Proceedings of the 2007 IEEE Symposium on Computational
Intelligence in Multicriteria Decision Making (MCDM 2007)

Multiobjective Genetic Algorithm for Extracting Subgroup Discovery Fuzzy Rules

María José del Jesus, Pedro González, and Francisco Herrera

Abstract— **This paper presents a multiobjective genetic algorithm for obtaining fuzzy rules for subgroup discovery. This kind of fuzzy rules lets us represent knowledge about patterns of interest in an explanatory and understandable form which can be used by the expert.**

The multiobjective algorithm proposed in this paper defines three objectives. One of them is used as a restriction on the rules in order to obtain a Pareto front composed of a set of quite different rules with a high degree of coverage over the examples. The other two objectives take into account the support and the confidence of the rules. The use of the mentioned objective as restriction allows us the extraction of a set of rules which describe more complete information on most of the examples.

Experimental evaluation of the algorithm, applying it to a market problem shows the validity of the proposal obtaining novel and valuable knowledge for the experts.

I. Introduction

KNOWLEDGE Discovery in Databases (KDD) is defined as the non trivial process of identifying valid, original, potentially useful patterns which have comprehensible data [1]. Within KDD process the data mining stage is responsible for high level automatic knowledge discovery from information obtained from real data.

A data mining algorithm can discover knowledge using different representation models and techniques from two different perspectives:

- *Predictive induction*, whose objective is the discovery of knowledge for classification or prediction [2].
- *Descriptive induction*, whose main objective is the discovery of interesting knowledge from data. In this area, attention can be drawn to the discovery of association rules following an unsupervised learning model [3], subgroup discovery [4] [5] and other approaches to non-classificatory induction.

In a subgroup discovery algorithm the objective is, given a set of data and a specific property of them we are interested in, find population subgroups that are statistically "most interesting", e.g., are as large as possible and have the most unusual distributional characteristics with respect to the property of interest.

The induction of rules describing subgroups can be considered as a multi-objective problem rather than a single objective one, in which the measures used for evaluating a rule can be thought of as different objectives of the subgroup discovery rule induction algorithm. In this sense, MOEAs are adapted to solve problems in which different objectives must be optimized. In the specialized bibliography several evolutionary proposals for multiobjective optimization can be found [6] [7]. Recently the MOEAs have been used in the extraction of knowledge in data mining [8] [9].

This paper describes a new proposal for the induction of fuzzy rules which describe subgroups based upon a multiobjective evolutionary algorithm (MOEA) which combines the approximated reasoning method of the fuzzy systems with the learning capacities of the genetic algorithms (GAs).

The multiobjective algorithm proposed in this paper defines three objectives. Two of them take into account the support and the confidence of the rules. The third objective is used as a restriction for directing the evolutionary process towards the obtaining of rules which describe information related to examples not described by the previously obtained rules. The use of the last objective promotes the obtaining of rules belonging to different parts of the search space.

The paper is arranged in the following way: Section 2 describes some preliminary concepts. The multiobjective evolutionary approach to obtain subgroup discovery descriptive fuzzy rules is explained in Section 3. Finally, Section 4 shows the experimentation carried out and the analysis of results and Section 5 outlines the conclusions and further research.

II. Preliminaries

A. Subgroup Discovery

Subgroup discovery represents a form of supervised inductive learning in which, given a set of data and having a property of interest to the user, attempts to locate subgroups which are statistically "most interesting" for the user. In this sense, a subgroup is interesting if it has an unusual statistical

Manuscript received October 30, 2006. This work was supported by the Spanish Ministry of Science and Technology and by the European Fund. FEDER under Projects TIC-2005-08386-C05-01 and TIC-2005-08386-C05-03, and the nets TIN2004-20061-E and TIN2004-21343-E.

M.J. Del Jesus is with the Department of Computer Science, University of Jaén, Las Lagunillas s/n, 23071 Jaén (Spain) (e-mail: mjjesus@ujaen.es).

P. González is with the Department of Computer Science, University of Jaén, Las Lagunillas s/n, 23071 Jaén (Spain) (e-mail: pglez@ujaen.es).

F. Herrera is with the Department of Computer Science and Artificial Intelligence, University of Granada, C/ Periodista Daniel Saucedo Aranda, s/n. 18071 Granada (Spain) (e-mail: herrera@decsai.ugr.es).

distribution respect of the property of interest. The methods for subgroup discovery have the objective of discover interesting properties of subgroups obtaining *simple* rules (i.e. with an understandable structure and with few variables), *highly significant* and with *high support* (i.e. covering many of the instances of the target class).

An induced subgroup description has the form of an implication,

$$R^i: Cond^i \rightarrow Class_j$$

where the property of interest for subgroup discovery is the class value $Class_j$ that appears in the rule consequent, and the rule antecedent $Cond^i$ is a conjunction of features (attribute-value pairs) selected from the features describing the training instances.

The concept of subgroup discovery was initially formulated by Klösgen in the rule learning algorithm EXPLORA [4] and by Wrobel in the algorithm MIDOS [5]. In the specialized bibliography, different methods have been developed which obtain descriptions of subgroups represented in different ways and using different quality measures, as SD [10], CN2-SD [11] or APRIORI-SD [12] among others.

One of the most important aspects of any subgroup discovery algorithm is the quality measures to be used, both to select the rules and to evaluate the results of the process. It can be distinguished between objective and subjective quality measures. Some of the most used objective quality measures for the descriptive induction process are:

- *Coverage for a rule* [11]: measures the percentage of examples covered on average by one rule R^i of the induced rule set.

$$Cov(R^i) = Cov(Cond^i \rightarrow Class_j) = \\ = p(Cond^i) = \frac{n(Cond^i)}{n_s} \quad (1)$$

  where $n(Cond^i)$ is the number of examples which verifies the condition $Cond^i$ described in the antecedent (independently of the class to which belongs), and n_s is the number of examples.

- *Support for a rule*: considers the number of examples satisfying both the antecedent and the consequent parts of the rule. Lavrac et al. compute in [11] the support as:

$$Sup(R^i) = Sup(Cond^i \rightarrow Class_j) = \\ = p(Class_j.Cond^i) = \frac{n(Class_j.Cond^i)}{n_s} \quad (2)$$

  where $n(Class_j.Cond^i)$ is the number of examples which satisfy the conditions for the antecedent ($Cond^i$) and simultaneously belong to the value for the target variable ($Class_j$) indicated in the consequent part of the rule.

- *Significance for a rule* [4]): indicates how significant is a finding, if measured by the likelihood ratio of a rule.

$$Sig(R^i) = Sig(Cond^i \rightarrow Class_j) = \\ = 2 \cdot \sum_{j=1}^{n_c} n(Class_j.Cond^i) \cdot \log \frac{n(Class_j.Cond^i)}{n(Class_j) \cdot p(Cond^i)} \quad (3)$$

  where n_c is the number of values for the target variable and $p(Cond^i)$, computed as $n(Cond^i)/n_s$, is used as a normalized factor.

- *Unusualness for a rule*: is defined as the *weighted relative accuracy* of a rule [13].

$$WRAcc(Cond^i \rightarrow Class_j) = \\ = \frac{n(Cond^i)}{n_s} \cdot \left(\frac{n(Class_j.Cond^i)}{n(Cond^i)} - \frac{n(Class_j)}{n_s} \right) \quad (4)$$

  The WRAcc of a rule can be described as the balance between the coverage of the rule ($p(Cond^i)$) and its accuracy gain ($p(Class_j.Cond^i) - p(Class_j)$).

B. Disjunctive Normal Form Fuzzy Rules

In the proposal presented in this paper, fuzzy rules in disjunctive normal form (DNF fuzzy rules) are used as description language to specify the subgroups, which permit a disjunction for the values of any variable present in the antecedent part. Bellow, the notation used in this paper is described. We consider a data mining problem with:

- a set of features:

$$\{X_m \ / \ m = 1,..,n_v\}$$

  used to describe the subgroups, where n_v is the number of features. These variables can be categorical or numerical;

- a set of values for the target variable:

$$\{Class_j \ / \ j = 1,\ldots,n_c\}$$

  where n_c is the number of values for the target variable considered;

- a set of examples:

$$\{E^k = (e_1^k, e_2^k, \ldots, e_{n_v}^k, class_j) / k = 1,\ldots,n_s\}$$

  where $class_j$ is the target variable value for the sample E^k (i.e., the class for this example) and n_s is the number of examples for the descriptive induction process;

- a set of linguistic labels for the numerical variables. The number of linguistic labels and the definition for the corresponding fuzzy sets depend on each variable:

$$X_m : \{LL_m^1, LL_m^2, \ldots, LL_m^{l_m}\}$$

  In this expression the set of linguistic labels for the variable X_m is represented, which has l_m different linguistic labels to describe its domain in an understandable way.

Then, a DNF fuzzy rule can be expressed as:

$$R^1 : \text{If } X_1 \text{ is } LL_1^1 \text{ or } LL_1^3 \text{ and } X_7 \text{ is } LL_7^1 \text{ then } Class_j$$

where LL_7^1 is the linguistic label number *1* of the variable *7*.

The fuzzy sets corresponding to the linguistic labels for a linguistic variable m, ($LL_m^1 \ldots LL_m^{l_m}$), are specified by means of the corresponding membership functions which can be defined by the user or defined by means of a uniform partition if the expert knowledge is not available. In this algorithm, uniform partitions with triangular membership functions are used, as it is shown in Fig. 1 for a variable m with 5 linguistic labels.

It must be noted that any subset of the complete set of variables (with any combination of linguistic labels related to the operator OR) can take part in the rule antecedent. In this way a subgroup is a compact and interpretable description of patterns of interest in data.

For this kind of fuzzy rule, it must be considered that

- an example E^k *verifies the antecedent part of a rule* R^i if:

$$APC(E^k, R^i) = T(TC(\mu_{LL_1^1}(e_1^k), \ldots, \mu_{LL_1^{l_1}}(e_1^k)), \ldots, TC(\mu_{LL_{n_v}^1}(e_{n_v}^k), \ldots, \mu_{LL_{n_v}^{l_{n_v}}}(e_{n_v}^k))) > 0 \qquad (5)$$

  where:
  - *APC* (Antecedent Part Compatibility) is the degree of compatibility between an example and the antecedent part of a fuzzy rule, i.e., the degree of membership for the example to the fuzzy subspace delimited by the antecedent part of the rule,
  - LL_m^p is the linguistic label number p of the variable m,
  - $\mu_{LL_m^p}(e_m^k)$ is the degree of membership for the value of the feature m for the example E^k to the fuzzy set corresponding to the linguistic label p for this variable,
  - T is the t-norm selected to represent the meaning of the AND operator –the fuzzy intersection–, in our case the minimum t-norm, and
  - TC is the t-conorm selected to represent the meaning of the OR operator –the fuzzy union– which in our case is the maximum t-conorm.
- an example E^k is *covered by a rule* R^i if:

$$APC(E^k, R^i) > 0 \text{ AND } E^k \in Class_j$$

This means that an example is covered by a rule if the example has a degree of membership higher than 0 to the fuzzy input subspace delimited by the antecedent part of the fuzzy rule, and the value indicated in the consequent part of the rule agrees with the value of the target feature for the example. For the categorical variables, the degrees of membership are 0 or 1.

C. Multiobjective Genetic Algorithms

GAs are general purpose search algorithms which use principles inspired by natural genetics to evolve solutions to problems [14]. In the area of subgroup discovery any rule induction algorithm must optimize simultaneously several objectives. The more suitable way to approach them is by means of multiobjective optimization algorithms in which we search a set of optimal alternative solutions (fuzzy rules in this proposal) in the sense that no other solution within the search space is better than it in all the considered objectives. The expert will use the set of rules obtained to select all or a set of them for the description of the subgroups based on the particular preference information of the problem.

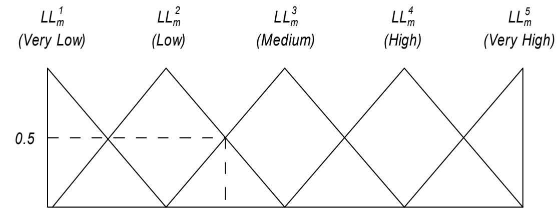

Fig. 1. Example of fuzzy partition for a continuous variable.

In a formal way, a multiobjective optimization problem can be defined in the following way:

$$\min/\max\ \vec{y} = f(\vec{x}) = (f_1(\vec{x}), f_2(\vec{x}), \ldots, f_n(\vec{x})) \qquad (6)$$

where $\vec{x} = (x_1, x_2, \ldots x_m)$ is the decision vector and $\vec{y} = (y_1, y_2, \ldots, y_n)$ is the objective vector (a tuple with n objectives). The objective of any multiobjective optimization algorithm is to find all the decision vectors for which the corresponding objective vectors can not be improved in a dimension without degrading another, which is denominated Pareto-optimal front.

Most multi-objective optimization algorithms use the concept of domination to obtain the Pareto-optimal front. In these algorithms, two decision vectors (solutions) are compared on the basis of whether one dominates the other or not. A solution a is said to *dominate* a solution b if the next two conditions are both true:

- The solution a is no worse than b in all the objectives.
- The solution a is strictly better than b in al least one objective.

If any of the above conditions is violated, the solution a does not dominate the solution b. It is intuitive that if a solution a dominates another solution b, the solution a is better than b. Since the concept of domination allows a way to compare solutions with multiple objectives, most multi-objective optimization methods use this domination concept to search for non-dominated solutions, as the proposal presented in this paper.

In the last two decades an increasing interest has been developed in the use of GAs for multiobjective optimization. There are multiple proposals of multiobjective GAs [6] [7] as the algorithms MOGA [15], NSGA II [16] or SPEA2 [17] for instance.

III. A Multiobjective Evolutionary Approach to Obtain Descriptive Fuzzy Rules

This section describes the data mining algorithm *MESDIF* (Multiobjective Evolutionary Subgroup DIscovery Fuzzy rules), a multiobjective GA for the extraction of rules which describe subgroups.

The algorithm extracts rules whose antecedent represents a conjunction of variables and whose consequent is fixed. This way, each run of the evolutionary multiobjective algorithm obtains a variable number of different rules expressing information on a single value of the target variable. As the objective is to obtain a set of rules which describe subgroups for all the values of the target feature, the evolutionary algorithm must be run so many times as different values has the target feature. This assures the knowledge extraction in all the classes.

This algorithm can generate fuzzy and/or crisp DNF rules, for problems with continuous and/or nominal variables.

The multiobjective GA is based on the SPEA2 approach [17], and so applies the concepts of elitism in the rule selection (using a secondary or elite population) and search of optimal solutions in the Pareto front.

In order to preserve the diversity at a phenotypic level our algorithm uses a niches technique that considers the proximity in values of the objectives and an additional objective based on the novelty to promote rules which give information on examples not described by other rules of the population.

Fig. 2 shows the scheme of the proposed model.

Once outlined the basis of the model, some important topics about its components are described in detail.

```
Step 1. Initialization:
  Generate an initial population P0 and create
  an empty elite population P'0 = Ø. Set t = 0.
Repeat
  Step 2. Fitness assignment:
    calculate fitness values of the individuals
    in Pt and P't.
  Step 3. Environmental selection:
    copy all non-dominated individuals in Pt
    and P't to P't+1. As the size of P't+1 must
    be exactly the number of individuals to
    store (N), we may have to use a truncation
    or a filling function.
  Step 4. Mating selection:
    perform binary tournament selection with
    replacement on P't+1 applying later
    crossover and mutation operators in order
    to fill the mating pool (obtaining Pt+1).
  Step 5. Increment generation counter
     (t = t+1)
While stop condition is not verified.
Step 6.
  Return the non-dominated individuals in P't+1.
```

Fig. 2. Scheme of the proposed algorithm

A. Chromosome Representation

The genetic representation of the solutions is the most determining aspect of the characteristics of any GA proposal. In this sense, the proposals in the specialized literature follow different approaches in order to encode rules within a population of individuals:

- The *"Chromosome = Rule"* approach, in which each individual codifies a single rule.
- The *"Chromosome = Set of rules"*, also called the Pittsburgh approach, in which each individual represents a set of rules.

In [18] a detailed description of these approaches can be found.

In a subgroup discovery task, a number of descriptive features and a single target feature of interest are considered. For this purpose, the "*Chromosome = Rule*" approach is more suited because the objective is to find a reduced set of rules in which the quality of each rule is evaluated independently of the rest, and it is not necessary to evaluate jointly the set of rules. This is the encoding approach used in the following evolutionary proposal, and so each individual codifies a single rule, and a set of rules is codified by a subset of the complete population.

As we mentioned previously the multiobjective GA discovers DNF fuzzy rules whose consequent is prefixed to one of the possible values of the target feature. Therefore, all the individuals of the population are associated with the same value of the target variable, and so the chromosome only represents the antecedent of the rule.

All the information relating to a rule is contained in a fixed-length chromosome with a binary representation in which, for each feature it is stored a bit for each of the possible values of the feature; in this way, if the corresponding bit contains the value 0 it indicates that the corresponding linguistic label or discrete value of the variable is not used in the rule, and if the value is 1 it indicates that the corresponding value is included. If a rule contains all the bits corresponding to a feature with the value 1, this indicates that this feature has no relevance for the information contributed in the rule (all the values or the feature verify the rule condition), and so this feature is ignored.

This binary representation model defines chromosomes with so many genes by variable as possible values exist for the same one. The set of possible values for the categorical features is that indicated by the problem, and for continuous variables is the set of linguistic terms determined heuristically or with expert information.

B. Definition of the Objectives of the Algorithm

The rule induction process tries to get rules with high predictive accuracy, comprehensible and interesting. In our proposal, three objectives are defined, and the algorithm tries to maximize all the defined objectives.

- *Confidence*. Determines the relative frequency of examples satisfying the complete rule among those satisfying only the antecedent. In this paper we use an adaptation of Quinlan's accuracy expression in order to generate fuzzy classification rules [20]: the sum of the degree of membership of the examples of this class (the examples covered by this rule) to the zone determined by the antecedent, divided the sum of the degree of membership of all the examples that verifies the antecedent part of this rule (irrespective of their class) to the same zone:

$$Conf\ (\mathrm{R}^i) = \frac{\sum_{E^k \in E \,/\, E^k \in Class_j} APC(E^k, R^i)}{\sum_{E^k \in E} APC(E^k, R^i)} \qquad (7)$$

where *APC* (Antecedent Part Compatibility), as defined in (5), is the compatibility degree between an example and the antecedent part of a fuzzy rule, i.e., the degree of membership for the example to the fuzzy subspace delimited by the antecedent part of the rule.

- *Support*. This measure is different from the Support defined in (2) and measures the degree of coverage that the rule offers to examples of that class, calculated as the quotient between the number of examples belonging to the class which are covered by the rule and the total number of examples from the same class:

$$Sup_c(R^i) = \frac{n(Class_j.Cond^i)}{n(Class_j)} \quad (8)$$

- *Original support*. This objective is a measure of the originality level of the rule compared with the rest of rules. Unlike the other two objectives, which are calculated for each individual rule, for the calculation of this objective the other rules of the population are considered. It is computed adding, for each example belonging to the antecedent of the rule, the factor *1/k*, where *k* is the number of rules of the population that describe information on that example. This measure promotes diversity in the population at a phenotypic level covering instances not described by other rules of the population.

The last objective, Original Support, is a restriction in the rules in order to obtain a Pareto-optimal front with a high degree of global coverage. It is related with the cooperation between rules, directing the evolutionary process towards the obtaining of rules describing information on examples not described by the previously obtained rules. The use this objective is important because the proposed algorithm is not a covering one and the obtained rules can be overlapped. Then, this objective tries to promote the obtaining of rules belonging to different parts of the search space, avoiding the convergence of the population to a part of this search space.

C. Fitness Assignment

To avoid the fact that individuals dominated by the same individuals of the elite population have identical fitness values, for each individual both dominating and dominated individuals are taken into account for the computation of the fitness. The fitness assignment for the rules extracted is performed in the following way:

1. For each individual in both the population and the elite population is computed the value for the three objectives (confidence, support and original support), and these values are used to compute what individuals are dominated, and how many individuals dominate each non-dominated individual. Then, strength of each individual is computed as the number of individuals that it dominates.
2. The raw fitness of each individual is determined as the sum of the strength of its dominators (even in the population as in the elite population).
3. Additional density information is now computed to discriminate between individuals with the same values of raw fitness. It is necessary due to the computation of the raw fitness offers a niching mechanism based in the concept of Pareto dominance, but it can fail when much of the individuals are non-dominated. The density estimation technique taken is the inverse of the distance to the *k*-th nearest neighbor, computed as.

$$D(i) = \frac{1}{\sigma_i^k + 2}$$

where σ_i^k denotes the distance of individual *i* to the *k*-th nearest individual (taking into account individuals in population and elite population), and *2* is added in the denominator to ensure that its value is greater than zero and that $D(i) < 1$.
4. The fitness value of each individual is the sum of its raw fitness value and its density.

D. Environmental Selection

This algorithm establishes a fixed length for the elite population, so it is necessary to define a truncation and a fill function.

The truncation function allows eliminating the non-dominated solutions of the elite population if it exceeds the defined size. For this purpose it is used a niche schema defined around the density measured by the distance to its k-nearest neighbor, in which, in an iterative process, in each iteration it is eliminated from the elite population the individual that is nearest of others respect of the values of the objectives.

The fill function allows adding dominated individuals from the population and the elite population until the exact size of the set is reached (ordering the individuals according to their fitness values).

E. Reproduction Model and Genetic Operators

The following reproduction model is used in the proposal:

- Join the original population with the elite population. Determine the non-dominated individuals of the joining of these populations.

- Apply a binary tournament selection on the non-dominated individuals.
- Apply recombination to the resulting population by a two point cross operator and a biased uniform mutation operator in which half the mutations carried out have the effect of eliminating the corresponding variable, in order to increase the generality of the rules.

IV. A CASE STUDY IN MARKETING: KNOWLEDGE DISCOVERY IN TRADE FAIRS

In the area of marketing, and specifically in the trade fairs planning, it is important to extract conclusions of the information on previous trade fairs to determine the relationship between the trade fair planning variables and the success of the stand. This problem over the extraction of useful information on trade fairs has been analyzed in the Department of Organization and Marketing of the University of Mondragón, Spain [21].

Businesses consider trade fairs to be an instrument which facilitates the attainment of commercial objectives such as contact with current clients, the securing of new clients, the taking of orders, and the improvement of the company image amongst others [22]. One of the main inconveniences in this type of trade fair is the elevated investment which they imply in terms of both time and money. This investment sometimes coincides with a lack of planning which emphasises the impression that trade fairs are no more than an "expense" which a business must accept for various reasons such as tradition, client demands, and not giving the impression that things are going badly, amongst other factors [23]. Therefore convenient, is the automatic extraction of information about the relevant variables which permit the attainment of unknown data, which partly determines the efficiency of the stands of a trade fair.

A questionnaire was designed to reflect the variables that better allow explaining the trade fair success containing 104 variables. 7 of these variables are continuous and the rest are categorical. The possible values for the categorical variables have been obtained as a result of a discretization carried out by experts in the field of marketing.

Then, the stand's global efficiency is rated as *high*, *medium* or *low*, in terms of the level of achievement of objectives set for the trade fair. The data contained in this dataset were collected in the Machinery and Tools biennial held in Bilbao in March 2002 and offers information on 228 exhibitors.

For this real problem, the data mining algorithm should extract information of interest about each efficiency group. The rules generated will determine the influence which the different fair planning variables have over the results obtained by the exhibitor, therefore allowing fair planning policies to be improved.

TABLE I
RESULTS FOR LOW, MEDIUM AND HIGH EFFICIENCY

Efficiency	*# Var.*	*Sup_c*	*Conf*	*Cov*	*Sup*	*Sig*	*Wracc*
Low	8	0.079	0.820	0.026	0.013	5.026	0.007
	4	0.026	1.000	0.004	0.004	3.584	0.001
	5	0.395	0.724	0.127	0.066	25.684	0.042
	6	0.289	0.759	0.088	0.048	19.672	0.031
Medium	6	0.088	0.892	0.658	0.057	6.623	0.008
	1	0.959	0.657	0.947	0.623	0.605	0.004
	2	0.574	0.802	0.469	0.373	12.104	0.065
	2	0.845	0.676	0.811	0.548	3.447	0.017
	4	0.182	0.750	0.158	0.118	2.441	0.011
High	5	0.095	0.595	0.031	0.017	6.565	0.010
	3	0.024	1.000	0.004	0.004	3.383	0.001
	4	0.047	0.722	0.013	0.009	3.812	0.004

A. Results of the Experimentation on the Marketing Dataset

As our proposal is a non-deterministic approach, the experimentation is carried out with 5 runs for each class of the categorical target variable: *low*, *medium* and *high* efficiency. The parameters used in this experimentation are:

- Population size: 100.
- Elite population size: 5.
- Maximum number of evaluations of individuals in each GA run: 10000.
- Cross probability: 0.7.
- Mutation probability: 0.01.
- Number of linguistic labels for the continuous variables: 3

A small size for the elite population has been selected to direct the evolution towards a limited Pareto set, with the objective of obtain a small set of rules.

Table I shows the best results obtained for all the classes of the target variable (*low*, *medium* and *high* efficiency). In this table, for each rule obtained it is shown:

- the number of variables involved *(# Var)*,
- the *Support* (*Sup_c*) as defined in (8) and used in our proposal,
- the *Confidence* (*Conf*) of each rule as defined in (7),
- the *Coverage (Cov)* as defined in (1),
- the *Support* (*Sup*) as defined in (2),
- the *Significance* (*Sig)* as defined in (3), and
- the *Unusualness* (*Wracc*) of the rule as computed in (4).

It must be noted that high values in support (*Sup_c*, expression (8)) means that the rule covers most of the examples of the class, and high values in confidence (*Conf*, expression (7)) means that the rule has few negative examples.

The experimentation shows that:

- The rules generated have adequate values of confidence (*Conf*, expression (7)) and support (*Sup_c*, expression (8)).
- The algorithm induces set of rules with a high confidence (higher than the minimum confidence value).

- Nevertheless, the rule support, except for some rules, is low. The market problem used in this work is a difficult real problem in which inductive algorithms tend to obtain small disjuncts (specific rules which represent a small number of examples). However, the small disjunct problem is not a determining factor in the induction process for subgroup discovery because partial relations, i.e., subgroups with interesting characteristics, with a significant deviation from the rest of the dataset, are sufficient.
- The results show that *Low* and *High* efficiency classes are the more interesting for the subgroup discovery task, but also the more difficult.
- The knowledge discovered for each one of the target variable values is understandable by the user due to the use of DNF fuzzy rules, and the low number of rules and conditions in the rule antecedents (below 10% of the 104 variables). Moreover, the rules obtained with the *MESDIF* algorithm are very simple.

In Tables II, III and IV the extracted rules for the three levels of efficiency (*low*, *medium* and *high*) are shown.

As it has been previously indicated, the dataset used in this experiment contains continuous and categorical variables. Continuous variables (e.g. "Stand size") are considered linguistic variables with using linguistic labels, defined by means of uniform partitions with triangular membership functions. Categorical variables (e.g."Number of annual fairs") can take the values previously defined and discretized in the questionary by the experts.

Marketing experts from Department of Organization and Marketing of the University of Mondragón (Spain) analysed the results obtained and indicated that:

- The exhibitors who obtained worse results were those with a medium or high size of the stand, not using indicator flags in it and with a low or medium valuation of the assembly and disassemble services.
- The companies which obtain medium efficiency are those with none or high satisfaction with the relation maintained with the clients, and medium, high or very high global satisfaction.
- Finally, the exhibitors who obtained better results (high efficiency) are big or huge companies using telemarketing with the quality contacts.

V. Conclusions

In this paper an evolutionary multiobjective model for the descriptive induction of fuzzy rules which describe subgroups applied to a real knowledge extraction problem in trade fairs is described.

The use of a subgroup discovery algorithm for this problem is well suited because in subgroup discovery task the objective is not to generate a set of rules which cover all the dataset examples, but individual rules that, given a property of interest of the data, describe in an interpretable

TABLE II
Rules for *Low* efficiency

#Rule	*Rule*
1	IF (Publicity utility = None OR Medium OR High) AND (Number of annual fairs = 2-5 OR 6-10 OR 11-15 OR >15) AND (Use of consultants = NO) AND (Importance improvement image of the company = None OR Low OR Medium) AND (Addressees if only clients = NO) AND (Stand size = Medium OR High) AND (Valuation assembly/disassembly = Low OR Medium) AND (Indicator flags = NO) THEN Efficiency = Low
2	IF (Stand size = Medium OR High) AND (Telemarketing = ALL OR Only quality) AND (Gifts = NO) AND (Indicator flags = NO) THEN Efficiency = Low
3	IF (Use of consultants = NO) AND (Importance improvement image of the company = None OR Low OR Medium) AND (Stand size = Medium OR High) AND (Valuation assembly/disassembly = Low OR Medium) AND (Indicator flags = NO) THEN Efficiency = Low
4	IF (Publicity utility = None OR Low OR High) AND (Importance improvement image of the company = None OR Low OR Medium) AND (Addressees if only clients = NO) AND Stand size = Medium OR High) AND (Valuation assembly/disassembly = Low OR Medium) AND (Indicator flags = NO) THEN Efficiency = Low

TABLE III
Rules for *Medium* efficiency

#Rule	*Rule*
1	IF (Satisfaction relation clients = None OR High) AND (Importance public relations = Very high) AND (Global satisfaction = Medium OR High OR Very high) AND (Quality visitors valuation = Low OR High) AND (Gifts = NO) AND (Inserts = NO) THEN Efficiency = Medium
2	IF (Previous promotion = YES) THEN Efficiency = Medium
3	IF (Satisfaction relation clients = None OR High) AND (Global satisfaction = Medium OR High OR Very high) THEN Efficiency = Medium
4	IF (Global satisfaction = Medium OR High OR Very high) AND (Inserts = NO) THEN Efficiency = Medium
5	IF (Satisfaction relation clients = None OR High) AND (Previous promotion = YES) AND (Company advertising mention = YES) AND (Inserts = NO) THEN Efficiency = Medium

TABLE IV
Rules for *High* efficiency

#Rule	*Rule*
1	IF (Importance new contacts = Low OR Medium OR Very High) AND (Visitor information valuation = Medium OR High) AND (Gratefulness letter = All OR Only quality) AND (Telemarketing = None OR Only quality) AND (Little gifts before fair = YES) THEN Efficiency = High
2	IF (Employees = 251-500 OR >500) AND (Follow-up modality = Only quality) AND (Telemarketing = NO OR Only quality) THEN Efficiency = High
3	IF (Employees =251-500 OR >500) AND (Visitor information valuation = Medium OR High) AND (Gratefulness letter = All OR Only quality) AND (Telemarketing = NO OR Only quality) THEN Efficiency = High

way the more interesting subgroups for the user.

In spite of the characteristics of the problem (elevated number of variables and lost values, low number of examples and few continuous variables) this multiobjective approach to the problem allows to obtain sets of rules, with an appropriate balance between the quality measures specified in the algorithm that are easily interpretable, and with a high level of confidence and support.

DNF fuzzy rules contribute a more flexible structure to the rules, allowing each variable to take more than one value, and facilitating the extraction of more general rules. In this kind of fuzzy rules, fuzzy logic contributes to the interpretability of the extracted rules due to the use of a knowledge representation nearest to the expert, also allowing the use of continuous features without a previous discretization.

The inclusion in the multiobjective algorithm of an objective (original support) that is a restriction, allows us to obtain rules with high values on the other objectives (confidence and support), also describing information on examples from which the other rules did not obtain information. This allows the inclusion of an additional level of promotion of the diversity in the multiobjective algorithm.

As future work, we will study the inclusion in the *MESDIF* algorithm of different quality measures (and combinations of them) as objective functions in order to obtain fuzzy subgroup discovery rules with better properties. We will also study new definitions for the support and original support, taking into taking into account the degree of membership of the example to the class of the rule.

REFERENCES

[1] U. M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery: An Overview," in Advances in Knowledge Discovery and Data Mining, U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, Eds.: AAAI Press, 1996, pp. 1–30.

[2] D. Michie, D. J. Spiegelhalter, and C. C. Taylor, Machine learning, neural and estatistical classification: Ellis Horwood, 1994.

[3] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and I. Verkamo, "Fast Discovery of Association Rules," in Advances in Knowledge Discovery and Data Mining, U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, Eds. Menlo Park, Calif.: AAAI Press, 1996, pp. 307–328.

[4] W. Klösgen, "Explora: A Multipattern and Multistrategy Discovery Assistant," in Advances in Knowledge Discovery and Data Mining, U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, Eds. Menlo Park, Calif.: AAAI Press, 1996, pp. 249–271.

[5] S. Wrobel, "An algorithm for multi-relational discovery of subgroups," in Principles Of Data Mining And Knowledge Discovery, vol. 1263, Lecture Notes In Artificial Intelligence, 1997, pp. 78-87.

[6] K. Deb, Multi-Objective Optimization using Evolutionary Algorithms: John Wiley & Sons, 2001.

[7] C. A. Coello, D. A. Van Veldhuizen, and G. B. Lamont, Evolutionary Algorithms for Solving Multi-Objective Problems: Kluwer Academic Publishers, 2002.

[8] A. Ghosh and B. Nath, "Multi-objective rule mining using genetic algorithms," Information Sciences, vol. 163, pp. 123-133, 2004.

[9] H. Ishibuchi and T. Yamamoto, "Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining," Fuzzy Sets and Systems, vol. 141, pp. 59-88, 2004.

[10] D. Gamberger and N. Lavrac, "Expert-guided subgroup discovery: Methodology and application," Journal Of Artificial Intelligence Research, vol. 17, pp. 1-27, 2002.

[11] N. Lavrac, B. Kavsec, P. Flach, and L. Todorovski, "Subgroup discovery with CN2-SD," Journal of Machine Learning Research, vol. 5, pp. 153-188, 2004.

[12] B. Kavsek, N. Lavrac, and V. Jovanoski, "APRIORI-SD: Adapting association rule learning to subgroup discovery," in Advances In Intelligent Data Analysis V, vol. 2810, Lecture Notes In Computer Science, 2003, pp. 230-241.

[13] N. Lavrac, P. Flach, and B. Zupan, "Rule evaluation measures: A unifying view," in Inductive Logic Programming, vol. 1634, Lecture Notes In Artificial Intelligence, 1999, pp. 174-185.

[14] D. E. Goldberg, Genetic algorithms in search, optimization and machine learning: Addison-Wesley, 1989.

[15] C. M. Fonseca and P. J. Fleming, "Genetic algorithms for multiobjective optimization: formulation, discussion and generalization," in *Proc. Fifth International Conference on Genetic Algorithms (ICGA)*, San Mateo, CA, 1993, pp. 416-423.

[16] K. Deb, A. Pratap, A. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," IEEE Transactions on Evolutionary Computation, vol. 6, pp. 182-197, 2002.

[17] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength pareto evolutionary algorithm for multiobjective optimisation," in Evolutionary methods for design, optimisation and control, K. Giannakoglou, D. Tsahalis, F. Periaux, K. Papailiou, and T. Fogarty, Eds.: CIMNE, 2002, pp. 95-100.

[18] O. Cordón, F. Herrera, F. Hoffmann, and L. Magdalena, Genetic fuzzy systems: evolutionary tuning and learning of fuzzy knowledge bases: World Scientific, 2001.

[19] M. L. Wong and K. S. Leung, Data Mining using Grammar Based Genetic Programming and Applications: Kluwer Academics Publishers, 2000.

[20] O. Cordón, M. J. del Jesus, and F. Herrera, "Genetic Learning of Fuzzy Rule-based Classification Systems Co-operating with Fuzzy Reasoning Methods," International Journal of Intelligent Systems, vol. 13, pp. 1025-1053, 1998.

[21] M. Mesonero, "Hacia un modelo efectivo de planificación ferial basado en algoritmos genéticos" (in spanish). Ph.D. dissertation, Dept. Organización y Marketing. Mondragón: Universidad de Mondragón, 2004.

[22] S. Gopalakrishna, G. L. Lilien, J. D. Williams, and I. K. Sequeira, "Do trade shows pay off," Journal of Marketing, vol. 59, pp. 75-83, 1995.

[23] S. Millar, How to get the most of the trade shows: NTC Publishing Group, 2003.