A Hybrid Solution for Advice in the Knowledge Management Field

Álvaro Herrero[1], Aitor Mata[2], Emilio Corchado[1], and Lourdes Sáiz[1]

[1] Department of Civil Engineering, University of Burgos
C/ Francisco de Vitoria s/n, 09006 Burgos, Spain
{ahcosio,escorchado,lsaiz}@ubu.es
[2] Department of Computing Science and Automatic, University of Salamanca
Plaza de la Merced, s/n, 37008 Salamanca, Spain
aitor@usal.es

Abstract. This paper presents a hybrid artificial intelligent solution that helps to automatically generate proposals, aimed at improving the internal states of organization units from a Knowledge Management (KM) point of view. This solution is based on the combination of the Case-Based Reasoning (CBR) and connectionist paradigms. The required outcome consists of customized solutions for different areas of expertise related to the organization units, once a lack of knowledge in any of those has been identified. On the other hand, the system is fed with KM data collected at the organization and unit contexts. This solution has been integrated in a KM system that additionally profiles the KM status of the whole organization.

1 Introduction

Knowledge Management (KM) enables organizations to capture, share, and apply the collective experience and know-how (knowledge) of their staff. Ever-growing volumes of data 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 [1]. These knowledge infrastructures consist of three main dimensions: people, organizational and technological systems.

In recent years, the deployment of information technology has become a crucial tool for enterprises to achieve a competitive advantage and organizational innovation [2]. 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.

This paper proposes the application of the Case-Based Reasoning (CBR) [3] and connectionist paradigms to profile the KM status of an enterprise and then automatically generate improvement proposals. The underlying idea is to produce specific and customized suggestions without human intervention to improve the KM status of the analyzed organization. The inputs of this hybrid advising solution are KM data gathered from the analyzed organization by surveys [4]. To process such data, a connectionist projection model (See Section 2) is used.

The paper is structured in the following way. Section 2 introduces the unsupervised neural projection model applied in this work, while section 3 describes the CBR paradigm. Section 4 presents the proposed hybrid solution applying the two previously introduced AI paradigms. Section 5 describes the application of the proposed solution to a KM system. Finally, Section 6 presents the conclusions and some proposals for future work in the same field.

2   Connectionist Projection Model

The identification of patterns that exist across dimensional boundaries in high dimensional data sets is a challenging task. Such patterns may become visible if changes are made to the spatial coordinates. Projection models perform such change by projecting high-dimensional data onto a lower dimensional space in order to identify "interesting" directions in terms of any specific index or projection. Such indexes or projections are, for example, based on the identification of directions that account for the largest variance of a data set –as is the case of Principal Component Analysis (PCA) [5], [6], [7] or the identification of higher order statistics such as the skew or kurtosis index -as is the case of Exploratory Projection Pursuit (EPP) [8]. Having identified the most interesting projections, the data is then projected onto a lower dimensional subspace plotted in two or three dimensions, which makes it possible to examine its structure with the naked eye.

The combination of this technique together with the use of scatter plot matrixes constitutes a very useful visualization tool to investigate the intrinsic structure of multidimensional data sets, allowing experts to study the relations between different components, factors or projections, depending on the applied technique.

The solution proposed in this paper applies an unsupervised neural model called Cooperative Maximum Likelihood Hebbian Learning (CMLHL) [9], [10]. It is based on Maximum Likelihood Hebbian Learning (MLHL) [9], [11]. Considering 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), then CMLHL can be expressed as:

1. Feed-forward step:

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

(1)

2. Lateral activation passing:

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

(2)

3. Feedback step:

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

(3)

4. Weight change:

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

(4)

Where: η is the learning rate, τ is the "strength" of the lateral connections, b the bias parameter, p a parameter related to the energy function [9], [10], [11] and A a symmetric matrix used to modify the response to the data [10]. The effect of this matrix is based on the relation between the distances separating the output neurons.

3   CBR Paradigm

Case-Based Reasoning [12] origins are in knowledge based systems. CBR systems solve new problems acquiring the needed knowledge from previous situations [13]. The main element of a CBR system is the case base, a structure that stores the information used to generate new solutions.
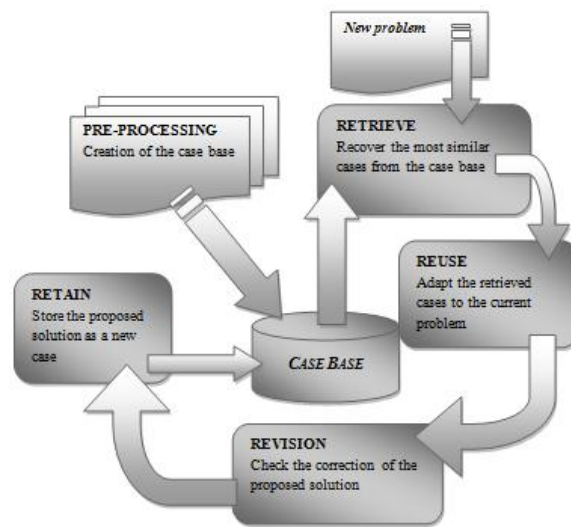

Fig. 1. Basic structure of a CBR system, including the four main stages of the CBR cycle

The learning capabilities of CBR systems are due to its own structure, composed of four main stages [14]: *retrieve*, *reuse*, *revision* and *retain*. These stages are depicted in Fig. 1. The first stage is called *retrieve*, and consists in finding the cases (from the case base) that are most similar to the new problem. Once a set of cases is extracted from the case base, they are *reused* by the system. In this second stage (*retain*), the selected cases are adapted to fit in the new problem. After applying the new solution to the problem, that solution is *revised* to check its performance. If it is an acceptable solution, then it is *retained* by the system and could eventually serve as a solution to future problems.

As a methodology [12], CBR has been used to solve a great variety of problems. It is a cognitive structure that can be easily applied to solve problems such as those related with soft computing, since the procedures used by CBR are quite easy to

assimilate by soft computing approaches. CBR has also helped to create applications related to quite different environments, such as health sciences [15], eLearning [16], planning [17], Intrusion Detection Systems [18] and oceanographic matters [19]. Applying CBR to solve a problem usually requires other AI techniques. It is not only a simple way of structuring the reutilization of information, but also a model that can combine different techniques to improve their individual results. Different kinds of neural networks such as ART-Kohonen [20] or Growing Cell Structures [21] have been combined with CBR to automatically create the inner structure of the case base. Some effort has also been devoted to the case-based maintenance issue [22].

4   A Hybrid Advising Solution

As it is previously stated, data processing and subsequent conclusion extraction are challenging tasks in the KM field. This paper proposes a hybrid artificial intelligent solution aimed at supporting KM managers by advice. This solution operates in two different steps (See Fig. 2 below):

- KM Profiling: KM data can be available in some enterprises or organizations. To know the KM status of an organization (company, department within a company, etc) from such data, we propose a neural processing phase. The CMLHL model in unison with a KM expert is able to profile the KM status of the different elements (specific knowledge of departments, working groups, employees, etc.) within the analyzed organization (See Section 4.1 for further details).
- Proposal Generation: once the KM status of those elements is known, coherent proposals must be generated for the worst cases. That is, the elements whose situation is critical or alarming (from a KM standpoint) must take corrective action. That action is proposed by a CBR system (See Section 4.2 for further details).

The proposed hybrid artificial system facilitates the organization under analysis to access the knowledge it requires (at the right time) to develop its activities in a satisfactory way.
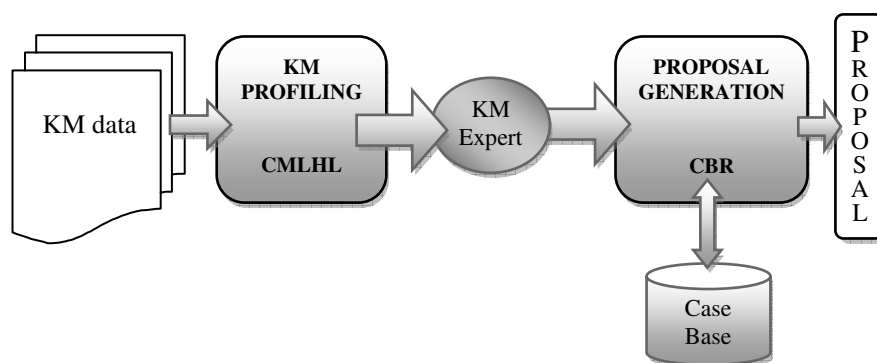

Fig. 2. Schema of the hybrid solution

4.1   KM Profiling

High-volume KM data can be gathered in an organization. This does not guarantee that an in-depth profitable analysis could be subsequently performed. To do so, the application of the CMLHL model is proposed in this paper.

In order to show the KM status of the different units (departments, working groups, employees, etc.) within an organization, some data must be collected. KM experts must choose and define the procedures and methodologies to do it according to the context. Some different techniques can be applied: interviews, surveys, database mining, and so on.

Once these data are captured, CMLHL is applied to visualize the KM status of the different units. CMLHL, as a topology preserving mapping due to the use of lateral connections, clusters projected similar samples of the dataset. The projection is then interpreted by the KM expert to generate a KM profile matrix. According to this matrix, every single department is characterized by 2 features:

- *Temporal restrictions*: urgency of improving the situation.
- *Required knowledge*: amount of knowledge needed to get to a proper situation.

These two features are included in the case definition of the Proposal Generation as described below.

4.2   Proposal Generation

The case base stores information about the previously solved problems together with the decision taken by the expert (supervised phase) or by the system (unsupervised phase) to solve a new problem. This is the main idea underlying the CBR methodology: it is possible to reuse past information in order to solve new problems.

The proposed CBR system is trained until the case base grows up to a big and valid enough state. Then it will generate customized solutions without the intervention of the KM expert. Information about problems and related solutions is stored in the case base that follows the structure described in Table 1. The first eleven parameters refer to the problem to be solved, including data related to the analyzed organization. The other two parameters defining a case (KM Profiling Features) are obtained in the first step of this solution (KM Profiling). All these thirteen features were selected to define cases as the proposal to be generated strongly depends on them. The values of these parameters compose a vector that determines the location of the cases in the case base. When a new case is introduced in the case base, a vector is assigned to that new case. Values of vectors will be similar if the cases they represent are similar. Additionally, these vectors will determine the cases to be recovered from the case base in order to reuse them for a solution. When a new problem comes to the system, those cases that are most similar to the new problem are recovered from the case base. This similarity is calculated by the vector identifying the new problem and those associated to stored cases.

The final solution (proposal) to the problem must be in the set of solutions shown in Table 2. There may be one or several solutions for a certain problem, depending on the parameters of the organization under analysis. As the case base is enlarged by introducing the decisions taken by the expert, the CBR system generates its own automatic decisions. To know when the human expert is not needed, the solution

(set of proposals) given by the expert is compared with the eventual solution given by the CBR system from the data available until the moment. When the number of cases in which there is any difference between them is under a certain threshold (never greater than 10% of situations; i.e. the system chooses the right solution in 90% of the cases at least), human supervision is not required. Then, the expert may only be consulted in special situations. If the system is not able to produce a proper solution and it is necessary to check the human expert opinion, the expert will be offered a set of available solutions (those collected in Table 2), if none of them fits the problem, then the expert will introduce in the system a proper solution.

Table 1. Description of the parameters defining a case. The table shows the different parameters (organized in three different levels: Initial Parameters, KM Profiling Features, Solution) and their possible values.

Level	Parameter	Possible Values
Initial Parameters	General Environment – Stability	Stable (1), dynamic (0).
Initial Parameters	General Environment – Complexity	Simple (1), complex (0).
Initial Parameters	General Environment – Hostility	Favourable (1), hostile (0).
Initial Parameters	General Environment – Diversity	Integrated (1), diverse (0).
Initial Parameters	Economic sector	Emerging (4), growing (3), mature (2), declining (1).
Initial Parameters	Developing methods	Internal (1), external (2), cooperation (3), internalization (4).
Initial Parameters	Organizational structure	Simple (1), functional (2), divisional (3), matrix structure (4).
Initial Parameters	Number of employees	Number of employees of the analysed organization.
Initial Parameters	Employees average age	Average age of all the employees.
Initial Parameters	Type of work	Individual (1), collaborative (2).
Initial Parameters	Scope of application	Local (1), regional (2), national (3), continental (4), international (5).
KM Profiling Features	Temporal restrictions	A lot of urgency (3), during this year (2), later (1).
KM Profiling Features	Level of required knowledge	Wide (3), medium (2) or basic (1).
Solution	Proposal	One or more of the solutions in *Table 2*.

Table 2 shows a predefined sample set of possible solutions that may be applied to an organization after being analyzed by the system. The solutions range from easy internal collaboration solutions (first one) to intensive abstract working (last one). The system will be adapted to the specific characteristics of the analyzed organizations. To do so, the expert decisions will increase this set by proposing new solutions (when the existing ones are not appropriate enough). Already generated solutions can be used in the future to different organizations in new proposals, even if they have never been used in such a knowledge field.

Table 2. Solutions that may be proposed by the system

Solution	Description
1	Experimented employees should collaborate with the new ones.
2	Staff should be trained in their specialized knowledge in deep.
3	Staff should swap their responsibilities from time to time.
4	Experts in a certain area should be employed.
5	Collaborative work should be done.
6	Collaborate with other enterprises in the same field.
7	Take into account suggestions and opinions of clients and providers.
8	Temporally hire an expert in some specific area.
9	Encourage the proposition of new ideas and solutions.
10	Enrol in external knowledge communities.
11	Report detected successes, fails and mistakes.
12	Identify the external and internal knowledge networks.
13	Identify the knowledge experts.
14	Evaluate the learning time in a certain field.
15	Estimate the probability of losing an expert.
16	Estimate the needed time to transfer the knowledge from an expert.
17	Describe the potential uses of the available knowledge.
18	Define internal knowledge communities.

4.3 A Real-Life Case Study

The proposed hybrid solution was applied to a real-life case study: companies from the wall painting sector in the Spanish autonomous region of Castilla y León [4]. The data selected for the KM profiling were taken from a staff survey. A total of 68 records from 39 different companies were generated. 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: taking values from 2 (lowest level of knowledge) to 8 (highest level of knowledge).
- Willingness to acquire new knowledge: binary value.
- Interest in updating the knowledge held: binary value.
- Interest in sharing the knowledge held: binary value.

As an example, the application of the proposed solution to two employees (E_1 and E_2) is described in the following paragraphs.

4.3.1 A KM Profiling Example

As previously stated, 88 questions were answered by employees E_1 and E_2. Thus, it is not possible to present all the acquired data, although some information may be supplied. The answers from the questions on spray varnishing (one of the 21 painting techniques) were as shown in Table 3.

Table 3. Sample data for the KM Profiling step. The table shows some questions included in the survey and the answers from the two sample employees.

	Value of the answer	
Question	Employee E_1	Employee E_2
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

The 88-feature data were projected by means of CMLHL. The obtained projection is shown in Fig. 3.

Cluster 1:
later, basic level
Cluster 2:
later, medium level
Cluster 3:
later, wide level
Cluster 4:
during this year, basic level
Cluster 5:
during this year, medium level
Cluster 6:
during this year, wide level
Cluster 7:
a lot of urgency, basic level
Cluster 8:
a lot of urgency, medium level
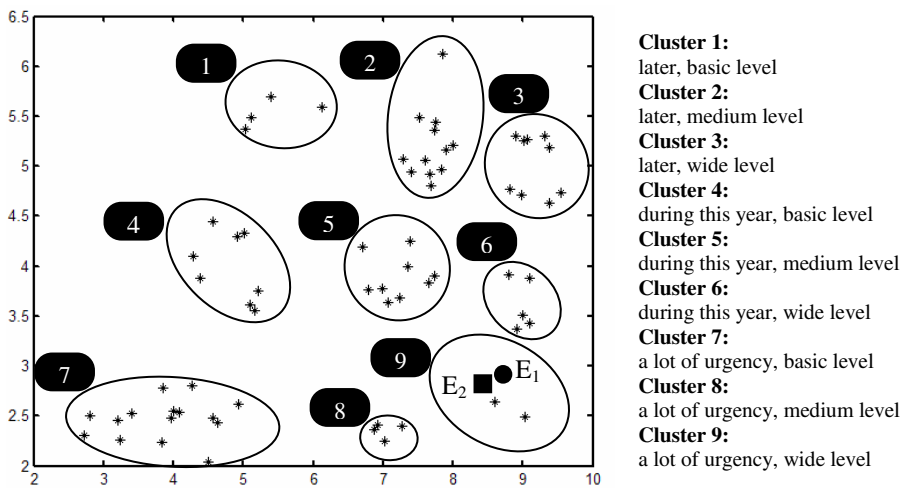**Cluster 9:**
a lot of urgency, wide level

Fig. 3. a) CMLHL factor pair 1-2 projection of the wall-painting case study. b) KM profiling values assigned (temporal restrictions and required knowledge) by the KM expert to each cluster.

Employees E_1 and E_2 were diagnosed as having a "Knowledge Deficit" that is borne out by their belonging to one of the worst clusters. The following values were given to the KM profiling features:
- Temporal restrictions: a lot of urgency (3).
- Required knowledge: wide level (3).

For the sake of brevity, the rest of this section is only describing the E_1 case.

4.3.2 A Proposal Generation Example
Once the KM profiling features of E_1 were known, values were given to the other case parameters as shown in Table 4. These values allow the system to recover (from the case base) those cases that have similar values in their parameters. There is at least one proposal associated to any of the cases recovered. The system measures the

Table 4. E_1 case definition

Parameter	Discrete Values of Parameters
General Environment – Stability	Dynamic (0)
General Environment – Complexity	Simple (1)
General Environment – Hostility	Favourable (1)
General Environment – Diversity	Diverse (0)
Economic sector	Mature (2).
Developing methods	Internal (1).
Organizational structure	Functional (2).
Number of employees	6
Employees average age	49
Type of work	Collaborative (2)
Scope of application	Regional (2)
Temporal restrictions	A lot of urgency (3)
Required knowledge	Wide (3)

similarity between the values of the recovered cases and the analyzed values of E_1 and consider those most similar and their solutions. If there is a common proposal (or even a mainly common proposal) for those cases, then that is the proposal offered by the system as solution to E_1.

Taking this information into account, the CBR system generates the following proposals for the E_1 case:

- 2.- Staff should be trained in their specialized knowledge in deep.
- 11.- Report detected successes, fails and mistakes.

These solutions were also checked by the KM expert, being verified that they are the best solutions for such situation.

5   Enhancing DIPKIP

DIPKIP [4] is a KM system that responds to the need for information management and knowledge flows within a KM organization. It 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. It is named after its 4 steps: Data acquisition, Intelligent Processing, Knowledge Identification and Proposal that can be briefly described in the following way:

- **First Step - Data Acquisition**: it aims to capture information about the organization in which DIPKIP is to be applied. Information can be acquired through interviews, surveys, database mining, a combination of these, and so on.
- **Second Step - Intelligent Processing**: the data obtained in the first step is analyzed through CMLHL. This model provides a visualization of the internal structure of the data set. CMLHL was selected as it provided the clearest projections of the case studies for subsequent expert analysis.
- **Third Step - Knowledge Identification**: a KM expert, based on the data projection generated in the second step, catalogues the analyzed organization 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.

- **Fourth Step - Proposal**: based on the previous step, DIPKIP sets out proposals relating to the following KM processes: creation/acquisition, transference/ distribution and putting into practice/updating. Once the required KM processes have been identified, the KM expert has to decide on the specific actions to implement these processes. DIPKIP outputs must be customized by considering the situation of the analyzed organization. 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.

In the original version of DIPKIP, the KM expert was in charge of analyzing the data projection generated in the third step: Knowledge Identification. This analysis was intended to determine the DIPKIP proposals for the analyzed organization. In the proposed enhanced version of DIPKIP presented in this paper, it is extended by means of the CBR paradigm as previously described in section 4: a CBR system is included in DIPKIP last step in order to automate the proposal generation. Thus, the KM expert intervention is only required in the most difficult situations.

The initial version of the DIPKIP system required expert intervention to identify the knowledge (third step) and to generate a proposal (fourth step). In this version, the fourth step is split in two different phases to eliminate the need of a human expert. The first phase relies on supervised training: the CBR system is trained by storing in the case base the relation between the clustered data and the decision taken. To store information in the case base, it is necessary to analyze the output of the third step.The case base stores information related to the previously solved problems together with the decision taken by the expert (or by the system on its own) to solve the proposed situation.

A schema about the first (supervised) phase of the extended fourth step is depicted in Fig. 4.a. The decisions taken by the KM expert make the case base grow, until the CBR system can work in an independent way. After that, the CBR system is then autonomous and the KM expert is only consulted, in this fourth step, when the CBR
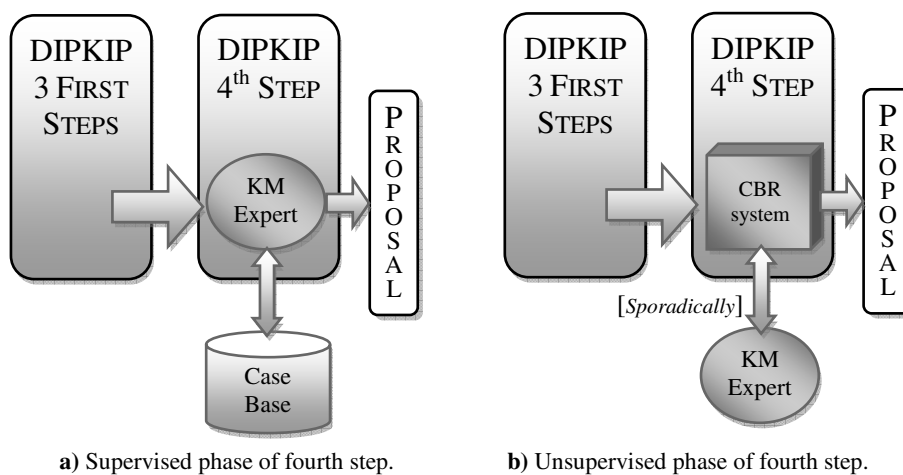

a) Supervised phase of fourth step.   **b)** Unsupervised phase of fourth step.

Fig. 4. Two phases of the proposed DIPKIP system extension

system is not sure enough to propose a good solution (Fig. 4.b). If the expert is consulted by the system, when working in an autonomous way, new solutions may be stored in the system.

6   Conclusions and Future Work

This study presents a novel KM hybrid artificial intelligent solution that automates the proposal generation for overcoming KM deprived situations. To do so, the Case-based Reasoning and connectionist paradigms have been applied. Additionally, this solution has been applied to a four-step KM system called DIPKIP, in which the last step has been split into two different phases. This upgrading has been performed by means of a CBR system allowing an automatic proposal generation. The new model has been applied successfully to a real-case study.

Future work will focus on the application of this hybrid solution to different case studies.

References

1. Sivan, Y.Y.: Nine Keys to a Knowledge Infrastructure: A Proposed Analytic Framework for Organizational Knowledge Management. In: Proceedings of WebNet 2000 - World Conference on the WWW and Internet, pp. 495–500. AACE (2000)
2. Shu-Mei, T.: The Effects of Information Technology on Knowledge Management Systems. Expert Systems with Applications: An International Journal 35(1-2), 150–160 (2008)
3. Aamodt, A., Plaza, E.: Case-Based Reasoning - Foundational Issues, Methodological Variations, and System Approaches. AI Communications 7(1), 39–59 (1994)
4. Herrero, A., Corchado, E., Sáiz, L., Abraham, A.: DIPKIP: A Connectionist Knowledge Management System to Identify Knowledge Deficits in Practical Cases. Computational Intelligence (in press, 2009)
5. Hotelling, H.: Analysis of a Complex of Statistical Variables into Principal Components. Journal of Education Psychology 24, 417–444 (1933)
6. Pearson, K.: On Lines and Planes of Closest Fit to Systems of Points in Space. Philosophical Magazine 2(6), 559–572 (1901)
7. Oja, E.: Neural Networks, Principal Components, and Subspaces. Int. Journal of Neural Systems 1, 61–68 (1989)
8. Friedman, J.H., Tukey, J.W.: A Projection Pursuit Algorithm for Exploratory Data-Analysis. IEEE Transactions on Computers 23(9), 881–890 (1974)
9. Corchado, E., MacDonald, D., Fyfe, C.: Maximum and Minimum Likelihood Hebbian Learning for Exploratory Projection Pursuit. Data Mining and Knowledge Discovery 8(3), 203–225 (2004)

10. Corchado, E., Fyfe, C.: Connectionist Techniques for the Identification and Suppression of Interfering Underlying Factors. Int. Journal of Pattern Recognition and Artificial Intelligence 17(8), 1447–1466 (2003)
11. Fyfe, C., Corchado, E.: Maximum Likelihood Hebbian Rules. In: Proc. of the 10th European Symposium on Artificial Neural Networks (ESANN 2002), pp. 143–148 (2002)
12. Watson, I.: Case-based Reasoning is a Methodology Not a Technology. Knowledge-Based Systems 12(5-6), 303–308 (1999)
13. Aamodt, A.: A Knowledge-Intensive, Integrated Approach to Problem Solving and Sustained Learning. Knowledge Engineering and Image Processing Group. University of Trondheim (1991)
14. Aamodt, A., Plaza, E.: Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. AI Communications 7(1), 39–59 (1994)
15. Corchado, J.M., Bajo, J., Abraham, A.: GERAmI: Improving the Delivery of Health Care. IEEE Intelligent Systems. Special Issue on Ambient Intelligence (2008)
16. Althoff, K.D., Mänz, J., Nick, M.: Maintaining Experience to Learn: Case Studies on Case-Based Reasoning and Experience Factory. In: Proc. 6th Workshop Days of the German Computer Science Society (GI) on Learning, Knowledge, and Adaptivity (LWA 2005) (2005)
17. Cox, M.T., Muñoz-Avila, H., Bergmann, R.: Case-based Planning. The Knowledge Engineering Review 20(3), 283–287 (2006)
18. Herrero, A., Corchado, E., Pellicer, M.A., Abraham, A.: MOVIH-IDS: A Mobile-Visualization Hybrid Intrusion Detection System. Neurocomputing (in press, 2009)
19. Fdez-Riverola, F., Corchado, J.M.: FSfRT: Forecasting System for Red Tides. Applied Intelligence 21(3), 251–264 (2004)
20. Yang, B.S., Han, T., Kim, Y.S.: Integration of ART-Kohonen Neural Network and Case-based Reasoning for Intelligent Fault Diagnosis. Expert Systems With Applications 26(3), 387–395 (2004)
21. Diaz, F., Fdez-Riverola, F., Corchado, J.M.: Gene-CBR: A Case-based Reasoning Tool for Cancer Diagnosis using Microarray Data Sets. Computational Intelligence 22(3/4), 254–268 (2006)
22. Liu, C.-H., Chen, L.-S., Hsu, C.-C.: An Association-based Case Reduction Technique for Case-based Reasoning. Information Sciences 178(17), 3347–3355 (2008)