

On the Improvement of Knowledge Management Status through Case-Based Reasoning in a Hybrid Approach

Emilio Corchado and Aitor Mata

Department of Computing Science and Automatic
University of Salamanca
Salamanca, Spain
{escorchado, aitor}@usal.es

Álvaro Herrero and Lourdes Sáiz

Department of Civil Engineering
University of Burgos
Burgos, Spain
{ahcosio, lsaiz}@ubu.es

Abstract— From an enterprise point of view, Knowledge Management (KM) enables organizations to capture, share, and apply the collective experience and know-how (knowledge) of their staff. Up to now, little effort has been devoted to apply Artificial Intelligent techniques to KM systems. This paper proposes the application of case-based reasoning, in combination with a neural model, to develop a KM system. This combined approach profiles the KM status of the whole organization and automatically generates proposals, aimed at improving the KM situation of organization units. The system is fed with KM data collected at the organization and unit contexts. The 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.

Case-Based Reasoning, Artificial Neural Networks, Unsupervised Learning, Knowledge Management

I. INTRODUCTION

Ever-growing volumes of data are increasingly viewed as important and essential sources of information that may eventually be turned into knowledge. Knowledge Management (KM), a relatively new and fast-growing discipline, 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 datasets, 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] to automatically generate improvement proposals, producing specific and customized suggestions without human intervention to improve the KM status of an enterprise. To do so, CBR is combined with neural paradigms to profile 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]. The purpose of the proposed hybrid solution is to

support decision making that relates to knowledge acquisition, sharing and updating processes that are key to KM processes in the company.

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

II. A HYBRID ADVISING FRAMEWORK

Data processing and subsequent conclusion extraction are challenging tasks in the KM field. This paper proposes a hybrid AI framework aimed at supporting KM managers by advice. This solution operates in two different steps (see Fig. 1 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, it has been proposed a neural processing phase [4]. The neural model for KM status profiling is comprehensively described in section III.
- **Proposal Generation:** once the KM status 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 remedial action is proposed by applying the CBR paradigm, described in section IV.

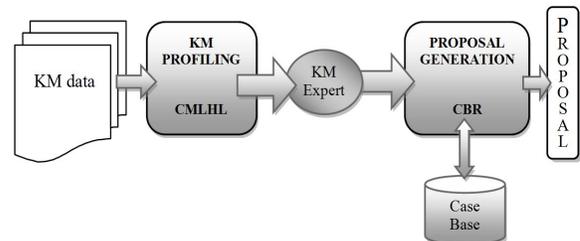


Figure 1. Schema of the hybrid framework.

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.

III. KM STATUS PROFILE

In order to generate improvement proposals (from a KM standpoint), the KM status must be known. To profile the situation of an organization, KM data are required. Further information on the KM data gathering can be found in [4]. Once these data are available, an unsupervised neural approach is proposed, relying on the identification of patterns in the KM dataset.

Patterns that exist across dimensional boundaries in high dimensional datasets 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 are the skew or kurtosis index in the case of Exploratory Projection Pursuit (EPP) [5]. Having identified the most interesting projections, the data is then projected onto a lower dimensional subspace plotted in 2D or 3D, which makes it possible to examine its structure with the naked eye.

The solution proposed in this paper applies a neural model called Cooperative Maximum Likelihood Hebbian Learning (CMLHL) [6, 7], that is based on Maximum Likelihood Hebbian Learning (MLHL) [6, 8]. This unsupervised neural model has been chosen because it has already proved to outperform other models in the target task [4][9]. 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 follows. Feed-forward step:

$$y_i = \sum_{j=1}^N W_{ij} x_j, \forall i \quad (3)$$

Lateral activation passing:

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

Feedback step:

$$e_j = x_j - \sum_{i=1}^M W_{ij} y_i, \forall j \quad (5)$$

Weight change:

$$\Delta W_{ij} = \eta \cdot y_i \cdot \text{sign}(e_j) |e_j|^{p-1} \quad (6)$$

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 [7], [6] and A is a symmetric matrix used to modify the response to the data [7].

IV. AUTOMATIC PROPOSAL GENERATION

Once the required KM processes have been identified, the KM expert has to decide on the specific actions to implement these processes. Thus, decision-support tools are key elements within a KM system. To support KM manager when deciding, the CBR paradigm is presented in this work to automatically generate proposals. These proposals are related to the following KM processes: creation/acquisition, transference/ distribution and putting into practice/ updating.

The CBR paradigm is briefly presented in the following subsection, while the different strategies and techniques applied during its stages are described in the following subsections.

A. CBR Foundations

Case-Based Reasoning [10] origins are in knowledge based systems. CBR systems solve new problems acquiring the needed knowledge from previous situations [11]. The main element of a CBR system is the case base, a structure that stores the information about previous problems and their applied solutions.

The learning capabilities of CBR systems are due to its own structure, composed of four main stages [12] (as can be seen in Fig. 2): *retrieve*, *reuse*, *revise* and *retain*. 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

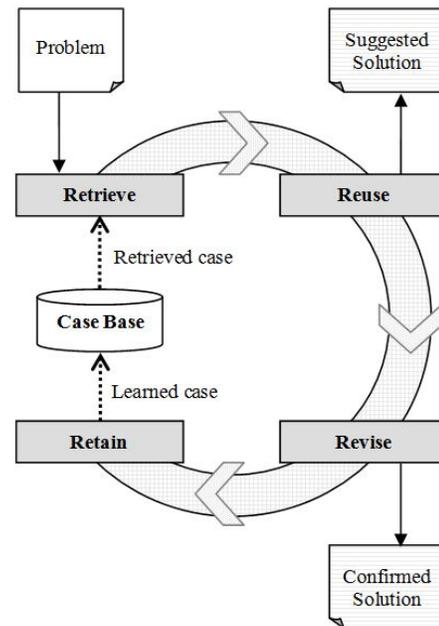


Figure 3. CBR cycle.

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 [10], 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 [13], eLearning [14], planning [15], Intrusion Detection Systems [16] and oceanographic matters [17]. 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.

B. Generating Proposals through CBR

For the proposed hybrid solution, the case base stores information about the previously solved problems together with the decision taken to solve a new problem. The 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 I. 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 a previously listed set of solutions [9]. 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 II), if none of them fits the problem, then the expert will introduce in the system a proper solution.

TABLE I. DESCRIPTION OF THE PARAMETERS DEFINING A CASE, ORGANIZED IN THREE DIFFERENT LEVELS: INITIAL PARAMETERS, KM PROFILING FEATURES, AND SOLUTION.

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

The possible solutions that may be applied to an organization range from easy internal collaboration solutions to intensive abstract working. 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.

The four stages within the CBR cycle are performed as follows:

1) *Retrieve*: when a new problem (P) comes into the system to be solved, the CBR sub-system looks for its relevant counterparts in the case library. The matching of the target problem (P) and each case (C) in the case base is guided by a similarity metric, which is the *Hamming distance* between these two elements. Cases are retrieved from the case base when its similarity value relative to P is not above a predefined threshold (α). As a result, a set of retrieved cases (R) is obtained according to:

$$R = \{C \in CB \mid Hamm(P, C) \geq \alpha\} \quad (7)$$

where CB is the case base and $Hamm(P, C)$ denotes the degree of similarity between C and P , according to the Hamming distance.

2) *Reuse*: once the similar cases are retrieved from the case base, they are used to generate the new solution to the presented problem (P). The solutions of the retrieved cases are aggregated. The similarities of cases being regarded as estimates of the appropriateness of their solutions for solving the new problem. Thus, cases with higher similarity degrees will have more influence in determining the final solution. The outcome of the target problem P is predicted to be a weighted average of the outcomes of the retrieved cases as given by:

$$Out(P) = \frac{\sum_{C_i \in R} Hamm(P, C_i) \cdot Out(C_i)}{\sum_{C_i \in R} Hamm(P, C_i)} \quad (8)$$

where $Out(P)$ and $Out(C_i)$ are the output values for target P and case C_i respectively. As a result, the output generated to a problem P (its solution) depends on the output of the retrieved cases (C_i). The influence of the output of the retrieved cases in the final output of the problem will be higher in those cases that are more similar to the problem to be solved (which is determined by calculating the distance between the problem P and the retrieved cases, C_i).

3) *Revise*: when the system has generated a solution for a target problem, it must be checked to validate its correction. This process will also follow two different phases. At the beginning, while the case base is growing from the expert decisions, it is the expert who checks the correction of the proposed solution. If the expert considers a solution automatically generated as a valid one, the system measures the distance between the solution proposed and the solutions stored for the retrieved cases, and calculates its average. For each solution, accepted or rejected, the system calculates that value. Once the supervised phase has finished, the system calculates the average of the distances of the accepted solutions, and establish that value as a threshold for accepting future solutions. And then, after determining the threshold value for revising the automatic solutions, the CBR sub-system can work in a fully unsupervised mode, with no external aid from the expert.

4) *Retain*: if an automatically generated solution is considered as a valid one by the expert or by the system, then it is stored in the case base. That new solution will be kept in the case base to be used in future problem requests as well as the solutions previously revised by the expert in the supervised phase of the training of the system. The case base is formed by a series of combinations of the parameters in Table I, always with a *solution* value, corresponding to

the optimal output given either by the expert in the training phase or by the system when autonomously working.

After those four main phases, the system has successfully generated a solution for the requested problem. It mainly consist on an specific action to be taken in order to improve the KM status of the analyzed organization. That solution is an adaptation of the previously existing ones by taking into account the characteristics of the introduced problem. In the following section, an experimental study of the proposed system is comprehensively described.

V. EXPERIMENTAL STUDY

The proposed hybrid solution was applied to a real-life case study: a set of companies from the wall painting sector in the Spanish autonomous region of Castilla y León [4]. This section describes how the previously introduced techniques and steps are applied to all the samples in this real-life problem. However, for the sake of brevity, detailed data concerning only one of the samples (E_1 employee) are provided in this section.

A. KM Profiling

The data selected for the KM profiling were taken from a staff survey, generating 68 records from the staff of 39 different companies. The information contained in the 88-feature dataset relates to 21 painting techniques (brush painting, spray varnishing, fibreglass installation, plaster or stucco work, etc...). For each one of those techniques, the survey measured the four 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.

TABLE II. SAMPLE DATA FOR EMPLOYEE E_1 REGARDING GENERAL ISSUES, SPRAY VARNISHING AND STUCCO WORK.

| Technique | Factor | Value |
|------------------|--|-------|
| General Issues | Scaffolding fixing | 1 |
| | Receiving information about labour risk | 1 |
| | Receiving information about environmental protection | 0 |
| | Receiving information about other issues | 0 |
| Spray Varnishing | Knowledge level held | 6 |
| | Willingness to acquire new knowledge | 0 |
| | Interest in updating the knowledge held | 0 |
| | Interest in sharing the knowledge held | 0 |
| Stucco Work | Knowledge level held | 4 |
| | Willingness to acquire new knowledge | 1 |
| | Interest in updating the knowledge held | 1 |
| | Interest in sharing the knowledge held | 0 |

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 dataset. The proposed solution was applied to this dataset. That is, after projecting the data by means of CMLHL and stating the KM profiling features of each case, 67 cases were added to the case base. The remaining case was selected as a validation example, and the application of the proposed solution to this employee (E_1) is described in this section. According to the designed survey, 88 questions were answered by employee E_1 . As it is not possible to present the 88 features for this case, only the answers to the questions on general issues, spray varnishing and stucco work (2 of the 21 painting techniques) are shown in Table II.

After projecting the 88-feature data of E_1 by means of CMLHL, this employee was diagnosed as having a "Knowledge Deficit" that is borne out by their belonging to one of the worst groups in the projection. The following values were given to the KM profiling features:

- Temporal restrictions: a lot of urgency (3).
- Required knowledge: wide level (3).

B. Proposal Generation

Once the KM profiling features of E_1 were known, the case associated to this new problem was defined by considering some characteristics of the enterprise for the other case parameters. After defining the new case, the CBR cycle was started by retrieving from the case base the most similar cases according to the Hamming distance. After checking the retrieved cases for different new situations, the following conclusion arises: most of the recovered cases are projected by CMLHL in the same group. In the case of the E_1 employee, the case representing the E_2 employee (from a different company) was recovered, whose answers from the questions on general issues, spray varnishing and stucco work (two of the 21 painting techniques) are shown in Table III.

TABLE III. SAMPLE DATA FOR EMPLOYEE E_2 REGARDING GENERAL ISSUES, SPRAY VARNISHING AND STUCCO WORK.

| Technique | Factor | Value |
|------------------|--|-------|
| General Issues | Scaffolding fixing | 1 |
| | Receiving information about labour risk | 1 |
| | Receiving information about environmental protection | 0 |
| | Receiving information about other issues | 0 |
| Spray Varnishing | Knowledge level held | 6 |
| | Willingness to acquire new knowledge | 1 |
| | Interest in updating the knowledge held | 0 |
| | Interest in sharing the knowledge held | 0 |
| Stucco Work | Knowledge level held | 4 |
| | Willingness to acquire new knowledge | 1 |

| Technique | Factor | Value |
|-----------|---|-------|
| | Interest in updating the knowledge held | 1 |
| | Interest in sharing the knowledge held | 0 |

The system measures the similarity between the values of the recovered cases and the analyzed values of E_1 and consider those most similar and their solutions. There is at least one proposal (case solution) associated to any of the recovered cases. 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 these issues into account, the CBR system reuses the solutions of previous cases and 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.

At the revision stage, these proposals were checked by the KM expert to validate the proposal. The KM expert analyzes the proposed solutions, to check their validity. If the KM expert concludes that those solutions are good for the situation of employee E_1 , then the proposed solutions are stored in the case base for being eventually used in future situations.

VI. 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 CBR and connectionist paradigms have been applied. This proposal has been applied successfully to a real-case study. It has been proved that the proposed solution can be applied to companies with very different characteristics.

Future work will focus on the application of this hybrid solution to different case studies. Additionally, new models and techniques will be analysed to be applied in the different stages of the CBR cycle.

ACKNOWLEDGMENT

This research is partially supported through the Junta de Castilla and León project BU006A08, Business intelligence for production within the framework of the *Instituto Tecnológico de Castilla y León* (ITCL) and the *Agencia de Desarrollo Empresarial* (ADE), projects of the Spanish Ministry of Education and Innovation project CIT-020000-2008-2 and CIT-020000-2009-12 (funded by the European Regional Development Fund), project of the Spanish Ministry of Science and Innovation TIN2010-21272-C02-01 (funded by the European Regional Development Fund). The authors would also like to thank the vehicle interior manufacturer, Grupo Antolin Ingeniería S.A., within the framework of the project MAGNO2008 - 1028.- CENIT Project funded by the Spanish Government.

REFERENCES

- [1] Y. Y. Sivan, "Nine keys to a knowledge infrastructure: A proposed analytic framework for organizational knowledge management," in

- WebNet 2000 - World Conference on the WWW and Internet, 2000, AACE pp. 495-500.
- [2] T. Shu-Mei, "The effects of information technology on knowledge management systems," *Expert Systems with Applications: An International Journal*, vol. 35, pp. 150-160, 2008.
- [3] A. Aamodt and E. Plaza, "Case-based reasoning - foundational issues, methodological variations, and system approaches," *AI Communications*, vol. 7, pp. 39-59, Mar 1994.
- [4] Á. Herrero, E. Corchado, L. Sáiz, and A. Abraham, "Dipkip: A connectionist knowledge management system to identify knowledge deficits in practical cases," *Computational Intelligence*, vol. 26, pp. 26-56, 2010.
- [5] J. H. Friedman and J. W. Tukey, "A projection pursuit algorithm for exploratory data-analysis," *IEEE Transactions on Computers*, vol. 23, pp. 881-890, 1974.
- [6] E. Corchado, D. MacDonald, and C. Fyfe, "Maximum and minimum likelihood hebbian learning for exploratory projection pursuit," *Data Mining and Knowledge Discovery*, vol. 8, pp. 203-225, May 2004.
- [7] E. Corchado and C. Fyfe, "Connectionist techniques for the identification and suppression of interfering underlying factors," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 17, pp. 1447-1466, Dec 2003.
- [8] C. Fyfe and E. Corchado, "Maximum likelihood hebbian rules," in *10th European Symposium on Artificial Neural Networks (ESANN 2002)*, 2002, pp. 143-148.
- [9] Á. Herrero, A. Mata, E. Corchado, and L. Sáiz, "A hybrid solution for advice in the knowledge management field," in *Hybrid Artificial Intelligence Systems*, 2009, Springer, Heidelberg LNAI (5572) pp. 157-168.
- [10] I. Watson, "Case-based reasoning is a methodology not a technology," *Knowledge-Based Systems*, vol. 12, pp. 303-308, 1999.
- [11] A. Aamodt, "A knowledge-intensive, integrated approach to problem solving and sustained learning," *Knowledge Engineering and Image Processing Group*. University of Trondheim, 1991.
- [12] A. Aamodt and E. Plaza, "Case-based reasoning: Foundational issues, methodological variations, and system approaches," *AI Communications*, vol. 7, pp. 39-59, March 1994 1994.
- [13] J. M. Corchado, J. Bajo, and A. Abraham, "Gerami: Improving the delivery of health care," *IEEE Intelligent Systems*. Special Issue on Ambient Intelligence, 2008.
- [14] K. D. Althoff, J. Mänz, and M. Nick, "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)* Saarland University, Germany, 2005.
- [15] M. T. Cox, H. Muñoz-Avila, and R. Bergmann, "Case-based planning," *The Knowledge Engineering Review*, vol. 20, pp. 283-287, 2006.
- [16] Á. Herrero, E. Corchado, M. A. Pellicer, and A. Abraham, "Movih-ids: A mobile-visualization hybrid intrusion detection system," *Neurocomputing*, vol. 72, pp. 2775-2784, 2009.
- [17] F. Fdez-Riverola and J. M. Corchado, "Fsfrit: Forecasting system for red tides," *Applied Intelligence*, vol. 21, pp. 251-264, 2004.