

---

# Evolutionary Induction of Descriptive Rules in a Market Problem

M.J. del Jesus<sup>1</sup>, P. González<sup>1</sup>, F. Herrera<sup>2</sup>, and M. Mesonero<sup>3</sup>

<sup>1</sup> Department of Computer Science, University of Jaén, 23071 – Jaén, Spain  
mjjesus@ujaen.es, pglez@ujaen.es

<sup>2</sup> Department of Computer Science and Artificial Intelligence, University of  
Granada., 18071 – Granada, Spain  
herrera@decsai.ugr.es

<sup>3</sup> Department of Department of Organisation and Marketing, University of  
Mondragón, 20560 – Oñati (Gipuzkoa), Spain  
mmesoner@eteo.muni.es

**Abstract.** Nowadays, face to face contact with the client continues to be fundamental to the development of marketing acts. Trade fairs are, in this sense, a basic instrument in company marketing policies, especially in Industrial Marketing. Due to the elevated investment in term of both time and money it is necessary the automatic extraction of relevant and interesting information which helps to improve fair planning policies.

In this paper, we analyse this problem and the kind of knowledge the user is interested in. We study the use of Soft Computing methodologies, specifically Fuzzy Logic and Genetic Algorithms, in the design of the Data Mining algorithms most proper to this problem, descriptive induction algorithms for subgroup discovery. Then we present an evolutionary model for the descriptive induction of fuzzy or crisp rules which describe subgroups. The proposal includes a GA in an iterative model which extracts rules while examples are left uncovered and the rules obtained surpass an specified confidence level.

**Key words:** Data Mining algorithms, Fuzzy Logic, Genetic Algorithms, descriptive induction, Subgroup Discovery, Market problems

## 1 Introduction

In the field of business an interesting problem is the study of the influence that the planning variables of a trade fair have over the successful achievement of its objectives. Faced with a real problem of this type a data mining algorithm should extract relevant interesting information concerning each of the effectiveness groups by which the stands have been arranged. The information obtained must be open to interpretation so as to be useful for the policies of trade fair planning.

In this chapter this problem is approached with an algorithm based on Soft Computing methodologies, specifically using an evolutionary data mining algorithm which allows for the extraction of a set of descriptive fuzzy rules.

The term Soft Computing applies to a methodology whose main characteristic is the tolerance to inaccuracy and uncertainty which gives it the capacity to adapt and allows us to solve problems comprehensively in a changeable environment and at low cost. The basic underlying principle in Soft Computing is the hybridization of techniques in the development of computing methods which give us an acceptable solution at low cost through the search for an approximate solution to a problem which has been set precisely or imprecisely. Within Soft Computing methodologies such as Fuzzy Logic, Neural Networks and Evolutionary Computing are included [54].

In the process of Knowledge Discovery in Databases and specifically in the process of Data Mining there are different tasks which can be treated and solved as search and optimisation problems. Evolutionary algorithms imitate the principles of natural evolution in carrying out global search and optimisation processes and are applicable to the development of data mining algorithms [26].

In Data Mining one of the objectives to be considered, besides predictive accuracy and interest, is the user's potential comprehension of the results. In this respect, Fuzzy Logic represents a tool for knowledge representation which allows us to adjust for uncertainty and inaccuracy simply and easily interpretable for the user.

This chapter is arranged in the following way: In Sect. 2, the market problem and the kind of knowledge the user is interested in are dealt with. In Sect. 3 the type of data mining algorithm to be applied to the problem is defined, descriptive rule induction for subgroup discovery. In addition some proposals carried out in the area of subgroup discovery will be analysed and the rule quality measures that are most frequently used will be described. In Sect. 4 one of the methodologies used in our proposal, Genetic Algorithms, are described and some proposals in specialised literature for rule induction with Genetic Algorithms will be analysed. The second soft methodology used in our method, the Fuzzy Logic and its use in Data Mining algorithms is explained in Sect. 5. The evolutionary approach to obtain descriptive fuzzy rules is described in Sect. 6, and in Sect. 7 the experimentation carried out and the analysis of results are explained. Finally, the conclusions and further research are outlined.

## 2 The Extraction of Useful Information on Trade Fairs

This study deals with a market problem studied in the Department of Organisation and Marketing of the University of Mondragón, Spain: the extraction of useful information on trade fairs [45].

In today's increasingly virtual world, face to face contact with the client continues to be fundamental to the development of marketing acts. Trade fairs are, in this sense, a basic instrument in company marketing policies, especially those which are run in industrial surroundings, or Industrial Marketing. AFE (Spanish Trade Fairs Association)'s data shows that to be so: during 2002, 419 events took place in Spain, in which participated 114.646 exhibitors, with an attendance of 15.739.986 visitors.

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 [32]. 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 [48]. 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.

Trade Fair planning models indicate the convenience of dividing participation into three moments in time: before, during and after the fair [32, 48].

- Pre-Fair. The exhibitor should reflect, amongst other things, on the following aspects:
  - The trade fair which s/he will attend
  - Objectives set
  - Budget
  - Visitors present at the fair
  - Stand design
  - Company personnel to attend
  - Pre-advertising campaign
- During the fair. The exhibiting company should analyse the way in which contact will be made with visitors and the communicative acts to be developed at the stand itself, amongst other things.
- Post-fair. The close of the fair is the moment to harvest the fruit of the effort made during the fair. To this end, it is necessary to evaluate participation, follow up on contacts made and measure the results achieved.

In the Machinery and Tools biennial held in Bilbao in March 2002, information was collected on all these aspects. To be precise, 104 variables of 228 exhibitors were analysed. Of these variables, 7 are continuous and the rest are categorical features, result of an expert discretization. Additionally, for each exhibitor, based on various marketing criteria, the stand's global efficiency was rated as high, medium or low, in terms of the level of achievement of objectives set for the trade fair.

For this real problem, the data mining algorithm should extract information of interest about each of the three efficiency groups of the stands. The extracted information should be interpretable so that it can be used the policies of fair planning. This makes the application of a data mining algorithm appropriate for the extraction of simple rules. 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.

### 3 Descriptive Rule Induction Processes

Knowledge discovery through a process of descriptive rule induction is included within the area of Knowledge Discovery in Databases (KDD) which is defined as the non trivial process of identifying valid, original, potentially useful patterns which have comprehensible data [21]. KDD is a wide ranging process which covers distinct stages: the comprehension of the problem, the comprehension of the data, pre-processing (or preparation) of the data, Data Mining and post-processing (assessment and interpretation of the models). The data mining stage is responsible for automatic knowledge discovery of a high level and from information obtained from real data [22].

Data Mining 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 [13, 14, 46]. Under this focus, models are generated with the objective of achieving a high degree of prediction from the data base. In this area, attention can be drawn to supervised learning processes of classification rules.
- *Descriptive induction*, whose fundamental objective is the discovery of interesting knowledge from the data. In this area, attention can be drawn to the discovery of association rules following an unsupervised learning model [1], subgroup discovery [36, 58], and other approaches to non classificatory induction.

Considering the characteristics of the problem to be solved, the obtention of simple rules which provide conclusive information about the efficiency of the stands in trade fairs, the most suitable approach is descriptive induction.

A subdivision of descriptive induction algorithms which has recently received a lot of attention from researchers is subgroup discovery which, given a set of data and having a property of interest to the user, attempts to locate subgroups which are statistically “most interesting”. The discovered subgroups must be as large as possible and have the most unusual statistical (distributional) characteristics with respect to the property of interest. Subgroup discovery methods take relations between independent (explaining) variables

and a dependent (target) variable into account. The concept was initially formulated by Klosgen in his rule learning algorithm EXPLORA [36], and by Wrobel in the algorithm MIDOS [58, 59].

In the majority of proposals in the specialised literature, subgroup discovery is described as a task half way between predictive induction and descriptive induction. It could be said that subgroup discovery aims to generate models based on rules whose objective is descriptive, using a predictive perspective in the majority of cases in order to obtain them [44]. It is a form of descriptive induction which induces rules from labelled training instances with the objective of categorising subgroups of the population which fulfil the given interest property. This suggests the use of classic learning techniques in order to solve the task. But the adaptation of this type of tasks should not lose sight of the final aim of subgroup discovery, which is to find individual rules or interest patterns, simply represented, which allows their effective use by potential users of the information.

Subgroup discovery is a descriptive induction mechanism that focuses not necessarily on finding complete relations between the target and the explaining variables. For some problems, partial relations, i.e., subgroups with “interesting” characteristics, with a significant deviation from the total population, are sufficient. This means that the rules generated do not have to fulfil high support criteria, which is necessary for other data mining approaches, e.g. the induction of association rules. Neither do they have to verify common demands for algorithms for obtaining classification systems, for example that the rules have to have a given degree of accuracy or that the final set of rules is as different as possible in order to cover all the areas of the search space. In some subgroup discovery problems, it is interesting to obtain overlapped rules, as these can show information on a single area of examples from different perspectives.

In this section we will describe some subgroup discovery algorithms presented in the KDD literature and the quality measures used frequently in this kind of inductive algorithms.

### 3.1 Approaches for Subgroup Discovery

The first algorithms developed for subgroup discovery are EXPLORA [36] and MIDOS algorithms [58, 59]. Both use a rule-extraction model based on decision trees, in order to obtain the best subgroups among the population. In order to evaluate the subgroups evaluation measurements are defined which determine the interest of an expression through a combination of unusualness [42] and size. MIDOS tackles, within this same approach, the problem of discovery in multi-relational databases. Another system which uses a multi-relational hypothesis space is the SubgroupMiner system [37].

In addition to these first proposals, different methods have been developed which obtain descriptions of subgroups represented in different ways and using different quality measurements:

- The SD algorithm [28] induces rules guided by expert knowledge. In this proposal, instead of defining an optimal measurement for the search for and the automatic selection of subgroups as do the EXPLORA and MIDOS algorithms, the objective is to help the expert to carry out flexible and effective searches over a wide range of optimal solutions. In consequence, the final decision over which subgroups will be selected is left to the expert.
- The CN2-SD algorithm [43] induces subgroups in the form of rules using as quality measurement the relationship between true positive rate and false positive rate. This algorithm is applied in [44] to a marketing problem, specifically to a relational database which stores information about interviews with potential clients. Based on the clients' responses to whether they recognise, use and appreciate given brands, the objective of the descriptive induction algorithm is to discover which brands can potentially be more used by clients, and to direct a specific marketing campaign towards these clients.
- In data analysis with high uncertainty it is useful to present a subgroup of the population by listing its support factors, instead of using a subgroup discovery approach which obtains descriptions of subgroups in the form of rules. In [11], this approach is applied to a marketing problem in which the objective is to identify the characteristics of clients who do not recognise and/or use a given brand of non-alcoholic drink.

What the described proposals have in common is the adaptation of classification rule extraction models for the subgroup discovery task. Currently, interest is starting to be shown in the development of subgroup discovery approaches which modify association rule extraction algorithms [41].

### 3.2 Rule Quality Measurements

Any rule induction algorithm uses a measurement of rule quality or of the set of rules which guides the data mining process. In addition, other quality measurements are calculated which evaluate the result of the induction process as a whole. In both cases, we can distinguish between objective and subjective quality measurements [52]. In order to solve subgroup discovery tasks, both objective and subjective measurements must be considered and the more appropriate quality criteria will depend on the application. Obviously, for automatic rule induction only objective quality criteria can be applied. However, to evaluate the quality of descriptions of induced subgroups and their use in decision making, subjective criteria are more important, although also more difficult to evaluate. In this section we will centre exclusively on objective criteria, which are those which are included in data mining algorithms.

Among the objective quality measurements, some of those most used in the literature are:

- *Size* of the rule or set of rules.
- *Predictive accuracy*, defined as the percentage of correctly classified instances.
- *Support* or relative frequency of correctly classified covered examples.
- *Significance*, called “*evidence*” in [36], which indicates how significant a subgroup is compared to the null hypothesis of statistical independence [13].
- *Unusualness*, defined as the weighted relative accuracy of a rule [42]. The average relative accuracy can be described as the balance between the coverage of a rule (a measure of generality, computed as the relative frequency of all the examples covered by the rule) and its accuracy gain: the higher a rule’s unusualness, the more relevant is it. Unusualness measures the novelty in the distribution of the subgroups, considering its coverage.
- *Interest*, which can be determined objectively or subjectively, and which can be found defined in various ways in the literature. In our proposal (see Sect. 6) we apply an interest measurement based on information gain.
- *Accuracy gain* is the difference between the accuracy of a rule and the default accuracy [44].

The measurements for the evaluation of each individual rule can be complemented with variations which calculate the average of the set of subgroups obtained. This allows the comparison of different subgroup discovery algorithms [41].

The appropriateness of the quality measurements used in a problem will determine to a great extent the effectiveness of the mechanism of descriptive rule induction used for the discovery of subgroups. This means that the study of appropriate measurements of the quality of rules is an aspect which is currently being researched.

## 4 Genetic Algorithms in Rule Induction Processes

In any knowledge extraction process in databases, and specifically in a data mining process there are different tasks or problems which can be approached and solved as optimisation and search problems. Evolutionary Algorithms [6] imitate the principles of natural evolution in order to form procedures for searching and overall optimisation and can be used as data mining algorithms, as pre- or post-processing algorithms or as tools for the optimisation of the parameters of other algorithms [26]. Among the different types of Evolutionary Algorithms, currently the most widely used in the field of Data Mining are Genetic Algorithms and Genetic Programming. In this chapter we will centre on Genetic Algorithms and their application to descriptive rule induction.

In this section we will briefly describe Genetic Algorithms and we will analyse their use for rule induction in the specialised literature.

#### 4.1 Genetic Algorithms

Genetic Algorithms (GAs) are optimisation and search algorithms inspired in natural evolution processes and initially defined by Holland [30, 34].

Stated simply, they work as follows: the system starts with an initial population of individuals who encode, through a form of genetic representation, candidate solutions for the proposed problem. This population of individuals (called chromosomes) evolves in time through a process of competition and controlled variation. Each chromosome in the population is associated with a fitness function in order to determine which chromosomes will be selected to form part of the new population in the competition process. The new population will be created using genetic operators of crossover and mutation. This evolutionary cycle continues until a given stop condition is reached: a given maximum number of evaluations of individuals is calculated, the population reaches a given maximum number of generations, a solution with a given fitness function value is reached, or the population does not generate new individuals for a given number of generations, among other possible stop conditions.

Bäck et al. in [6] gives a complete description of Genetic Algorithms as well as other examples of Evolutionary Algorithms.

The application of a GA to solve a problem must determine:

- A genetic representation (called genotype) of the solutions of the problem (called phenotype).
- A way to create an initial population of solutions.
- An evaluation function (called fitness function) which provides a quality value to each chromosome.
- Operators which modify the genetic composition of the descendants during reproduction.
- Values for the parameters used (population size, probabilities of application of genetic operators, etc.).

#### 4.2 Rule Induction Using Genetic Algorithms

GAs have several advantages as a rule induction method:

- They tend to cope well with attribute interaction because they usually evaluate a rule as a whole via fitness function, rather than evaluating the impact of adding/removing one condition to/from a rule.
- They have the ability to scour a search space thoroughly and the ability to allow arbitrary fitness functions in the search [19]. The fitness function can contain different criteria such as the ability to penalise overlap among rules or rule sets with too many rules or a problem-specific quality measure, etc.
- In addition, the genetic search performs implicit backtracking in its search of the rule space, thereby allowing it to find complex interactions that other non-backtracking searches would miss.



- An additional advantage over other conventional rule-learning algorithms is that the search is carried out among a set of competing candidate rules or rule sets.

However, this is not to say that GAs are inherently superior to rule induction algorithms as no rule discovery algorithm is superior in all cases [20, 47].

In KDD literature different GA proposals have been presented with predictive or descriptive aims. As we indicated in the previous section, rule induction algorithms for subgroup discovery (the aim of which is fundamentally descriptive) share characteristics with algorithms which guide the induction process using predictive quality measurements. In this section we will describe some of the principle GA proposals for rule induction, no matter what is their final aim.

In the design of any rule induction GA, the following aspects, among others, must be determined:

1. A genetic representation of the solutions of the problem (rule or set of rules).
2. A fitness function that provides a quality value for each rule or set of rules.
3. Operators that modify the genetic composition of the descendant during reproduction.

The first aspect is perhaps the most determining of the rest of the characteristics of any proposal. The GAs follows two 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. GABIL [17] and GIL [35] are examples of classification GAs which use this representation model.

In turn, within the “*Chromosome = Rule*” approach, there are three generic proposals:

- The *Michigan* approach in which each individual codifies a single rule but the final solution will be the final population. In this case, it is necessary to evaluate the behaviour of the whole set of rules and what the individual rules add to it. ZCS [54] and XCS [55] algorithms are examples of GAs with this type of representation.
- The *IRL* (Iterative Rule Learning) approach, in which each chromosome represents a rule, but the GA solution is the best individual obtained and the global solution is formed by the best individuals obtained when the algorithm is run multiple times. In [15] and [31] two proposals with this model are described.
- The *cooperative-competitive* approach, in which the complete population or a subset of it codifies the rule base. COGIN [33] and REGAL [29] are examples of GAs with this type of representation.

The choice of representation model depends, among other aspects, on the task to be carried out by the data mining algorithm and therefore on the type of rule to be discovered.

The most noteworthy systems based on GAs in the specialised literature, grouped according to their approach, are described below.

### The “*Chromosome = Rule*” Approach

- GA-NUGGETS. Freitas in [25] presents a GA designed to discover descriptive rules. The algorithm uses a special codification which depends on the problem, in which each individual represents the antecedent of a rule. The consequent is calculated in the evaluation of the individual such that the predictive capacity of the rule and the interest of the chosen consequent are maximised. The authors propose in [49] a modification of this algorithm, in which the fitness function consists of two parts: the first part measures the degree of interest of the rule (computed by an information-theoretical measurement, the information gain, and refers to the antecedent and to the consequent of the rule) and the second part, its predictive accuracy.
- GA-PVMINER is a parallel GA developed by Araujo, Lopes and Freitas [2] that uses Parallel Virtual Machine to discover descriptive rules for dependence modelling. In GA-PVMINER the fitness function is based on the J-measure proposed by Smyth and Goodman [53] which measures the degree of interest of a rule. The algorithm uses a crossover operator based on that used in GA-MINER, two types of mutation and a removal operator which acts in the consequent part.
- Fidelis et al. in [22] proposes a GA for predictive rule induction with a codification model which is very appropriate for the treatment of quantitative variables. The antecedent of a rule is represented in each individual and the consequent is fixed in each run of the GA.
- GLOWER, developed by Dhar et al. [19]. The design of this GA was motivated by financial prediction problems, and incorporates ideas from tree induction and rule learning.
- Carvalho and Freitas present in [8, 9] a hybrid decision tree/genetic algorithm approach for a predictive rule induction process, which follows the following model: examples belonging to large disjuncts are classified by rules produced by a decision-tree algorithm while examples belonging to small disjuncts are classified by rules produced by a GA specifically designed for this task. In the GA each individual represents a small-disjunct rule identified by a decision tree leaf node. The genome of an individual consists of the conditions in the antecedent part of the rule and the consequent part of the rule, which specifies the predicted class, is not represented in the genome. Rather, it is fixed for a given GA run, so that all individuals have the same rule consequent during the whole of that run, as occurs in [22].

The antecedent of the rule, with variable length, is represented in a fixed-length genome equal to  $n - k$ , where  $n$  is the total number of variables and  $k$  the number of ancestor nodes of the decision tree leaf node identifying the small disjunct in question. The GA uses a new operator specially designed to improve the comprehensibility of the rules. The authors propose in [10] some modifications to the original GA. The most significant change is the use of a sequential niching method [7] to foster population diversity and avoid the GA convergence to a single rule.

- ARMGA. Yan, Zhang y Zhang propose in [60] a GA with codification of the consequent for descriptive induction model for association rules without the need for previous specification by the user of the minimum levels of confidence and support.

### The “*Chromosome = Set of Rules*” Approach

- GA-RULE. Pei, Goodman y Punch in [50] propose a GA to induce predictive rules in disjunctive normal form. In the genotype both the antecedent and the consequent of the rules are encoded.
- dAR is a GA developed by Au y Chan [4] the objective of which is the predictive induction of association rules. For this, each individual represents a set of rules only through their antecedents, since the rule consequents are determined in the fitness function computation, in the same way as in GA-NUGGETS. The authors describe in [5] DEML, a GA for predictive rule induction with applications for churn prediction. The algorithm uses a codification scheme similar to that used in dAR but the mutation operator includes a hill-climbing search procedure to search for improvements even when premature convergence occurs.
- GA-MINER, developed by Flockart and Radcliffe [23]. The GA-MINER system incorporates domain information to form a hybrid genetic algorithm using explicit representation of rules. The algorithm, apart from using the conventional genetic operators, experiments with adding two new operators, the addition and the subtraction operators and with different expressions for the fitness function. This algorithm is designed for running GAs on large scale parallel data mining.
- EDRL-MD (Evolutionary Decision Rule Learner with Multivariate Discretization). In this GA, developed by Kwedlo and Kretowski [38], a single chromosome encodes a set of rules in a variable length genome. The consequent is not coded, it remains fixed for each iteration of the GA (as in [8, 9, 10, 22]). As well as the standard genetic operators it includes four specific for rule induction: one operator which modifies a chromosome in order to include an additional positive example, another which modifies the chromosome in order to eliminate a negative example, a rule elimination operator and a rule inclusion operator. This predictive induction GA is modified in [39] so that the algorithm minimises misclassification cost rather than the number of classification errors.

Wang et al. present in [55] an algorithm to generate predictive rules which incorporates both approaches. The proposed algorithm consists of two phases: knowledge integration, which integrates rules from different sources using a GA with *Pittsburgh* approach; and knowledge refinement which improves the rule set performance using a GA with the “*Chromosome = Rule*” approach.

## 5 Fuzzy Logic in Rule Induction Processes

Fuzzy Logic allows us to model inaccurate and quantitative knowledge, as well as to handle uncertainty and deal naturally to a reasonable extent with human reasoning. Ever since it was proposed in 1965 by [61], it has been applied to many areas of research, fundamentally because of its proximity to human reasoning and because it provides an effective way of capturing the approximate and inexact nature of the real world.

As we know, the principle objective of any process of Data Mining is the identification of interesting patterns and their description in a concise and significant manner. The use of Fuzzy Logic in Data Mining is sensible because fuzzy models represent a description of the data directed towards the user through a set of qualitative models which establish significant and useful relationships between variables. Fuzzy sets allow us to establish flexible limits between the different levels of meaning, without ignoring or overemphasising the elements closest to the edges, in the same way as human perception does. In any process of knowledge extraction there is a component of human interaction and fuzzy sets allow us to represent knowledge in a linguistic form, to easily incorporate previous knowledge and to provide interpretable solutions.

In rule induction processes, Fuzzy Logic is included in such a way that the models extracted are fuzzy rules. In the most interpretable type of fuzzy rules, linguistic fuzzy rules, and therefore the most appropriate for Data Mining, the continuous variables are defined as linguistic variables; that is, variables which take as possible values linguistic labels, the semantics of which are represented by an associated fuzzy set [62]. The use of Fuzzy Logic in rule induction processes with quantitative variables eases the interpretability of the knowledge which is finally extracted, the incorporation of qualitative knowledge of the problem, the treatment of lost values and classes with limits which are not well defined, and the processing of noise in variables which are the result of real measurements [3, 62].

One of the fundamental aspects when working with fuzzy rules is the definition of membership functions associated with the fuzzy sets used. In Kouk’s algorithm [40] for the extraction of fuzzy rules the expert needs to give the algorithm the continuous variables and their corresponding membership functions. In this case, the quality of the results obtained by the algorithm depends on the suitability of the fuzzy sets. For many applications it is very difficult to know from the outset which fuzzy sets will be the most appropriate, and so algorithms such as that in [27] have been described, which determine the fuzzy sets and generate the membership functions through clustering techniques.

However, in order to increase the interpretability of the results obtained in some proposals such as [3], knowledge of the problem is introduced in the initial definition of the fuzzy sets, such that the rules obtained are based on these fuzzy sets. Our proposal is centred on this approach.

The interpretability of fuzzy rules can be increased by the use of linguistic hedges. Chen and Wei in [12] propose an algorithm for the extraction of association fuzzy rules with linguistic hedges. These operators allow the knowledge extracted to be more comprehensible and closer to the user, and also enrich the semantics of the rules; the rules are therefore more detailed.

## 6 An Evolutionary Approach to Obtain Descriptive Fuzzy Rules

In the evolutionary model of extraction of fuzzy rules for subgroup discovery which we present, two components can be distinguished:

- An iterative model of extraction of fuzzy rules for the description of attributes of subgroups supported by different areas (not necessarily apart) of the instance space. This model includes the hybrid GA described below.
- A hybrid genetic algorithm for the extraction of one fuzzy rule that is simple, interpretable, and has an adequate level of support and confidence.

Both elements will be described in the next subsections.

### 6.1 Iterative Model for the Extraction of Descriptive Fuzzy Rules

The objective of the model for the extraction of descriptive fuzzy rules is to obtain a set of rules which give information on the majority of available examples for each value of the target variable.

The proposal follows the IRL (Iterative Rule Learning) approach: it includes a hybrid GA which generates a rule in an iterative plan. The iterative model allows new rules to be obtained while the generated rules reach a minimum level of confidence and give information on areas of search space in which examples which are not described by the rules generated by the previous iterations, remain.

The complete algorithm diagram is as follows:

```

START
  RuleSet  $\leftarrow$   $\emptyset$ 
  REPEAT
    Execute the GA obtaining rule R
    Local Search (R)
    RuleSet  $\leftarrow$  RuleSet + R
    Modify the set of examples
  WHILE confidence(R)  $\geq$  minimum confidence and
    R represents new examples
END

```

The repetition mechanism promotes the generating of different rules (in the sense that they give information on different groups of examples). This is achieved by penalizing – once a rule is obtained – the set of examples represented by the same in order to generate future rules. It is important to point out that this penalization does not impede the extraction of concealed rules. In subgroup discovery algorithms, the possibility of extracting information on described examples is not eliminated since redundant descriptions of subgroups can show the properties of groups from a different perspective.

As can be seen in the extraction model diagram, in each iteration the confidence of the obtained rule must be higher than a previously specified minimum value. In the algorithms of descriptive induction, one of the fundamental problems, and partly significant to the quality of the obtained results, is the specification of minimum support (minimum confidence) required of the rules to be extracted. This value depends greatly on the problem to be solved and its solution is a problem which is still not completely resolved. Zhang et al. describes in [63] a method based on fuzzy logic for the setting of the minimum confidence level.

## 6.2 Genetic Algorithm for the Extraction of a Descriptive Fuzzy Rule

In order to obtain the best fuzzy rule, a hybrid GA which, following the evolutionary obtainment of the fuzzy rule, applies a stage of post-processing, a hill-climbing process is used.

The elements of the GA will be described below.

### Chromosome Representation

The objective of the GA is to discover rules whose consequent is formed by a target variable which has been defined previously. The rules generated will be fuzzy or crisp, according to whether the variables involved are continuous or categorical, and they are coded according to the “*Chromosome = Rule*” approach.

Only the antecedent is represented in the chromosome and all the individuals in the population are associated with the same value of the target feature. This form of categorizing the target feature, used in algorithms for classification rule extraction [35] means that the evolutionary algorithm must be run many times in order to discover the rules of the different classes. As we commented in Sect. 4.2, in the specialized literature other proposals have been considered, such as codifying within the individual’s genome [17] (in this way consequent evolution is also possible) or the deterministic selection for each rule of the value of the target feature [27, 49]. In problems such as the one we present, the focus we have adopted is suitable because it is necessary to describe all the values of the target feature, and the two alternatives mentioned above do not ensure information extraction relating to all the classes.

Some of the variables of the problem are continuous variables which are treated as linguistic variables with linguistic labels. The fuzzy sets corresponding to the linguistic labels are defined by a uniform fuzzy partition with triangular membership functions, as shown in Fig. 1.

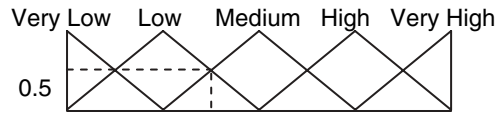


Fig. 1. Example of fuzzy partition for a continuous variable

All the information relating to a rule is contained in a fixed-length chromosome for which we use an integer representation model (the  $i$ -th position indicates the value adopted by the  $i$ -th variable). The set of possible values for the categorical features is that indicated by the problem plus an additional value which, when it is used in a chromosome, indicates that the corresponding variable does not take part in the rule. For continuous variables the set of values is the set of linguistic terms determined heuristically or with expert information, plus the value indicating the absence of the variable. In Fig. 2 we show an example of the rule and the chromosome which codifies it.

Zone	Employees	Sector	...	Bar
2	7	2	...	1

Fig. 2. Whole encoding model of a rule

In this example, the variable “Employees” does not influence the rule because the possible values are from 1 to 6, and in the corresponding gene the value is 7, which indicates the absence of this variable in the rule.

**Fitness Function**

In this process of rule discovery the objective is to obtain rules with high predictive capacity, and which are understandable and interesting. This objective can be achieved by using a weighted lineal combination of these three measurements [49]. In our proposal, the fitness function reflects these three quality factors by using a weighted lineal combination thus:

$$\text{fitness}(c) = \frac{\omega_1 \cdot \text{Support}(c) + \omega_2 \cdot \text{Interest}(c) + \omega_3 \cdot \text{Confidence}(c)}{\omega_1 + \omega_2 + \omega_3}$$

- *Confidence*. This measurement determines the accuracy of the rule, in that it reflects the degree to which the examples within the zone of the space

marked by the antecedent verify the information indicated in the consequent of the rule. In order to calculate this factor we use an adaptation of Quinlan's accuracy expression [51] in order to generate fuzzy classification rules [16]: the sum of the degree of membership of example of this class to the zone determined by the antecedent, divided the sum of the degree of membership of all the examples (irrespective of their class) to the same zone. In order to calculate these membership degrees, we use triangular membership functions and the minimum t-norm. In the case of non-fuzzy rules, the degrees of membership correspond to the classic sets, 0 or 1.

- *Support*. This is the measurement of the degree of coverage that the rule offers to examples of that class. It is calculated as the quotient between number of new examples in the class covered by the rule and the number of examples that are not covered. This way of measuring support is sensible, when using the GA within an iterative process, in order to obtain different rules each time the GA is run. From the second iteration rules which cover examples belonging to zones delimited by previously obtained rules are penalised, because the support factor only considers examples which have not been described by already-obtained rules. No distance function is used as differences are penalised on a phenotypical level.
- The *interest* of a rule can be determined objectively (based on data) or subjectively (based on the user). In the specialised literature there are proposals of both forms, depending on the specific problem to which a data mining algorithm is applied, and neither of the two approaches has a significant advantage over the other. However, it seems evident that in practice both approaches should be used: the objective criteria as a filter to select potentially interesting rules and the subjective criteria in order that the final user can determine which rules are really interesting [24].

We follow this focus in our proposal, and in the GA the degree of interest is assessed objectively. To this end we use the interest criteria provided by Noda et al. in a dependence modelling process in [49] where the level of interest of a rule is determined by two terms, one referring to the antecedent and the other to the consequent. In our proposal the consequent is prefixed and only the antecedent is used for the interest calculation, based on an information measurement given by the following expression:

$$\text{Interest} = 1 - \left( \frac{\sum_{i=1}^n \text{Gain}(A_i)}{n \cdot \log_2(|\text{dom}(G_k)|)} \right)$$

Where Gain is information gain,  $n$  is the number of variables which appear in the antecedent of the rule and  $|\text{dom}(G_k)|$  is the cardinality of the objective variable (the number of values possible for the variable which is considered as a class). The denominator term is introduced to standardise the global value.

As discussed in [24] variables with high information gain are suitable for predicting a class when they are considered individually. However, from the point of view of the interest of a rule, it is understood that the user already



knows which are the most predictive variables for a specific application domain, and therefore the rules which contain these variables are less interesting, as they are less surprising and carry less information. Therefore, it is understood that the antecedent of a rule is more interesting if it contains attributes with a small quantity of information.

The overall objective of the evaluation function is to direct the search towards rules which maximise accuracy, minimising the number of negative and not-covered examples.

### Reproduction Model and Genetic Operators

A steady-state reproduction model [6] is used: the original population is only modified through the substitution of the worst individuals by individuals resulting from crossover and mutation.

The genetic operators used are a multi-point crossover operator and a random mutation operator which is biased such that half the mutations carried out have the effect of eliminating the corresponding variable, in order to increase the generality of the rules.

### Post-Processing Phase of the Genetic Algorithm: Local Search Algorithm

The post-processing phase follows the algorithm bellow:

```

START
  Best_Rule ← R; Best_support ← support(R); Better ← True
  REPEAT WHILE Better
    Better ← False
    FOR (i=1 to gene_number)
      R'i ← R without considering variable i
      IF (support (R'i) >= support (R))
        Better ← True
        IF (support (R'i) > Best_support)
          Best_support ← support (R'i)
          Best_Rule ← R'i
    END FOR
    IF (Better AND support(Best_Rule) >= min_supp)
      Return Best_Rule
    ELSE
      Return R
  END WHILE
END

```

It improves the rule obtained by a hill-climbing process, modifies the rule while increasing the degree of support. To accomplish this, in each iteration

a variable is determined such that when it is eliminated, the support of the resulting rule is increased; in this way more general rules are obtained. Finally, the optimised rule will substitute the original only if it overcomes minimum confidence. The diagram is as follows:

## 7 Experimentation

The experimentation is carried out with the market dataset obtained in the Machinery and Tools biennial held in Bilbao in March 2002 (see Sect. 2). From total set of 104 variables, marketing experts have made a selection of variables which reduces the total set to a subset of 18 variables (in Table 1, a brief definition of these variables is described). The evolutionary rule induction algorithm has been applied to this set of variables.

**Table 1.** Variables description

Name	Description
Efficiency	Target variable: Global efficiency for the stands stated by <i>high</i> , <i>medium</i> or <i>low</i>
Zone	Geographic zone of the company
Sector	Sector to which the exhibitor belongs
Fair utility	Utility provided by the fairs
Annual fair number	Number of fairs participating annually as exhibitor
Written objectives	Existence of objectives for the BIEMH in writing
Previous promotion	Accomplishment of previous promotion to the fair
Promotion listings	Listings of clients to inform of the presence in the fair
Monitoring promotion	Monitoring of the previous promotion campaign
Operation after fair	Importance granted to the operations conducted after the fair
Client Contacts	Importance granted to the number of contacts made with present clients
Quality	Importance granted to the quality of the contacts made
Visitors number importance	Importance granted to the number of visitors in stand
Telephone calls	Accomplishment of telephone calls
Machinery demo type	Type of demonstrations of machinery
New features	Presence of new features in stand
Stewardesses	Presence of stewardesses
Stand at entrance	Stand in pavilion entrance
Near of stairs	Stand near the stairs

Parameters of the experimentation:

- The proposed algorithm is run five times for each one of the target variable values (global efficiency *high*, *medium* or *low*).
- Number of chromosomes in the GA: 100.
- Maximum number of evaluations of individuals in each GA run: 5000.
- Fitness function weights. Support: 0.4; confidence: 0.3; interest: 0.3.
- Minimum confidence value: 60.

In Table 2 the best results obtained are described. Here, for each value of the target variable the confidence, support and interest corresponding to each rule induced are shown (by means of three real numbers belonging to [0,100]). In Tables 3, 4 and 5 the rule expressions are described.

**Table 2.** Quality measurements of the rules extracted

Class	Rule	Support	Confidence	Interest
1	1	10,526	100,000	61,282
	2	13,158	100,000	60,663
	3	18,421	100,000	58,341
	4	7,895	100,000	58,248
	5	7,895	100,000	59,971
	6	5,263	100,000	57,806
	7	5,263	100,000	53,024
2	1	10,811	100,000	59,112
	2	10,135	100,000	55,906
	3	6,081	100,000	58,062
	4	3,378	100,000	61,805
	5	6,081	100,000	59,567
	6	3,378	100,000	57,870
	7	4,730	100,000	59,923
	8	3,378	100,000	60,617
	9	2,027	100,000	60,929
	10	3,378	100,000	59,232
	11	95,946	64,840	62,340
	12	0,676	100,000	60,977
3	1	4,762	100,000	62,110
	2	9,524	100,000	59,904
	3	11,905	100,000	59,045
	4	4,762	100,000	59,845
	5	7,143	100,000	60,580

We can observe that the algorithm induces set of rules with a high confidence (higher than the minimum confidence value) and interest level, around 60 in most cases. This high level of interest, according to the definition of the interest measurement used (see Sect. 6.2.2.) indicates that the variables

**Table 3.** Rules for *medium* efficiency

---

1	IF Zone = <i>North</i> and Fairs utility = <i>Low</i> AND Visitors number importance = <i>Medium</i> AND Stand at pavilion entrance = <i>Yes</i> THEN Efficiency = <i>Medium</i>
2	IF Zone = <i>North</i> AND Importance of contacts quality = <i>High</i> AND Telephone calls = <i>Yes</i> AND New features = "Catalogue" THEN Efficiency = <i>Medium</i>
3	IF Sector = <i>Rest</i> AND Importance of operations after the fair = <i>Medium</i> AND New features = <i>Product improvement</i> THEN Efficiency = <i>Medium</i>
4	IF Sector = <i>Starting+Deformation</i> AND Number of annual fairs = <i>More than 11</i> THEN Efficiency = <i>Medium</i>
5	IF Previous promotion = <i>Yes</i> AND Visitors number importance = <i>Low</i> AND Stand at pavilion entrance = <i>Yes</i> THEN Efficiency = <i>Medium</i>
6	IF Sector = <i>Rest</i> AND Importance of operations after the fair = <i>Low</i> AND Visitors number importance = <i>High</i> THEN Efficiency = <i>Medium</i>
7	IF Zone = <i>North</i> AND Sector = <i>Starting+Deformation</i> AND Fairs utility = <i>Low</i> AND Previous promotion = <i>Yes</i> AND Importance of contacts quality = <i>Medium</i> THEN Efficiency = <i>Medium</i>
8	IF Importance of contacts quality = <i>Medium</i> AND Stewardesses = <i>Yes</i> THEN Efficiency = <i>Medium</i>
9	IF Previous promotion = <i>No</i> AND Importance of contacts quality = <i>High</i> AND Stand at pavilion entrance = <i>Yes</i> THEN Efficiency = <i>Medium</i>
10	IF Sector = <i>Rest</i> AND Importance of operations after the fair = <i>Low</i> AND Importance of contacts quality = <i>Medium</i> THEN Efficiency = <i>Medium</i>
11	IF Number of annual fairs = <i>Less than 11</i> THEN Efficiency = <i>Medium</i>
12	IF Number of annual fairs = <i>More than 11</i> AND Importance of contacts quality = <i>Medium</i> THEN Efficiency = <i>Medium</i>

---

which intervene in the general rules are variables with low information gain value, more surprising to the user and they carry more information. We must note that variables with high information gain are suitable for predicting a class when they are considered individually. However, from the point of view of the interest of a rule, it is understood that the user already knows which are the most predictive variables for a specific application domain, and therefore the rules which contain these variables are less interesting, as they are less surprising and carry less information. Therefore, it is understood that the

**Table 4.** Rules for *low* efficiency

---

1	IF Sector = <i>Starting+Deformation</i> AND Written objectives = <i>No</i> AND Previous promotion = <i>No</i> THEN Efficiency = <i>Low</i>
2	IF Written objectives = <i>No</i> AND Importance of present clients contacts = <i>Low</i> AND Importance of contacts quality = <i>High</i> AND Stand at pavilion entrance = <i>No</i> AND Near of stairs = <i>No</i> THEN Efficiency = <i>Low</i>
3	IF Zone = <i>North</i> AND Sector = <i>Starting+Deformation</i> AND Written objectives = <i>No</i> AND Telephone calls = <i>Yes</i> AND New features = <i>Product improvement</i> AND Stand at pavilion entrance = <i>No</i> THEN Efficiency = <i>Low</i>
4	IF Importance of present clients contacts = <i>Low</i> AND Importance of contacts quality = <i>Low</i> THEN Efficiency = <i>Low</i>
5	IF Zone = <i>East</i> AND Written objectives = <i>No</i> AND Existence of promotion listings = <i>No</i> AND Importance of operations after the fair = <i>High</i> AND Stand at pavilion entrance = <i>No</i> AND Near of stairs = <i>No</i> THEN Efficiency = <i>Low</i>
6	IF Zone = <i>North</i> AND Fairs utility = <i>Low</i> AND Importance of present clients contacts = <i>Medium</i> AND New features = <i>Product improvement</i> THEN Efficiency = <i>Low</i>
7	IF Sector = <i>Starting+Deformation</i> AND Promotion campaign monitoring = <i>No</i> AND Importance of present clients contacts = <i>High</i> AND Machinery demonstrations type = <i>Sporadic operation</i> AND Stewardesses = <i>Yes</i> THEN Efficiency = <i>Low</i>

---

antecedent of a rule is more interesting if it contains attributes with a small quantity of information, as the rule induces here.

The rule support, except for some rules, is low. The model induces, for this problem, specific rules which represent a small number of examples. The market problem used in this work is a difficult real problem in which inductive algorithms tend to obtain small disjuncts, more common in datasets than one might think at first glance. However, the small disjunct problem, as we described in Sect. 3, is not a determining factor in the induction process for subgroup discovery. This is because partial relations, i.e., subgroups with interesting characteristics, with a significant deviation from the rest of the dataset, are sufficient.

The knowledge discovered for each one of the target variable values is understandable by the user due to the use of Fuzzy Logic, and the small number of rules and conditions in the rule antecedents.

**Table 5.** Rules for *high* efficiency

---

1	IF Written objectives = <i>Yes</i> AND Stewardesses = <i>No</i> AND Stand at pavilion entrance = <i>Yes</i> AND Near of stairs = <i>Yes</i> THEN Efficiency = <i>High</i>
2	IF Sector = <i>Rest</i> AND Number of annual fairs = <i>More than 11</i> AND New features = <i>Authentic newness</i> THEN Efficiency = <i>High</i>
3	IF Zone = <i>East</i> AND Sector = <i>Rest</i> AND Fairs utility = <i>High</i> AND Importance of contacts quality = <i>High</i> AND New features = <i>Authentic newness</i> THEN Efficiency = <i>High</i>
4	IF Zone = <i>East</i> AND Sector = <i>Rest</i> AND Number of annual fairs = <i>Less than 11</i> AND Existence of promotion listings = <i>Yes</i> AND Importance of operations after the fair = <i>High</i> AND Importance of contacts quality = <i>Medium</i> AND Stand at pavilion entrance = <i>No</i> THEN Efficiency = <i>High</i>
5	IF Fairs utility = <i>High</i> AND Written objectives = <i>Yes</i> AND New features = <i>Authentic newness</i> AND Stand at pavilion entrance = <i>No</i> AND Near of stairs = <i>No</i> THEN Efficiency = <i>High</i>

---

## 8 Concluding Remarks

The area of Soft Computing provides a set of tools which, independently or together, are being successfully used in knowledge extraction tasks.

Fuzzy Logic allows the user to incorporate directly linguistic knowledge into the data mining process, to mix this knowledge with non-linguistic information and to treat appropriately incomplete data or data with noise. But perhaps one of the characteristics which is most important for the use of fuzzy logic in this type of algorithm is its ability to represent knowledge in a linguistic form which is directly interpretable, through fuzzy rules.

Genetic Algorithms carry out a global search which is independent of the domain. This makes them a strong tool which can be applied to various stages of the knowledge extraction process.

In this chapter we describe an evolutionary model for the descriptive induction of fuzzy or crisp rules which describe subgroups. The proposal includes a GA in an iterative model which extracts rules when some examples are left uncovered, and the rules obtained surpass a given confidence level which is specified by the user.

We have applied this proposal to a real knowledge extraction problem in trade fairs. The experiment carried out has determined a simple set of rules which use few variables and therefore has a simple structure. The information extracted is comprehensible for and usable by the final user.

In future studies, we will examine the use of a more flexible structure for the rule and the study of an appropriate interest measurement for this

structure. Moreover, we are working in the development of a multiobjective version of this model [18].

## Acknowledgments

This work was supported by the Spanish Ministry of Science and Technology and by the European Fund. FEDER under Projects TIC-04036-C05-01 and TIC-04036-C05-04.

## References

1. Agrawal R, Mannila H, Srikant R, Toivonen H, Verkamo I (1996) Fast discovery of association rules. In: Fayyad UM, Piatetsky-Shapiro G, Smyth P, Uthurusamy R (eds) *Advances in Knowledge Discovery and Data Mining*. AAAI Press, pp. 307–328 [270](#)
2. Araujo DLA, Lopes HS, Freitas AA (1999) A parallel genetic algorithm for rule discovery in large databases. In: *Proceedings IEEE Conference on Systems, Man and Cybernetics*, vol III, pp. 940–945 [276](#)
3. Au WH, Chan KCC (1998) An effective algorithm for discovering fuzzy rules in relational databases. In: *Proceedings of IEEE International Conference on Fuzzy Systems (Fuzz IEEE'98)*, pp. 1314–1319 [278](#), [279](#)
4. Au WH, Chan KCC (2002) An evolutionary approach for discovering changing patterns in historical data. In: *Proceedings of 2002 SPIE 4730, Data Mining and Knowledge Discovery: Theory, Tools and Technology IV*, pp. 398–409 [277](#)
5. Au WH, Chan KCC, Yao X (2003) A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE Transactions on Evolutionary Computation* 7 (6): 532–545 [277](#)
6. Bäck T, Fogel D, Michalewicz, Z (1997) *Handbook of Evolutionary Computation*, Oxford University Press, Oxford [273](#), [274](#), [283](#)
7. Beasley D, Bull DR, Martin RR (1993) A sequential niche technique for multimodal function optimization. *Evolutionary Computation* 1: 101–125 [277](#)
8. Carvalho DR, Freitas AA (2000) A genetic algorithm-based solution for the problem of small disjuncts. In: Zighed DA, Komorowski J, Zytkow J (eds) *Principles of Data Mining and Knowledge Discovery (Proc. 4th European Conf. PKDD-2000, Lyon, France)*. Springer-Verlag, pp. 345–352 [276](#), [277](#)
9. Carvalho DR, Freitas AA (2002) A genetic algorithm for discovering small-disjunct rules in data mining. *Applied Soft Computing* 2: 75–88 [276](#), [277](#)
10. Carvalho DR, Freitas AA (2002) A genetic algorithm with sequential niching for discovering small-disjunct rules. In: Langdon WB, Cantu-Paz E, et al. (eds) *Proceedings Genetic and Evolutionary Computation Conference (GECCO-2002)*. Morgan Kaufmann, New York San Francisco, pp. 1035–1042 [277](#)
11. Cestnik B, Lavrac N, Zelezny F, Gamberger D, Todorovski L, Kline M (2002) Data mining for decision support in marketing: A case study in targeting a marketing campaign. In: *Proceedings of the ECML/PKDD-2002 Workshop on Integration and Collaboration Aspects of Data Mining, Decision Support and Meta-Learning*, pp. 25–34 [272](#)

12. Chen G, Wei Q (2002) Fuzzy association rules and the extended mining algorithms. *Information Sciences* 147: 201–228 [279](#)
13. Clark P, Niblett T (1989) The CN2 induction algorithm. *Machine Learning*, 3(4): 261–283 [270](#), [273](#)
14. Cohen W (1995) Fast effective rule induction. In *Proceedings of the Twelfth International Conference on Machine Learning*. Morgan Kaufmann, pp. 115–123 [270](#)
15. Cordón O, Herrera F (1997) A three-stage evolutionary process for learning descriptive and approximative fuzzy logic controller knowledge bases from examples. *International Journal of Approximate Reasoning* 17 (4): 369–407 [275](#)
16. Cordón O, del Jesus MJ, Herrera F (1998) Genetic Learning of Fuzzy Rule-based Classification Systems Co-operating with Fuzzy Reasoning Methods. *International Journal of Intelligent Systems* 13 (10/11): 1025–1053 [282](#)
17. De Jong KA, Spears WM, Gordon DF (1993) Using genetic algorithms for concept learning. *Machine Learning* 13 (2–3): 161–188 [275](#), [280](#)
18. Deb K (2001) *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley & Sons [289](#)
19. Dhar V, Chou D, Provost F (2000) Discovering interesting patterns for investment decision making with Glower-a Genetic Learner Overlaid With Entropy Reduction. *Data Mining and Knowledge Discovery* 4: 251–280 [274](#), [276](#)
20. Domingos P (1998) Occam’s two razors: the sharp and the blunt. In: *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD’98)*, pp. 37–43 [275](#)
21. Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) From data mining from knowledge discovery in databases. In: Fayyad UM, Piatetsky-Shapiro G, Smyth P, Uthurusamy R (eds): *Advances in Knowledge Discovery & Data Mining*. AAAI/MIT, pp. 1–34 [270](#)
22. Fidelis MV, Lopes HS, Freitas AA (2000) Discovering compressible classification rules with a genetic algorithm. In: *Proceedings on Evolutionary Computation (CEC’2000)*, pp. 805–810 [270](#), [276](#), [277](#)
23. Flockhart IW, Radcliffe NJ (1995) GA-MINER: Parallel data mining with hierarchical genetic algorithms (Final Report by the University of Edinburgh, UK, EPCC-AIKMS-GA-Miner-Report 1.0) [277](#)
24. Freitas AA (1999) On Rule Interestingness Measures. *Knowledge-Based Systems* 12: 309–315 [282](#)
25. Freitas AA (1999) A genetic algorithm for generalized rule induction. In: Roy R, Furuhashi T, Chawdhry PK (eds) *Advances in Soft Computing – Engineering Design and Manufacturing (Proc. WSC3, 3rd on-line world conf., hosted on the internet)*, pp. 340–353 [276](#)
26. Freitas AA (2002) *Data Mining and Knowledge Discovery with Evolutionary Algorithms*, Springer Verlag [268](#), [273](#)
27. Fu AW, Wong MH, Sze SC, Wong WC, Wong WL, Yu WK (1998) Finding fuzzy sets for the mining of fuzzy association rules for numerical attributes. In: *Proceedings of the First International Symposium on Intelligent Data Engineering and Learning (IDEAL’98)*, pp. 263–268 [278](#), [280](#)
28. Gamberger D, Lavrac N (2002) Expert guided subgroup discovery: Methodology and application. *Journal of Artificial Intelligence Research*, 17: 501–527 [272](#)
29. Giordana A, Neri F (1995) Search-intensive concept induction. *J Evolutionary Computation*, 3 (4): 375–416 [275](#)



30. Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning. Addison-Wesley 274
31. González A, Pérez R (1998) Completeness and consistency conditions for learning fuzzy rules. *Fuzzy Sets and Systems* 96: 37–51 275
32. Gopalakrishna S, Lilien GL, Williams JD, Sequeira IK (1995) Do trade shows pay off. *Journal of Marketing* 59: 75–83 269
33. Greene DP, Smith SF (1993) Competition-based induction of decision models from examples. *Machine Learning* 3: 229–257 275
34. Holland JH (1975) *Adaptation in natural and artificial systems*. University of Michigan Press 274
35. Janikow CZ (1993) A knowledge-intensive genetic algorithm for supervised learning. *Machine Learning* 13: 189–228 275, 280
36. Klösgen W (1996) Explora: a multipattern and multistrategy discovery assistant. In: Fayyad V, Piatetsky-Shapiro G, Smyth P, Uthurusamy R. (eds) *Advances in Knowledge Discovery and Data Mining*. MIT Press pp. 249–271 270, 271, 273
37. Klösgen W (2002) *Handbook of data mining and knowledge discovery*. Oxford University Press 271
38. Kwedlo W, Kretowski M (1999) An evolutionary algorithm using multivariate discretization for decision rule induction. In: *Proceedings of the 3rd European Conference on Principles of Data Mining and Knowledge Discovery (PKDD'99)*, pp. 392–397 277
39. Kwedlo W, Kretowski M (2001) An evolutionary algorithm for cost-sensitive decision rule learning. *Lecture Notes in Artificial Intelligence* 2167: 288–299 277
40. Kuok C, Fu A, Wong M (1998) Mining fuzzy association rules in databases. *ACM SIGMOD Record*, 27: 41–46 278
41. Lavrac N, Cestnik B, Gamberger D, Flach P (2004) Decision support through subgroup discovery: three case studies and the lessons learned. *J Machine Learning* 57 (1–2): 115–143 272, 273
42. Lavrac N, Flach P, Zupan B (1999) Rule evaluation measures: A unifying view. In: *Proceedings of the Ninth International Workshop on Inductive Logic Programming*. Springer, pp. 74–185 271, 273
43. Lavrac N, Flach P, Kavsek B, Todorovski L (2002) Adapting classification rule induction to subgroup discovery. In: *Proceedings of the Second IEEE International Conference on Data Mining*. IEEE Computer Society, pp. 266–273 272
44. Lavrac N, Kavsek B, Flach P, Todorovski L (2004) Subgroup discovery with CN2-SD. *Machine Learning Research* 5: 153–188 271, 272, 273
45. Mesonero M (2004) *Toward an effective trade show planning model based on genetic algorithms*. Ph.D. thesis (in Spanish), Mondragon Unibertsitatea. 268
46. Michalski RS, Mozetic I, Hong J, Lavrac N (1986) The multi-purpose incremental learning system AQ15 and its testing application on three medical domains. In: *Proceedings of the Fifth National Conference on Artificial Intelligence*. Morgan Kaufmann, pp. 1041–1045 270
47. Michie D, Spiegelhalter DJ, Taylor CC (1994) *Machine learning, neural and estatistical classification*. Ellis Horwood 275
48. Millar S (2003) *How to get the most of the trade shows*. NTC Publishing Group 269
49. Noda E, Freitas AA, Lopes HS (1999) Discovering Interesting Prediction Rules with a Genetic Algorithm. In *Proceedings of the Congress on Evolutionary Computation* 2, pp. 1322–1329 276, 280, 281, 282

50. Pei M, Goodman ED, Punch III WF (1997) Pattern discovery from data using genetic algorithms. In: Proceedings of the First Pacific-Asia Conference on Knowledge Discovery and Data Mining. KDD: Techniques and applications, pp. 264–276 [277](#)
51. Quinlan JR (1987) Generating production rules Machine Learning. Morgan Kaufmann [282](#)
52. Silberschatz A, Tuzhilin A (1995) On subjective measures of interestingness in knowledge discovery. In: Proceedings of the First International Conference on Knowledge Discovery and Data Mining, pp. 275–281 [272](#)
53. Smyth P, Goodman RM (1991) Rule induction using information theory. In: Piatetsky-Shapiro G, Frawley J (eds.) Knowledge Discovery in Databases. MIT Press, Cambridge, pp. 159–176 [276](#)
54. Tettamanzi A, Tomassini M (2001) Soft Computing. Integrating Evolutionary, Neural and Fuzzy Systems. Springer [268](#), [275](#)
55. Wang CH, Hong TP, Tseng SS (1998) A hybrid genetic knowledge-integration strategy. In: Proceedings of the 1998 IEEE International Conference on Evolutionary Computation, pp. 587–591 [275](#), [278](#)
56. Wilson SW, Goldberg DE (1994) ZCS: a zeroth order classifier system. Evolutionary Computation 2: 1–18
57. Wilson SW (1995) Classifier system based on accuracy. Evolutionary Computation 3 (2): 149–175
58. Wrobel S (1997) An algorithm for multi-relational discovery of subgroups. In: Proceeding of the First European Conference on Principles of Data Mining and Knowledge Discovery, pp. 78–87 [270](#), [271](#)
59. Wrobel S (2001) Inductive logic programming for knowledge discovery in databases. In: Dzeroski S, Nada Lavrac N (eds) Relational Data Mining, Springer: 74–101 [271](#)
60. Yan X, Zhang C, Zhang S (2003) A database-independent approach of mining association rules with genetic algorithms. In: Proceedings of IDEAL 2003, pp. 882–886 [277](#)
61. Zadeh LA (1965) Fuzzy Sets. Inform. Control. 9: 338–352 [278](#)
62. Zadeh LA (1975) The concept of a linguistic variable and its applications to approximate reasoning, Parts I, II, III. Information Sciences 8–9: 199–249, 301–357, 43–80 [278](#)
63. Zhang S, Lu J, Zhang C (2004) A fuzzy logic based method to acquire user threshold of minimum-support for mining association rules. Information Sciences 164: 1–16 [280](#)