOL

A key problem in the analysis of high-dimensional data sets lies in the identification of patterns that exist across dimensional boundaries. Such patterns may become visible if changes are made to the spatial coordinates; however, an *a priori* decision as to which parameters will reveal most patterns requires prior knowledge of unknown patterns.

When researchers originally investigated high-dimensional and complex information such as spectroscopic data sets, they were looking for intrinsic structure by generating a scatter plot matrix in which they plotted one dimension of the data against another. This technique rapidly became less viable as the dimensionality of the data increased. Investigators later used some other techniques such as Principal Component Analysis (PCA) to provide a single projection in an effort to provide as much information as possible.

Projection methods project data points onto lower dimensions identified as “interesting” directions in terms of any one specific index. Such indexes are, for example, based on the variance of a data set (such as PCA) or higher-order statistics such as skewness or kurtosis, as in the case of Exploratory Projection Pursuit (EPP). Having identified the most interesting projections, the data is then projected onto a lower dimensional subspace plotted in two or three dimensions, which allows its structure to be examined with the naked eye. The remaining dimensions are discarded as they mainly relate to a very small percentage of the information or the data set structure. Thus, the structure is identified through a multivariable data set and may be visually analyzed with greater ease. The combination of EPP together with the use of scatter plot matrices constitutes a very useful visualization tool to investigate the intrinsic structure of multidimensional data sets. It enables experts to study the relations between different components, factors or projections, depending on the technique that is used. In this research, the results are interpreted by KM experts.

Projection techniques are powerful and proven tools in the development of KM systems such as DIPKIP. Section V describes the way in which this particular system is used to categorize the requirements for the acquisition, transfer, and updating of knowledge in two different real-life case studies. The following subsections review certain projection models applied in the experimental study of this work for comparison purposes.

A. Principal Component Analysis

PCA (Hotelling 1933), (Pearson 1901) describes the variation in multivariate data in terms of a set of uncorrelated variables, in decreasing order of importance, each of which is a linear combination of the original variables. Using PCA it is possible to find a smaller group of underlying variables that describe the data, with the result that the first few components of such a group might explain most of the variation in the original data. It should be noted that even if we are able to characterize the data with a few variables, it does not follow that an interpretation will ensue. This statistical technique may be performed by using connectionist models (Oja 1989), (Sanger 1989), (Fyfe 1997).

B. Exploratory Projection Pursuit

EPP (Friedman and Tukey 1974) is a statistical technique for solving the complex problem of identifying structure in high-dimensional data. It involves low-dimensional data projections in which structure is identified by eye and requires an index of “interestingness” by which each projection is measured. Subsequently, the data is transformed by optimizing this index, in order to examine the projections of greatest interest in greater detail. From a statistical point of view the most interesting directions are those which are as non-Gaussian as possible. Typical random data set projections are usually Gaussian (Diaconis and Freedman 1984), so identification of the most interesting features in the data calls for further investigation of these “interesting” directions. As in the case of PCA, this statistical technique may be implemented by using connectionist models (Hyvärinen 1998), (Hyvarinen 2001), (Corchado, MacDonald, and Fyfe 2004), (Fyfe, Baddeley, and McGregor 1994).

While PCA is focused on the identification of the largest variance directions, EPP looks for higher order statistics, such as skewness or kurtosis.

C. Self-Organizing Map

The Self-Organizing Map (SOM) (Kohonen 1990) was developed as a visualization tool for high dimensional data on a low dimensional display. It is also based on the use of unsupervised learning, but it is a topology-preserving-mapping model rather than a projection architecture. Composed of a discrete array of *L* nodes arranged on an *N*-dimensional lattice, it maps these nodes into a *D*-dimensional data space while preserving their ordering. The dimensionality of the lattice (*N*) is normally less than that of the data, in order to perform the dimensionality reduction. An input vector is presented to the network and a winning node *c* is chosen, whose weight vector has the smallest Euclidean distance from the input. Thus, the SOM quantizes data vectors to the reference vector in the map that is closest to the input vector. The weights of the winning node and the nodes close to it are then updated to move closer to the input vector. When this algorithm is sufficiently iterated, the map self-organizes to produce a topology-preserving mapping of the lattice of weight vectors to the input space based on the statistics of the training data. It is applied here for comparative purposes, as it is one of the most widely used unsupervised neural models for visualizing structure in high-dimensional data sets.

D. Curvilinear Component Analysis

Curvilinear Component Analysis (CCA) (Demartines and Herault 1997) is a nonlinear dimensionality reduction method that was developed as an improvement on the SOM. It tries to circumvent the limitations inherent in previous linear models, such as PCA. CCA is a self-organized neural network that performs two tasks: a vector quantization of the submanifold in the data set (input space) and a neural nonlinear projection of these quantizing vectors toward an output space, providing a revealing view of the way in which the submanifold unfolds. The projection part of CCA is similar to other nonlinear mapping methods; it minimizes a cost function based on interpoint distances in both input and output spaces. Quantization and nonlinear mapping are separately performed by two layers of connections. Firstly, the input vectors are forced to become prototypes of the distribution using a vector quantization (VQ) method. Then, the output layer builds a nonlinear mapping of the input vectors by considering Euclidean distances. In the empirical part of this work, a slight modification of standard CCA has been used. Instead of the Euclidean metric, the linear mapping built by the output layer considers the cosine distance metric, which takes account of the differences between the angles of the vectors obtained by VQ. Some other distances were tested in this work, but CCA based on cosine distance achieved the best results.

III. DIPKIP, A NOVEL KNOWLEDGE MANAGEMENT SYSTEM

This paper proposes a novel KM system (see Fig. 1) to support decisions that enable efficient KM in a company. The DIPKIP system identifies expertise and experts, ascertains whether knowledge is put to good use, calculates the amount of knowledge that is used, and identifies lack of knowledge. Once the diagnosis is ready, it proposes a set of solutions to improve these situations that relate to the absence of certain types of knowledge, its updating or knowledge sharing.

Hence, it allows the diagnosis and improvement of administrative or operating units (staff members, departments, divisions, etc.). The system is directed at knowledge managers whose goal is to improve the KM situation of a company.

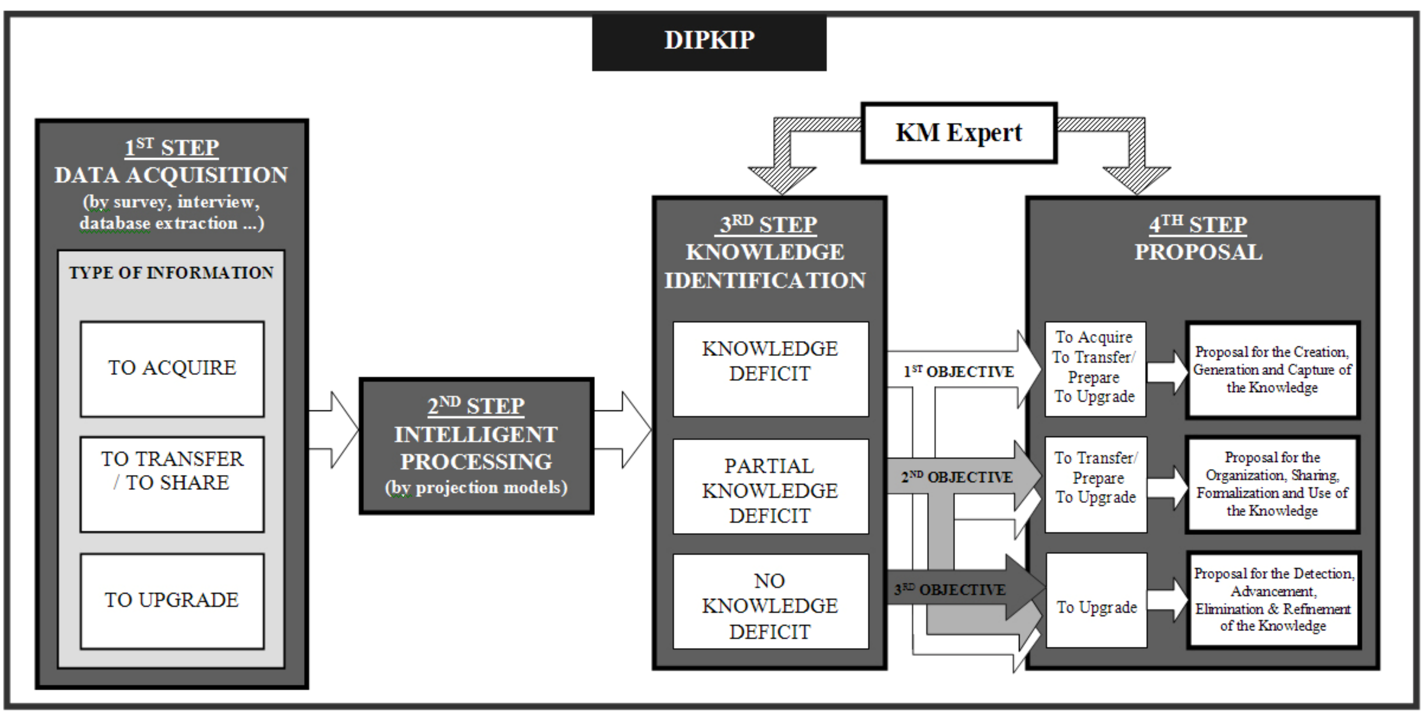


Fig. 1. DIPKIP: the KM system proposed in this study.

The four DIPKIP steps shown in Fig. 1 are described below in detail.

A. First Step: Data Acquisition

The first step aims to capture information concerning the situation in which DIPKIP is to be applied. As it is a general solution to identify the KM status of a company, the knowledge manager is responsible for designing how to gather the information that must be fed into the model, which could be acquired through interviews, surveys, database mining, a combination of these, and so on.

The key issue concerning this step is to decide which information is useful and which is irrelevant when determining the KM situation of the company under study.

B. Second Step: Intelligent Processing

The second step is Intelligent Processing, by which the data obtained in the first step is analyzed through Cooperative Maximum-Likelihood Hebbian Learning (CMLHL). This model (described in section IV) provides a visualization of the internal structure of the data set, which allows the KM expert to move on the third step of this KM system: 'Knowledge Identification'. At this stage, PCA, CCA and SOM, were also tested as alternative unsupervised neural models for comparison purposes. The CMLHL projections were selected for this second step, as this method provided the clearest projections of the case studies for subsequent expert analysis (see section V).

C. Third Step: Knowledge Identification

The output in the second step is a projection of the data acquired in the first step. The projection models embedded in the second step can be generalized as visualization tools. These tools rely on human expertise to process visual information, by performing exploratory browsing in a search for patterns and exceptions (Ahlberg and Shneiderman 1994).

Then, at this stage of the DIPKIP system, the KM expert, based on the data projection, catalogues the company into one of three classes, according to the situations that can arise in the field of strategic knowledge - knowledge deficit, partial knowledge deficit and no knowledge deficit.

D. Fourth Step: Proposal

Finally, the fourth step of the DIPKIP model is the Proposal, in which the following knowledge processes are proposed: creation/acquisition, transference/distribution and putting into practice/updating. The processes to be addressed are related to an underlying knowledge model based on the following knowledge life-cycle:

1. Knowledge creation/acquisition.
2. Knowledge transference/distribution.
3. Knowledge update.
4. Return to step 1 and restart the life-cycle.

Once the company is catalogued into one of the three areas (knowledge deficit, partial knowledge deficit and no knowledge deficit), the third step of DIPKIP sets out proposals relating to the above-mentioned KM processes. The KM expert should know which processes are applicable for each company under analysis:

- For knowledge deficit situations, the objective is to acquire or create the necessary knowledge. Before that, the shortcomings and the level or specificity of knowledge that is required must be detected and identified. At this point, some other issues must be considered: the way of acquiring the knowledge, quantifying its cost, and estimating its urgency.
- A partial knowledge deficit in any one area indicates that knowledge is only available to experts and has neither been made explicit nor been widely communicated within the organization. The organization holds critical knowledge, but it is not accessible to everybody that needs it. Faced with this situation, the knowledge must be communicated and shared on a case-by-case basis. In this process, experts and potential usages are identified, and a search is conducted for the means to express the knowledge and to make it available.
- No knowledge deficit implies that people and the organization have mastered the required know-how and that it is available to those who need it.

It is worth emphasizing that previous contributions to KM (Wiig 1994; Ruggles 1997) (Heinrichs and Lim 2005), (Nonaka and Toyama 2005), (Collins and Smith 2006), (Nielsen and Michailova 2007), (Donate-Manzanares and Guadamillas-Gomez 2007), (Kautz and Kjærgaard 2007), (Soo, Devinney, and Midgley 2007) have focused on knowledge creation, acquisition and transference. DIPKIP goes one step further by proposing a knowledge update process, which consists of increasing existing knowledge and examining it in depth to remove obsolete and useless knowledge. This in itself is a novel contribution, which to the best of our knowledge has not been considered elsewhere. Permanent knowledge updates increase the availability of the latest knowledge on the market and add value to the business.

Once DIPKIP has identified the required KM processes, the knowledge manager has to decide on the specific actions to implement these processes. Some examples for each of the KM processes are detailed below:

- Knowledge creation/acquisition: to plan specific training, identify problems and solutions, to promote group work with experienced colleagues, to request information and help from clients and suppliers, etc.
- Knowledge transference/distribution: to document the knowledge held, to design knowledge maps, to encourage people to take part in open discussion forums, to manage the optimum size of work groups, etc.
- Knowledge update: to foster personal creativity, to stimulate innovative mechanisms, to participate in knowledge communities including people from other companies, to enable the use of Communication and Information Technologies (CIT), etc.

It is worth emphasizing that the DIPKIP outputs must be customized by taking the particularities of each different situation into account. The knowledge manager supervising DIPKIP must adapt the proposals to the actual situation of the department, which is defined by such aspects as: number of employees, average age, business setting (clients, market, competitors, facilities to access resources, impact of the external business environment…), modifications/changes to critical/key knowledge of the firm.

IV. A Cooperative Neural Projection Method

The main architecture used in this study is the connectionist model known as Cooperative Maximum-Likelihood Hebbian Learning (CMLHL) (Corchado and Fyfe 2003). It is based on Maximum-Likelihood Hebbian Learning (MLHL) (Corchado, MacDonald, and Fyfe 2004) and introduces lateral connections (Corchado, Han, and Fyfe 2003). This connectionist model has been chosen because it reduces data dimensionality while preserving the topology of the original data set.

MLHL is a family of learning rules that is based on maximizing the likelihood of the residual from a negative feedback network whenever such residuals are deemed to come from a distribution in the exponential family. The main advantage of this model is that by maximizing the likelihood of the residual with respect to the actual distribution, we are matching the learning rule to the probability density function of the residual by applying different values of the "p" parameter specified in the learning rule.

For an N-dimensional input vector (x) and an M-dimensional output vector (y) with W_{ij} being the weight linking input j to output i, MLHL consists of the following stages:

The output of the network is calculated in the feedforward step:

$$y_i = \sum_{j=1}^{N} W_{ij} \cdot x_j , \forall i . \qquad (1)$$

The activation (e_j) is fed back through the same weights and subtracted from the input:

$$e_j = x_j - \sum_{i=1}^{M} W_{ij} \cdot y_i , \forall j . \qquad (2)$$

And finally, the weights are updated:

$$\Delta W_{ij} = \eta \cdot y_i \cdot sign\left(e_j\right) \cdot \left|e_j\right|^{p-1} . \qquad (3)$$

Where:

- η represents the learning rate.
- p is a parameter related to the MLHL energy function that is used to match the probability density

function and the learning rule.

The extended connectionist model, known as CMLHL, includes lateral connections acting after the feedforward, but prior to the feedback step. The resulting learning scheme is therefore as follows: there is a feedforward step (Equation 1) followed by the lateral activation step:

$$y_i(t+1) = [y_i(t) + \tau(b - Ay)]^+ . \qquad (4)$$

Then, the feedback step, in equation 2, is followed by the weight change defined by equation 3, in which:

- τ represents the “strength” of the lateral connections.
- b is the bias parameter.
- A is a symmetric matrix used to modify the response to the data based on the relation between the distances between the output neurons.

A. Lateral Connections

Lateral connections have been derived from the Rectified Gaussian distribution (Seung, Socci, and Lee 1998) which 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 standard Gaussian distribution may be defined by:

$$p(y) = Z^{-1} \cdot e^{-\beta \cdot E(y)}. \qquad (5)$$

$$E(y) = \frac{1}{2} \cdot y^T \cdot Ay - b^T \cdot y . \qquad (6)$$

in which, the quadratic energy function $E(y)$ 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. The factor Z normalizes the integral of $p(y)$ to unity.

The cooperative distribution is chosen as its modes are closely spaced along a non-linear continuous manifold. The energy functions that can be used are those that block the directions in which the energy diverges towards negative infinity. Thus, the matrix has to fit the following property:

$$y^T \cdot Ay > 0, \forall y : y_i > 0, i = 1...N \,. \qquad (7)$$

in which, N is the dimensionality of y.

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

$$A_{ij} = \delta_{ij} + \frac{1}{N} - \frac{4}{N} \cdot \cos\left(\frac{2\pi}{N}(i-j)\right) \text{ and} \qquad (8)$$

$$b_i = 1. \qquad (9)$$

in which, δ_{ij} is the Kronecker delta, and i and j, the output neuron identifiers.

Matrix A modifies the response to the data based on the relation between the distances between the outputs. The projected gradient method is used (Corchado, Han, and Fyfe 2003), consisting of a gradient step followed by a rectification as specified in equation 4, in which the rectification $[\;]^+$ is necessary to ensure that the y-values remain within the positive quadrant. If the step size (τ) is chosen correctly, this algorithm will probably be shown to converge to a stationary point of the energy function (Bertsekas 1999). In practice, this stationary point is generally a local minimum.

The distribution mode can be approached by gradient descent on the derivative of the energy function (see equation 6) with respect to y:

$$\Delta y \propto -\frac{\partial E}{\partial y} = -(Ay - b) = b - Ay \,. \qquad (10)$$

The resulting model (CMLHL) can expose the independent factors of a data set in a way that captures some type of global ordering in the data set and displays it with greater sparsity than other models.

Several versions of this model have successfully been applied to different datasets. Some of them are artificial, such as the well-known bars dataset (Corchado and Fyfe 2003), (Földiák 1992) while others are real, such as datasets on banking (Corchado, MacDonald, and Fyfe 2004), asteroids (Corchado, MacDonald, and Fyfe 2004), (Howell, Merényi, and Lebofsky 1994), (Cetin and Lewandowski 1991) and algae (Corchado, MacDonald, and Fyfe 2004). In the present research, this model is applied in the second step of the DIPKIP KM system (Intelligent Processing), as presented in previous section, to categorize the needs relating to knowledge acquisition, transfer and updating.

B. Fine Tuning

The CMLHL fine-tuning process is based on the effect of changing the τ parameter, which is the strength of the lateral connections between the output neurons. Experiments were conducted (Corchado and Fyfe 2003) using the bars data set proposed by Földiák (Földiák 1992) which adds noise in a graduated manner across the outputs. These experiments showed that altering the strength of the lateral connection parameter modulated the ability of the neural network to “gather” features together on the outputs. As predicted, a low τ value allows the neural model to code horizontal and vertical bars around a mode. An increase in the τ value means that the weak correlations between horizontal and vertical bars begin to have an impact on the learning. As the strength of the lateral connections becomes stronger, the bars are still learned around a mode but at the same time orientations start to separate. Subsequently, a separation emerges between the two different orientations, which is an interesting issue since all the data inputs to the network consist of both horizontal and vertical bars.

Increasing the τ value further will force the network to learn only one orientation of bars. However, if the lateral connections are too strong, then the coding of the bars may be squashed into an area of the output space that is too small for all of the bars to be coded individually. The reason why one orientation of bars is suppressed is due to the pixel overlap between different orientations of bars. If the lateral excitation between the output neurons is strong enough, a single output neuron may be able to switch its preference from a horizontal bar to a vertical one. That orientation identification was considered (Corchado and Fyfe 2003) to be a precursor of the creation of the concept of horizontal/vertical in animals inhabiting a mixed environment.

V. EXPERIMENTAL DOMAIN

The DIPKIP model was applied to two real-life case studies: a multinational company in the automotive industry and wall painting companies. This section describes these two case studies and sets out the results of having applied the proposed model.

These different case-studies were conducted at the request of companies that had previously expressed an interest in understanding the current situation and the future evolution of their knowledge assets.

Two very different data sets capture a departmental perspective in a multinational company (Automotive Industry case study) and a worker's perspective in small companies (Wall Painting case study) based in the autonomous region of Castilla y León – Spain. This diversity and the complexity of the cases in which DIPKIP has satisfactorily functioned underline the versatility of the proposed model.

Despite the differences between each case study, the model performed an empirical test of the knowledge situations, which resulted in proposals for knowledge acquisition, transference, and updating. This last Proposal step outlined actions that led to successful and effective KM in the company. The parametric values specified in the experimental setup were obtained after a fine-tuning process following detailed criteria described in previous studies (Corchado and Fyfe 2002), (Corchado, Han, and Fyfe 2003). The main parameters to be tuned were the p factor and the strength of the lateral connections (τ).

The CMLHL projections are compared at the end of each case study with those of other dimensionality-reduction models (PCA, MLHL, SOM and CCA). Several experiments were required to tune the SOM to different options and parameters: grid size, batch/online training, initialization, number of iterations and distance criterion, among others. In the case of CCA, other parameters, such as initialization, epochs and distance criterion were tuned. Only the best results (from the standpoint of the projection), which were obtained after tuning the models, are included in this work.

A. First Case Study: Automotive Industry

This case study is related to a previous line of research (Corchado et al. 2005), (Corchado, Corchado et al. 2004), (Corchado, Fyfe et al. 2004) that analyzed a multinational market leader in the design and the manufacture of a wide range of components for the automotive industry. It was an opportune moment for such a choice as the managerial environment of this company welcomed the introduction of KM. The multinational company was undergoing organizational change and was facing high growth and expansion that required rapid adaptation to the demands of the sector. At the same time, it was handling greater resources, all of which entailed more imminent transference and more accurate forecasting of knowledge. It had a pressing need to capitalize on these factors by using and sharing them within the company. The design of the preliminary theoretical model of KM was based on three components:

- Organization: concerning the strategy and people.
- Processes: creation/acquisition, transference/distribution, and putting into practice/updating.
- Technology: technological aids (on the basis of which the proposals in the model are defined).

The DIPKIP Data Acquisition step in this case study consisted in interviewing the managers of the following company departments: New Business, Purchases, Marketing, Project Management, Improved Industrial Practices, Protection of Design and Technology, Finance, Human Resources, Quality, Organization, and Information Systems. The interviewees evaluated several areas of expertise: “Human Resources Management Abroad”, “Personnel Selection Processes”, “International Mobility”, “Languages”, “Presentation Techniques”, “Patents Management”, “Positive Evaluation by Clients”, and “Environmental Strategy”, among others. Five features relating to each area were considered and measured in the following way:

- Current level of knowledge: absent (1), partial (3), enough (5) and expert (7).
- Importance of the knowledge: important (3), very important (6) and essential (9).
- Urgency to acquire knowledge: within the present year (9), next year (6) and later (1).
- Level of knowledge that may be needed: basic (3), medium (6) and high (9).

- Degree of the knowledge held by other departments within the company: non-existent (3), existent but not shared (6) and existent and shared (9).

They were converted into numerical values (in parentheses), for the application of the neural models.

The total population amounted to 140 records. The different knowledge levels depicted each department's situation with regard to their assigned tasks or to the activities that had to be implemented. Valuable data was also obtained on the importance of this knowledge to the company, which helped to identify a knowledge deficit (that had to be overcome to perform the activity). This enabled the right decision to be taken in relation to the way in which the knowledge should be acquired, and the time and cost required to do so. It was also possible to specify knowledge that was not usefully employed, either because the employee did not use it to the full, or because it also had additional value and a potential use within other departments.

Furthermore, the analysis also covered the expected evolution of the knowledge level, so as to detect new knowledge, to eliminate obsolete knowledge and to validate new needs, among other aspects.

1) Visualization and Discussion

The projections shown in Fig. 2 were obtained after tuning the CMLHL model in order to apply the second step of the DIPKIP KM system: Intelligent Processing. The vertical and horizontal axes forming this projection are combinations of the features contained in the original datasets.

The final values of the different parameters processed by CMLHL were as follows: number of iterations = 10,500, learning rate = 0.13, p parameter = 0.7, and τ parameter = 0.0015. Fig. 2 depicts all the possible combinations of the factor pairs obtained through CMLHL. Factor pairs under the diagonal are not shown as they provide no extra information. The main results obtained by CMLHL (factor pairs 1-2 and 1-3 from Fig. 2) are analyzed in depth in this section.

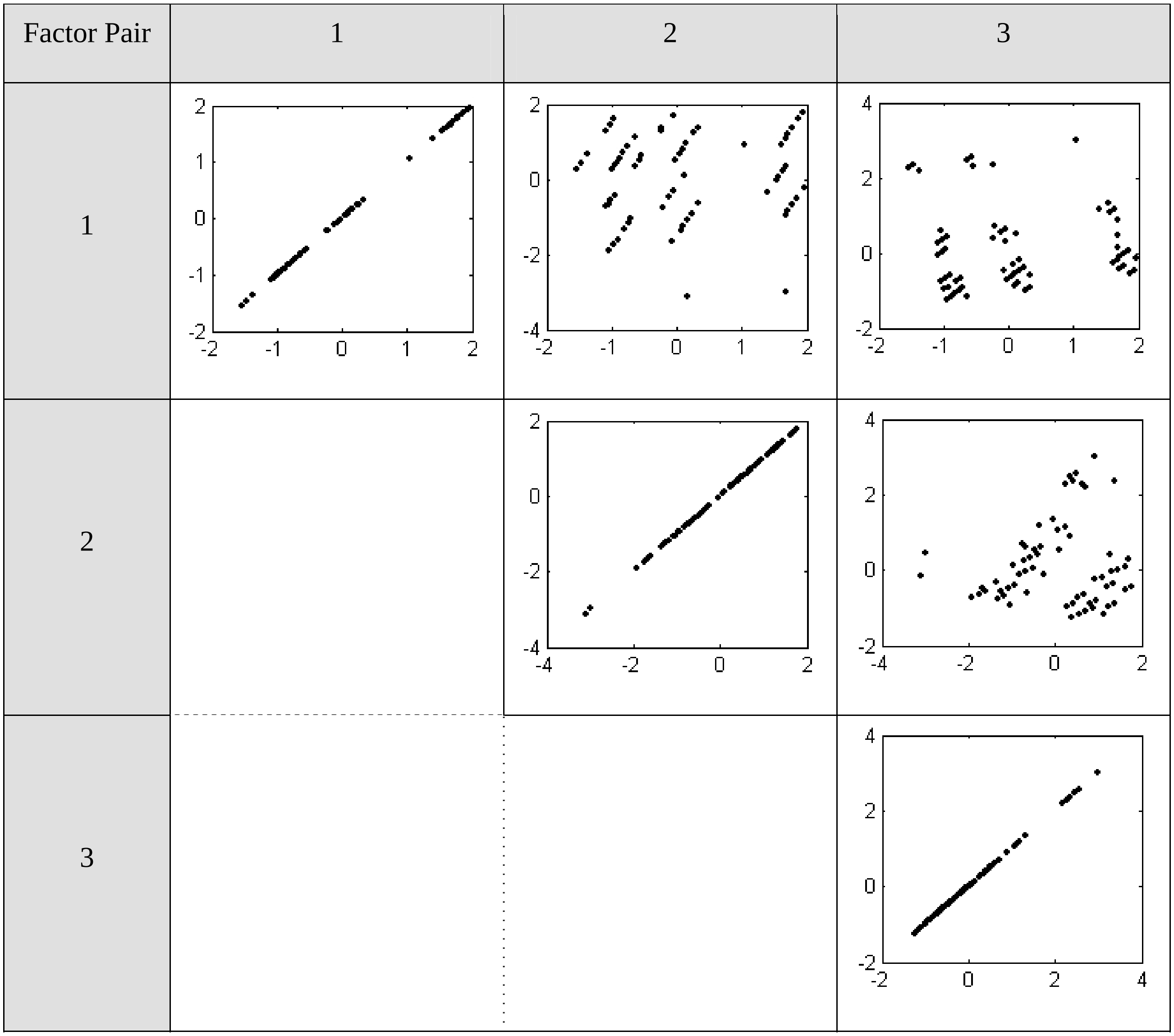


Fig. 2. CMLHL projections of the first 3 factors - KM Automotive Industry dataset.

Second step of the DIPKIP KM system: Intelligent Processing.

Fig. 3 presents the CMLHL projection of factor pair 1-2 taken from Fig. 2, which identified an intrinsic structure consisting of 9 groups (clouds). These groups are labeled (1A to 3C) and classified, as indicated in Fig. 4. The classification was obtained from the results of the third DIPKIP step: Knowledge Identification.

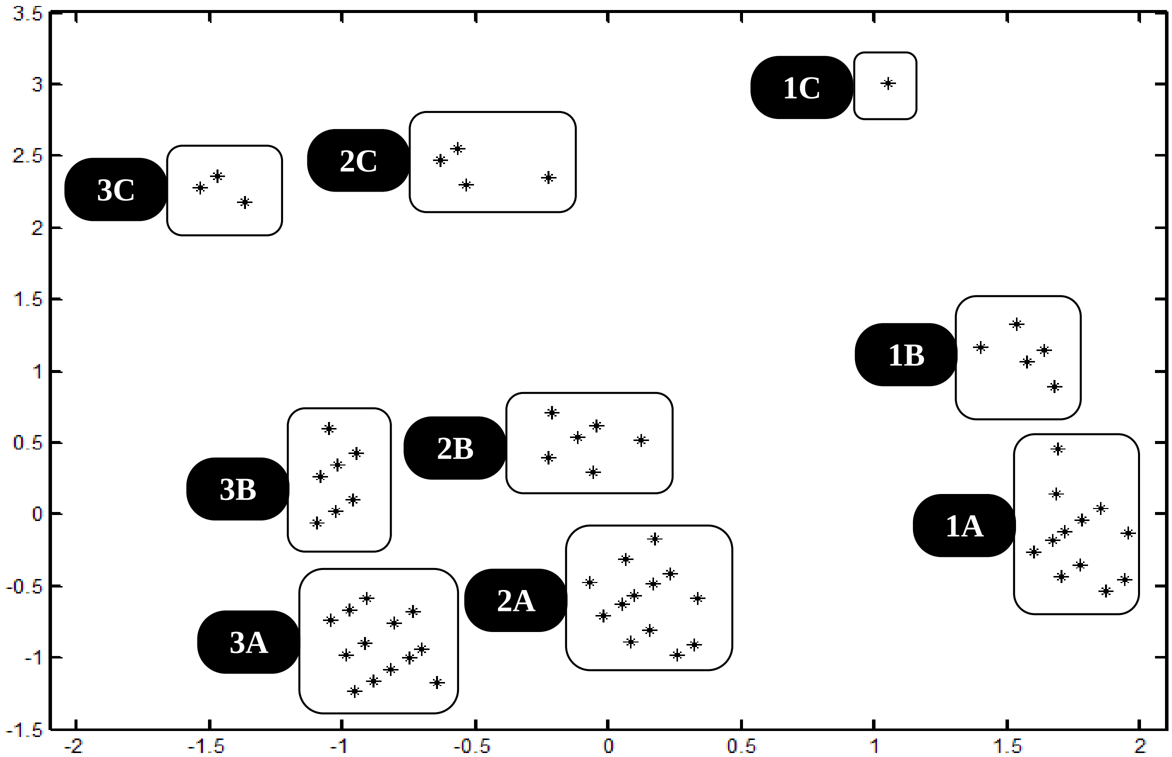


Fig. 3. CMLHL factor pair 1-3 projection (from Fig. 2) - Automotive Industry case study.

CLOUD 3C **GROWTH STRATEGY** *later* *wide level*	*CLOUD 2C* **IMPROVEMENT STRATEGY** *later* *medium level*	*CLOUD 1C* **GOOD** *later* *basic level*
CLOUD 3B **ALARM** *during this year* *wide level*	*CLOUD 2B* **ALARM** *during this year* *medium level*	*CLOUD 1B* **ALMOST GOOD** *during this year* *basic level*
CLOUD 3A **CHAOS** *a lot of urgency* *wide level*	*CLOUD 2A* **CRITICAL** *a lot of urgency* *medium level*	*CLOUD 1A* **CRITICAL** *a lot of urgency* *basic level*

Fig. 4. Schematic diagram - third DIPKIP step (Knowledge Identification) - Automotive Industry case study.

The terms “later”, “during this year” and “a lot of urgency” refer to temporal restrictions on the acquisition of new knowledge. The terms “wide”, “medium” and “basic” refer to the level of required knowledge.

Cloud 1C is formed by samples that reflect a “GOOD” situation. In this case, all the records (departments) belonging to this group were classified as being in a “GOOD” situation because the required knowledge level is low, hence the assessment of knowledge acquisition was not a priority. Furthermore, the presence of only one record underlines that the knowledge the company has to acquire is limited to only one specific area. In contrast, in the area occupied by Cloud 3A, there is great urgency to acquire knowledge at a wide level. This situation is labeled as “CHAOS”. Similarly, in the areas occupied by Clouds 1A and 2A, there is an urgent need to acquire knowledge at a medium and a basic level. In these cases, it might be said that knowledge is being withheld, a situation which could place the company in a “CRITICAL” situation, since it may influence all those parameters that somehow help to generate activity within the company, such as the concession of new projects, the incorporation of new clients, and so on.

The points within Cloud 2C suggest that the company may acquire knowledge at a later stage in this area, but at a medium level, to improve its knowledge (“IMPROVEMENT STRATEGY”). The proposal arising from Cloud 3C is that knowledge should be acquired in the medium-to-long term but at a broader level, suggesting that the company should think about enlarging and growing, both in terms of new processes and products (“GROWTH STRATEGY”).

Cloud 1B identifies an “ALMOST GOOD” situation, because knowledge is needed urgently and at a basic level. Clouds 2B and 3B identify an “ALARM” situation, as there is no urgency and knowledge is needed at a medium level.

The CMLHL neural model clearly captures some kind of topological order. If we start at Cloud 3A and move to the right of Fig. 3, the urgency of knowledge acquisition is the same, but the level decreases. On the contrary, if we move up through Fig. 3, the level is the same but the urgency for knowledge acquisition decreases.

Fig. 5 shows the CMLHL projection based on factor pair 1-2 (from Fig. 2), and Fig. 6 shows a diagram of this projection. The analysis of factor pair 1-2 (Fig. 5) provides a viewpoint that complements the analysis of factor pair 1-3 (Fig. 3), and it helps us to relate the data in terms of shared characteristics or parameters. It is then a straightforward task to chart the knowledge situation of the company and to select the most appropriate strategy. The diagram from Fig. 6 is analyzed by the KM expert in the third step, to identify the

KM situation of the departments under analysis.

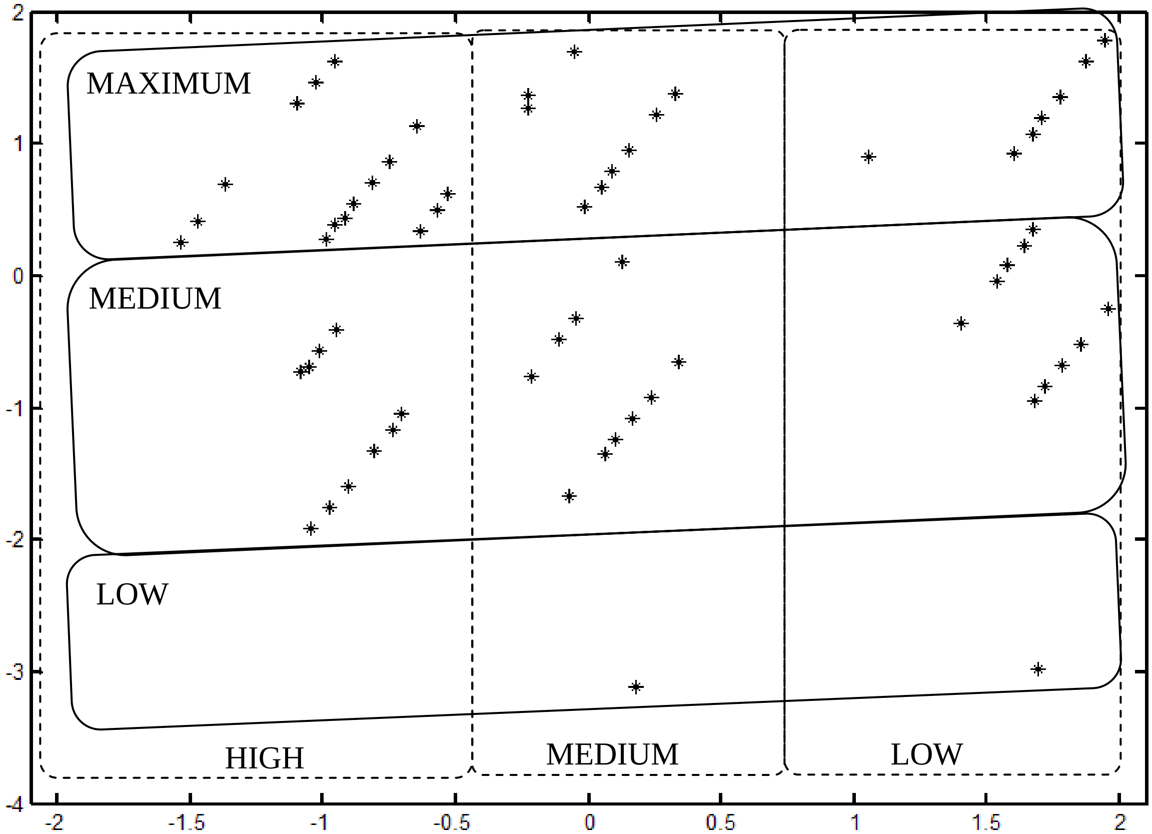


Fig. 5. CMLHL factor pair 1-2 projection (from Fig. 5) - Automotive Industry case study.

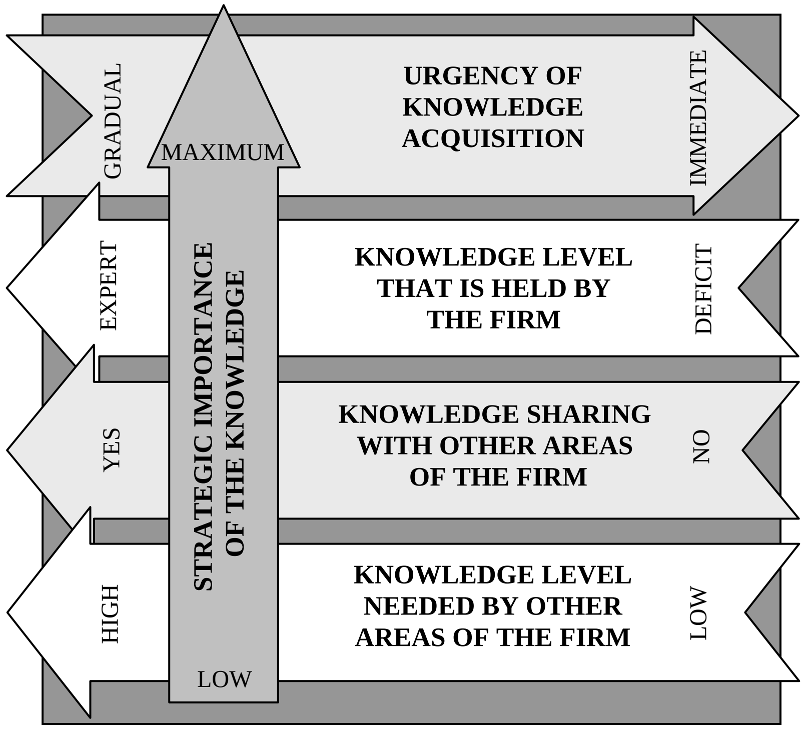


Fig. 6. Diagram of factor pair 1-2 (from Fig. 5).

The different parameters shown in Fig. 6 facilitate a visual analysis of Fig. 5:

- **Importance of the knowledge**: it may be seen that the points in the upper-right section of Fig. 5 represent indispensable knowledge for the company, whereas points in the central positions represent less important knowledge, and those in the lower third represent the least important ones. In this case study, most of the knowledge identified by the company is strategically important. This identification of critical knowledge is a prior condition in order to reach the KM objectives.
- **Urgency of acquisition**: the knowledge that has to be updated within the company or the extra knowledge that has to be acquired is described from a temporal perspective, ranging from most urgent (right half of Fig. 5) to least urgent (left half of Fig. 5). Half of the examined knowledge needs immediate acquisition or updating. On the basis of these projections, it is relatively easy to decide on the most efficient strategies to propose in the fourth step - acquisition and updating of the relevant knowledge - by employing specialized consultants and/or by providing staff training courses.
- **Knowledge level**: the right side and the centre of Fig. 5 contain the lowest levels of knowledge

(knowledge deficit), while expert knowledge is found in the upper left corner. Most of the knowledge is held at a medium level, while there is a slight tendency to move towards higher levels. In this case, the company must pay special attention to training and staff development policies that will manage to overcome the knowledge deficit, as well as organizational restructuring and redesign.

- **Sharing or diffusion of knowledge with other departments**: the knowledge located towards the left-hand side of Fig. 5 represents very broad needs, and its necessity decreases as we move from right to left. Distribution of the knowledge required at medium and lower levels by other business areas is worse than that needed at a higher level. The fourth step of DIPKIP proposes resources and measures to enable knowledge sharing and diffusion in order to improve cooperation and mutual trust between workers and groups.

2) *Step-by-step Sample*

Two departments (E and F) within the company were chosen to present each of the four DIPKIP steps in detail, subject to standard business confidentiality constraints.

First Step: Data Acquisition

In response to the detailed interviews, the manager of departments E and F gave the following answers for a specific expertise:

Question	Department E	Department F
Current level of knowledge	partial (3)	absent (1)
Importance of the knowledge	essential (9)	essential (9)
Urgency to acquire knowledge	within the present year (9)	within the present year (9)
Level of knowledge that may be needed	basic (3)	basic (3)
Degree of knowledge held by other departments	non-existent (3)	non-existent (3)

Table 1. Data Acquisition step for departments E and F - Automotive Industry case study.

Second Step: Intelligent Processing

After training the CMLHL model, department E was identified as belonging to Group 1A in Fig. 3 while department F was located in group 2A, representing an even worse situation than the one in group 1A.

Third Step: Knowledge Identification

Groups 1.A and 2.A from Fig. 3 were identified as “CRITICAL” situations. They represent two of the worst situations from a KM point of view. Both departments were diagnosed as having a “Knowledge Deficit”.

Fourth Step: Proposal

Having identified knowledge deficit situations in these two departments, the proposal was to acquire, create and generate the necessary knowledge. The outcomes of this step were specified in the following way:

- Department E: transfer the knowledge possessed by this department to the people that need it, providing incentives to those that already possess it. Moreover, it was recommended that the person who received the knowledge should document it.
- Department F: specific training on the relevant expertise and a 15-day training period in the company.

Finally, the detailed proposals were applied to validate the model. These departments (E_1 and F_1) were originally identified as belonging to Groups 1A and 2A respectively. Once the employees had been trained, the surveys were once again presented to the same personnel. The new situations of these departments following implementation of the DIPKIP proposals are depicted in Fig. 7 as E_2 and $F_{2,}$ respectively. As shown below, the situation of these departments was improved by applying the outcomes of DIPKIP in the following way:

- Department E: moving from group 1A to group 1B. For the situations represented by group 1A, there is a set of pieces of knowledge that could place the company in a critical state. Such knowledge is essential for adjudicating new projects, incorporating new clients and other parameters that generate value for the company. By documenting the knowledge, it is available for new users. Thus, the urgency of acquiring it is reduced, which moves Department E to group 1B, a semi-optimal location.

- Department F: moving from group 2A to group 1A. The new situation of this department represents a slight improvement as the proposed training maintains the same level of urgency but reduces the required level of knowledge.

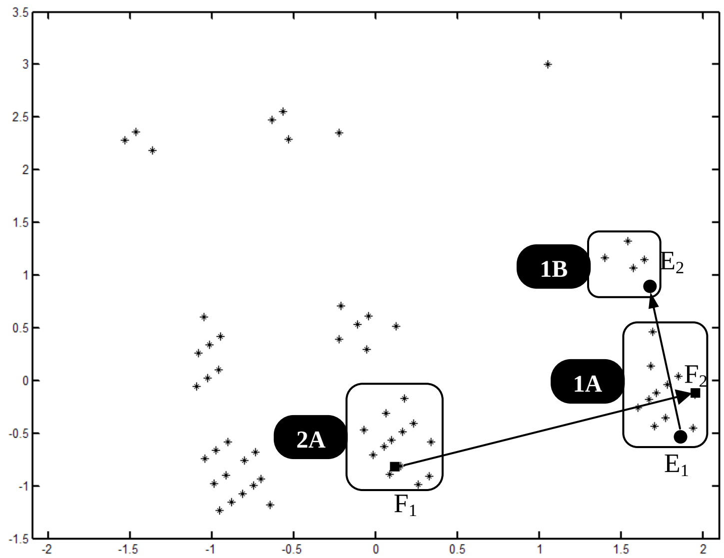


Fig. 7. CMLHL projection (factor pair 1-2) - Automotive Industry case study

Fourth DIPKIP KM step: corrective action.

3) *Comparison with other Unsupervised Methods*

Fig. 8 shows the results obtained by using PCA. In this case, the neural implementation of PCA has identified a clear internal structure, based on six groups, but the resulting clustering is not as sparse as that obtained by CMLHL (Fig. 3), which provides a more structured and interesting analysis by identifying nine clearly defined and ordered groups.

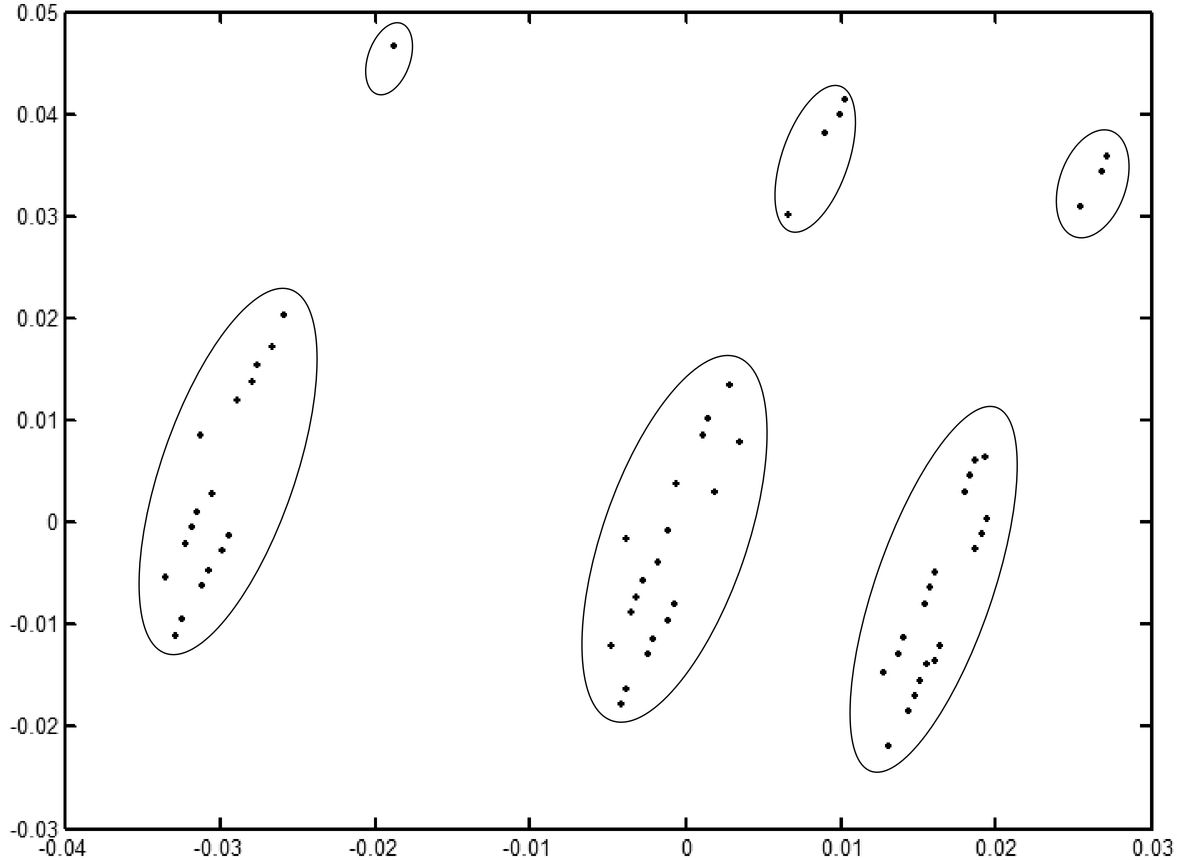


Fig. 8. PCA projection - Automotive Industry case study.

Fig. 9 shows the results obtained by using MLHL. This neural model has almost identified the same groups as CMLHL. The main advantage of the CMLHL projection is that the result is more spread out (it is easier to identify subclusters if there are any), which may be explained by the inherent advantages of lateral connections.

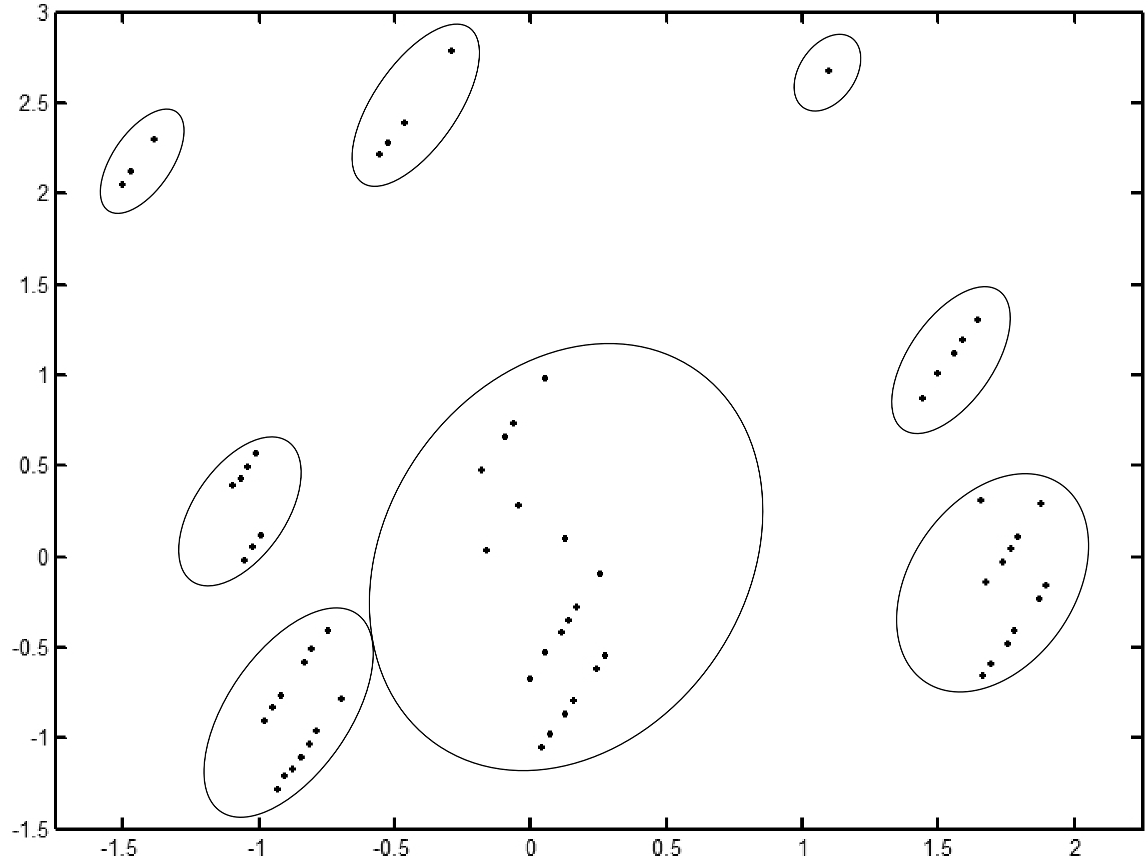
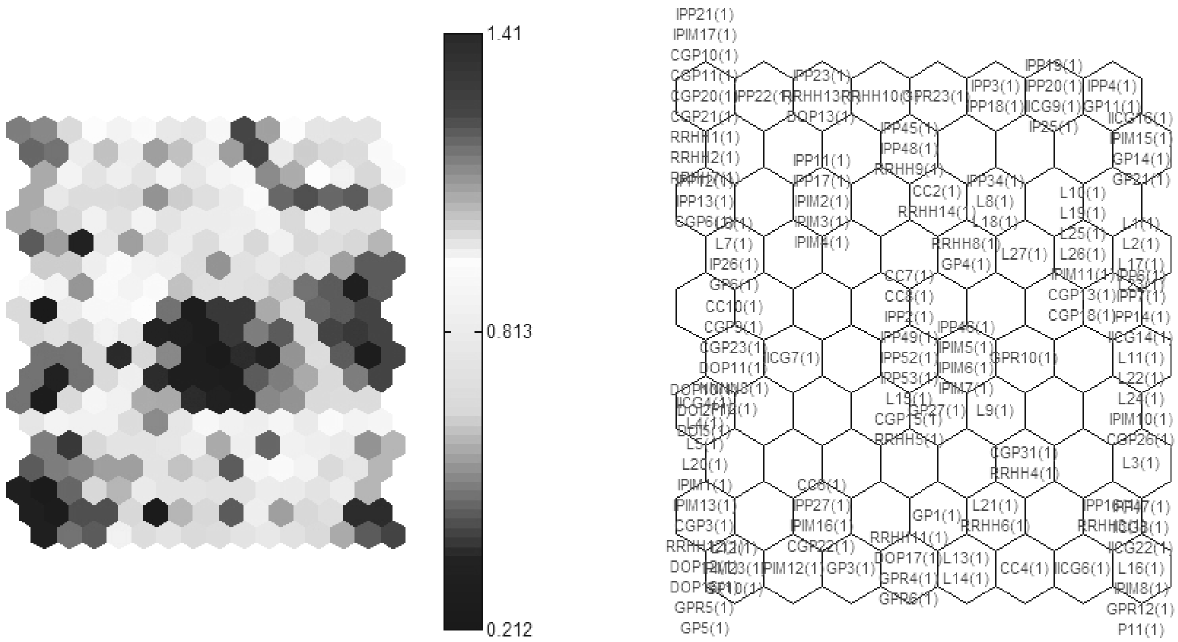


Fig. 9. MLHL factor pair 1-2 projection - Automotive Industry case study.

Fig. 10.a shows the U-matrix of the best projection generated by the SOM, which was unable to visualize any interesting internal structures.



(a) (b)

Fig. 10. SOM mapping - Automotive Industry case study: (a) U-matrix (b) lattice.

Having analyzed the CCA projection shown in Fig. 11, we may say that, in general terms, this mapping only takes account of the information on the importance of the knowledge, the urgency in its acquisition and the degree to which it is held by other departments within the company. The other variables (current level of knowledge and level of knowledge that may be needed) are not taken into account at all. Group 1 in Fig. 11 contains classes of knowledge of low importance and with a low urgency of acquisition. These features are related to classes of knowledge that are not crucial for the company and that are held at a medium level by other departments.

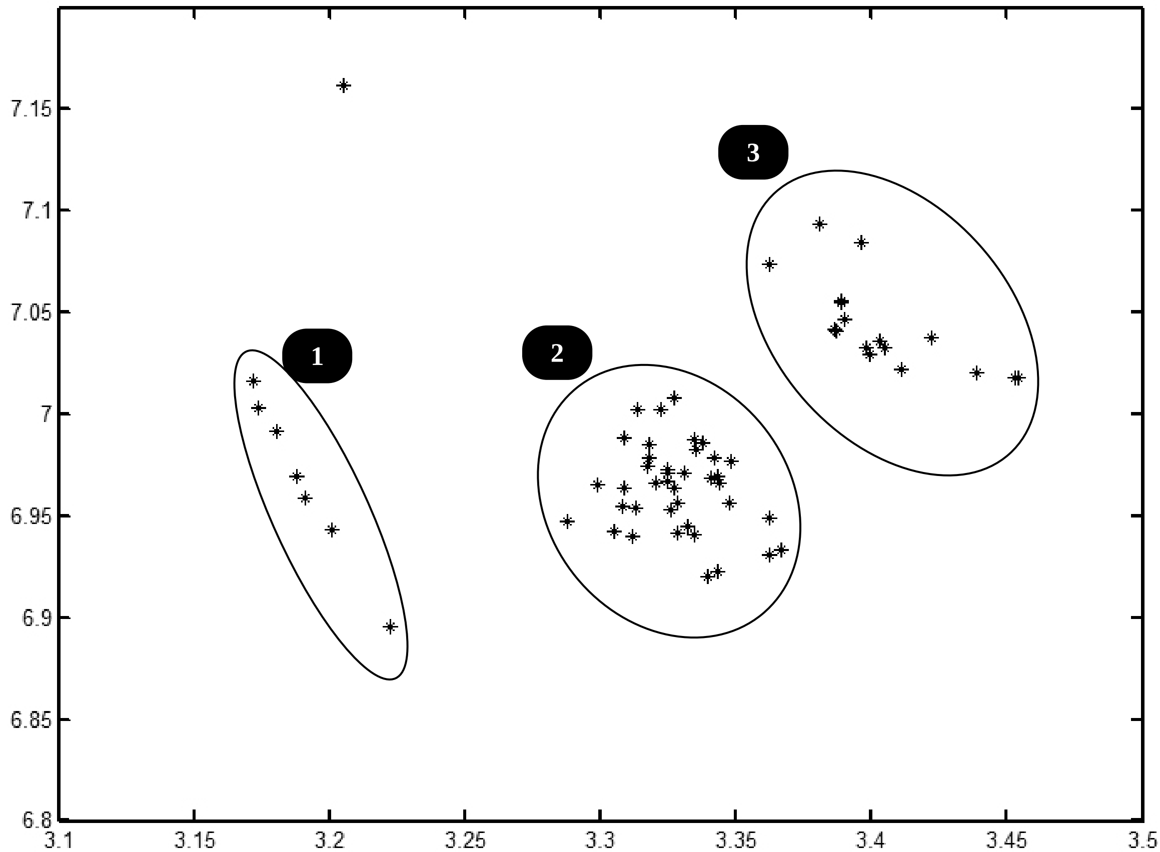


Fig. 11. CCA projection - Automotive Industry case study.

Group 2 (Fig. 11) shows classes of knowledge with a medium level of importance and urgency.

These are strategic classes of knowledge in the generation of competitive advantages. They are held at a medium level by other departments. Finally, those records contained in group 3 (Fig. 11) have the highest level of importance and urgency. There is no order within this group when taking the current level of these classes of knowledge into account. Taking these points into consideration, the diagram in Fig. 12 describes the projection shown in Fig. 11.

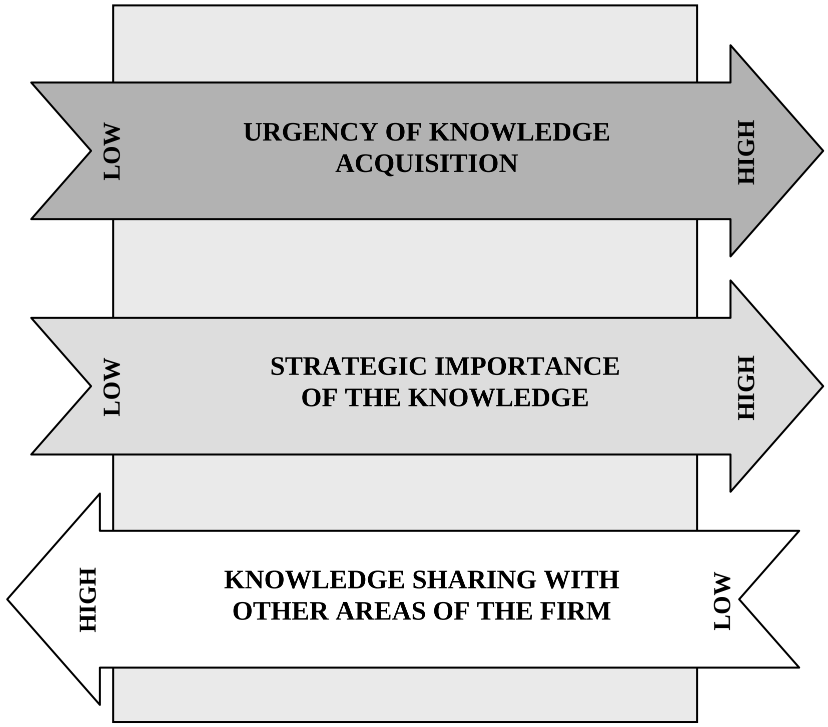


Fig. 12. Diagram of CCA projection (Fig. 11).

B. Second Case Study: Wall Painting

The second case study examined the knowledge situation of various companies from the wall painting sector in the Spanish autonomous region of Castilla y León. The data under analysis were taken from a survey of their staff. In this case study, the DIPKIP Data Acquisition step may be described as follows: a total of 68 records (interviews with workers) from 39 different companies were surveyed. The information contained in the 88-feature data set relates to 21 painting techniques (brush painting, spray varnishing, plaster or stucco work, etc...). For each one of these techniques, the survey measured the 4 following factors:

- Knowledge level held.
- Willingness to acquire new knowledge.
- Interest in updating the knowledge held.
- Interest in sharing the knowledge held.

The first of these factors takes a value that ranges from 2 (lowest level of knowledge) to 8 (highest level of knowledge), while the other 3 factors take values of either 0 or 1. In addition to these 84 technique-related features (4 features per 21 techniques), 4 further features (concerning general issues such as training and protection) were added to each record to form the 88-feature data set.

1) Visualization and Discussion

Fig. 13 shows the best CMLHL projection (factor pair 1-2), which allows us to identify 9 different groups (labeled as 1A, 1B, 1C, 2A, 2B, 2C, 3A, 3B and 3C). The final values introduced into CMLHL were: number of iterations = 8,000, learning rate = 0.0162, p parameter = 1.36, and τ parameter = 0.00802.

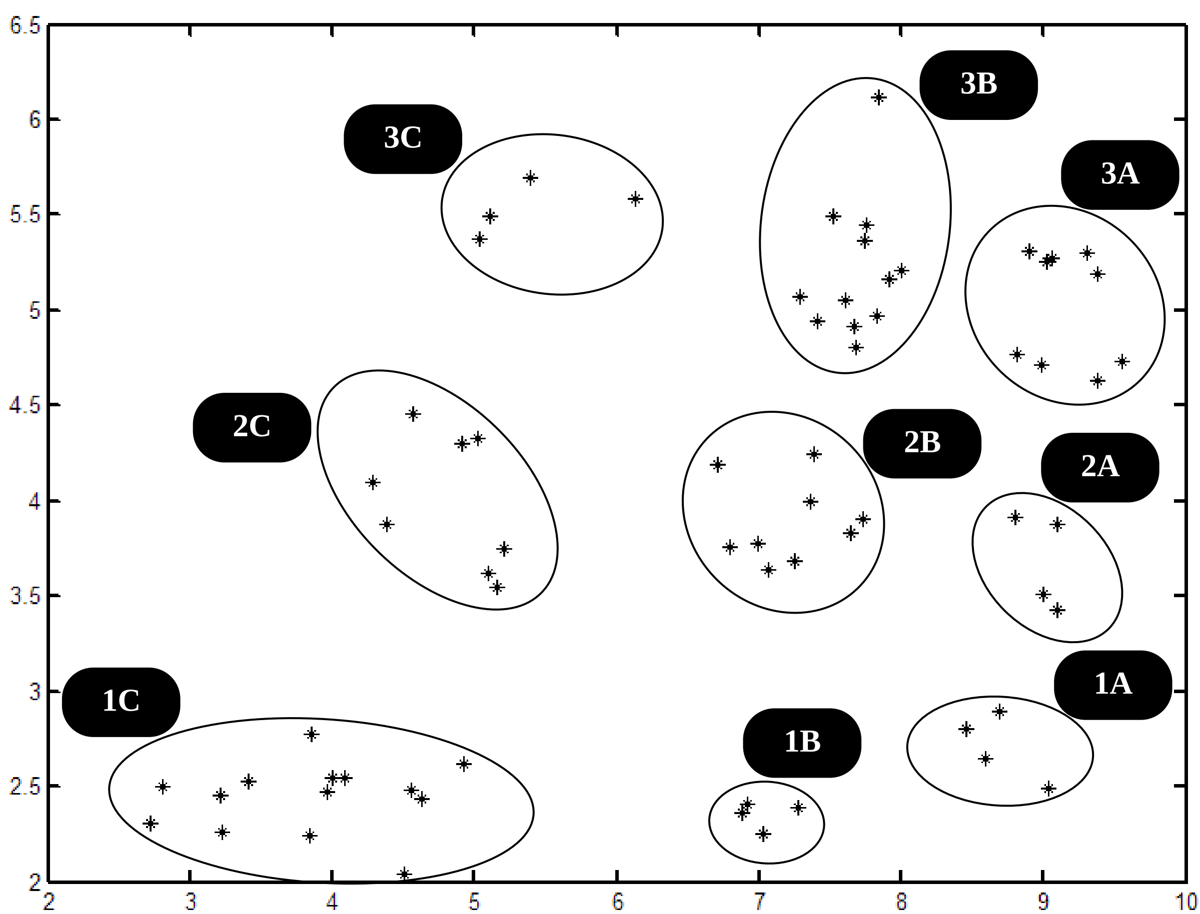


Fig. 13. CMLHL factor pair 1-2 projection - Wall Painting case study.

Third step of DIPKIP: Knowledge Identification.

Fig. 14 presents a diagram of Fig. 13 showing the classifications for this second case study.

An in-depth analysis (from a KM perspective) of the CMLHL projection of factor pair 1-2 (Fig. 13) led to the following conclusions:

- Group 3C: this group contains the best records (in general terms). All these records reflect the highest levels of knowledge among workers, the greatest interest in updating their knowledge and a willingness to share their knowledge with other colleagues. The convergence of these characteristics reflects responsible KM practices.
- Group 2C: the necessary skills for the job are known at a medium-high level, but this level is lower than the level held by group 3C. There is great interest in acquiring new knowledge within this group. These workers are aware of the importance of updating the knowledge. However, their interest in sharing the knowledge they hold with other companies is at a medium level.
- Group 1C: the level of knowledge held by workers is very low and some of the most important skills are not present. The same is true with respect to the updating of knowledge: it reaches the lowest levels or is completely absent. Nevertheless, as in previous groups, workers exhibit a high interest in knowledge updating.
- Group 3B: the only difference between this group and group 3C is that group 3B has a medium interest in dealing with these issues, while group 3C has the highest one.
- Group 2B: most of the workers within this group have medium levels of knowledge in all the related areas. Their lowest level relates to sharing the knowledge.
- Group 1B: this group is similar to group 1C, but its situation is even worse. The workers in this group are not really interested in sharing the low level of knowledge they hold.
- Group 3A: these records reflect good knowledge of painting skills. However, it does not appear that the workers consider it important to update their knowledge. The best feature of this group is a high interest in sharing the knowledge.
- Group 2A: a medium-high level of knowledge exists within this group, but its members express little interest in updating and acquiring it.
- Group 1A: the worst situations (from a KM point of view) are included in this group. This reflects the

lowest levels of knowledge and interest in acquiring and updating current knowledge.

These conclusions are summarized in Fig. 14. By moving from right to left, it may be seen how the level of interest in updating the knowledge increases. On the other hand, the “Knowledge level held” and the “Willingness to share the knowledge held” increase from the bottom to the upper-left-hand side of the graph.

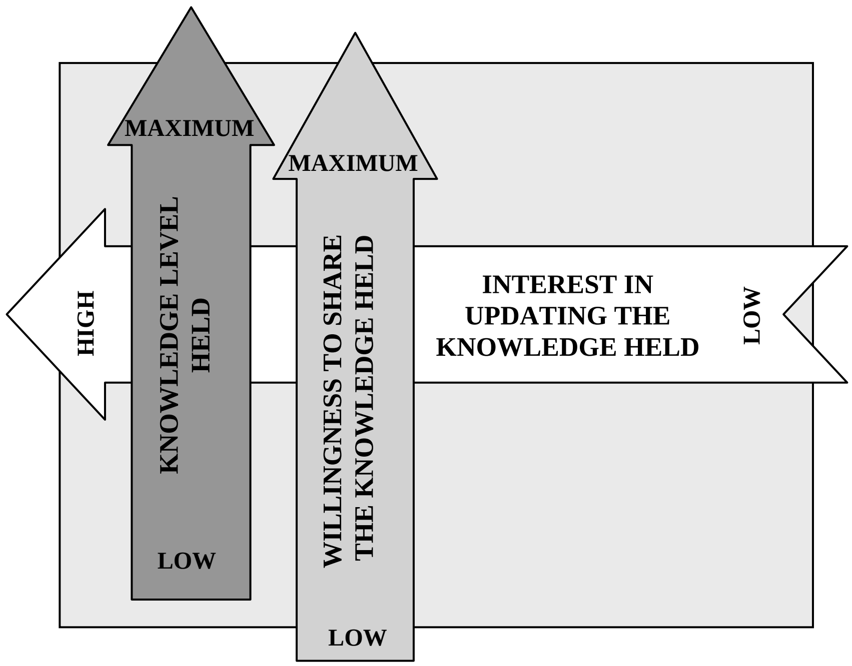


Fig. 14. Diagram of factor pair 1-2 (from Fig. 13).

An additional CMLHL factor pair projection was also investigated. The second-best projection is associated with factor pair 1-3 and is shown in Fig. 15. Having analyzed factor pair 1-3, the conclusion was that 3 main groups could be identified. The data is grouped (in a general way) according to the variables “Current level of knowledge” and “Interest in updating the knowledge held”. Group 3 contains the workers with the highest levels of the previously mentioned variables, while group 1 contains the companies with the lowest interest in updating the knowledge. It may be concluded that factor pair 1-3 maps the Wall Painting data set in a complementary way to the CMLHL projection of factor-pair 1-2.

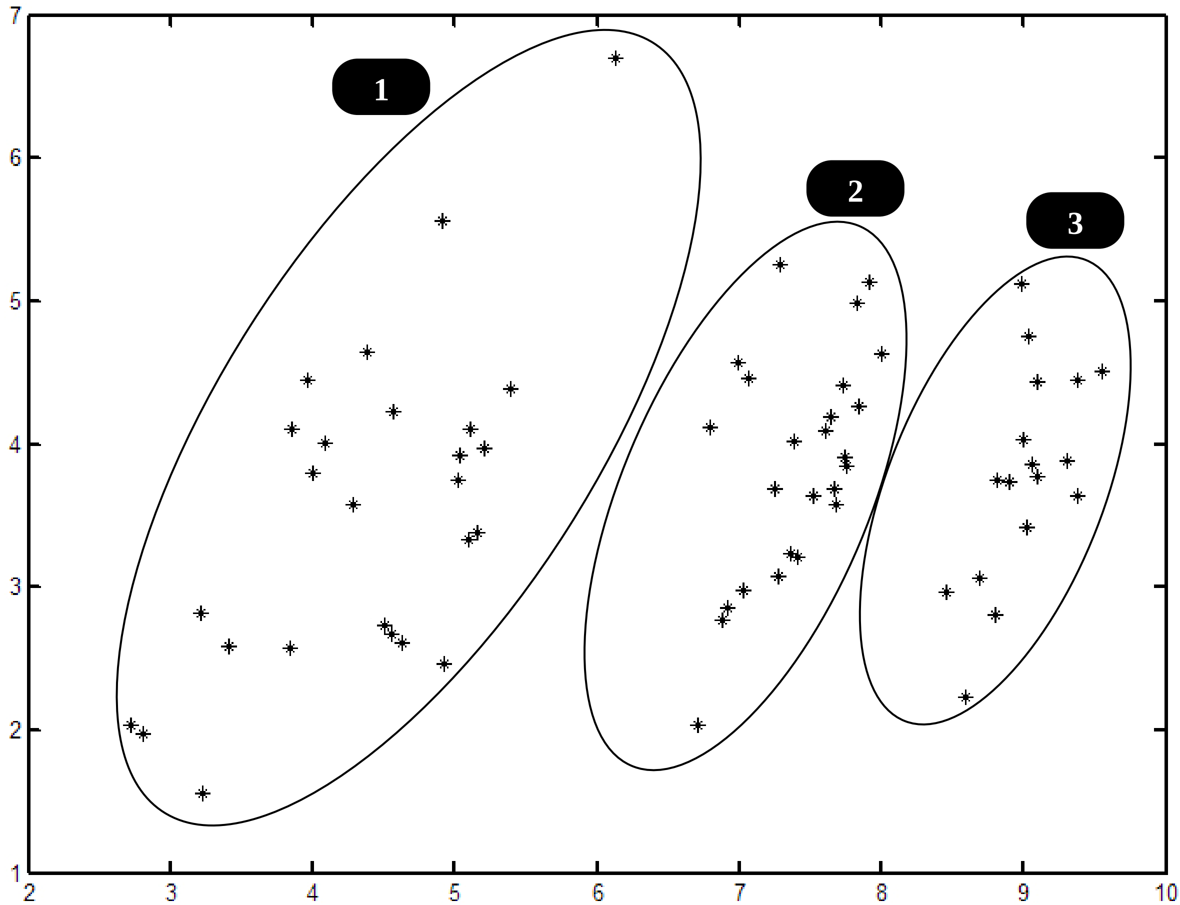


Fig. 15. CMLHL factor pair 1-2 projection - Wall Painting case study.

2) Step-by-step Sample

As in the previous case study, a step-by-step simulation of DIPKIP was conducted with two employees (R and Q) from different companies.

First Step: Data Acquisition

Once the detailed interviews were designed, responses were collected from the two employees. As previously mentioned in the case-study description, 88 questions were answered by each employee. Thus, it is not possible to present all the acquired data, although some information may be supplied. The answers from employees Q and R to the questions on spray varnishing (one of the 21 painting techniques) were as follows:

Question	Employee Q	Employee R
Knowledge level held	6	6
Willingness to acquire new knowledge	1	0
Interest in updating the knowledge held	0	0
Interest in sharing the knowledge held	0	0

Table 2. Data Acquisition Step for employees Q and R - Wall Painting case study.

Second Step: Intelligent Processing

After training the CMLHL model for this dataset, both employees were identified as belonging to Group 1A in Fig. 13. As previously mentioned, this group reflects the lowest levels of knowledge and interest in both acquiring and updating current knowledge.

Third Step: Knowledge Identification

These two situations were diagnosed as having a "Knowledge Deficit" that is borne out by their belonging to group 1A.

Fourth Step: Proposal

Having been identified as Knowledge Deficit situations, the proposal for these two departments is to acquire, create and generate the necessary knowledge. The outcomes of this step were specified in the following way:

- Employee Q: reduce the cost entailed in sharing knowledge (above all with regard to time and commitment) for the employee that holds it. Put mechanisms in place that will allow the employee to stay abreast of the latest developments in the profession.
- Employee R: create work spaces that are shared between various people.

As in the previous case study, the detailed proposals were applied in order to validate the model. These two employees were originally identified (Q_1 and R_1) as belonging to Group 1A. A second follow up evaluation took place after a period of 3 months, which had elapsed following implementation of the DIPKIP corrective actions. The new situations of these employees after following the DIPKIP proposals are depicted in Fig. 16

as Q_2 and R_2 respectively. As shown below, the situation of these employees is improved (from group 1A to groups 1C and 2A) by applying the outcomes of DIPKIP.

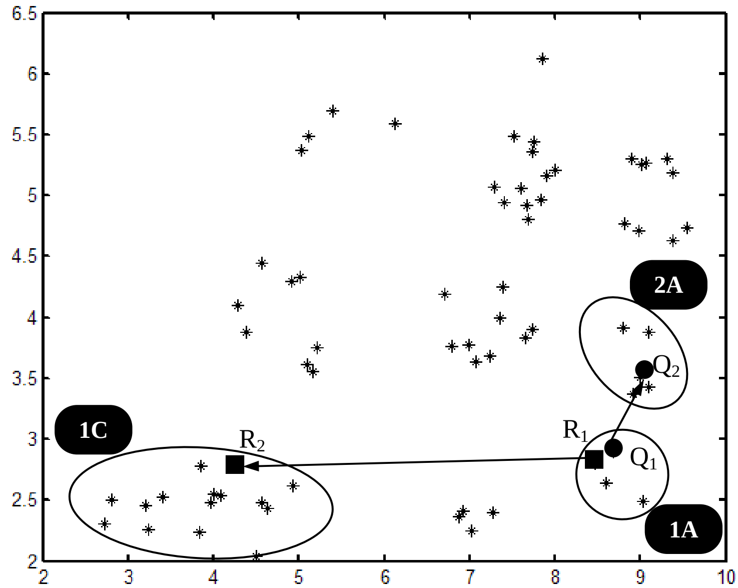


Fig. 16. CMLHL factor pair 1-2 projection - Wall Painting case study

Fourth DIPKIP step: corrective action.

Company R (labeled as R_1 in the original data set) is now included in group 1C as R_2. As in group 1A, the level of knowledge held by the workers in group 1C is very low. On the contrary, records in group 1C exhibit a high interest in knowledge updating, while group 1A represents the lowest levels of interest in acquiring new knowledge and updating existing ones.

Company Q (labeled as Q_1 in the original data set) is now found in group 2A as Q_2. The main difference between this group and the initial one (group 1A) relates to the level of knowledge held for different techniques.

3) Comparison with other Unsupervised Methods

As is evident in Fig. 17, the two first principal components were unable to identify the inner structure of the dataset very clearly. Fig. 18 depicts the results obtained by using MLHL and as can be seen, the mapping obtained for the Wall Painting data set is not as sparse as that obtained by CMLHL (Fig. 13).

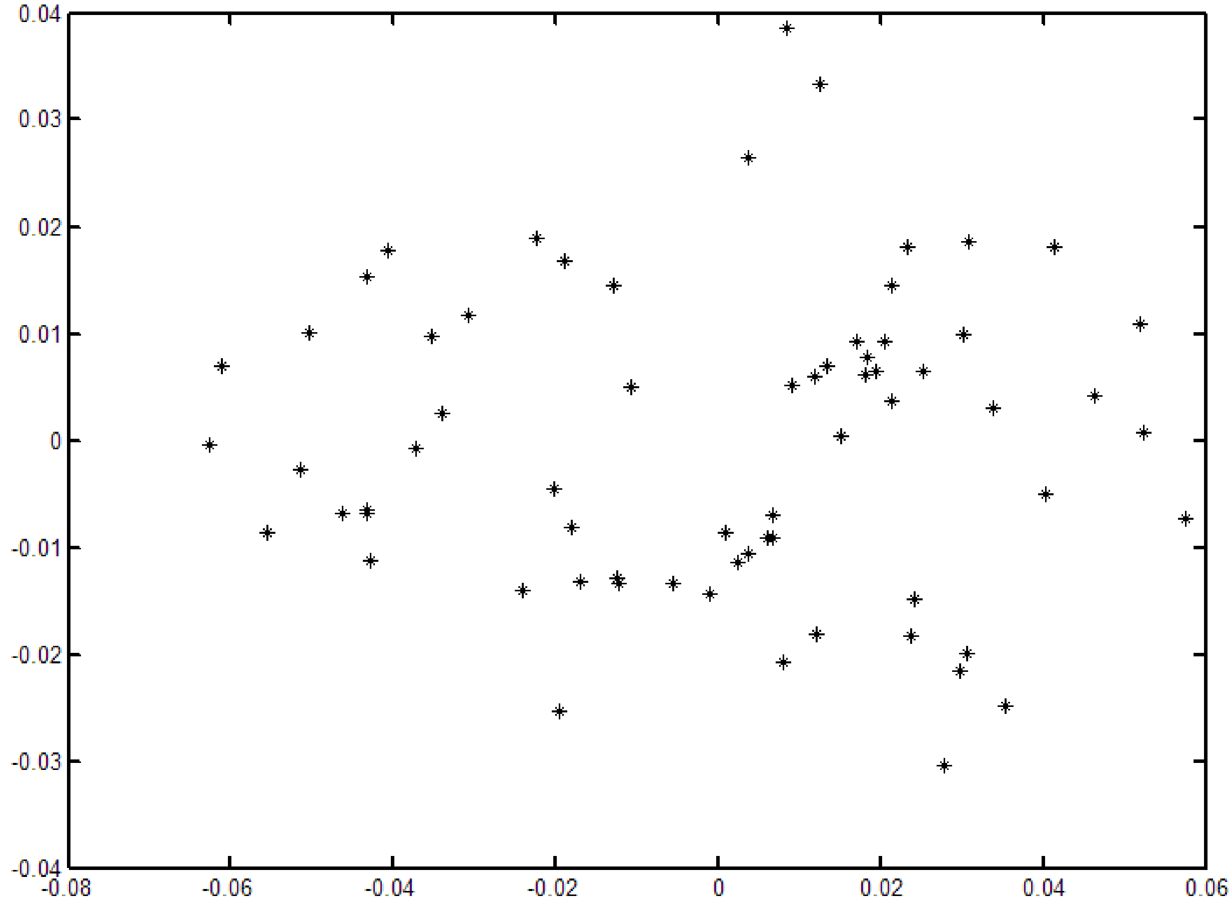


Fig. 17. PCA projection - Wall Painting case study.

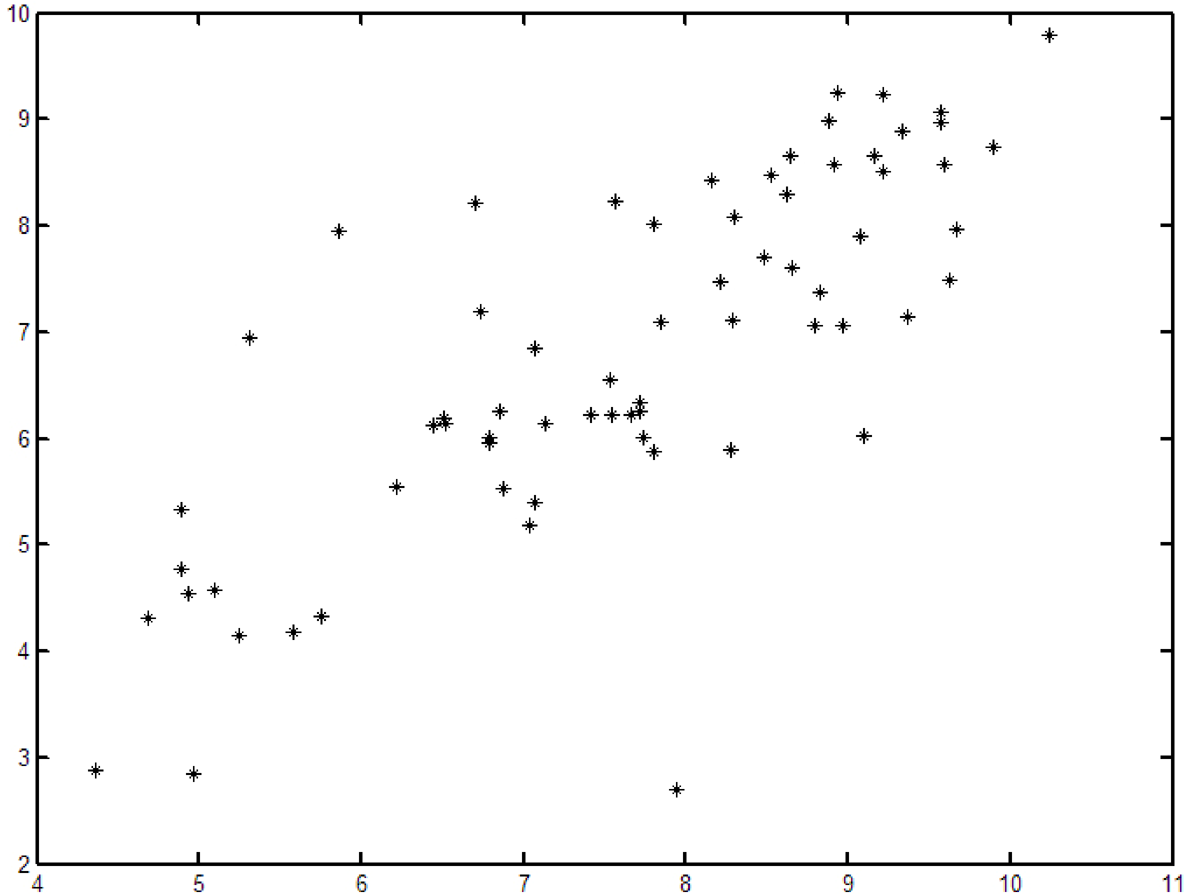
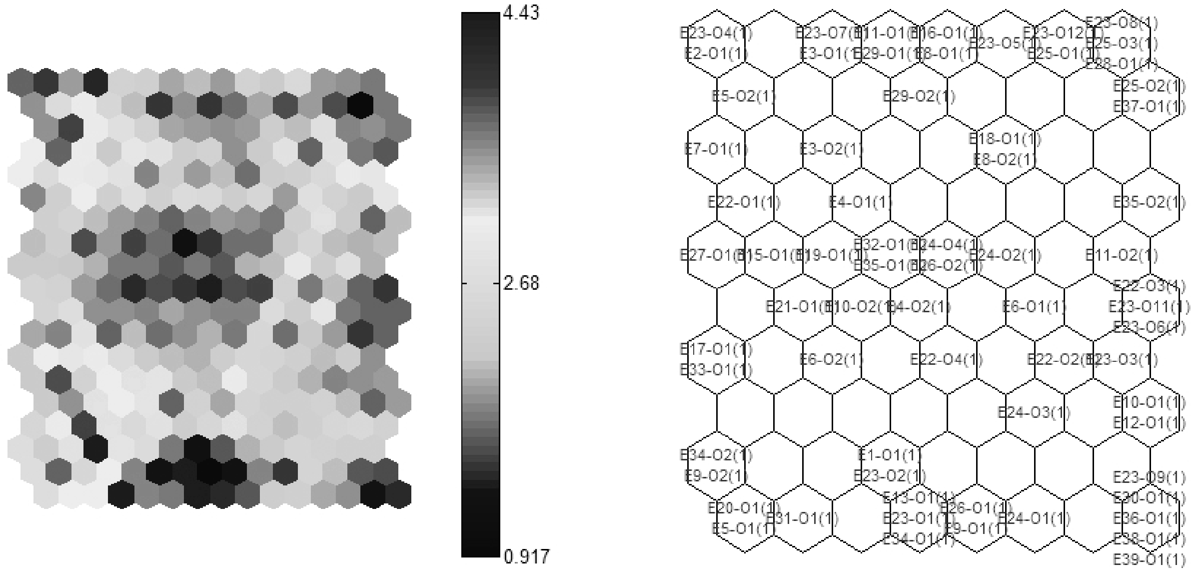


Fig. 18. MLHL factor pair 1-2 projection - Wall Painting case study.

In this data set, neither SOM (Fig. 19), nor CCA (Fig. 20) were able to identify interesting structures for the application of the third step of DIPKIP.



(a) (b)

Fig. 19. SOM mapping - Wall Painting case study: (a) U-matrix (b) lattice.

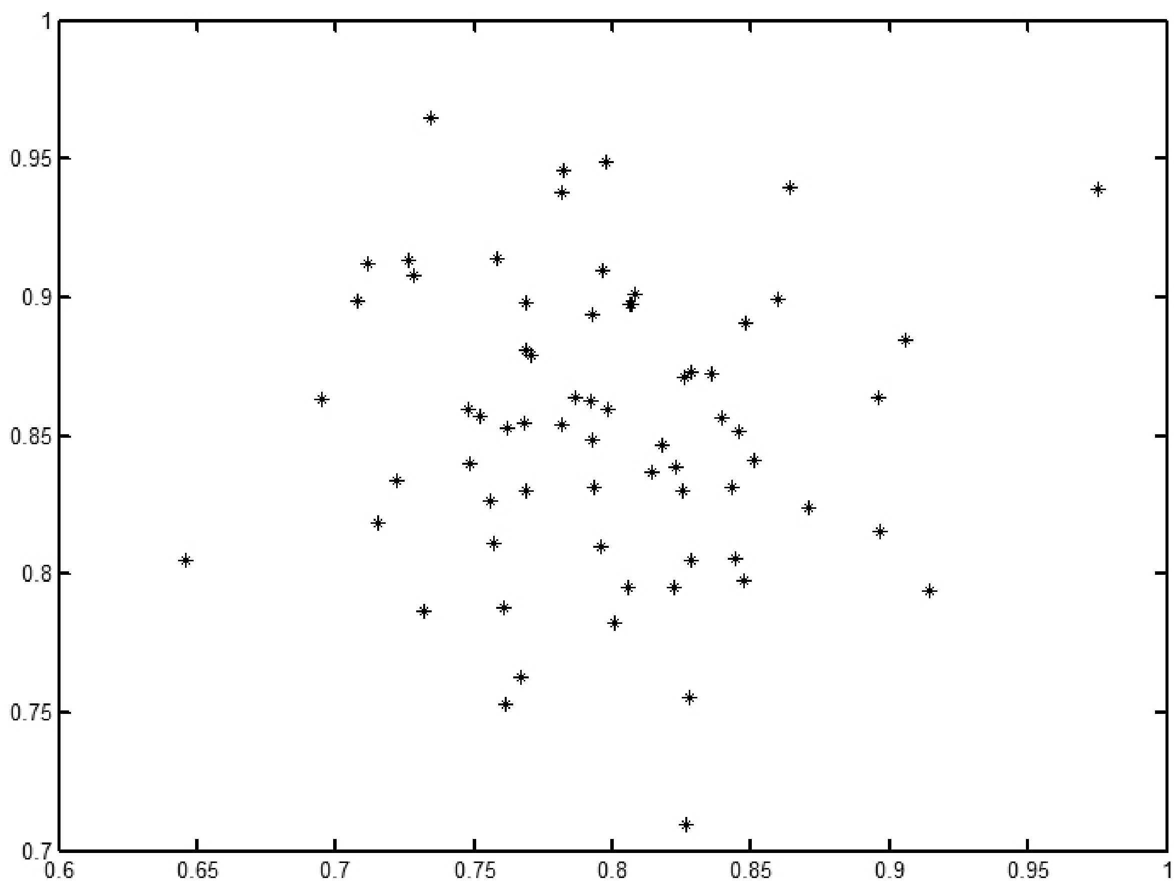


Fig. 20. CCA projection - Wall Painting case study.

VI. CONCLUSIONS AND FUTURE WORK

In this study, KM is understood as a system that integrates its specific functions and processes to create/acquire, transfer/distribute and put into practice/update the ideas and knowledge held by a company and its personnel. By doing so, KM, in its various forms, allows people to achieve greater levels of creativity, ensures permanent training and recycling in their specialist areas and helps them to share and pass on the benefits of their knowledge to other workers, who are also willing to integrate the knowledge held by their colleagues into their own work. Thus, DIPKIP (Data Acquisition, Intelligent Processing, Knowledge Identification and Proposal), a novel, neural KM system, has been applied in two different real-life case studies. This model has demonstrated its ability to analyze knowledge situations and to propose improvements from the standpoint of KM. The research has also analyzed the impact of the proposals, which were applied in the case studies to demonstrate the way in which DIPKIP improved the KM level.

In both cases, DIPKIP proved itself to be a robust tool for the analysis and identification of critical situations that enable companies to take decisions in the field of KM, concerning the acquisition, transfer and updating of the knowledge. The updating of knowledge processes enables a company to make use of the latest knowledge, to access the most recent innovations and to keep the KM system updated with all of its attendant benefits.

One of the main objectives propelling this study was the desire to introduce greater rigor and robustness into the field of KM research, thereby bridging the gap between theoretical formulations and satisfactory practical applications. The model ensures a complete diagnosis of the current situation, and subsequently an appropriate decision-making process for the effective application of the actions that it proposes in response to such situations.

The application of the neural projection architecture to the new KM system presented in this study not only enabled data to be grouped together and ordered, but it also produced results of great interest that may be used in decision-making. Among these results, the following may be highlighted.

The location of the knowledge is clearly identified according to the particular parameters used in the study, such as importance, level of urgency or diffusion of knowledge. This technique enables the KM situation in any company to be mapped out in a very short space of time and it is consequently of use in the related decision-making process. More specifically, the results verify:

- whether the knowledge or knowledge deficit of the company is critical to the implementation of its strategy.
- the level at which the knowledge is held or should be held.
- whether its acquisition or updating is urgent.
- whether such knowledge is shared between the people in the company who require it.

The data sets provide information on the type of knowledge held by a company or companies. Additionally, they can also justify a variety of actions, on a case by case basis, to situate knowledge in the best possible spaces, to move it between areas and to abandon knowledge that is no longer advisable or is ineffective.

In short, the KM system described in this research may be used to identify the knowledge held by a company in an easy and accurate manner and to map out actions for progress in the future.

Alternative methods such as PCA, MLHL, SOM or CCA were evaluated when performing the second step of DIPKIP (Intelligent Processing). CMLHL was shown to provide sparser projections and to capture some type of global ordering in the data sets.

In conclusion, we may say that a novel system for KM is proposed in this research, which responds to the need for information management and knowledge flows within a KM organization, through the incorporation in the KM system of knowledge updating processes.

Future work will be based on the study of different distributions and learning rules to improve the architecture as a whole, and it is anticipated that this neural architecture will eventually be embedded inside a more complex hybrid KM system.

An upgraded version of DIPKIP is also envisaged in the future. The model will be extended in order to improve its outcomes by cataloguing various classes of knowledge in a map of knowledge lines. Once the strategic lines of knowledge are identified, DIPKIP could then propose corrective actions concerning different classes of knowledge and different departments.

Acknowledgments.

This research has been partially funded through project BU006A08 of the JCyL.

References

Agndal, H., and U. Nilsson. 2006. Generation of Human and Structural Capital: Lessons from Knowledge Management. *Electronic Journal of Knowledge Management* 4 (2):91-98.

Ahlberg, Christopher, and Ben Shneiderman. 1994. Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays. Paper read at Proceedings of the SIGCHI conference on Human Factors in Computing Systems, at Boston, Massachusetts, United States.

Beckman, T. 1997. A Methodology for Knowledge Management. *Proceedings of the IASTED International Conference on AI and Soft Computing*.

Bertsekas, D. P. 1999. *Nonlinear Programming*: Athena Scientific Belmont, Mass.

Cetin, H., and D. Lewandowski. 1991. Interactive Classification and Mapping of Multi-dimensional Remotely Sensed Data using N-dimensional Probability Density Functions. *Photogrammetric Engineering and Remote Sensing* 57:1579-1587.

Collins, C. J., and Keng Smith. 2006. Knowledge Exchange and Combination: the Role of Human Resource Practices in the Performance of High-Technology Firms. *The Academy of Management Journal (AMJ)* 49 (3):544-560.

Corchado, E., J. M. Corchado, L. Saiz, and A. Lara. 2004. Constructing a Global and Integral Model of Business Management Using a CBR System. Paper read at CDVE 2004.

Corchado, E., and C. Fyfe. 2002. Maximum and Minimum Likelihood Hebbian Rules for Exploratory Projection Pursuit. In *ICONIP'02: Proc. of the 9th Int. Conf. on Neural Information Processing - Computational Intelligence for the E-Age*, edited by L. Wang, J. C. Rajapakse, K. Fukushima, S. Y. Lee and X. Yao.

———. 2003. Connectionist Techniques for the Identification and Suppression of Interfering Underlying Factors. *Int. Journal of Pattern Recognition and Artificial Intelligence* 17 (8):1447-1466.

Corchado, E., C. Fyfe, L. Saiz, and A. Lara. 2004. Development of a global and integral model of business management using an unsupervised model. Paper read at IDEAL 2004.

Corchado, E., Y. Han, and C. Fyfe. 2003. Structuring Global Responses of Local Filters Using Lateral Connections. *Journal of Experimental & Theoretical Artificial Intelligence* 15 (4):473-487.

Corchado, E., D. MacDonald, and C. Fyfe. 2004. Maximum and Minimum Likelihood Hebbian Learning for Exploratory Projection Pursuit. *Data Mining and Knowledge Discovery* 8 (3):203-225.

Corchado, Emilio, Juan Corchado, Lourdes Sáiz, and Ana Lara. 2005. A Beta-Cooperative CBR System for Constructing a Business Management Model. Paper read at Advances in Data Mining.

Chen, Liming, Nigel Richard Shadbolt, Feng Tao, Carole Goble, Colin Puleston, and Simon James Cox. 2005. Semantics-assisted Problem Solving on the Semantic Grid. *Computational Intelligence* 21 (2):157-176.

Davenport, Elisabeth, and Hazel Hall. 2002. Organizational Knowledge and Communities of Practice. *Annual Review of Information Science and Technology* 36 (1):170-227.

Demartines, P., and J. Herault. 1997. Curvilinear Component Analysis: A Self-Organizing Neural Network for Nonlinear Mapping of Data Sets. *IEEE Transactions on Neural Networks* 8 (1):148-154.

Diaconis, P., and D. Freedman. 1984. Asymptotics of Graphical Projection Pursuit. *The Annals of Statistics* 12 (3):793-815.

Donate-Manzanares, M., and F. Guadamillas-Gomez. 2007. The Relationship Between Innovation and Knowledge Strategies: its Impact on Business Performance. *International Journal of Knowledge Management Studies* 1 (3):388-422.

Dustdar, Schahram. 2002. Collaborative Knowledge Flow - Improving Process-Awareness and Traceability of Work Activities. Paper read at Proceedings of the 4th International Conference on Practical Aspects of Knowledge Management.

Edvinsson, Leif, and Michael Malone. 1997. *Intellectual Capital: Realizing Your Company's True Value by Finding Its Hidden Brainpower*: HarperBusiness.

Földiák, P. 1992. Models of Sensory Coding, PhD thesis, University of Cambridge.

Friedman, J. H., and J. W. Tukey. 1974. A Projection Pursuit Algorithm for Exploratory Data-Analysis. *IEEE Transactions on Computers* 23 (9):881-890.

Fyfe, C., R. Baddeley, and D.R. McGregor. 1994. Exploratory Projection Pursuit: an Artificial Neural Network Approach. In *Research Report/94/160, University of Strathclyde*.

Fyfe, Colin. 1997. A Neural Network for PCA and Beyond. *Neural Processing Letters* 6 (1-2):33-41.

Grevesen, C. W., and F. Damanpour. 2007. Performance Implications of Organisational Structure and Knowledge Sharing in Multinational R&D Networks. *International Journal of Technology Management* 38 (1):113-136.

Heinrichs, John H., and Jeen-Su Lim. 2005. Model for Organizational Knowledge Creation and Strategic Use of Information: Research Articles. *Journal of the American Society for Information Science and Technology* 56 (6):620-629.

Holsapple, C. W, and K.D. Joshi. 2002. Knowledge Management: A Three-Fold Framework. *The Information Society* 18 (1):47-64.

Hotelling, H. 1933. Analysis of a Complex of Statistical Variables Into Principal Components. *Journal of Education Psychology* 24:417-444.

Howell, E. S., E. Merényi, and L. A. Lebofsky. 1994. Using Neural Networks to Classify Asteroid Spectra. *Journal of Geophysical Research* 99 (E5):10847-10865.

Hyperwave. *Knowledge Management Software* [cited 10/16/2008]. Available from http://www.hyperwave.com.

Hyvarinen, A. 2001. Complexity pursuit: Separating interesting components from time series. *Neural Computation* 13 (4):883-898.

Hyvärinen, A. 1998. New Approximations of Differential Entropy for Independent Component Analysis and Projection Pursuit. Paper read at Proceedings of the 1997 Conference on Advances in Neural Information Processing Systems.

IBM. *Lotus, Knowledge Management Software* [cited 10/16/2008]. Available from http://www-01.ibm.com/software/lotus/.

Kautz, K., and A. Kjærgaard. 2007. Towards an Integrated Model of Knowledge Sharing in Software Development: Insights from a Case Study. *International Journal of Knowledge Management* 3 (2):91-117.

Kohonen, T. 1990. The Self-Organizing Map. *Proceedings of the IEEE* 78 (9):1464-1480.

Maier, Ronald, and Ulrich Remus. 2002. Defining Process-oriented Knowledge Management Strategies. *Knowledge and Process Management* 9 (2):103-118.

Marguardt, M. J. 1996. *Building the Learning Organization: A Systems Approach to Quantum Improvement and Global Success*: McGraw-Hill.

Maurer, H., and K. Tochtermann. 2002. On a New Powerful Model for Knowledge Management and its Applications. *Journal of Universal Computer Science* 8 (1):85-96.

Nielsen, Bo Bernhard, and Snejina Michailova. 2007. Knowledge Management Systems in Multinational Corporations: Typology and Transitional Dynamics. *Long Range Planning* 40 (3):314-340.

Nonaka, I., and R. Toyama. 2005. The Theory of the Knowledge-creating Firm: Subjectivity, Objectivity and Synthesis. *Industrial and Corporate Change* 14 (3):419-436.

Oja, Erkki. 1989. Neural networks, principal components, and subspaces. *Int. Journal of Neural Systems* 1:61-68.

OPENTEXT. *Livelink, Knowledge Management Software* [cited 10/16/2008]. Available from http://www.opentext.com.

Pearson, K. 1901. On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine* 2 (6):559-572.

Prescott, E. C., and M. Visscher. 1980. Organization Capital. *The Journal of Political Economy* 88:1–13.

Rollett, H. 2003. *Knowledge Management: Processes and Technologies*: Kluwer Academic Publishers.

Ruggles, R. L. 1997. *Knowledge Management Tools*: Butterworth-Heinemann.

Sanger, D. 1989. Contribution Analysis: a Technique for Assigning Responsibilities to Hidden Units in Connectionist Networks. *Connection Science* 1 (2):115-138.

Senge, Peter, C. Otto Scharmer, Joseph Jaworski, and Betty Sue Flowers. 2004. *Presence: Human Purpose and the Field of the Future*: Cambridge, Mass.: SoL, Society for Organizational Learning, .

Seung, H. S., N. D. Socci, and D. Lee. 1998. The Rectified Gaussian Distribution. *Advances in Neural Information Processing Systems* 10:350-356.

Shu-Mei, Tseng. 2008. The Effects of Information Technology on Knowledge Management Systems. *Expert Systems with Applications: An International Journal* 35 (1-2):150-160.

Sivan, Y.Y. 2000. Nine Keys to a Knowledge Infrastructure: A Proposed Analytic Framework for Organizational Knowledge Management. Paper read at Proceedings of WebNet 2000 - World Conference on the WWW and Internet.

Soo, C. W., T. M. Devinney, and D. F. Midgley. 2007. External Knowledge Acquisition, Creativity and Learning in Organisational Problem Solving. *International Journal of Technology Management* 38 (1):137-159.

Strohmaier, M. 2003. A Business Process Oriented Approach for the Identification and Support of Organizational Knowledge Processes. *Proceedings of 4. Oldenburger Fachtagung Wissensmanagement, Potenziale-Konzepte-Werkzeuge*.

Taylor, H. 2007. Tacit Knowledge: Conceptualizations and Operationalizations. *InternatIonal Journal of Knowledge Management* 3 (3):60-73.

Wang, Z. 2004. Knowledge Systems Engineering: A New Discipline of Knowledge Management and Enabling. *International Journal of Knowledge and Systems Sciences* 1 (1):9-16.

Wei-Wen, Wu. 2008. Choosing Knowledge Management Strategies by Using a Combined ANP and DEMATEL Approach. *Expert Syst. Appl.* 35 (3):828-835.

Weissor, Shiue, Li Sheng-Tun, and Chen Kuan-Ju. 2008. A Frame Knowledge System for Managing Financial Decision Knowledge. *Expert Systems with Applications: An International Journal* 35 (3):1068-1079.

Wiig, K. M. 1994. *Knowledge Management Foundations: Thinking about Thinking-how People and Organizations Represent, Create, and Use Knowledge*: Schema Press, Limited.

Woitsch, Robert, and Dimitris Karagiannis. 2002. Process-oriented knowledge management systems based on KM-services: the PROMOTE approach. *International Journal of Intelligent Systems in Accounting, Finance & Management* 11 (4):253-267.

Wu, Liang-Chuan, Chorng-Shyong Ong, and Yao-Wen Hsu. 2008. Knowledge-based Organization Evaluation. *Decision Support Systems* 45 (3):541-549.