Divisible Rough Sets Based on Self-organizing Maps

Rocío Martínez-López and Miguel A. Sanz-Bobi

Computer Science Department, Escuela Técnica Superior de Ingeniería-ICAI,
Universidad Pontificia Comillas, 28015 Alberto Aguilera 25, Madrid, Spain
{rmartinez, masanz}@dsi.icai.upcomillas.es

Abstract. The rough sets theory has proved to be useful in knowledge discovery from databases, decision-making contexts and pattern recognition. However this technique has some difficulties with complex data due to its lack of flexibility and excessive dependency on the initial discretization of the continuous attributes. This paper presents the divisible rough sets as a new hybrid technique of automatic learning able to overcome the problems mentioned using a combination of variable precision rough sets with self-organizing maps and perceptrons. This new technique divides some of the equivalence classes generated by the rough sets method in order to obtain new certain rules under the data which originally were lost. The results obtained demonstrate that this new algorithm obtains a higher decision-making success rate in addition to a higher number of classified examples in the tested data sets.

1 Introduction

The rough sets theory [1], [2] is a powerful tool belonging to the inductive machine learning area which allows the recognition of hidden patterns in data. Additionally, it is capable of assigning uncertainty to the obtained knowledge and of identifying partial or total dependences in databases. The rough sets theory can also be applied to classification problems and to cases where elimination of redundant data is needed.

This paper proposes an improvement of the knowledge extraction technique based on rough sets. The aim is to overcome its inflexibility when dealing with inconsistent data. This is achieved by means of a new hybrid technique, "divisible rough sets", which combines variable precision rough sets [3] with other machine learning techniques such as neural networks, and, in particular with self-organizing maps (SOMs) [4]. Also, the modified Chi2 algorithm [5] is applied in the discretization process.

Previously, some hybrids have been developed involving rough sets and other statistical or machine learning methods improving both the performance and results [6], [7] and connections between rough sets and neural networks have been studied [8]. Specifically, in [9] another combination of SOMs and rough sets, where the first ones extract knowledge from the reducts obtained using rough set theory, is proposed.

The structure of this paper is as follows. First, the theoretical concepts of the divisible rough sets are described. Once the new technique has been detailed, its efficiency in knowledge extraction is tested in some real data sets and its results are then compared with those obtained with variable precision rough sets and self-organizing maps methods. Finally, some conclusions are presented.

2   Divisible Rough Sets Method Definition

Knowledge extraction based on rough sets suffers from some disadvantages, fundamentally, when it deals with very inconsistent data sets. Among them are the lack of flexibility and the excessive dependence on the chosen intervals in the discretization process. The suggested algorithm tries to overcome them improving the learning process, essentially using the information included in the equivalence classes belonging to the boundary region that do not produce any certain rule with the application of the classic rough sets method. In order to achieve this, these equivalence classes are split into several subsets after the clustering of the examples that compose them.

In addition, the divisible rough sets algorithm adds the notion of centre of an equivalence class, obtained from its examples. These centres are useful in the other possible equivalence class division, achieved from new examples examined once the rough sets regions are established, since it is considered that if a certain number of these examples are far from the centre of their class, it will be convenient to divide it.

These two possible equivalence class divisions create equivalence subclasses defined by the discrete values of the available attributes and by new attributes generated by means of mathematical equations where these attributes are numerically expressed. These new subsets will be able to create both certain and uncertain rules.

This analysis of the boundary region is similar to the one done with hierarchy-structured probabilistic tables [10], although here its reduction is carried out by means of new numerical attributes which come from the distances to the centres.

3   Divisible Rough Sets Algorithm

The divisible rough sets method consists of the following steps:

− Step 1: Creation of the decision table: data set examples are distributed in a table.
− Step 2: Initial knowledge extraction: the rough sets method is applied and the results are refined dividing the equivalence classes with clustering.
− Step 3: Knowledge update: via hyperplanes, the examples closer to other equivalence classes than to their own are separated.
− Step 4: Test: rules obtained in the previous steps are tested with new examples.

Now, these steps of the algorithm are briefly described.

3.1   Step 1: Creation of the Decision Table

The aim is to express the examples of the data set from which it is intended to extract knowledge so that they can be processed in the following step of the algorithm. First of all, the decision attribute, which classifies the examples, and the condition attributes, which are the factors a priori considered to be appropriate to make this classification, are selected. The final purpose will be to determine the value of the decision attribute from the information provided by the condition attributes, that is, to get the underlying knowledge rules which govern the relation between these attributes.

3.2   Step 2: Initial Knowledge Extraction

The objective of this step is the discovery of the rules hidden in the data set. In order to do this, the variable precision rough sets method and then clustering processes through self-organizing maps are applied to obtain more certain rules.

The initial knowledge extraction is divided into two phases:

− In the first one, the variable precision rough sets model is applied with a low error.
− In the second one, a clustering process is made in each one of the equivalence classes not included in the positive region in order to achieve new certain rules from equivalence classes that had only generated uncertain rules in the first phase.

At the end of both phases, each example has one more attribute: the distance to the centre of the equivalence class or group which it belongs to. Furthermore, the centres of each group or class will be useful in the following step: the knowledge updating.

The second phase, which constitutes the core of the new method, will be explained. At the end of the first one, the equivalence classes X_i of the partition A^*, which were induced by the equivalence relation of all condition attributes, were assigned to the positive or boundary region of the partition B^* induced by the decision attribute. Depending on the region, each of these equivalence classes will be treated as follows:

Case 1: Class X_i is included in the positive region. A single centre M average of the examples that constitute the equivalence class is calculated.

Case 2: Class X_i is included in the boundary region. One or several centres are calculated from its examples by means of a self-organizing map. Once the appropriate centres of the class X_i are calculated, they will be used to divide X_i in groups.

Let $P = \{p_1, p_2, …, p_n\}$ be the obtained centres set, being each p_j a centre point defined by its values in the condition attributes. Each class X_i will be divided into as many groups as centres make the set P. Let $G = \{G_1, G_2, …, G_n\}$ be the groups set composed from X_i, being each G_j a group which contains a subset of examples of the class X_i with an associated centre point p_j.

In order to assign all the examples belonging to X_i to the different groups of G, the Euclidean distance measure is applied. That is, an example $x \in X_i$ is included in the group G_j with the closer associated centre p_j to x. Therefore, the index j of the group which the example x should belong to, is calculated:

$$j(x) = \arg_j \min \|x - p_j\| \tag{1}$$

Once the examples of the set X_i are distributed into the groups G_j, it is observed that G is a set of non-empty subsets of X_i, $G_1 \cup G_2 \cup … \cup G_n = X_i$ and $G_j \cap G_j = \{\}$ for $i \neq j$. Therefore, G has the required characteristics to be considered a partition of the equivalence class X_i, and consequently G_j can be treated as a block of the partition G. The elements belonging to one of these blocks are characterized by the fact that they have the same discretized values in all the condition attributes and in the added attribute (being closer to the centre associated to the group than to the centres of the rest of the partition blocks). This last attribute makes some examples of an equivalence class to be equivalent and become members of the same group or "equivalence subclass".

Next, for each of the blocks of partition G it is checked to see if it can be included in the positive or boundary region of the partition B^*. The blocks of G belonging to the positive region, will produce certain rules, while the ones of the boundary region, will produce uncertain rules with a confidence factor given by:

$$\alpha_{Y_j}(G_j) = \frac{|G_j \cap Y_j|}{|G_j|} \tag{2}$$

All these rules are more complex because the concept of nearness to the respective centre is added. As a result, a rule created from the group G_j of an equivalence class X_i is: *"if attribute c_1 is xx, attribute c_2 is yy, ..., and attribute c_n is zz, and also the closest centre of the equivalence class to the example is p_j, then attribute d is dd"*.

Centres estimation with self-organizing maps. A SOM is created to determine the centres of the groups of class X_i, which belongs to the boundary region.

The map of neurons is set up with its weight vectors randomly initialized and then each example of X_i is presented. Once the training process ends, the network is used as a classifier of the examples of X_i quantifying the number of examples that activate each neuron. Next, the neuron with the greatest amount of associated examples is selected, and then, also the neurons with at least half this amount. The reference vectors of the selected neurons will be part of the final centres set P used in this phase.

3.3 Step 3: Knowledge Update

The aim of this step is to obtain more certain rules besides those that were produced in the previous step. Thus, the equivalence class areas which are closer to groups of examples of neighbouring equivalence classes are studied with new update examples.

Let $Q = \{x_1, x_2, ..., x_n\}$ be the data set kept aside for knowledge update. For each $x_i \in Q$, with an unknown value for the decision attribute, this process is applied:

First, the equivalence class $X_i \in A^*$ which the example x_i belongs to is determined. Also, if this equivalence class has been divided in the previous phase, then the group G_j, where the example should be included, will be the one with its closest centre.

If the example does not belong to an equivalence class or group included in the positive region, then its Euclidean distances to the centres of the neighbouring classes or groups are calculated. If the nearest centre is closer than the one of its own equivalence class or group, the case is counted.

When a given rate of update examples are closer to another class or group than to their own one, then the original training examples of the class or group will be divided by means of the hyperplane equidistant to the centre of the own set and to the one which the update examples are closest to. This rate is estimated considering the training examples that compose the equivalence class or group.

This hyperplane is equivalent to that defined by a perceptron neuron. Henceforth, in the classes or groups divided like that, an example will be assigned to one of the two subsets depending on the output obtained when it is presented to the perceptron.

Next, it is verified if each of the generated subsets belongs to the positive or to the boundary region and the average centre of those subsets are calculated.

3.4   Step 4: Test

Let $T = \{x_1, x_2, \ldots, x_n\}$ be a set of examples not used in the previous steps. For each example $x_i \in T$ the following process will be carried out:

The first to be determined is the equivalence class $X_i \in A^*$ which the example x_i belongs to. Also, if X_i has been divided in the initial knowledge extraction phase, then the group where the example should be included in will be that with its closest centre. Likewise, if in phase 3 X_i or one of its groups have been divided by one or more hyperplanes, then the example must be presented to the perceptron.

If the set which the example belongs to is included in the positive region, the decision is directly assigned to the example. If it is in the boundary region, its decision will be the one with the highest confidence factor. If the test example does not belong to any equivalence class, the decision will be the one of the equivalence class or group belonging to the positive region with the nearest centre.

4   Experimental Results

In order to test the new proposed algorithm, its results have been compared with those obtained with VPRS and self-organizing maps. Five quite different real data sets from the UCI repository of machine learning databases [11] were used:

Table 1. Analysed Data Sets

Data sets	Characteristics			Success rate		
	Examples	Attributes	Classes	VPRS	SOM	Div. RS
Iris	150	4	3	92.00%	94.00%	94.00%
Liver Disorders	345	6	2	54.78%	61.74%	66.09%
Abalone	4177	4	9	55.46%	55.39%	54.53%
Glass	214	7	6	57.75%	63.38%	66.20%
Diabetes	768	7	2	59.77%	72.66%	70.31%

Each data set is randomly divided into the training set (two thirds), and the test set (one third). The divisible rough sets algorithm requires the additional division of the training set into initial training set (two thirds) used in step 2, and update set (the rest).

In order to carry out the discretizations, a variation of the Modified Chi2 algorithm [5] was applied. The variation proposed allows for the lowering of the initial consistency level so as to obtain equivalence classes with more examples. Other approaches to discretization, e.g. using hyperplanes or soft cuts, are expounded in [12].

The increase of the success rate with regard to VPRS is due to:

1. Test examples not belonging to any equivalence class created in the training process are assigned to the decision of the equivalence class or group belonging to the positive region with the nearest centre.
2. There are more certain rules, which are generated after the equivalence class divisions. As a result, more examples pertain to certain rules and this involves an increase of the confidence in the decision assignment, since the success rate in this case is greater than the one obtained when examples pertained to uncertain rules.

5   Conclusions

This paper proposes a new hybrid technique named "divisible rough sets", which combines a variable precision rough sets method with self-organizing. Additionally, in the data preprocessing phase a variation of the modified Chi2 algorithm was applied. The aim of this variation is the production of more relevant rules.

The objective of this new technique is to obtain certain rules from the uncertain rules provided by the rough sets method. Thus, it tries to increase the number of new examples presented to the knowledge database which can be assigned to a decision considered as certain. In addition, the class assignment to the examples that pertain to uncertain rules or to the examples that do not pertain to any rules (because there was no example similar to them in the training phase) has been allowed.

Finally, as a future extension of the present work the use of Radial Basis Function Networks (RBFN) is suggested instead of the perceptron applied in the knowledge update phase. These networks would allow a non linear division of the examples included in an equivalence class belonging to the boundary region, bringing a subsequent improvement of the success rate in complex data sets.

References

1. Pawlak, Z., "Rough Sets". International Journal of Computer and Information Sciences, vol. 11, no. 5, pp. 341-356. Oct. 1982.
2. Pawlak, Z., Grzymala-Busse, J., Slowinski, R. & Ziarko, W., "Rough Sets". Communications of the ACM, vol. 38, no. 11, pp. 89-95. Nov. 1995.
3. Ziarko, W., "Variable Precision Rough Set Model. Journal of Computer and System Sciences", vol. 46, no. 1, pp. 39-59. Feb. 2003.
4. Kohonen, T., "Self-organizing Formation of Topologically Correct Feature Maps". Biological Cybernetics, vol. 43, no. 1. pp. 59-69. 1982.
5. Shen, L., Tay, F.E.H., "A Discretization Method for Rough Sets Theory". Intelligent Data Analysis, vol. 5, pp. 431-438. 2001.
6. Browne, C., Düntsch, I., Gediga, G., "IRIS Revisited, A Comparison of Discriminant and Enhanced Rough Set Data Analysis". In: Polkowski, L., Slowron, A. (Eds.), "Rough Sets in Knowledge Discovery", Heidelberg: Physica Verlag, vol. 2, pp. 345-368. 1998.
7. Banarjee, M., Mitra, S., Pal, S.K., "Rough Fuzzy MLP: Knowledge Encoding and Classification". IEEE Transactions on Neural Networks, vol. 9, pp. 1203-1216. 1998.
8. Nguyen H. S., Szczuka, M.S., Slezak, D., "Neural Networks Design: Rough Set Approach to Continuous Data". Principles of Data Mining and Knowledge Discovery. First European Symposium, PKDD '97. Proceedings. Berlin, Germany: Springer-Verlag, pp. 359-366. 1997.
9. Pal, S. K., Dasgupta, B., Mitra, P., "Rough Self Organizing Map". Applied Intelligence, vol. 21, no. 3, pp 289-299. 2004.
10. Ziarko, W., "Acquisition of Hierarchy-Structured Probabilistic Decision Tables and Rules from Data". Expert Systems, vol. 20, no. 5, pp. 305-310. Nov. 2003.
11. Merz, C.J., Murphy, P.M., UCI repository of machine learning databases. http://www.ics.uci.edu/~mlearn/MLRepository.html. Irvine, CA: University of California, Department of Information and Computer Science. 1996.
12. Nguyen, H. S., Nguyen, S. H., "Discretization Methods in Data Mining". In: Polkowski, L., Slowron, A. (Eds.), "Rough Sets in Knowledge Discovery", Heidelberg: Physica Verlag, vol. 1, pp. 451-482. 1998.