Fuzzy Modelling using Fuzzy Mathematical Programming

J.M. Cadenas **M.C. Garrido** **J.J. Hernández**
Dpto. Ingeniería de la Información y la Comunicaciones
Facultad de Informática. Universidad de Murcia
30071-Espinardo. Murcia (Spain)
jcadenas@dif.um.es mgarrido@dif.um.es jjha1@alu.um.es

Abstract

In this paper we will describe several methods that solve, using mathematical programming, different problems that, although can be used in others fields in the fuzzy modelling context, we have used them in the of fuzzy classification context. These problems are: feature selection, rule selection and clustering. In order to evaluate the accuracy of these methods, a lot of data sets have been used. Two of these have been: the classical Iris data set and the Wisconsin Breast Cancer data set.

Keywords: Mathematical Programming, Fuzzy Mathematical Programming, Fuzzy Modeling.

1 Introduction

Mathematical programming is a broad discipline that has been applied to a greet variety of problems. In the context of Artificial Intelligence, one of these problems is the fuzzy rules learning. In this paper we will describe three methods, solved using mathematical programming, which will be applied inside a process of fuzzy rules learning. Figure 1 summarize this process.

In this process we will first apply the clustering method. This method solves the problem of doing a particional fuzzy clustering. We will formulate a fuzzy mathematical problem, using an objective function different to the function used by the Fuzzy C-Means (FCM) methods, which give us a set of fuzzy clusters. The optimization techniques used to find optimal solution will let to find solutions better than FCM can find. Once we have found the clusters we will use them to generate the fuzzy rules. The rules will be as: "If Antecedent Then conclusion", where the antecedent will be the fuzzy cluster and the conclusion will be the class which the cluster is labelled. These rules will be used by the another two proposed methods, which are situated in the of rule optimization context. One of them solves the problem of selecting features from fuzzy rules. We will formulate a mathematical programming problem that will be able to select features that the Mangasarian's method [6] isn't able to select.

Finally, we will apply the last proposed method, which will be able to remove rules that aren't important to the classification task. We will formulate a fuzzy mathematical problem, which give us information about what rules are really important to the classification task.

This paper is organized in the followed way: section 2 presents the clustering method, section 3 presents the feature selection method and section 4 presents the rule selection method. Section 5 will show the results obtained with Iris and Wisconsin Breast Cancer and finally the paper conclude with the conclusions and future work in section 6.

2 Fuzzy clustering

In this section we are going to carry out a fuzzy clustering upon a data set. The same objective function that FCM uses could have been used here, but we have used a different one, removing the fuzziness index characteristic of this function (which is a very little intuitive parameter) and introducing a parameter called proximity distance.

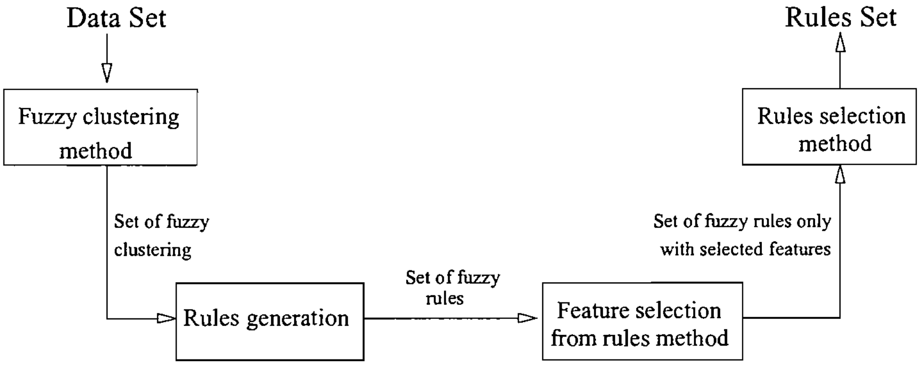

Figure 1: Fuzzy rules learning process

The proximity distance extends the idea of classic k-mean algorithm. In this algorithm only is added the distance of the example from the nearest center.

The idea that we propose is that every example must add with a center if the distance of this centre from the example is smaller than the distance of the example from the nearest centre added with the proximity distance. These distances have a membership value, which describes the degree of belonging of the example to the cluster. These degrees of belonging are normalized.

The fuzzy mathematical programming problem that solves this is:

$$
\begin{aligned}
&Min \quad \sum_{i=1}^{m}\sum_{j=1}^{k} d_{ij}^2 \frac{p_{ij}}{n_i} \\
&s.t.: \\
&\left.\begin{aligned}
&n_i = \sum_{j=1}^{k} p_{ij} \\
&p_{ij} = max(0, (min_i - d_{ij}) + t_{ij}) \\
&t_{ij} \gtrsim 0
\end{aligned}\right\} \forall\, i,j
\end{aligned}
$$

with $\min_i = \min_{j=1,..,k}\{d_{ij}^2\}$

Where d_{ij} is the distance of the i-th example to the j-th center, and p_{ij} is a membership value which describes the degree of belonging (still not normalized) of the i-th example to the j-th cluster. In order to normalize the degrees of belonging, these are divided by n_i. The fuzzy constraint controls the proximity distance value (t_{ij}). When this distance is 0 then we are in the case of classic k-mean problem. When this distance is bigger than 0 then we are in the case of fuzzy clusters.

Using the solution of the problem, the fuzzy clusters are generated. This clusters are defined by a mean and a covariance matrix. Using the fuzzy clusters the rules are generated. The rules are as: "If Antecedent Then conclusion", where the antecedent is the fuzzy cluster and the conclusion is the class which the cluster is labelled.

3 Feature selection from rules

In this section we are going to select features from the rules generated in the previous section. The rules that we have used have been Gaussian defined by a mean (m) and a covariance matrix. However, the idea that we propose can be generalized to every type of rule. We select features looking for a linear discriminant between each pair of rules that have a different conclusion, that is, labelled with a different class, and that use as few of the features as possible. At least, the features that use the discriminant will be the features that we will select. The error that a discriminant commits separating two rules is the degree of belonging of the discriminant's point that have the biggest degree of belonging to any of the two rules. In order to determine this, we have to transform the Gaussian to Gaussian with mean equal to 0 and variance equal to 1. This transforms the points from original space to the destiny space in the following way, $y = \Lambda^{1/2}V^t(x-m) = \Theta(x-m)$, where x represents the points of the original space and y the points of the destiny space. Λ is the eigenvalues matrix of the covariance matrix of the

Gaussian, and V is the eigenvectors matrix. Calling C_i to the center of the fuzzy set A_i of the rule R_i, w_{ij} and γ_{ij} to the discriminant between the sets A_i and A_j, and θ_i to the matrix θ of the fuzzy set A_i, which can be computed before beginning the optimization process, the mathematical programming problem can be formulated in the following way:

$$
\begin{aligned}
Min \quad & (1-\lambda)\left[\sum_{i=1}^{n-1} \sum_{\substack{j=i+1 \\ class(j) \neq class(i)}}^{n} e^{\frac{-min\left(\frac{\gamma 1_{ij}^2}{||w1_{ij}||^2}, \frac{\gamma 2_{ij}^2}{||w2_{ij}||^2}\right)}{2}}\right] \\
+ \quad & \lambda \left[(I - e^{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} v_{ij}}) I^t \right]
\end{aligned}
$$

$s.t.:$

$$
\left.
\begin{array}{l}
\gamma 1_{ij} = \gamma_{ij} - w_{ij}C_i, \\
\gamma 2_{ij} = \gamma_{ij} - w_{ij}C_j, \\
-w_{ij}C_i + \gamma_{ij} + 1 \leq 0 \\
w_{ij}C_i - \gamma_{ij} + 1 \leq 0 \\
-v_{ij} \leq w_{ij} \leq v_{ij}
\end{array}
\right\} \begin{array}{l} \forall \text{ i,j, } \; j>i \\ \text{class(j)} \neq \text{class(i)} \end{array}
$$

Where class(i) is the class that represents the conclusion of the rule R_i, I is the identity matrix and $\lambda \in [0, 1)$.

4 Rule selection

We select rules looking for a reduced set of fuzzy rules that have so much cover, or almost so much cover, as the original set of fuzzy rules. The fuzzy mathematical programming problem that solves this is:

$$
\begin{aligned}
Min \quad & I^T(1 - e^{\alpha v}) \\
s.t.: & \\
& maxcover - \sum_{j=1}^{n} \max_{j=1,..,k}\{(1 - e^{\alpha v_j})\mu_{ij}\} \lesssim 0 \\
& v \geq 0
\end{aligned}
$$

The first constraint is the fuzzy constraint. Maxcover represents the cover of the whole set of rules, and $\sum_{i=1}^{n} \max_{j=1,..,k}\{(1 - e^{-\alpha v_j})\mu_{ij}\}$ represents the cover of the selected rules, where μ_{ij} is the degree of belonging of the i-th example to the j-th cluster and $\alpha > 0$ (a good value is α=5).

5 Experiments

We have applied the learning process with two sets: the classical Iris data set (IRIS) and the Wisconsin Breast Cancer data set (WBC). A table that summarize the results obtained is:

	Our Process	FCM	GC-Anfis
IRIS	3 rules, 2 features → 5 errors 6 rules, 2 features → 4 errors	4 rules → 15 errors	4 rules → 3 errors
WBC	3 rules,4 features → 22 errors 8 rules,4 features → 20 errors		13 rules → 7 errors

	FCM+DE	GC-Anfis+HD+GA	MFGN [1]
IRIS			6 rules → 3 errors
WBC	104 rules → 25 errors	9 rules → 16 errors	

Note 1: The mixture components of MFGN model, [7], are consideres as rules in order to compare results.

6 Conclusions

In this paper we have described several methods that can be used in the learning context. These methods solve three types of problems in this area: feature selection, rule selection and clustering. In order to find solutions to these problems, we have used fuzzy mathematical programming. The described solutions show that we can solve problems within the cognitive learning context in an effective way, using the formalism that fuzzy mathematical programming give us. With this formalism we only have to be worried about defining the optimization function, since the searching solutions process is implemented by the problem resolution technique that we are using. It saves a lot of design and implementation time. We can see that the results obtained have been good. Methods like GC-Anfis obtain better results than ours because they carry out a supervised adjust process that we don't carry out.

Acknowledgements

The authors thank the Comision Interministerial de Ciencia y Tecnologa (CICyT) for the support given to this work under the projects 1FD97-0255-C03-01 and TIC2000-0062-P4-03.

References

[1] J.M. Cadenas, J.L. Verdegay (1999). Modelos de Optimización con datos Imprecisos. Servicio de Publicaciones. Universidad de Murcia. 1999.

[2] M. Delgado, A. Gómez-Skarmeta, F. Martín (1995). Un enfoque aproximativo para la generación de reglas mediante análisis cluster. V Congreso Español sobre Tecnologías y lógica fuzzy, pp.43-48. 1995.

[3] A. Flores, J.M. Cadenas, F. Martín (1998). A local geometrical propierties application to fuzzy clustering. Int. Jour. Fuzzy Sets and Systems 100, pp. 245-256. 1998.

[4] A. Flores, J.M. Cadenas, F. Martn (1999). Membership Function in the Fuzzy C-Means algorithm. Int. Jour. Fuzzy Sets and Systems 101 (1), pp. 49-58. 1999.

[5] H. Liu, H. Motoda (1998). Feature selection for knowledge discovery and data mining. Kluwer Academic Publishers. 1998.

[6] O.L. Mangasarian (1997). Mathematical Programming in Data Mining. Data Mining and Knowledge Discovery 1, 183-201. Kluwer Academic Publishers. 1997.

[7] A. Ruiz, P.E. Lopez de Teruel, M.C. Garrido (1999). Probabilistic Inference from Arbitrary Uncertainty using Mistures of Factorized Generalized Gaussians. Jorunal of Artificial Intelligent Research, vol. 9, pp. 167-217. 1998.