Knowledge Extraction from Unsupervised Multi-topographic Neural Network Models

Shadi Al Shehabi, Jean-Charles Lamirel

Knowledge Extraction from Unsupervised Multi-Topographic Neural Network Models

Shadi Al Shehabi and Jean-Charles Lamirel

LORIA, Campus Scientifique, BP 239
54506 Vandoeuvre-lès-Nancy Cedex, France
{Shadi.Al-Shehabi, Jean-Charles.Lamirel}@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. One of the most powerful features of this model is its generalization mechanism that allows rule extraction to be performed. The extraction of association rules 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. A first experimental illustration of rule extraction on documentary data constituted by a set of patents issued form a patent database is presented.

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 [4]. 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 in 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 matching an explicative structure issued from a symbolic classification to an unsupervised numerical self-organizing map (SOM) [7]. 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 MultiSOM model [6]. This model represents a significant extension of the SOM model, in which 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 [10]. The advantage of the multi-viewpoint analysis provided by MultiSOM as compared to the global analysis provided by SOM [5] has been clearly demonstrated for precise mining tasks like patent analysis [8]. 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 [1] to use the neural gas (NG) model as a basis for extending the MultiSOM model to a MultiGAS model. Hence, NG model [11] is known as more efficient and homogeneous than SOM model for classification tasks where explicit visualization of the data analysis results is not required.

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 specifically exploit the generalization mechanism of the model. We also make use of our own recall and precision measures that derive from the Galois lattice theory and from Information Retrieval (IR) domain [9]. The first section of the paper presents the symbolic approach for rules extraction. The second section presents the rule extraction principles based on the MultiGAS model. The experiment presented in the last section shows how our method can be used both for controlling the rules inflation that is inherent to symbolic methods and for extracting the most significant rules.

2 The Symbolic Model and Association Rules Extraction

The symbolic approach to Database Contents Analysis is mostly based on the Galois lattice model (see [2] and [12]). A Galois lattice, *L(D,P)*, is a conceptual hierarchy built on a set of data *D* which are described by a set of properties *P* also called the intention (Intent) of the concept of the lattice. A class of the hierarchy, also called "formal concept", is defined as a pair *C=(d,p)* where *d* denotes the extension (Extent) of the concept, i.e. a subset of *D*, and *p* denotes the intention of the concept, i.e. a subset of *P*. The lattice structure implies that it exists a partial order on a lattice such that:

$$\forall\, C_1, C_2 \in L,\ C_1 \leq C_2 \Leftrightarrow \text{Extent}(C_1) \subseteq \text{Extent}(C_2) \Leftrightarrow \text{Intent}(C_1) \supseteq \text{Intent}(C_2)$$

Association rules are one of the basic types of knowledge extraction from large databases. Given a database, the problem of mining association rules consists in generating all association rules that have certain user-specified minimum support and confidence. An association rule is an expression $A \rightarrow B$ where A and B are conjunctions of properties. It means that if an individual data possesses all the properties of A then he necessarily possesses all the properties of B. The support of the rule is $supp(A \cup B)$, and the confidence: $Conf = supp(A \cup B)/supp(A)$. An approach proposed by [12] shows that a subset of association rules can be obtained following the direct links of heritage between the concepts in the Galois lattice. Even if no satisfactory solution regarding rule computation time have been given, some attempt to solve the rule selection problem by combining rules evaluation measures is also proposed in [2].

3. MultiGAS Model for Rule Extraction

A reliable unsupervised neural model, like a gas, 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. We will rely on our own class quality criteria for extracting rules from the classes of the original gas and its generalizations, that is the *Precision* and *Recall* measures based on the properties of class members, which are defined in [9]. 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. We have demonstrated in [9] 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. As compared to classical inertia measures, averaged measures of *Recall* and *Precision* present the main advantages to be independent of the classification method. They can thus be used both for comparing classification methods and for optimizing the results of a method relatively to a given dataset. In this paper we will focus on peculiar properties of the classes and on local measures of *Precision* and *Recall* associated to single classes. Hence, as soon as these informations can be fruitfully exploited for generating explanations on the contents of individual classes, they also represent a sound basis for extracting rules from these latter classes. The general form of the extraction algorithm follows:

Let C being a class, P_C being the set of properties associated to the members of C, and P_C^* being the set of peculiar properties of C, with $P_C^* \subseteq P_C$:

$\forall\ p_1, p_2 \in P_C^*$
1) ***If*** (Rec(p_1, p_2) = Prec(p_1, p_2) = 1) ***Then*** there is an equivalence rule: $p_1 \leftrightarrow p_2$
2) ***ElseIf*** (Rec(p_1, p_2) = Prec(p_2) = 1) ***Then*** there is an association rule: $p_1 \rightarrow p_2$
3) ***ElseIf*** (Rec(p_1, 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 represent the local *Precision* and *Recall* measures, respectively.

The optional step 4) (*) can be used for increasing the number of extracted rules. In this step the constraint of peculiarity is relaxed for the most general property.

The gas generalization principle consists in summarizing the contents of an original gas by progressively reducing its number of neurons. A triangle-based strategy for gas generalization has been successfully tested in [1]. Its main advantage is to produce homogeneous gas generalization levels while ensuring the conservation of the topographic properties of the gas codebook vectors on each level. A basic rule extraction strategy consists in applying the above described extraction algorithm both on an original gas and on its generalizations. The expected result of this strategy is to be able to control the rule number and the rule quality by the choice of a proper generalization level.

4. Experimental Results

Our test database is a database of 1000 patents that has been used in some of our preceding experiments [8]. For the viewpoint-oriented approach the structure of the patents has been parsed in order to extract four different subfields corresponding to fours 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. Only, the Use viewpoint will be considered in our experiment. This viewpoint generates itself a description space of size 234. Our experiment is initiated with an optimal gas generated thanks to an optimization algorithm based on the quality criteria [9]:

- Original gas of 100 neurons (optimal) is firstly generated for the Use viewpoint.
- Generalized gases of 79, 62, 50, 40, 31, 26, 16 and 11 neurons are generated for this latter viewpoint by applying the generalization mechanism to the 100 neurons original gas.

Our 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). The results are presented at figure 1. 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%)

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 concerning the amount of extracted rules. 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) are also 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. 1b). 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. Another interesting result is the behavior of our extraction algorithm when it is used without its optional step. The fig. 1a 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 levels 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.

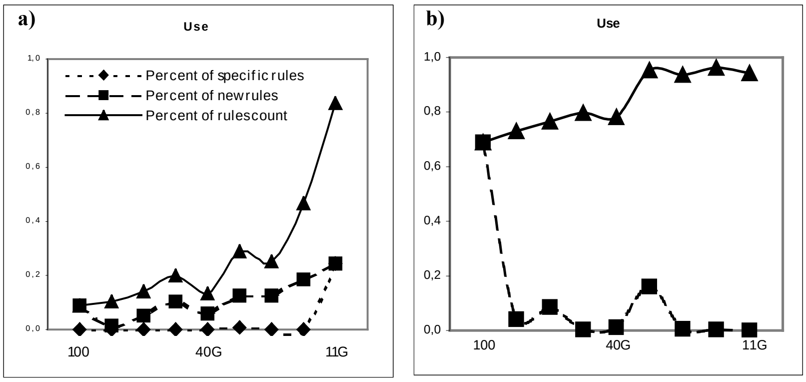

Fig. 1. Rule extraction curves for Use viewpoint. a) extraction algorithm without optional step. b) the same with optional step. New rules: rules that are found in a given level but not in the preceding ones. Specific rules: rules which are found only in a given level. Rules count: is the total number of rules that are extracted from all levels. ((xG): represents a level of generalization of x neurons).

		Use
Symbolic model	Total rule count	536
	Average confidence	100%
	Global rule count	2238
	Average confidence	59%
MultiGAS model (9 levels)	Peculiar rule count	251
	Average confidence	100%
	Extended rule count	536
	Average confidence	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%). The rules generated by the MultiGAS model on the 9 levels are only total rules. The peculiar rule count is obtained with the standard version of the extraction algorithm. The extended rule count is obtained with the extended version of the extraction algorithm including the optional step.

5. 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. It takes benefit of the generalization mechanism that is 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 adapt this model to the multi-viewpoint context of the MultiGAS model that represents itself a powerful context for knowledge extraction. Furthermore, we plan to test our model on a reference dataset on genome. Indeed, these dataset has been extensively used for experiments of rule extraction and selection with symbolic methods [2].

References

1. S. Al Shehabi, J.C. Lamirel. Multi-Topographic Neural Network Communication and Generalization for Multi-Viewpoint Analysis. International Joint Conference on Neural Networks - IJCNN'05. (Montréal, Québec, Canada). 2005.
2. H. Cherfi. Étude et réalisation d'un système d'extraction de connaissances à partir de textes. Thèse de l'Université de Nancy 1, Henri Poincaré, 2004.
3. B. Hammer, A. Rechtien, M. Strickert, T. Villmann, Rule extraction from self-organizing maps, in: J.R.Dorronsoro (Ed.), Artificial Neural Networks -- ICANN 2002, Springer, 877-882, 2002.
4. J. Han, M. Kamber, A. K. H. Tung. (2001). Spatial clustering methods in data mining: A survey, H. Miller and J. Han (eds.), Geographic Data Mining and Knowledge Discovery, Taylor and Francis.
5. T. Kohonen, Self-Organizing Maps. 3rd ed. Springer Verlag, Berlin, 2001.
6. J.C. Lamirel, Application d'une approche symbolico-connexionniste pour la conception d'un système documentaire hautement interactif. Thèse de l'Université de Nancy 1, Henri Poincaré, 1995.
7. J.C. Lamirel, Y. Toussaint, S. Al Shehabi. A Hybrid Classification Method for Database Contents Analysis. In The 16th International FLAIRS Conference - FLAIRS 2003. (St. Augustine, Florida). 2003.
8. J.C. Lamirel, S. Al Shehabi, M. Hoffmann, C. Francois. Intelligent patent analysis through the use of a neural network : experiment of multi-viewpoint analysis with the MultiSOM model. Proceeding of ACL, Sapporo, Japan. 2003.
9. J.C. Lamirel, S. Al Shehabi, C. Francois, M. Hoffmann. New classification quality estimators for analysis of documentary information: application to web mapping. Scientometrics, vol. 60, no. 3, pp. 445-462, Feb2004.
10. J.C. Lamirel, S. Al Shehabi, C. François, X. Polanco. Using a compound approach based on ealborated neural network for Webometrics: an example issued from the EICSTES Project. Scientometrics, Vol. 61, No. 3 (2004), pp. 427-441.
11. T. Martinetz, K. Schulten. A "neural-gas" network learns topologies. In T. Kohonen, K. Mäkisara, O. Simula, and J. Kangas, editors, Artificial neural networks, North-Holland, Amsterdam, 1991, pp. 397-402.
12. A. Simon and A. Napoli. Building Viewpoints in an Object-based Representation System for Knowledge Discovery in Databases, Proceedings of IRI'99, Atlanta, Geogia, S. Rubin editor, The International Society for Computers and Their Applications, ISCA, pages 04-108, 1999.