EFFICIENT KNOWLEDGE EXTRACTION USING UNSUPERVISED NEURAL NETWORK MODELS

Jean-Charles Lamirel and Shadi Al Shehabi
LORIA, Campus Scientifique, BP 239
54506 Vandoeuvre-lès-Nancy Cedex, France
{Jean-Charles.Lamirel, Shadi.Al-Shehabi}@loria.fr

Abstract – *This paper presents a new approach whose aim is to extent the scope of numerical models by providing them with knowledge extraction capabilities. The basic model which is considered in this paper is a multi-topographic neural network model. The powerful features of this model are its generalization mechanism and its mechanism of communication between topographies. These two mechanisms allow rule extraction to be performed whenever a single viewpoint or multiple viewpoints on the same data are considered. The association rule extraction is itself based on original quality measures which evaluate to what extent a numerical classification model behaves as a natural symbolic classifier such as a Galois lattice.*

Keywords – knowledge extraction, unsupervised learning, neural gas (NG), MultiGAS model, MultiSOM model, symbolic model, association rules, multi-viewpoint analysis

1 Introduction

Data mining or knowledge discovery in database (KDD) refers to the non-trivial process of discovering interesting, implicit, and previously unknown knowledge from large databases. Such a task implies to be able to perform analyses on high-dimensional input data. The most popular models used in KDD are the symbolic models. Unfortunately, these models suffer of very serious limitations. Rule generation is a highly time-consuming process that generates a huge number of rules, including a large ratio of redundant rules. Hence, this prohibits any kind of rule computation and selection as soon as data are numerous and they are represented by very high-dimensional description space. This latter situation is very often encountered with documentary data. To cope with these problems, preliminary KDD trials using numerical models have been made. An algorithm for knowledge extraction from self-organizing network is proposed in [3]. This approach is based on a supervised generalized relevance learning vector quantization (GRLVQ) which is used for extracting decision trees. The different paths of the generated trees are then used for denoting rules. Nevertheless, the main defect of this method is to necessitate training data. On our own side, we have proposed a hybrid classification method for mapping an explicative structure issued from a symbolic classification into an unsupervised numerical self-organizing map (SOM) [6]. SOM map and Galois lattice are generated on the same data. The cosine projection is then used for associating lattice concepts to the SOM classes. Concepts properties act as explanation for the SOM classes. Furthermore, lattice pruning combined with migration of the associated SOM classes towards the top of the pruned lattice is used to generate explanation of increasing scope on the SOM map. Association rules can also be produced in such a way. Although it establishes interesting links between numerical and symbolic worlds this approach necessitates the time-

consuming computation of a whole Galois lattice. In a parallel way, in order to enhance both the quality and the granularity of the data analysis and to reduce the noise which is inevitably generated in an overall classification approach, we have introduced the multi-viewpoint analysis based on a significant extension of the SOM model, named MultiSOM [5]. The viewpoint building principle consists in separating the description of the data into several sub-descriptions corresponding different property subsets. In MultiSOM each viewpoint is represented by a single SOM map. The conservation of an overall view of the analysis is achieved through the use of a communication mechanism between the maps, which is itself based on Bayesian inference [9]. The advantage of the multi-viewpoint analysis provided by MultiSOM as compared to the global analysis provided by SOM [4] has been clearly demonstrated for precise mining tasks like patent analysis [7]. Another important mechanism provided by the MultiSOM model is its on-line generalization mechanism that can be used to tune the level of precision of the analysis. Furthermore, we have proposed in [2] to use the neural gas (NG) model as a basis for extending the MultiSOM model to a MultiGAS model. NG model [10] is known as more efficient than SOM model for classification tasks where explicit visualization of the data analysis results is not required. Hence, thanks to the loss of topographic constraints as compared to SOM, NG tends to better represent the structure of the data, yielding better classification results [2].
In this paper we propose a new approach for knowledge extraction that consists in using our MultiGAS model as a front-end for unsupervised extraction of association rules. In our approach we exploit both the generalization and the intercommunication mechanisms of the model. We also make use of our original recall and precision measures that derive from the Galois lattice theory and from Information Retrieval (IR) domains [8]. The first section presents the MultiGAS model. The second section presents the rule extraction principles based on the MultiGAS model. The experiment that is presented on the last section shows how our method can be used both to control the rules inflation that is inherent to symbolic methods and for extracting the most significant rules.

2 MultiGAS Model

The principle of the MultiGAS model is to be constituted by several gases that have been generated from the same data. Each gas is itself issued from a specific data description subspace. The relation between gases is established through the use of two main mechanisms: the inter-gas communication mechanism and the generalization mechanism.

The inter-gas communication mechanism enables to highlight semantic relationships between different topics belonging to different viewpoints related to the same data. In MultiGAS, this communication is based on the use of the data that have been projected onto each gas as intermediary neurons or activity transmitters between gases. The inter-gas communication is established by standard Bayesian inference network propagation algorithm which is used to compute the posterior probabilities of target gas's neuron T_k which inherited of the activity (evidence Q) transmitted by its associated data neurons. This computation can be carried out efficiently because of the specific Bayesian inference network topology that can be associated to the MultiGAS model. Hence, it is possible to compute the probability $P(ac_mt|T_k,Q)$ for an activity of modality act_m on the gas neuron T_k which is inherited from activities generated on the source gas. This computation is achieved as follows [9]:

$$P(act_m|T_k,Q) = \frac{\sum_{d \in act_m,T_k} Sim(d,S_d)}{\sum_{d \in T_k} Sim(d,S_d)} \qquad (1)$$

Such that S_d is the source neuron to which the data d has been associated, $Sim(d, S_d)$ is the cosine correlation measure between the codebook vector of the data d and the one of its source neuron S_d and $d \in act_m,T_k$ if it has been activated with the modality act_m from the source gas.

The neurons of the target gas getting the highest probabilities can be considered as the ones who include the topics sharing the strongest relationships with the topics belonging to the activated neurons of the source gas.

The main roles of the generalization mechanism are both to evaluate the coherency of the topics that have been computed on an original gas and to summarize the contents of this later into more generic topics. Our NG generalization mechanism [2] creates its specific link structure in which each neuron of a given level is linked to its 2-nearest neighbours (Fig. 1). For each new level neuron *n* the following codebook vector computation applies:

$$W_n^{M+1} = \frac{1}{3}\left(W_n^M + \sum_{n_k \in V_n^M} W_{n_k}^M \right) \qquad (2)$$

where V_n^M represents the 2-nearest neighbour neurons of the neuron *n* on the level *M* associated to the neuron *n* of the new generated level *M+1*. After codebook vector computation the repeated neurons of the new level (i.e. the neurons of the new level that share the same codebook vector) are summarized into a single neuron. Our generalization mechanism can be considered as an implicit and distributed form of a hierarchical classification method based on neighbourhood reciprocity. Its main advantage is to produce homogeneous generalization levels. It ensures the conservation of the topographic properties of the gas codebook vectors on each generalization level. Moreover, we have shown in [2] that this method produces more homogeneous results than the classical training approach while significantly reducing time consumption. Lastly, the inter-gas communication mechanism presented in the former section can be used on a given viewpoint between a gas and its generalizations as soon as they share the same projected data.

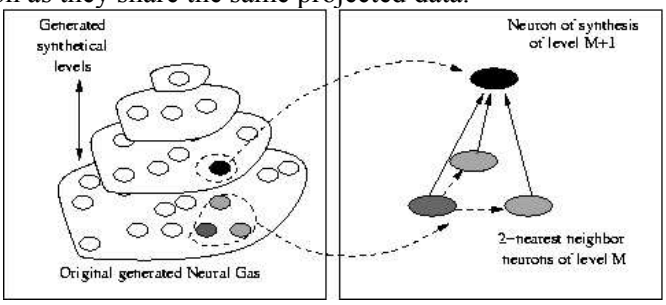

Fig 1. Gas generalization mechanism (2D representation of gas is used for the sake of clarity of the figure).

3 Quality of classification model

The classical evaluation measures for the quality of classification are based on the intra-class inertia and the inter-class inertia (see [8]). These measures are often strongly biased because they depend both on the pre-processing and on the classification methods. Therefore, we have proposed to derive from the Galois lattice and Information Retrieval (IR) domains two new quality measures, *Recall* and *Precision*. As compared to classical inertia measures, averaged measures of *Recall* and *Precision* present the main advantages to be independent of the classification method. The *Precision* and *Recall* measures are based on the properties of class members [8]. The *Precision criterion* measures in which proportion the content of the classes generated by a classification method is homogeneous. The greater the *Precision*, the nearer the intensions of the data belonging to the same classes will be one with respect to the other, and consequently, the more homogenous will be the classes. In a complementary way, the *Recall criterion* measures the exhaustiveness of the content of said classes, evaluating to what extent single properties are associated with single classes. The *Recall criterion* should be considered as a specific application of the statistical concept

of sensitivity (i.e. true positive rate) to class properties [1]. The *Recall* (Rec) and *Precision* (Prec) measures for a given property *p* of the class *c* are expressed as:

$$\underset{c}{\text{Rec}}\,(p) = \frac{\left|c_p^*\right|}{\left|C_p^*\right|}, \qquad \underset{c}{\text{Prec}}\,(p) = \frac{\left|c_p^*\right|}{\left|c\right|} \tag{3}$$

such that, C is a set of classes issued from a classification method applied on a set of documents D, $c \in C$, and

$$c_p^* = \left\{ d \in c, \quad \xi_d^p > 0 \right\} \tag{4}$$

where ξ_d^p is the weight of the property p for the data d.

We have demonstrated in [8] that if both values of *Recall* and *Precision* reach the unity value, the peculiar set of classes represents a Galois lattice. A class belongs to the peculiar set of classes of a given classification if it possesses peculiar properties. Finally, a property is considered as peculiar for a given class if it is maximized by the class members.
Averaged measures of *Recall* and *Precision* can be used for overall comparison of classification methods and for optimisation of the results of a method relatively to a given dataset. In this paper we will more specifically focus on peculiar properties of the classes and on local measures of *Precision* and *Recall* associated to single classes. Hence, as soon as this information can be fruitfully exploited for generating explanations on the contents of individual classes [6], it will also represent a sound basis for extracting rules from said classes.

4 Rules Extraction from MultiGAS model

An elaborated unsupervised neural model, like MultiGAS, represents a natural candidate to cope with the related problems of rule inflation and rule selection that are inherent to symbolic methods. Hence, its synthesis capabilities that can be used both for reducing the number of rules and for extracting the most significant ones. In the knowledge extraction task, the generalization mechanism can be specifically used for controlling the number of extracted association rules. The intercommunication mechanism will be useful for highlighting association rules figuring out relationships between topics belonging to different viewpoints.

4.1 Rules extraction by the generalization mechanism

We will rely on our own class quality criteria for extracting rules from the classes of the original gas and its generalizations. For a given class c, the general form of the extraction algorithm (**A1**) follows:

$\forall p_1, p_2 \in P_c^*$
1) ***If*** (Rec(p_1) = Rec(p_2) = Prec(p_1) = Prec(p_2) = 1) ***Then***: $p_1 \leftrightarrow p_2$ (equivalence rule)
2) ***ElseIf*** (Rec(p_1) = Rec(p_2) = Prec(p_2) = 1) ***Then***: $p_1 \rightarrow p_2$
3) ***ElseIf*** (Rec(p_1) = Rec(p_2) = 1) ***Then***
    If (Extent(p_1) $\subset$ Extent(p_2)) ***Then***: $p_1 \rightarrow p_2$
    If (Extent(p_2) $\subset$ Extent(p_1)) ***Then***: $p_2 \rightarrow p_1$
    If (Extent(p_1) $\equiv$ Extent(p_2)) ***Then***: $p_1 \leftrightarrow p_2$
$\forall p_1 \in P_c^*$, $\forall p_2 \in P_c - P_c^*$
4) ***If*** (Rec(p_1) = 1) ***If*** (Extent(p_1) $\subset$ Extent(p_2)) ***Then***: $p_1 \rightarrow p_2$ (*)

where Prec and Rec respectively represent the local *Precision* and *Recall* measures, Extent(p) represents the extension of the property p (i.e. the list of data to which the property p is associated), and P_c^* represent the set of peculiar properties of the class c.

The optional step 4) (*) can be used for extracting extended rules. For extended rules, the constraint of peculiarity is not applied to the most general property. Hence, the extension of this latter property can include data being outside of the scope of the current class c.

4.2 Rules extraction by the inter-gas communication mechanism

A complementary extraction strategy consists in making use of the extraction algorithm in combination with the principle of communication between viewpoints for extracting rules. The general form of the extraction algorithm (**A2**) between two viewpoints v_1 and v_2 will be:

$\forall p_1 \in P_c^*$, $\forall p_2 \in P_{c`}^*$ and $c \in v_1$, $c` \in v_2$
1) ***If*** (Rec(p_1) = Rec(p_2) = Prec(p_1) = Prec(p_2) = 1) ***Then*** *Test_Rule_Type*;
2) ***ElseIf*** (Rec(p_1) = Rec(p_2) = Prec(p_2) = 1) ***Then*** *Test_Rule_Type*;
3) ***ElseIf*** (Rec(p_1) = Rec(p_2) = Prec(p_1) = 1) ***Then*** *Test_Rule_Type*;
4) ***ElseIf*** (Rec(p_1) = Rec(p_2) = 1) ***Then*** *Test_Rule_Type*;

where *Test_Rule_Type* procedure is expressed as:

If ($\text{Extent}_{v1}(p_1) \subset \text{Extent}_{v2}(p_2)$) ***Then***: $p_1 \rightarrow p_2$
If ($\text{Extent}_{v2}(p_2) \subset \text{Extent}_{v1}(p_1)$) ***Then***: $p_2 \rightarrow p_1$
If ($\text{Extent}_{v1}(p_1) \equiv \text{Extent}_{v2}(p_2)$) ***Then***: $p_1 \leftrightarrow p_2$

Extended rules will be obtained as:
a) $\forall p_1 \in P_c^*$, $\forall p_2 \in P_{c`}$: Substituting respectively Rec(p_2) and Prec(p_2) by the *viewpoint-based measures* $\text{Rec}_{v1}(p_2)$ and $\text{Prec}_{v1}(p_2)$, related to the source viewpoint, in the previous algorithm.
b) $\forall p_1 \in P_c$, $\forall p_2 \in P_{c`}^*$: Substituting respectively Rec(p_1) and Prec(p_1) by the *viewpoint-based measures* $\text{Rec}_{v2}(p_1)$ and $\text{Prec}_{v2}(p_1)$, related to the destination viewpoint, in the previous algorithm.

5. Experimental results

Our test database is a database of 1000 patents that has been used in some of our preceding experiments [7]. For the viewpoint-oriented approach the structure of the patents has been parsed in order to extract four different subfields corresponding to four different viewpoints: Use, Advantages, Titles and Patentees. As it is full text, the content of the textual fields of the patents associated with the different viewpoints is parsed by a lexicographic analyzer in order to extract viewpoint specific indexes. Two viewpoints, Use and Advantages, will be considered in our experiment. The Use and Advantages viewpoints generate themselves description spaces of size 234 and 207 respectively. Each of our experiments is initiated with an optimal gas generated thanks to an optimization algorithm based on our quality criteria [8]:

- Original gases of 121 (optimal) and 100 (optimal) neurons for Advantages and Use viewpoints, respectively, are firstly generated.
- Generalized gases of 100, 83, 75, 64, 53, 44, 34, 28, 23, 18 and 13 neurons are generated by applying the generalization mechanism to the 121 original gas for Advantages viewpoint.
- Generalized gases of 79, 62, 50, 40, 31, 26, 16 and 11 neurons are generated by applying the generalization mechanism to the 100 neurons original gas for Use viewpoint.

Our first experiment consists in extracting rules from the single Use viewpoint. Both the original gas and its generalizations are used for extracting the rules. The algorithm is used once without its optional step, and a second time including this step (for more details, see algorithm A1). The results are presented at figure 2. Some examples of extracted rules are given hereafter.

Bearing of outdoor machines → *Printing machines* (supp = 2, conf = 100%)
Refrigerator oil → *Gear oil* (supp = 3, conf = 100%)

where conf of rule $A \rightarrow B$ is calculated as follows: conf = supp(A ∪ B)/supp(A), and supp(*A*) is the number of data to which the property A is associated.

For evaluating the complexity of our algorithm based on a numerical approach as compared to a symbolic approach we use the following complexity factor (CF) computation:

$$CF = (RC * MLC) / (MRC * LC) \quad (5)$$

where RC=rules count, MRC=maximum rules count (symbolic), LC=loops count, MLC=maximum loop count (symbolic).

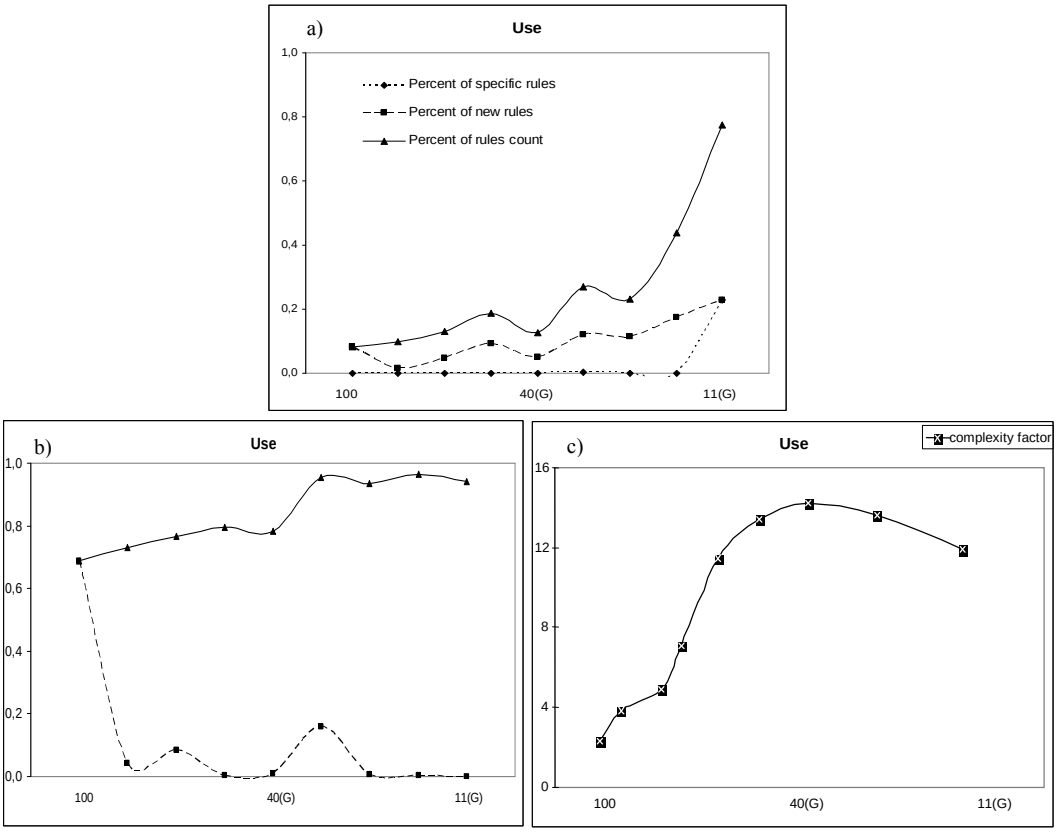

Fig. 2. Rule extraction curves for Use viewpoint. a) extraction algorithm without optional step. b) the same with optional step. c) complexity function for the algorithm including optional step. New rules: rules that are found in a given level but not in the preceding ones. Specific rules: rules which are found only in a given level. Rules count: is the total number of rules that are extracted from all levels. (x(G): represents a level of generalization of x neurons).

A global summary of the results is given in table 1. The table includes a comparison of our extraction algorithm with a standard symbolic rule extraction method as regards to the amount of extracted rules. In single viewpoint experiment, when our extraction algorithm is used with its optional step, it is able to extract the same number of rules as a classical symbolic model that basically uses a combinatory approach. Indeed, table 1 shows that all the rules of confidence 100%

(i.e. 536 rules) are extracted by the combination of gas levels. Moreover, a significant amount of rule can be extracted from any single level of the gas (see fig. 2b). Even if, in this case, no rule selection is performed, the main advantage of this version of the algorithm, as compared to a classical symbolic method, is the computation time. Indeed, as soon as our algorithm is class-based, the computation time it significantly reduced. Moreover, the lower the generalization level, the more specialized will be the classes, and hence, the lower will be the combinatory effect during computation (see fig. 2c). Another interesting result is the behaviour of our extraction algorithm when it is used without its optional step. The fig. 2a shows that, in this case, a rule selection process that depends of the generalization level is performed: the higher will be the generalization level, the more rules will be extracted. We have already done some extension of our algorithm in order to search for partial rules. Complementary results showed us that, even if this extension is used, no partial rules will be extracted in the low level of generalization when no optional step is used. This tends to prove that the standard version of our algorithm is able to naturally perform rule selection.
Our second experiment consists in extracting rules using the intercommunication mechanism between the Use and the Advantage viewpoints. The communication is achieved between the original gas of each viewpoint, and furthermore, between the same levels of generalization of each viewpoint. For each single communication step the extraction algorithm is applied is a bidirectional way. Some examples of extracted rules are given hereafter.

Natural oil (Advantages) → *Catapult oil* (Use) (supp = 2, conf = 100%)
Natural oil (Advantages) → *Drilling fluid* (Use) (supp = 2, conf = 100%)

The results of our multi-viewpoint experiment are similar to the ones of our single viewpoint experiment (see table 1). A rule selection process is performed when the standard version of our algorithm is used. The maximum extraction performance is obtained when *viewpoint-based Recall* and *viewpoint-based Precision* viewpoint are used (see algorithm A2).

		Use	Use ↔ Advantages
Symbolic model	Total rule count	536	649
	Average confidence	100%	100%
	Global rule count	2238	2822
	Average confidence	59%	45%
MultiGAS model (9 levels)	Peculiar rule count	251	250
	Average confidence	100%	100%
	Extended rule count	536	642
	Average confidence	100%	100%

Table 1. Summary of results. The table presents a basic comparison between the standard symbolic rule extraction method and the MultiGAS-based rule extraction method. The global rule count defined for the symbolic model includes the count of partial rules (confidence<100%) and the count of total rules (confidence=100%). In our experiments, the rules generated by the MultiGAS model on the 9 levels are only total rules. The peculiar rule count is the count of rules obtained with the standard versions of the extraction algorithms. The extended rule count is the count of rules obtained with the extended versions of the extraction algorithms including their optional steps.

6 Conclusion

In this paper we have proposed a new approach for knowledge extraction based on a MultiGAS model. Our approach makes use of original measures of recall and precision for extracting rules from gases. Thanks to the MultiGAS model, our experiments have been conduced on single viewpoint classifications as well as between multiple viewpoints classifications on the same data. They take benefit of the generalization and the inter-gas communication mechanisms that are embedded in the MultiGAS model. Even if complementary experiments must be done, our first results are very promising. They tend to prove that a neural model, as soon as it is elaborated enough, represents a natural candidate to cope with the related problems of rule inflation, rule

selection and computation time that are inherent to symbolic models. One of our perspectives is to more deeply develop our model in order to extract rules with larger context like the ones that can be obtained by the use of closed set in symbolic approaches. Another interesting perspective would be to adapt measures issued from information theory, like IDF or entropy, for ranking the rules. Furthermore, we plan to test our model on a reference dataset on genome. Indeed, these dataset has been already used for experiments of rule extraction and selection with symbolic methods. Lastly, our extraction approach can be applied in a straightforward way to a MultiSOM model, or even to a single SOM model, when overall visualization of the analysis results is required and less accuracy is needed.

References

[1] C. Ahn (1997), Statistical methods for the estimation of sensitivity and specificity of site-specific diagnostic tests, *Journal of Periodontal Research* 32: 351-354.

[2] S. Al Shehabi (2005), J.C. Lamirel. Multi-Topographic Neural Network Communication and Generalization for Multi-Viewpoint Analysis, To be appeared in: International Joint Conference on Neural Networks - IJCNN'05. (Montréal, Québec, Canada).

[3] B. Hammer, A. Rechtien, M. Strickert, T. Villmann (2002), Rule extraction from self-organizing networks, ICANN, Springer, 877-882.

[4] T. Kohonen (2001), *Self-Organizing Maps*, 3rd ed. Springer Verlag, Berlin.

[5] J.C. Lamirel, Application d'une approche symbolico-connexionniste pour la conception d'un système documentaire hautement interactif (1995), Thèse de l'Université de Nancy 1, Henri Poincaré.

[6] J.C. Lamirel, Y. Toussaint, S. Al Shehabi (2003), A Hybrid Classification Method for Database Contents Analysis, FLAIRS Conference, p. 286-292.

[7] J.C. Lamirel, S. Al Shehabi, M. Hoffmann, C. Francois (2003), Intelligent patent analysis through the use of a neural network : experiment of multi-viewpoint analysis with the MultiSOM model, Proceedings of ACL, Sapporo, Japan, p. 7-23.

[8] J.C. Lamirel, S. Al Shehabi, C. Francois, M. Hoffmann (2004), New classification quality estimators for analysis of documentary information: application to web mapping, *Scientometrics*, **Vol. 60**, No. 3, p. 445-462.

[9] J.C. Lamirel, S. Al Shehabi, C. François, X. Polanco (2004). Using a compound approach based on elaborated neural network for Webometrics: an example issued from the EICSTES Project. Scientometrics, **Vol. 61**, No. 3, p. 427-441.

[10] T. Martinetz, K. Schulten (1991). A "neural-gas" network learns topologies, In T. Kohonen, K. Mäkisara, O. Simula, and J. Kangas, editors, *Artificial neural networks*, North-Holland, Amsterdam, p. 397-402.