Obtaining a Fuzzy Classification Rule System from a Non-Supervised Clustering

Waldo Hasperué [1], Germán Osella Massa [2], Laura Lanzarini
III-LIDI (Institute of Research in Computer Sciences LIDI)
School of Computer Sciences, National University of La Plata, La Plata, 1900, Argentina
[1] Waldo Hasperué has a CIC Scholarship
[2] Germán Osella Massa has a Type 1 PhD Scholarship of CONICET.
{whasperue,gosella,laural}@lidi.info.unlp.edu.ar

Abstract. *The fuzzy classification systems have been broadly used to solve control and decision-making problem. However, its design is complex, even when having a human expert assistance. This paper presents a new strategy capable of automatically defining the corresponding Fuzzy Classification Rule System from a non-supervised clustering of the available data. Its application to three data sets of the UCI repository has given quite satisfactory results.*

Keywords. Fuzzy Classification Rule, Non-supervised Clustering, Competitive Dynamic Neuronal Nets.

1. Introduction

The systems based on fuzzy rules have been successfully used to solve classification problems in different areas. [11] [12] [14] [17]. Their capacity to handle uncertainty, allow them to model inaccurate knowledge, providing information close to human reasoning. These fuzzy models represent a description of user-oriented data through a set of qualitative rules that establish significant and useful relations among variables. The terms used in the rules are characterized by fuzzy sets. These sets allow establishing flexible limits among the different levels of meaning as it occurs with human perception, without ignoring or giving excessive emphasis to frontier elements [3]. However, it is not always easy to design a fuzzy system capable of answering in the desired manner. The choice of the linguistic concepts in which the range of each variable is divided into and their combination in the rules that constitute the basis of knowledge can be a complex task even when having a human expert assistance.

Several works describing extraction of classification rules using fuzzy sets can be found in the literature [1] [5] [7] [8] [9] [10] [13] [15] [18]. These works differ from one another both in the way they obtain the clusters from the original database and in the algorithm used for rule extraction.

This paper presents a new strategy capable of automatically defining the corresponding Fuzzy Classification Rule System from a non-supervised clustering of the available data. To make the non-supervised clustering a competitive dynamic neuronal network trained with the AVGSOM method is used [2]. Once said training is over, the strategy proposed in this paper uses the prototype vectors of each competitive neuron to define the fuzzy sets and to reduce the rule antecedent dimension.

2. Fuzzy system structure

The fuzzy system to be obtained is made up of such rules as:

R_i: *if* x_1 *is* A_{i1} *and ...* x_n *is* A_{in} *then* g_i ; $i=1:M$ (1)

where the rule antecedent is a diffuse description of the input space and the consequent is an indicator of a hard Cluster (non-fuzzy). The i-esimal rule activation degree, d_i , is calculated as the product of the belonging degree of each variable of the input pattern X in the respective fuzzy set as follows:

$$d_i(X) = \prod_{j=1}^{n} A_{ij}(x_{ij}) \qquad (2)$$

The experiences carried out showed that the expression indicated in (2) has a more accurate behavior than its pair expression - based on the minimal function – regarding the activation degree measuring of a diffuse rule.

The response of the fuzzy system will be the one corresponding to the rule bearing the highest activation degree as shown in (3).

Proceedings of the *ITI 2008 30th Int. Conf. on Information Technology Interfaces,* June 23-26, 2008, Cavtat, Croatia

$$y = g_{maxi}\ ;\ d_{maxi} = max(d_i)\quad with\ i=1:M \qquad (3)$$

3. Rule system generation

3.1. Clustering

To make the non-supervised clustering a competitive dynamic neuronal network trained with the AVGSOM [2] method is used whose main features are:

- The net structure is a graph formed by interconnected competitive neurons. The connection consists of a grid where each neuron has at most four neighbors.
- Each net neuron has its own prototype vector, which tries to represent a data set of similar input. The degree of similarity to use depends on the problem.
- Training is done through a competitive process where the neurons try to represent the input data. For each datum, its similarity with the prototype vector of each neuron is evaluated, the nearest being the winner. Adaptation is mainly applied to the winning neuron and in a lesser degree to its nearest environment. This allows correcting the structure so as to preserve topology.
- In each adaptation step, the local error information is accumulated in the winning neuron. Its goal is to prevent that the same net element accumulates the majority of the input pattern representation. The error estimation depends on the application.
- The accumulated error information is used to determine where the new units should be inserted in the net. When performing an insertion, the error information is locally redistributed, thus preventing new insertions in the same placer.

After the training, each competitive neuron becomes a representation of a set of input patterns and it will be used as such in the rule antecedent construction.

The number of clusters to be formed will have a direct relation to the error threshold set as boundary?

The use of a dynamic neuronal net trained with AVGSOM allows to accurately setting the initial clusters, a starting point to determine the classification rules. This refers not only to the prototype vector location, which is set following the relation existing among neurons in the net, but also to the number of clusters to be used.

This is a typical feature of Dynamic neuronal networks. This type of nets determines the size of the optimal structure via a constructive method arising from a minimal architecture and adding elements according to the measuring of errors that accumulate in successive interactions.

For further information on the way in which the neuronal net training is done, we recommend to consult [2].

3.2. Fuzzy set definition

As a result of the clustering process, the set $G= \{g_1, ...g_W\}$ containing the clustering of the input patterns is obtained. For each cluster g_k, the corresponding hypercube H_k shall be obtained in the following manner:

$$H_k = (H_{k1}, H_{k2}, \ldots, H_{km})$$
$$H_{kz} = (a_{kz}, b_{kz})\ \forall z \in 1..m$$

where m shows the input vector length and akz and b_{kz} contains the minimal and maximal values found in attribute z of the training patterns corresponding to cluster k, respectively.

From each k-dimensional H_i hypercube, a fuzzy set is generated for each input space variable. That is, the discourse universe of each variable j shall be covered by as many fuzzy sets as hypercubes presents, which may or may not overlap.

Be $\{[a_{1j}, b_{1j}], ..., [a_{ij}, b_{ij}], ...,[a_{mj},b_{mj}]\}$ the intervals set for the minimal and maximal values of each W hypercube in dimension j. Each one of these intervals, $[a_{ij}, b_{ij}]$, shall allow for the determination of a fuzzy set F_{ij} having a belonging function μ_{Fij} defined as follows:

$$if\ a_{ij}>b_{i-1j}\ then\ L_{ij}=b_{i-1j}\ or\ else\ L_{ij} = a_{i-1j}$$

$$if\ b_{ij}<a_{i+1j}\ then\ U_{ij}=a_{i+1j}\ or\ else\ U_{ij} = b_{i+1j}$$

$$\mu_{Fij} = \begin{cases} 0 & if\ x \le L_{ij} \\ (x-L_{ij})/(a_{ij}-L_{ij}) & if\ x \in (L_1, a_{ij}) \\ 1 & if\ x \in [a_{ij}, b_{ij}] \\ (U_{ij}-x)/(U_{ij}-b_{ij}) & if\ x \in (b_{ij}, U_{ij}) \\ 0 & if\ x \ge U_{ij} \end{cases}$$

3.3. Rule extraction

The algorithm for rule extraction consists in determining the most relevant variables when identifying the clusters. This is equivalent to determining the variables that have less overlapping among their fuzzy sets.

Being $[L_{ij}, a_{ij}, b_{ij}, U_{ij}]$ and $[L_{hj}, a_{hj}, b_{hj}, U_{hj}]$ the limits of fuzzy sets F_{ij} and F_{hj} respectively, the *exclusiveness rate* IE_{ihj} is defined between two fuzzy sets of the universe of variables j, F_{ij} and F_{hj}, as

$$C_{ihj} = \begin{cases} (b_{ij} - a_{hj}) / base & a_{hj} \in (a_{ij}, b_{ij})\,;\, b_{hj} > b_{ij} \\ (b_{hj} - a_{ij}) / base & b_{hj} \in (a_{ij}, b_{ij})\,;\, a_{hj} < a_{ij} \\ 1 & [a_{hj}, b_{hj}] \subseteq [a_{ij}, b_{ij}] \\ 0 & \text{another case} \end{cases}$$

$$IE_{ihj} = max(C_{ihj}, C_{hij}) \tag{4}$$

where

$$base = min((b_{ij} - a_{ij}), (b_{hj} - a_{hj})) \tag{5}$$

The rate defined in equation (4) allows to measure the overlapping degree between two fuzzy sets by taking values between 0 and 1, corresponding the highest value to the largest possible degree of overlapping.

The algorithm proposed in this paper begins construing the antecedent of the first rule from variable *j,* which allows to distinguish the largest number of possible clusters. That is to say, the rule having a fuzzy set F_{oj} with the largest number of IE_{ohj}=0 for $h \neq j$.

Thus, the first condition of the first rule begins its building up as follows

$$If\ (name\_var\_j\ is\ Vo) \tag{6}$$

where *name_var_j* is the name of the selected variable j and V_o is the linguistic value associated with the fuzzy set F_{oj}.

If the fuzzy set F_{oj} does not present overlapping with anyone of the other sets within the universe of discourse of variable *j*, the rule would be finished and the consequent should be the cluster that gave origin to set F_{oj} as shown in (7).

$$if\ (name\_var\_j\ is\ Vo)\ then\ Cluster(F_{oj}) \tag{7}$$

But this situation is rarely frequent, thus, it will be necessary to analyze the remaining variables of the clusters not yet properly identified by the antecedent of the rule under construction (6). The selection of the following variable to be use in the rule antecedent is done in a similar manner, considering only the remaining clusterings. The algorithm in fig. 1 illustrates this process.

In this way, a rule for each hypercube is at least obtained, that is, a rule for each cluster.

```
G = {g1, g2, …, gW} set of W clusters
      obtained from the neuronal net
      trained with AVGSOM.
Hk = (Hk1, Hk2,…, Hkm ) ∀gk∈G
i=0
Evaluate IEihj with (4)j=1..m;
                           i,h =1..W, i≠h
R1kj = Number of pairs (Fkj,Fhj) with
       h=1..W, h≠k, IEkhj=0; j,k=1..W
R2kj = sum(IEkhj);h=1:W,h≠k; j,k=1:W
Repeat
  I={Fst / R1st =min(R1kj);k,j=1:W}

  Obtain Fop / R2op = min(R2sj);Fsj∈I

  % begins generation of the rule
  % corresponding to the cluster o

  % Use Fop to generate the rule for
      cluster o
  G = G – {go}
  Z = {gz / IEoz = 0}
  Rem_Clusters = G – Z
  Cond=(feature_p is val_lin(Fop))
  while (#Rem_Clusters > 0)
    I={Fot / R1ot = min(R1oj); gj ∈
       Rem_Clusters }
    Obtain Foq / R2oq = min(R2oj);Foj∈I
    Rem_Clusters = Rem_Clusters – {gq}
    Cond = cond + "AND" + (feature_q is
       val_lin(Foq))
    end while
  Rules[++i] = "IF"+cond+" is
       cluster"+gq
Until ( #G=0 )
```

Figure 1. Algorithm of the proposed method.

4. Results

The rule extraction method presented in this paper was tested with two sets of data of the UCI's repository: Iris Plants Database and Wine recognition data [4].

The experiment consisted in training a dynamic competitive neural network with the AVGSOM method.

As the antecedents of the rules extracted by the proposed method are made up of arbitrary linguistic values of V_1, V_2, etc. type, in the results shown in the following tables, such values have been replaced by others, more suitable for the problem. For instance, in the case of Iris data base, *Low, Mid* and *High* values have been used.

Table 1 shows the results obtained for each set of data. In each case, both the number of input

Table 1. Classification using the proposed method.

Data bases and their classes	Training with the complete set		Training with 2/3 of the set: Training		Training with 2/3 of the set: Test	
	bc / N	%	bc / N	%	bc / N	%
Iris						
Iris versicolor	48/50	96%	27 / 28	96,43%	21/22	95,45%
Iris virginica	50/50	100%	38/38	100%	12/12	100%
Iris setosa	50/50	100%	34/34	100%	16/16	100%
Total	**148/150**	**98.67%**	**99/100**	**99%**	**49/50**	**98%**
Wine						
A	59/59	100%	38/38	100%	21/21	100%
B	70/71	98.6%	49/51	96%	19/20	95%
C	48/48	100%	29/29	100%	18/19	94.7%
Total	**177/178**	**94.94%**	**116/118**	**98,3%**	**58/60**	**96.7%**
Glass						
1	65/70	92.8%	39/46	84.8%	20/24	83.3%
2	75/76	98,7%	51/52	98.1%	23/24	95.8%
3	11/17	64.7%	8/11	72.7%	4/6	66.7%
5	13/13	100%	9/10	90%	3/3	100%
6	7/9	77,8%	4/4	100%	5/5	100%
7	29/29	100%	19/19	100%	10/10	100%
Total	**200/214**	**93.5%**	**130/142**	**91.5%**	**65/72**	**90.3%**

patterns and the number of patterns properly classified from the rules obtained have been indicated. Two tests were carried out; the first one used the complete set of input data for training AVGSOM and the same set of data for testing the rules. In the second test, two thirds of the set of input data were selected at random in order to carry out the training, using the remaining third for testing the rules obtained.

Table 2 shows the rules built from the found clusters for two of the sets of data using the proposed method.

Finally, Table 3 compares the results of the proposed method with those presented in [16] and [6] for the wine database. In each case, the best classification percentage, the number of rules used, and the number of variables per rule have been evaluated. As it can be seen, the proposed method permits to obtain a high percentage of properly classified data, by using a reduced number of rules. Note that [6] shows a similar value with 60 rules while the method proposed in this paper, though equal to [16], uses only 3 rules.

The difference between the method indicated in [16] and the one used in this paper lies in the use of a non-supervised clustering strategy, thus making input data labeling unnecessary.

Moreover, if the number of variables making up the antecedent of each fuzzy rule is observed, the method proposed here requires a lower number than [16].

5. Conclusions

A new strategy capable of automatically defining the corresponding fuzzy classification rule system from the data available has been presented. Its application to three data sets of the UCI repository has given quite satisfactory results.

The number of rules to obtain strongly depends on the number of initial clusters originating from the input data. This value is determined by the result of the non-supervised clustering based on a dynamic neuronal network trained with the AVGSOM method.

The comparison of the results obtained with [16] and [6] for the wine database illustrates its capacity to produce good results with a reduced number of rules. Besides, the building up of each rule antecedent with a reduced number of linguistic variables makes comprehension easier.

Table 2. Extracted Rules.

Base	Extracted Rules
Iris	If (Petal-length is low) then C1 Else If (Petal-width is low) and (Petal-Length id middle) then C2 Else C3
Wine	If (OD280/315 of diluted wines is high) and (Proline is high) and (Color intensity is middle) then C1 Else If (Alcohol is low) or (Hue is high) or (Flavanoids is high) or (Malic acid is middle) then C2 Else C3

Table 3. Comparison with other methods

Method	Best result	# rules	#variables/ rule
Ishibuchi	99.4 %	60	-
Roubos	99.4 %	3	4, 5, 5
This paper	99.4 %	3	3, 4, 4

6. References

[1] Chen J-M, Wu S-H, Lee H-M. The study of automatic insertion and deletion of fuzzy rules in fuzzy neural network models. Proceedings of the 1996 Asian Fuzzy Systems Symposium, 'Soft Computing in Intelligent Systems and Information Processing'; 1996 Dec 11-14. 1996. p. 248-253.

[2] Hasperué W, Lanzarini L. Dynamic Self-Organizing Maps. A new strategy to upgrade topology preservation. XXXI Congreso Latinoamericano de Informática CLEI; 2005 Oct 10-14; Cali, Colombia. p. 1081-1087.

[3] Hernández Orallo J, Ramírez Quintana MJ, Ferri Ramírez C. Introducción a la minería de datos. In Fayerman Aragón, D. Pearson, editors. Prentice Hall; 2004.

[4] Hettich S, Blake CL, Merz CJ. UCI Repository of machine learning databases. University of California, Irvine, Dept. of Information and Computer Sciences; 1998. http://archive.ics.uci.edu/ml/ [04/30/2008].

[5] Hong T-P, Lee Y-C. Mining coverage-based fuzzy rules by evolutional computation. Proceedings IEEE International Conference on Data Mining; 2001 Nov 29 – Dec 2. 2001. p. 218 – 224.

[6] Ishibuchi H, Nakashima T, Murata T. Performance evaluation of fuzzy classifier systems for multidimensional patter classification problems. IEEE Transactions on Systems, Man, and Cybernetics, Part B 1999; 29(5): 601-618.

[7] Ishibuchi H, Nakashima T. Effect of rule weights in fuzzy rule-based classification systems. The Ninth IEEE International Conference on Fuzzy Systems; 2000 May 7-10. 2000. p. 59-64; vol. 1.

[8] Ishibuchi H, Yamamoto T. Heuristic extraction of fuzzy classification rules using data mining techniques: an empirical study on benchmark data sets. IEEE International Conference on Fuzzy Systems; 2004 Jul 25-29. 2004. p. 161-166; vol. 1.

[9] Ishibuchi H, Nojima Y, Kuwajima I. Genetic Rule Selection as a Postprocessing Procedure in Fuzzy Data Mining. International Symposium on Evolving Fuzzy Systems; 2006 Sep. p. 286–291.

[10] Ishibuchi H. Evolutionary Multiobjective Design of Fuzzy Rule-Based Systems. IEEE Symposium on Foundations of Computational Intelligence; 2007 Apr 1-5. 2007. p. 9–16.

[11] Kitano H, Nakashima T, Ishibuchi H. Behavior analysis of futures trading agents using fuzzy rule extraction. IEEE International Conference on Systems, Man and Cybernetics; 2005 Oct 10-12. 2005 p. 1477-1481; Vol. 2.

[12] Koczy LT, Botzheim J. Fuzzy models, identification and applications. IEEE 3rd International Conference on Fuzzy models, identification and applications Computational Cybernetics; 2005 Apr 13-16. 2005. p. 13–19.

[13] Lee H-M, Chen J-M, Chang E-C. Efficiency validation of fuzzy domain theories using a neural network model. Tenth IEEE International Conference on Tools with Artificial Intelligence; 1998 Nov 10-12. 1998. p. 81-88.

[14] Mitra S, Hayashi Y. Neuro-fuzzy rule generation: survey in soft computing framework. IEEE Transactions on Neural Networks 2000. 11(3): 748–768.

[15] Pei Z. A Formalism to Extract Fuzzy If-Then Rules from Numerical Data Using Genetic Algorithms. International

Symposium on Evolving Fuzzy Systems; 2006 Sep. 2006. p. 143–147.
[16] Roubus H, Magne S, Janos A. Learning Fuzzy Classification Rules from Data. in RASC conference, Leichester, UK.
[17] Wong KW, Gedeon TD. Fuzzy rule interpolation for multidimensional input space with petroleum engineering application. IFSA World Congress and 20th NAFIPS International Conference; 2001 Jul 25-28. 2001. p. 2470–2475; Vol. 4.
[18] Zhou S-M, Gan JQ. Constructing L2-SVM-Based Fuzzy Classifiers in High-Dimensional Space with Automatic Model Selection and Fuzzy Rule Ranking. IEEE Transactions on Fuzzy Systems 2007; 15(3): 398–409.