Soft Comput
DOI 10.1007/s00500-010-0670-3

FOCUS

Evolutionary fuzzy rule extraction for subgroup discovery in a psychiatric emergency department

C. J. Carmona · P. González · M. J. del Jesus · M. Navío-Acosta · L. Jiménez-Trevino

Abstract This paper describes the application of evolutionary fuzzy systems for subgroup discovery to a medical problem, the study on the type of patients who tend to visit the psychiatric emergency department in a given period of time of the day. In this problem, the objective is to characterise subgroups of patients according to their time of arrival at the emergency department. To solve this problem, several subgroup discovery algorithms have been applied to determine which of them obtains better results. The multiobjective evolutionary algorithm MESDIF for the extraction of fuzzy rules obtains better results and so it has been used to extract interesting information regarding the rate of admission to the psychiatric emergency department.

Keywords Evolutionary fuzzy system · Subgroup discovery · Fuzzy rules extraction · Evolutionary algorithm · Psychiatric emergency

C. J. Carmona · P. González (✉) · M. J. del Jesus
Department of Computer Science,
University of Jaen, 23071 Jaen, Spain
e-mail: pglez@ujaen.es

C. J. Carmona
e-mail: ccarmona@ujaen.es

M. J. del Jesus
e-mail: mjjesus@ujaen.es

M. Navío-Acosta
Hospital Universitario 12 de Octubre,
CIBERSAM, Madrid, Spain
e-mail: mnavio.hdoc@salud.madrid.org

L. Jiménez-Trevino
Department of Psychiatry, University of Oviedo,
CIBERSAM, Oviedo, Spain
e-mail: jimenezluis@uniovi.es

1 Introduction

Recent data show an increase in psychiatric emergency department visits to the *Hospital Ramón y Cajal* in Madrid (Spain), but also to other hospitals in Madrid (Baca-Garcia et al. 2008). This increase in patients makes it particularly important to plan which resources are needed in every moment to care for patients properly. It would also help in knowing which patients are more likely to come to the psychiatric emergency service at any given time of day, or the characteristics and the type of disorder of the patients to optimise the organisation of the psychiatric emergency service. Other studies have addressed the relationship of heat waves with the activity of psychiatric emergency departments (Bulbena et al. 2009). There are several types of psychiatric emergencies: suicide attempts, substance dependence, alcohol intoxication, violent behaviour, acute depression, panic attacks, psychosis, personality disorders, or rapid changes in behaviour (Clercq et al. 1998).

Owing to the increasing number of patients and data stored for each of them, the use of data mining (DM) can provide better results to address this problem. DM consists of the automatic extraction of implicit and interesting patterns from large data collections (Fayyad et al. 1996). DM is included within the broader area of knowledge discovery in databases (KDD), which also involves preprocessing methods that facilitate the application of data mining algorithms and post-processing methods which refine and improve the knowledge extracted. 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 (Michie et al. 1994), and descriptive induction whose main objective is the extraction of interesting knowledge from the data.

Published online: 20 November 2010

As the objective of this study is to characterise subgroups of patients who tend to visit the psychiatric emergency department during a certain period of time, subgroup discovery (SD) is the most appropriate data mining technique. SD is a descriptive DM task, including some features of predictive DM. It can be considered that SD lies somewhere between the extraction of association rules and the obtaining of classification rules. The goal of SD is the discovery of interesting characteristics of subgroups with respect to a specific property which is of interest to the user (in this case, the time of arrival at the emergency facility). The interpretability of the results obtained is an important issue in SD because the goal of the SD task is to find significant, relevant and previously unknown information about groups of interest. In this sense, rules are a suitable tool for the representation of knowledge in the extraction of information describing subgroups.

In recent years, new algorithms for SD have been developed using soft-computing techniques like fuzzy rules and genetic algorithms (GAs) (del Jesus et al. 2007a, b). The conjunction of these techniques is called evolutionary fuzzy systems (EFSs) (Cordón et al. 2004; Herrera 2008), which has attracted considerable attention in the computational intelligence community. The use of GAs (Goldberg 1989) and fuzzy logic (Zadeh 1975) is interesting for the SD task. GAs explore the search space thoroughly and handle the relations between variables appropriately, and therefore develop searches particularly suited to rule extraction. Fuzzy logic, and particularly, the use of descriptive fuzzy rules allows us in representing and use knowledge in a similar way to human reasoning. With the use of fuzzy rules, we obtain more interpretable and actionable solutions in the field of SD, and in general, in the analysis of data to establish relationships and identify patterns (Hüllermeier 2005).

In this study, we apply several SD algorithms to patient data of the psychiatric emergencies department of the *Hospital Ramón y Cajal* in Madrid (Spain). The objective is to obtain rules which describe relationships between the different variables stored for each patient and the arrival time at the emergencies department. We will focus our attention in the use of EFS-based SD algorithms, SDIGA (del Jesus et al. 2007a) and MESDIF (del Jesus et al. 2007b), comparing their results with those obtained by the state of the art SD algorithm CN2-SD (Lavrac et al. 2004). We will present an experimental study showing that the MESDIF algorithm obtains better results in this medical problem. Then some of the rules extracted by this algorithm will be analysed.

This paper is organised in the following way: Sect. 2 describes the problem of discovering rules for the psychiatric emergency department data and surveys some specific work in this area. Section 3 introduces the subgroup discovery task, and Sect. 4 the evolutionary SD algorithms used in the study. Section 5 describes the experimental study, the analysis of results and the interpretation of some of the rules extracted. Finally, Sect. 6 outlines the conclusions of the study.

2 Knowledge extraction in the psychiatric emergency department problem

The organisation of resources in a psychiatric emergency department is critical for its proper functioning. In this sense, identifying which types of patients can visit the emergency department at which time can facilitate a better organisation of these resources. Thus, the objective stated in this work is the obtaining of information on rates of arrival to the psychiatric emergency department to determine what types of pathologies are more common depending on the time of admission. This is a complex medical problem, and a lot of patient-related data has been collected.

The following subsections describe the data set collected and the related work in data mining with medical data.

2.1 Data set description

Data for knowledge extraction in the psychiatric emergency problem were obtained from the psychiatric emergency department of *Hospital Ramón y Cajal* in Madrid (Spain). Information was collected on 72 variables with patient information (such as time of admission, consultation duration, reason for consultation, social-demographic data, personal history, previous treatments, medications consumed, drugs consumed, type of application, received diagnosis or intervention performed) was collected in a sample of 925 patients. Table 1 shows a brief description of the variables considered.

In this data set, the initial problem stated is whether there are variables which determine the time of admission to the emergency department (i.e. what types of patients attend consultations during the different periods). For this, the variable of interest is "*admission time*", which has been discretised to three time intervals according to the criteria of medical experts, who consider that these intervals as the most interesting for the problem:

1. Day (class 0), from 7:30 to 13:59, where there are a total of 221 patients.
2. Evening (class 1), from 14:00 to 20:59, with a total of 379 patients.
3. Night (class 2), from 21:00 to 7:29, with 313 patients.

Almost all the variables collected are categorical, except for three continuous variables:

Table 1 Description of the variables

N	Name	Description	Values
0	Derivation	Derivation	$0, 1, \ldots, 8$
1	Sex	Gender of the patient (male, female)	0,1
2	Age	Age of the patient	[9,90]
3	Education	Educational level	0,1,2,3
4	Occupation	Employment	$0, 1, \ldots, 8$
5	Coexist	Number of people living with the patient	$0, 1, \ldots, 5$
6	Reason for consultation	Reason why the patient comes for consultation	$0, 1, \ldots, 20$
7	Medical history	Medical history	$0, 1, \ldots, 20$
8	Psychiatric history	Psychiatric history	$0, 1, \ldots, 7$
9	Drugs	Admits the use of drugs	0,1
10	Alcohol	Admits drink alcohol	0,1
11	Cannabis	Admits the use of cannabis	0,1
12	Opiates	Admits the use of opiates	0,1
13	Cocaine	Admits the use of cocaine	0,1
14	Others substance	Admits the use of other substances	0,1
15	Self injuries	The patient shows signs of self injuries	0,1,2
16	Smoker	The patient smokes	0,1,2
17	Preventive treatment	Type of preventive treatment	$0, 1, \ldots, 5$
18	Psychotropic drugs treatment	Treatment with psychotropic drugs	0,1
19	BDZ	Treatment with benzodiazepines	0,1
20	Classic neuroleptics	Treatment with classic neuroleptics	0,1
21	Tricyclic neuroleptics	Treatment with tricyclic neuroleptics	0,1
22	Tricyclic antidepressant	Treatment with tricyclic antidepressant	0,1
23	SSRI	Treatment with selective serotonin reuptake inhibitor	0,1
24	IsRNA	Treatment with immunostimulatory RNA	0,1
25	Others antidepressants	Treatment with other antidepressants	0,1
26	Lithium	Treatment with lithium	0,1
27	Mood stabiliser	Treatment with mood stabiliser	0,1
28	Other treatments	Other treatments	0,1
29	Depot neuroleptics	Depot neuroleptics	0,1
30	Psychotherapy	Treatment with psychotherapy	0,1
31	Adhesion	Adherence to treatment	1,2,3
32	Previous psychiatric admissions	Number of previous psychiatric admissions	[0,30]
33	Previous medical admissions	Number of previous medical admissions	[0,12]
34	Requested analysis	Type of analysis requested	1,2,3,4
35	Accompanying	Accompanying person	1,2,3,4,5
36	Clinic initiation	Clinic initiation	$1, 2, \ldots, 6$
37	Previous consultation	Previous consultation	0,1,2,3,4
38	Consultation time	Consultation time	$0, 1, \ldots, 5$
39	Organic mental disorder	The patient presents organic mental disorder	0,1
40	Substance mental disorder	Presents mental disorder due to the use of substances	0,1
41	Psychotic disorder	Presents psychotic disorder	0,1
42	Affective disorders	Presents affective disorders	0,1
43	Neurotic disorders	Presents neurotic disorders	0,1
44	Physiological dysf. dis.	Presents behavioural disorders associated with physiological dysfunctions	0,1
45	Personality disorder	Presents personality disorder	0,1
46	Mental retardation	Presents mental retardation	0,1
47	Developmental disorder	Presents developmental disorder	0,1

Table 1 continued

N	Name	Description	Values
48	Childhood disorders	Presents childhood disorders	0,1
49	Eating disorders	Presents eating disorders	0,1
50	Autolytic behaviour	Presence of autolytic behaviour	0,1
51	Side effects	Side effects of the treatment	0,1
52	Psychopathology	Psychopathology	0,1
53	Emergency treatment	Emergency treatment	0,1
54	ET BDZ	Emergency treatment with benzodiazepines	0,1
55	ET Classic neuroleptics	Emerg. treat. with classic neuroleptics	0,1
56	ET Atypical neuroleptics	Emerg. treat. with atypical neuroleptics	0,1
57	ET Tricyclic antidepressant	Emerg. treat. with tricyclic antidepressant	0,1
58	ET SSRI	Emerg. treat. with selective serotonin reuptake inhibitor	0,1
59	ET IsRNA	Emerg. treat. with immunostimulatory RNA	0,1
60	ET Other antidepressants	Emerg. treat. with other antidepressants	0,1
61	ET Lithium	Emerg. treat. with lithium	0,1
62	ET Mood stabiliser	Emerg. treat. with mood stabiliser	0,1
63	ET Other treatments	Emerg. treat. with other treatments	0,1
64	ET Depot neuroleptics	Emerg. treat. depot neuroleptics	0,1
65	Discharge destination	Destination at discharge	$1, 2, \ldots, 8$
66	Intervention	Intervention	$1, 2, \ldots, 10$
67	Voluntary admission	Voluntary admission	0,1,2
68	Discharge type	Discharge type	0,1,2,3,4
69	Psychiatric family background	Psychiatric family background	$0, 1, \ldots, 7$
70	Relative degree	Degree of kinship	0,1,2
71	Psychiatric admissions	Previous psychiatric admissions	0,1
–	Admission time	Period of time in which the admission occurs	Day, Evening, Night

- *Age*, with integer values between 0 and 90.
- *Previous psychiatric admissions*, with integer values between 0 and 30.
- *Previous medical admissions*, with integer values between 0 and 12.

2.2 Related work

There are a number of fields where DM techniques have been applied to problems in medicine in general and in psychiatry in particular. The use of these techniques has been applied to develop a framework using data mining technologies that make it possible to automatically analyse huge clinical data sets and to discover patterns behind them (Masuda et al. 2002), to reanalyse a published study of which variables predicted psychiatrists' decisions to hospitalise suicide attempters, who were assessed in the emergency department (Baca-García et al. 2006), to provide an infrastructure for the highest quality research in psychiatric disorders, particularly in schizophrenia and schizo affective disturbances (Kielan et al. 2004), or to obtain rules to guide doctors towards a good relation with their patients to improve the results or the psychiatric treatments (Aguilar-Ruiz et al. 2004).

SD has been successfully applied in different medical domains, including the detection of patient groups with risk of atherosclerotic coronary heart disease (Gamberger and Lavrac 2003), breast cancer diagnosis (López et al. 2009; Mueller et al. 2009), brain ischaemia data analysis (Gamberger et al. 2007; Kralj et al. 2007), profiling examiners for sonographic examinations (Atzmueller et al. 2005), identification of interesting diagnostic patterns to supplement a medical documentation and consultation system (Atzmueller et al. 2004), or scrutinizing blood glucose management guidelines (Nannings et al. 2009).

Soft computing techniques for data mining have also been used in medicine and psychiatry (Fogel 2008; Yardimci 2009). Soft computing is a consortium of synergic methodologies and provides flexible information processing capability for handling ambiguous real life situations. Its aim is to exploit the tolerance for imprecision, uncertainty, approximate reasoning and partial truth to achieve tractability, robustness and low-cost solutions. Soft computing includes fuzzy logic, neural networks and GA methodologies, combining them into new methodologies, such as fuzzy logic and neural networks, neural networks and GA or fuzzy logic and GA. Among the large number of computational techniques used in medicine, soft computing provides unmatched utility because of its demonstrated strength in handling imprecise information and providing novel solutions to complicate problems.

Evolutionary algorithms have been applied to automatically and accurately determine unlabeled peptide quantities in cerebral stroke samples injected into a mass spectrometer (Valdés et al. 2008) to develop an evolutionary text-mining framework capable of inducing variable length patterns from unannotated psychiatry web resources (Yu et al. 2008) and to discover and select diagnostic risk factors in medicine (Mantzaris et al. 2009).

The use of fuzzy logic in psychiatry has been applied to characterise thyroid diseases with an alternative medical decision support system comprising knowledge extraction methods and fuzzy cognitive maps (Papageorgiou et al. 2008). A meta-learning algorithm which selects a subset from a previously generated set of fuzzy rules using evolutionary algorithms was presented in Drobics et al. (2007) to generate a model for predicting the presence/absence of hepatitis based on the laboratory results. In addition, a transparent and relatively accurate classifier was developed in (Ainon et al. 2009) using a hybrid soft computing technique. In this technique, the initial fuzzy model is first generated using a clustering method and the transparency and accuracy of the model are then simultaneously optimised using a multiobjective evolutionary technique.

3 Subgroup discovery

The concept of SD was initially formulated by Kloesgen (1996) and Wrobel (1997): given a population of individuals and a property of those individuals we are interested in, find population subgroups that are statistically "most interesting", e.g., are as large as possible and have the most unusual statistical characteristics with respect to the property of interest.

The main goal in SD is to discover characteristics of the subgroups by constructing simple rules with high support and significance. Because SD focusses its interest on partial relations instead of complete relations; small subgroups with interesting characteristics can be sufficient. Furthermore, the subgroups obtained must satisfy two conditions: first, the subgroups must be interpretable and second, the subgroups should be interesting according to the criteria of the user.

In SD, a rule R can be described as:

$$R : \text{Cond} \rightarrow \text{Class}$$

where the property of interest is the class value Class that appears in the consequent part of the rule and the antecedent part of the rule Cond is a conjunction of features (attribute–value pairs) selected from the features describing the training instances (Gamberger and Lavrac 2002; Lavrac et al. 2004).

3.1 Different approaches used in subgroup discovery

Below, some SD approaches that can be found in the specialized literature are reviewed:

- *Explora* (Kloesgen 1996) was the first approach for SD and was developed by Klösgen. It uses decision trees for the extraction of rules.
- *Midos* (Wrobel 1997) applies the Explora approach to multi-relational databases. It uses optimistic estimation and minimum support pruning.
- *Apriori-SD* (Kavsek and Lavrac 2006) is developed by adapting to subgroup discovery the classification rule-learning algorithm Apriori-C (Jovanoski and Lavrac 2001), a modification of the original Apriori association rule learning algorithm (Agrawal et al. 1993).
- *CN2-SD* (Lavrac et al. 2004) is based on the CN2 classification rule algorithm (Clark and Niblett 1989), and induces subgroups in the form of rules using as its quality measure the relation between true positives and false positives.
- *SD-MAP* (Atzmueller and Puppe 2006) is an exhaustive SD algorithm based on the well-known FP growth method (Han et al. 2000) for mining association rules with adaptations for the SD task. It allows the exploitation of background knowledge and the use of quality measures to evaluate the subgroups.
- *RSD* (Zelezny and Lavrac 2006) is an upgrade of the CN2-SD algorithm that enables relational SD.
- *MergeSD* (Grosskreutz and Rueping 2009) performs a depth-first search in the space of subgroup descriptions using a new pruning scheme allowing the algorithm to work with continuous variables.

3.2 Quality measures in subgroup discovery

One important aspect in SD is the quality measures used both to extract and evaluate the rules, but there is no consensus in the field about which are the most suitable measures for the SD process. For the evolutionary algorithms, the quality measures most widely employed in SD are

- *Number of rules* (n_r), a complexity measure computed as the number of induced rules.
- *Number of variables* (n_v), the number of variables of the antecedent. The number of variables for a set of rules is computed as the average of the variables for each rule of that set.

- *Support of a rule*, the frequency of correctly classified examples covered by the rule (Lavrac et al. 2004). It can be computed as

$$\mathrm{Sup}(R_i) = \frac{n(\mathrm{Class}_i \cdot \mathrm{Cond}_i)}{n_s} \quad (1)$$

where $n(\mathrm{Class}_i \cdot \mathrm{Cond}_i)$ is the number of examples which satisfy the conditions and also belong to the value for the target variable and n_s is the number of examples.

- *Support based on examples of the class* defined as the degree of coverage that the rule offers to examples of that class (del Jesus et al. 2007):

$$\mathrm{Sup}_c(R_i) = \frac{n(\mathrm{Class}_i \cdot \mathrm{Cond}_i)}{n(\mathrm{Class}_i)} \quad (2)$$

where $n(\mathrm{Class}_i)$ is the number of examples of the class. In this paper we use this measure as the support measure.

- *Confidence of a fuzzy rule* (del Jesus et al. 2007) determines the relative frequency of examples which verify the complete rule among those which satisfy only the antecedent part:

$$\mathrm{Cnf}(R_i) = \frac{\sum_{E^k \in E / E^k \in \mathrm{Class}_i} \mathrm{APC}(E^k, R_i)}{\sum_{E^k \in E} \mathrm{APC}(E^k, R_i)} \quad (3)$$

where the antecedent part compatibility (APC) is the degree of compatibility between an example and the antecedent part of a fuzzy rule:

$$\mathrm{APC}(E^k, R_i) = T(\mu_{LL_1^{l}}(e_1^k), \ldots, \mu_{LL_{n_v}^{l_{n_v}}}(e_{n_v}^k)) > 0 \quad (4)$$

where

- $\mu_{LL_{n_v}^{l_{n_v}}}(e_{n_v}^k)$ is the degree of membership for the value of the feature n_v for the example E^k to the fuzzy set corresponding to the linguistic label l_{n_v} for this variable (n_v).
- T is the $t-$ norm selected to represent the meaning of the AND operator (the fuzzy intersection) in our case the minimum $t-$ norm. It is important to consider that an example E^k is covered by a rule R_i if

$$\mathrm{APC}(E^k, R_i) > 0 \text{ AND } E^k \in \mathrm{Class}_i. \quad (5)$$

This means that an example is covered by a rule, i.e. 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 categorical variables, the degrees of membership are 0 or 1.

- *Significance of a rule* indicates the significance of a finding, if measured by the likelihood ratio of a rule (Kloesgen 1996).

$$\mathrm{Sig}(R_i) = 2 \cdot \sum_{k=1}^{n_c} n(\mathrm{Class}_k \cdot \mathrm{Cond}_i) \cdot \log \frac{n(\mathrm{Class}_k \cdot \mathrm{Cond}_i)}{n(\mathrm{Class}_k) \cdot p(\mathrm{Cond}_i)} \quad (6)$$

where $p(\mathrm{Cond}_i)$, computed as $n(\mathrm{Cond}_i)/n_s$, is used as a normalised factor, and n_c is the number of classes. It must be noted that although each rule is for a specific class value, the significance measures the novelty in the distribution impartially, for all the class values.

- *Unusualness of a rule*, defined as the weighted relative accuracy of a rule (Lavrac et al. 1999).

$$\mathrm{WRAcc}(R_i) = \frac{n(\mathrm{Cond}_i)}{n_s}\left(\frac{n(\mathrm{Class}_i \cdot \mathrm{Cond}_i)}{n(\mathrm{Cond}_i)} - \frac{n(\mathrm{Class}_i)}{n_s}\right) \quad (7)$$

The weighted relative accuracy of a rule can be described as the balance between the coverage of the rule $p(\mathrm{Cond})$ and its accuracy gain $p(\mathrm{Class}_i.\mathrm{Cond}_i) - p(\mathrm{Class}_i)$.

- *Coverage of a rule*, is the percentage of examples covered on average by one rule of the induced set of rules (Lavrac et al. 2004).

$$\mathrm{Cov}(R_i) = \frac{n(\mathrm{Cond}_i)}{n_s} \quad (8)$$

4 Evolutionary fuzzy systems for subgroup discovery

An EFS is essentially a fuzzy system enhanced by a learning process based on a evolutionary algorithm (Cordón et al. 2004; Herrera 2008). Currently, EFSs are being applied to a wide range of real-world problems. The research related to this area is growing, and a number of open problems and future directions can be found in (Ishibuchi 2007; Cordón et al. 2007; Casillas and Carse 2009).

The genetic representation of solutions is the most determinant aspect of any EFS proposal. In this sense, the proposals in the specialised literature follow two approaches to encode rules within a population of individuals (Cordón et al. 2004): The "Chromosome = Rule" approach, in which each individual codifies a single rule; and the "Chromosome = Set of Rules" approach, also called the Pittsburgh approach, in which each individual represents a set of rules.

There is a large body of literature which focuses on the extraction of fuzzy rules in descriptive data mining. This is widely applied to association rule extraction. The use of fuzzy sets in fuzzy rules extends the types of relationships that may be represented, facilitates the interpretation of rules in linguistic terms, and avoids unnatural boundaries in the partitioning of attribute domains. Proposals for the extraction of fuzzy association rules include (Hong et al.

2006; Kaya 2006; Alhajj and Kaya 2008; Chen et al. 2009; Alcalá-Fdez et al. 2009; Chen et al. 2009).

The use of fuzzy rules in an EFS contributes to the interpretability of the extracted rules since they allow us to obtain knowledge in a way approximate to human reasoning, but also attaining a balance between accuracy and interpretability (Gacto et al. 2009; Botta et al. 2009).

Fuzzy sets correspond to linguistic labels which are defined by means of their corresponding membership functions. These can be specified by the user or defined by means of a uniform partition, if the expert knowledge is not available. In this paper, uniform partitions with triangular membership functions are used.

To describe a fuzzy rule, we consider a SD problem with

- $\{X_m/m = 1, \ldots, n_v\}$, a set of features used to describe the subgroups, where n_v is the number of features. These variables can be categorical or numerical.
- $\{\text{Class}_j/j = 1, \ldots, n_c\}$, a set of values for the target variable, where n_c is the number of values.
- $\{E^k = (e_1^k, e_2^k, \ldots, e_{n_v}^k)/k = 1, \ldots, N\}$, a set of examples, where Class_j is the value of the target variable for the example E^k (i.e., the class for this example) and N is the number of examples for the descriptive induction process.
- $X_m : \{LL_m^1, LL_m^2, \ldots, LL_m^{l_m}\}$, 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: the variable X_m has l_m different linguistic labels to describe its domain in an understandable way.

A fuzzy rule describes a subgroup as

$$R : If X_1 \text{ is } LL_1^2 \text{ and } X_7 \text{ is } LL_7^1 \text{ then Class}_j \tag{9}$$

where Class_j is the value of the target variable corresponding to this rule, and the continuous variables are represented depending on the labels defined by themselves. In this way, when considering five linguistic labels in the previous rule for the continuous variables; for instance, the variable X_1 can take the following values $X_1 : \{LL_1^1, LL_1^2, \ldots, LL_1^5\}$.

Below two different SD proposals based on evolutionary fuzzy systems are studied: SDIGA (del Jesus et al. 2007a) and MESDIF (del Jesus et al. 2007b).

4.1 SDIGA: subgroup discovery iterative genetic algorithm

SDIGA is a mono-objective evolutionary algorithm for the extraction of fuzzy and/or crisp rules for SD task. This algorithm is described in detail in (del Jesus et al. 2007a) and applied to a marketing problem. In Romero et al. (2009), SDIGA is applied to a problem of e-learning.

SDIGA follows the iterative rule learning (IRL) 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. The core of SDIGA is a GA which uses a post-processing step based on a local search (a hill-climbing procedure). The hybrid GA extracts one simple and interpretable fuzzy rule. The post-processing step is applied to increase the generality of the extracted rule.

This model can use canonical or disjunctive normal form (DNF) fuzzy rules with a predefined set of linguistic labels. Each candidate solution is coded according to the "Chromosome = Rule" approach, in which only the antecedent is represented in the chromosome because all the individuals of the population are associated with the same value of the target feature. In this way, each run of SDIGA obtains a set of rules corresponding to a single value of the target variable. This allows us to ensure the extraction of rules describing subgroups for all the values of the target variable by simply running the algorithm with each one of the values of the target variable.

The objective of SDIGA is to obtain rules with high confidence, and which are understandable, interesting and general. This means that the problem has at least three objectives to maximise: support, confidence and unusualness of the rule. To achieve this, SDIGA uses the weighted sum method which weights a set of objectives into a single objective. Hence, this proposal uses a weighted lineal combination of different objectives:

- Support based on examples of the class (Eq. 2).
- Confidence (Eq. 3).
- Unusualness (Eq. 7).

The fitness function($f(c)$) of the GA combines these quality measures as an aggregation function as can be observed in Eq. 10:

$$f(\mathrm{c}) = \frac{\omega_1 \cdot \mathrm{Sup}_c(c) + \omega_2 \cdot \mathrm{Cnf}(c) + \omega_3 \cdot \mathrm{WRAcc}(c)}{\omega_1 + \omega_2 + \omega_3} \tag{10}$$

where ω_1, ω_2, and ω_3 are the weights considered for the different measures.

The genetic components used by SDIGA are a modified steady-state reproduction model, a two-point crossover and biased random mutation operators.

The post-processing phase, which improves the rule by means of a hill-climbing process, modifies the rule to increase the degree of support. The optimised rule will substitute the original one only if it overcomes a minimum confidence threshold.

It must be noted that the hybrid GA is included in an iterative process for the extraction of different rules in successive runs of the algorithm. For this purpose, when a run of the hybrid GA has finished and a fuzzy rule has been obtained, the positive instances of the rule (covered examples) are marked to prevent a new rule being obtained which covers exactly the same examples in the following GA runs. The iterative algorithm continues obtaining rules while the generated rules reach a minimum level of confidence and give information on areas of the search space in which there are examples not described by the rules generated in previous iterations.

4.2 MESDIF: multi-objective evolutionary subgroup discovery fuzzy rules

MESDIF is a multi-objective evolutionary algorithm (MOEA) for the extraction of rules which describe subgroups. This algorithm extracts rules whose antecedent represents a conjunction of variables and whose consequent is pre-fixed. The objective of this evolutionary process is to extract for each value of the target variable a variable number of different rules expressing information on the examples of the original set. The algorithm must be run as many times as the different values the target feature contains.

This algorithm can generate fuzzy and/or crisp DNF rules, for problems with continuous and/or nominal variables. The candidate solutions are coded according to the "Chromosome = Rule" approach, representing only the antecedent in the chromosome and associating all the individuals of the population with the same value of the target variable. Then, as in SDIGA, the extraction of rules describing subgroups of all the values of the target variable can be ensured by running the algorithm with each one of the values of the target variable.

The algorithm is based on the SPEA2 approach (Zitzler et al. 2002), 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 (Deb 2001).

To preserve the diversity at a phenotypic level our algorithm uses a niching 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. Therefore, in a run, we obtain a set of rules that provide us with knowledge about a property of interest.

The assignment of the fitness for the rules extracted is performed through the SPEA2 approach in MESDIF, and the rule induction process has the objective of obtaining rules with high predictive accuracy and which are comprehensible and interesting. In this proposal, three objectives have been defined, and the algorithm attempts to maximise all the defined objectives:

- Support based on the examples of the class (Eq. 2).
- Confidence (Eq. 3).
- Unusualness (Eq. 7).

This algorithm also uses a tournament selection operator, a two-point crossover operator, and a biased uniform mutation operator for the recombination of the population.

With respect to the environmental selection, MESDIF establishes a fixed length for the elite population. To maintain this fixed size of the elite population, the truncation and fill functions defined in SPEA2 are used. The truncation function allows the elimination of non-dominated solutions of the elite population if it exceeds the defined size (using a niche schema defined around the density measured by the distance to its k-nearest neighbour and, in an iterative process, eliminating in each iteration the individual that is nearest of others respect of the values of the objectives). The fill function allows to append 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).

At the end of the evolutionary process, a variable number of solutions are returned, composed of the the non-dominated individuals contained in the last elite population. This means that the algorithm can return at most as many rules as the size of the elite population.

5 Experimental study

The aim of this experimental study is, first, to determine which of the SD algorithms performs better on the psychiatric emergency department problem described in Sect. 2 Once determined which algorithm performs better, a new experiment is carried out using this algorithm to extract rules. These rules are then analysed in order to obtain interesting knowledge to characterise the type of patient more likely to visit the department depending on the time. So, Sect. 5.1 shows the comparison between the results obtained by the algorithms studied, and Sect. 5.2 analyses several subgroups obtained by the algorithm with the best results in the comparison.

5.1 Comparative study between SD algorithms

We have performed a comparative study between different SD algorithms to analyse the results obtained by them on the psychiatric emergencies problem. It must be noted that it has not been possible to experiment with Apriori-SD algorithm due to the dimensionality of the data set. The algorithms used for the comparison are a state-of-the-art algorithm and two evolutionary proposals:

- The classical SD algorithm CN2-SD (Lavrac et al. 2002).
- The evolutionary SD algorithm SDIGA (del Jesus et al. 2007a).
- The multi-objective evolutionary SD algorithm MESDIF (del Jesus et al. 2007b).

Because the CN2-SD algorithm is not able to handle continuous variables, a previous discretisation of these variables of the data set was needed. The discretisation process used is the Fayyad discretise (Fayyad and Irani 1993), the one used in the paper describing CN2-SD (Lavrac et al. 2002).

CN2-SD is a deterministic algorithm, whereas SDIGA and MESDIF are non-deterministic ones. So, for the non-deterministic algorithms SDIGA and MESDIF, five runs have been performed and the average values are used for each experiment.

To carry out the comparison, a tenfold cross-validation scheme has been used over the data set, and the average of the results obtained by MESDIF, SDIGA and CN2-SD for the test partitions are computed. We have developed the experiments according to the following procedure: for the deterministic algorithm CN2-SD we have performed a set of runs, varying the γ parameter (using the values 0.5, 0.7, 0.9 and additive); for the non-deterministic algorithm SDIGA, we have performed different runs using different values of minimum confidence (0.6, 0.7, 0.8 and 0.9, respectively); and for the non-deterministic algorithm MESDIF we have performed different runs using different sizes of the elite population (3, 4, 5 and 10, respectively). The parameters used for the experiments with the evolutionary algorithms SDIGA and MESDIF are shown in Table 2.

Table 2 Parameter specification for the evolutionary algorithms employed in the experimentation

Algorithm	Parameters
MESDIF	Linguistic labels = (3 and 5)
	Population size = 100
	Maximum evaluations = 10,000
	Crossover probability = 0.60
	Mutation probability = 0.01
SDIGA	Linguistic labels = (3 and 5)
	Weight for support = 0.4
	Weight for confidence = 0.3
	Weight for unusualness = 0.3
	Population size = 100
	Maximum evaluations = 10,000
	Crossover probability = 0.60
	Mutation probability = 0.01

Table 3 shows the average of the results obtained by MESDIF, SDIGA and CN2-SD for the test partitions, where

- *Granularity* is the number of linguistic labels used for the evolutionary algorithms, and *Disc* is the discretisation process used for CN2-SD.
- $Cnf_{\min}$ is the minimum confidence threshold used for SDIGA, Pop_{Eli} is the elite population size for MESDIF, and γ is the parameter used in the weighting scheme of the algorithm CN2-SD.
- n_{r} is the average of rules obtained.
- n_{v} is the average of variables for each rule.
- COV is the coverage, as defined in Eq. 8.
- SIG is the significance, as defined in Eq. 6.
- WRAcc is the unusualness, as defined in Eq. 7.
- SUP is the support, as defined in Eq. 1.
- CNF is the confidence, as defined in Eq. 3.

For each measure included in Table 3, the best result of each algorithm is highlighted in bold characters.

In any run of the algorithms, rules describing information corresponding to different values of *"admission time"*, the target variable, are obtained, but only SDIGA and MESDIF ensure the extraction of rules corresponding to all the values of this variable.

To analyse the results shown in Table 3, it is important to consider that an algorithm shows good behaviour for SD if it obtains good results with respect to the different quality measures, considering also a good relationship between support and confidence, and furthermore whether the algorithm obtains simple, general and accurate subgroups. Therefore, to select the algorithm with the best behaviour, it is necessary to analyse the number of rules and variables of each rule (a key aspect for the interpretability of the rules) and the results for each of the quality measures:

- With respect to the number of rules and variables of each rule, we are interested in discovering rules with few attributes in order to facilitate their comprehensibility. CN2-SD obtains rules describing subgroups with too many variables, which are not representative, and also rule sets which are too large; instead, MESDIF obtains subgroups which have few variables and are very representative, so obtaining better results.
- For the coverage measure(COV), the subgroups obtained by MESDIF cover more examples of the data set (with a best result of 0.549) than the subgroups of CN2-SD (best result is 0.247) and SDIGA (best result is 0.010). On an average, the rule sets obtained by MESDIF cover more than 50% of the examples.

Table 3 Results obtained by the algorithms from the data set

Algorithm	Granularity	Cnf_{min}	n_r	n_v	COV	SIG	WRAcc	SUP	CNF
SDIGA	3	0.6	3.13	1.97	0.006	0.420	0.000	0.712	**0.070**
		0.7	3.20	2.16	0.005	0.444	0.000	0.673	0.053
		0.8	3.33	2.38	**0.010**	**0.520**	0.000	0.633	0.056
		0.9	3.22	2.04	0.007	0.333	0.000	0.696	0.048
	5	0.6	3.18	2.05	0.009	0.285	0.000	0.695	0.059
		0.7	3.18	**1.96**	0.003	0.248	0.000	**0.775**	0.028
		0.8	3.42	2.27	0.006	0.358	0.000	0.664	0.055
		0.9	**3.09**	2.06	0.008	0.347	0.000	0.701	0.051
Algorithm	Granularity	Pop_{Eli}	n_r	n_v	COV	SIG	WRAcc	SUP	CNF
MESDIF	3	3	**9.00**	**4.15**	0.518	0.683	0.003	0.971	0.267
		4	12.00	4.40	0.524	0.834	0.005	0.977	0.301
		5	15.00	4.66	0.520	0.851	**0.007**	0.977	0.305
		10	29.92	5.52	0.494	1.003	0.006	0.953	**0.332**
	5	3	**9.00**	4.18	0.536	0.537	0.002	**0.995**	0.256
		4	12.00	4.33	**0.549**	0.765	0.006	0.993	0.302
		5	15.00	4.61	0.547	0.803	0.006	0.993	0.295
		10	30.00	5.38	0.544	**1.062**	**0.007**	**0.995**	0.189
Algorithm	Disc	γ	n_r	n_v	COV	SIG	WRAcc	SUP	CNF
CN2-SD	Fayyad	0.5	**23.00**	13.01	0.225	1.930	**0.003**	0.242	0.366
		0.7	30.40	13.00	0.227	1.814	0.002	0.242	0.330
		0.9	57.20	**12.86**	0.235	1.970	**0.003**	0.259	**0.381**
		Add	25.50	12.89	**0.247**	**2.021**	0.002	**0.270**	0.377

- In significance (SIG), the best results were obtained by CN2-SD (with values around 2.0), but MESDIF also obtains good results regarding this quality measure (with values around 1.0).
- For the unusualness measure (WRAcc), MESDIF obtains better results than the other algorithms (best values are around 0.007 against 0.003 for MESDIF and CN2-SD, respectively).
- In support (SUP), the results of MESDIF were excellent, covering almost 100% of the data, while SDIGA and CN2-SD cover around 70 and 25%, respectively).
- Respect to the confidence (CNF), CN2-SD and MESDIF obtain similar results (best results around 0.38 and 0.33, respectively). We have to note that this is a difficult, real problem in which it is very difficult to obtain high levels of confidence for the rules.

Taking into consideration the previous analysis, the algorithm which in general obtains the best results is MESDIF, because it obtains a better relationship between support and confidence, good results in unusualness and coverage, adequate results in significance, and also simple rules.

5.2 Analysis of rules obtained by MESDIF

Once determined that MESDIF is the algorithm which obtains better results for the psychiatric emergency data set, a new experiment is performed using the complete data set (without a cross-validation scheme) to analyse the rules obtained by MESDIF. Different runs have been performed using different sizes of the elite population: 3, 4, 5 and 10, respectively. The parameters used are the ones described in Table 2.

Table 4 shows the results obtained by MESDIF with respect to the different quality measures studied. In this experimental study with MESDIF, the best results were obtained using five linguistic labels for the continuous variables: very low, low, medium, high and very high. In this table, the best result obtained for each quality measure is highlighted in bold characters.

Table 5 shows some of the more representative rules obtained by MESDIF for each class, and their quality measures values: significance (SIG, Eq. 6), unusualness (WRAcc, Eq. 7), support (SUP, Eq. 2) and confidence (CNF, Eq. 3).

Table 4 Results obtained by MESDIF for the total examples of the data set

Algorithm	Granularity	Pop_{Eli}	n_r	n_v	COV	SIG	WRAcc	SUP	CNF
MESDIF	3	3	9.00	**4.11**	0.537	3.232	0.007	0.997	0.531
		4	12.00	4.25	0.532	3.738	0.010	0.971	0.509
		5	15.00	4.32	0.494	3.804	0.013	0.970	0.469
		10	30.00	5.48	0.503	**5.441**	**0.016**	0.938	0.438
	5	3	9.00	4.13	0.528	3.447	0.007	**0.999**	**0.566**
		4	12.00	4.20	**0.546**	3.533	0.011	0.990	0.491
		5	15.00	4.52	0.542	4.436	0.013	**0.999**	0.468
		10	30.00	5.35	0.532	5.200	**0.016**	0.993	0.431

Table 5 Rules obtained for MESDIF in the complete experimentation

♯	Rule	SIG	WRAcc	SUP_c	CNF
R_1	IF (others antidepressants=0 AND childhood disorders = 0 AND relative degree = 1) THEN admission time = day	4.044	0.017	0.449	0.289
R_2	IF (other substances=0 AND others antidepressants=0 AND relative degree = 0 AND childhood disorders = 1) THEN admission time = day	5.119	0.019	0.439	0.296
R_3	IF (age = high AND psychiatric history = 2 AND other substances = 0 AND childhood disorders = 0 AND autolytic behaviour = 0 AND side effects = 0 AND ET atypical neuroleptics = 0 AND mood stabiliser = 0) THEN admission time = day	20.463	0.014	0.149	0.450
R_4	IF (IsRNA=0 AND others antidepressants=0 AND childhood disorders=0) THEN admission time = day	0.539	0.008	0.968	0.250
R_5	IF (other substances = 0 AND tricyclic neuroleptics = 0 AND others antidepressants=0 AND personality disorder = 0 AND developmental disorder = 0) THEN admission time = day	4.223	0.025	0.796	0.277
R_6	IF (reason consultation = 2 AND developmental disorder = 0) THEN admission time = evening	10.419	0.022	0.258	0.512
R_7	IF (psychotropic drugs treatment = 1 AND developmental disorder = 0 AND autolytic behaviour = 0 AND discharge destination = 1) THEN admission time = evening	9.247	0.027	0.474	0.481
R_8	IF (developmental disorder = 0 AND autolytic behaviour = 0) THEN admission time = evening	1.471	0.009	0.884	0.427
R_9	IF (psychotropic drugs treatment = 1 AND developmental disorder = 0 AND autolytic behaviour = 0) THEN admission time = evening	5.558	0.020	0.705	0.447
R_{10}	IF (developmental disorder = 0 AND autolytic behaviour = 0 AND discharge destination = 1) THEN admission time = evening	4.804	0.023	0.579	0.460
R_{11}	IF (age = very low AND SSRI = 0 AND IsRNA = 0 AND depot neuroleptics = 0 AND previous consultation = 0 AND developmental disorder = 0) THEN admission time = night	8.186	0.008	0.231	0.550
R_{12}	IF (age = low AND SSRI = 0 AND IsRNA = 0 AND depot neuroleptics = 0 AND previous consultation = 0 AND developmental disorder = 0 AND ET tricyclic antidepressant = 0) THEN admission time = night	12.437	0.028	0.513	0.446
R_{13}	IF (previous consultation = 0 AND developmental disorder = 0 AND ET tricyclic antidepressant = 0) THEN admission time = night	8.616	0.037	0.731	0.401
R_{14}	IF (age = low AND cannabis = 0 AND SSRI = 0 AND IsRNA = 0 AND previous consultation = 0 AND developmental disorder = 0) THEN admission time = night	13.691	0.027	0.455	0.464
R_{15}	IF (lithium = 0 AND mood stabiliser = 0 AND autolytic behaviour = 1 AND SSRI = 0 AND ET depot neuroleptics = 0) THEN admission time = night	15.258	0.021	0.180	0.514

Analysing the obtained rules, the following can be highlighted regarding each of the periods corresponding to the different values of the variable of interest:

- The knowledge extracted for the subgroup of patients who visit the psychiatric emergency department during the day shows that this is an especially difficult subgroup. The rules obtained have a low level of confidence although some of them are very general and cover most of the patients. Rule R_3 is what may be of more interest to characterise the type of patients who visit the department during the day, although it is a more specific rule covering few patients.

- For the subgroup of patients attending the department during the evening, the algorithm obtains simpler rules (involving a fewer number of variables) and a priori easier to be interpreted by experts in psychiatry. In this subgroup, several rules include “no autolytic behaviour”, so it might be interesting for the characterisation of the patients. The rule that could provide more interesting knowledge is rule R_8. However, rule R_7 can provide interesting information for the characterisation of the subgroup.
- With respect to the rules extracted for the subgroup of patients attending the department during the night, they are more accurate than for the other subgroups. Rule R_{15} may be of more interest to characterise the patients who visit the department at this period of time. This rule can also be useful to characterise the patients with autolytic behavior.

The analysis conducted by experts of the rules obtained has allowed to establish some comments regarding the relationship between the available data of patients of the psychiatric emergency department and the time of arrival at the department:

- Although it is difficult to clarify, the rules obtained seem to show that some of the variables may have some influence on the time of arrival of the patients. The relationship between solar hours, weekend days, seasonality, weather conditions and psychiatric disorders has been a continuous subject of speculation due to contradictory findings. Several studies have tried to clarify this relationship (concretely between anxiety, depression, drugs, suicide and violent behaviour and temporal pattern) but results have been inconclusive. Beside social–environmental factors have been associated.
- Experts have special interest in the relationship that may exist between the time of admission and suicidality. Indeed, the rules obtained suggest a higher incidence of suicidal patients at night, but this requires a further study. Frequently, these studies have been performed in psychiatric emergency departments with clinic and organisational aims. In Bulbena et al. (2009) different results have been found by gender, day of week and hour of day. For example, the data indicate the existence of a seasonal pattern: in males, the peak day for attempted suicide was Monday, whereas in females, it was Sunday and Monday. Another example is that the majority of suicide attempts, especially by females, occurs late in the evening or early in the night. Experts have highlighted that the results obtained by MESDIF not only match the reported findings in Bulbena et al. (2009), but also allow to determine the characteristics of patients. This finding seems to be of particular importance for suicide prevention because it can contribute to the increase in the effectiveness of the organization of the work in the emergency department, and make therapists and patients’ families aware of the existence of periods of an increased suicide risk.

6 Conclusions

In this paper, a study on the rates of arrival at the psychiatric emergency department is developed. The main purpose is to discover relationships between the data of the patients and the arrival time at the department.

For this purpose, several SD algorithms have been used to determine which of them obtains better results: the classical CN2-SD algorithm, the evolutionary algorithm SDIGA and the multi-objective evolutionary algorithm MESDIF. The results show that MESDIF is the algorithm which obtains better results for this data set. It is further noted that since the extraction of rules with few variables is important to facilitate the comprehensibility of the rules in this field, evolutionary algorithms offer better solutions.

Once determined that the MESDIF algorithm is the one with better performance, experimentation has been carried out to extract subgroup description rules that allow extraction of interesting information with respect to the rate of admission of patients in the service and to characterise these patients.

This experimentation and the rules extracted have yielded useful information to be used in the organisation of the resources of the psychiatric emergency service. These results confirm that emergencies on suicide attempts are frequent at night although they are not exclusive in these hours.

As future work we consider to develop a new study including more variables and avoiding lost subjects according to the experts suggestions:

- It is necessary to include more factors and to obtain new data to actually determine the relationship between the different variables of interest. Therefore, they consider not only interesting to study the arrival time to the department, but also other variables such as season and other social and environmental factors. Assessing specific disorders instead of overall emergencies or other variables of a more general quality could shed new light on the relationship between weather conditions, social and environmental factors and psychiatric emergencies and its clinic and organisational impact.

- In addition, efforts to develop comprehensive models of suicidality should consider sleep problems as potentially independent indicators of risk. Chronic sleep problems are consistently associated with greater risk for suicidality, including suicidal ideation, planning and suicide attempt. Early morning awakening was associated with suicidal ideation, suicide planning and suicide attempt. Difficulty initiating sleep was a significant predictor of suicidal ideation and planning, while difficulty maintaining sleep during the night was a significant predictor of suicidal ideation and suicide attempt.

Acknowledgments This work was supported by the Spanish Ministry of Education, Social Policy and Sports under projects TIN-2008-06681-C06-02, and by the Andalusian Research Plan under project TIC-3928.

References

Agrawal R, Imieliski T, Swami A (1993) Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD international conference on management of data. ACM Press, pp 207–216

Aguilar-Ruiz J, Costa R, Divina F (2004) Knowledge discovery from doctor–patient relationship. In: Proceedings of the ACM symposium on applied computing, vol 1, pp 280–284

Ainon R, Lahsasna A, Wah T (2009) A transparent classification model using a hybrid soft computing method. In: AMS, pp 146–151

Alcalá-Fdez J, Alcalá R, Gacto MJ, Herrera F (2009) Learning the membership function contexts for mining fuzzy association rules by using genetic algorithms. Fuzzy Sets Syst 160(7):905–921

Alhajj R, Kaya M (2008) Multi-objective genetic algorithms based automated clustering for fuzzy association rules mining. J Intell Inform Syst 31(3):243–264

Atzmueller M, Puppe F (2006) SD-Map—a fast algorithm for exhaustive subgroup discovery. In: Proceedings of the 17th European conference on machine learning and 10th European conference on principles and practice of knowledge discovery in databases, vol 4213. Springer, Berlin, pp 6–17

Atzmueller M, Puppe F, Buscher HP (2004) Towards knowledge-intensive subgroup discovery. In: Proceedings of the Lernen—Wissensentdeckung—Adaptivität—Fachgruppe Maschinelles Lernen, pp 111–117

Atzmueller M, Puppe F, Buscher HP (2005) Profiling examiners using intelligent subgroup mining. In: Workshop on intelligent data analysis in medicine and pharmacology, pp 46–51

Baca-Garcfa E, Perez-Rodriguez M, Basurte-Villamor I, Saiz-Ruiz J, Leiva-Murillo J, Prado-Cumplido MD, Santiago-Mozos R, ArtTs-Rodrfguez A, Leon JD (2006) Using data mining to explore complex clinical decisions: a study of hospitalization after a suicide attempt. J Clin Psychiatry 67(7):1124–1132

Baca-Garcia E, Perez-Rodriguez M, et al (2008) Patterns of mental health service utilization in a general hospital and outpatient mental health facilities: analysis of 365,262 psychiatric consultations. Eur Arch Psychiatry Clin Neurosci 258(2):117–123

Botta A, Lazzerini B, Marceloni F, Stefanescu DC (2009) Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index. Soft Comput 13(5):437–449

Bulbena A, Sperry L, Garcia-Ribera C, Merino A, Mateu G, Torrens M, San-Gil J, Cunillera J (2009) Impact of the summer 2003 heat wave on the activity of two psychiatric emergency departments. In: Actas Esp. Psiquiatr., vol 37, pp 158–165

Casillas J, Carse B (2009) Special issue on genetic fuzzy systems: recent developments and future directions. Soft Comput 13(5): 417–418

Chen CH, Hong TP, Tseng VS (2009a) An improved approach to find membership functions and multiple minimum supports in fuzzy data mining. Expert Syst Appl 36(6):10,016–10,024

Chen CH, Hong TP, Tseng VS, Lee CS (2009b) A genetic-fuzzy mining approach for items with multiple minimum supports. Soft Comput 13(5):521–533

Clark P, Niblett T (1989) The CN2 induction algorithm. Mach Learn 3:261–283

Clercq MD, Lamarre S, Vergouwen H (1998) Emergency psychiatry and mental health policy: an international point of view. Elsevier, Amsterdam

Cordón O, Gomide F, Herrera F, Hoffmann F, Magdalena L (2004) Ten years of genetic fuzzy systems. Current framework and new trends. Fuzzy Sets Syst 14:5–31

Cordón O, Alcalá R, Alcalá-Fdez J, Rojas I (2007) Special issue on genetic fuzzy systems: what's next? Editorial. IEEE Trans Fuzzy Syst 15(4):533–535

Deb K (2001) Multi-objective optimization using evolutionary algorithms. Willey, New York

del Jesus MJ, González P, Herrera F, Mesonero M (2007a) Evolutionary fuzzy rule induction process for subgroup discovery: a case study in marketing. IEEE Trans Fuzzy Syst 15(4): 578–592

del Jesus MJ, González P, Herrera F (2007b) Multiobjective genetic algorithm for extracting subgroup discovery fuzzy rules. In: Proceedings of the IEEE symposium on computational intelligence in multicriteria decision making. IEEE Press, pp 50–57

Drobics M, Botzheim J, Koczy L (2007) Increasing diagnostic accuracy by meta optimization of fuzzy rule bases. In: IEEE international conference on fuzzy systems

Fayyad UM, Irani KB (1993) Multi-interval discretization of continuous-valued attributes for classification learning. In: 13th international joint conference on artificial intelligence, pp 1022–1029

Fayyad UM, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery: an overview. In: Advances in knowledge discovery and data mining. AAAI/MIT Press, pp 1–34

Fogel G (2008) Computational intelligence approaches for pattern discovery in biological systems. Brief Bioinform 9(4):307–316

Gacto MJ, Alcalá R, Herrera F (2009) Adaptation and application of multi-objective evolutionary algorithms for rule reduction and parameter tuning of fuzzy rule-based systems. Soft Comput 13(5):419–436

Gamberger D, Lavrac N (2002) Expert-guided subgroup discovery: methodology and application. J Artif Intell Res 17:501–527

Gamberger D, Lavrac N (2003) Active subgroup mining: a case study in coronary heart disease risk group detection. Artif Intell Med 28(1):27–57

Gamberger D, Lavrac N, Krstaic A, Krstaic G (2007) Clinical data analysis based on iterative subgroup discovery: experiments in brain ischaemia data analysis. Appl Intell 27(3):205–217

Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning. Addison-Wesley Longman Publishing Co, London

Grosskreutz H, Rueping S (2009) On subgroup discovery in numerical domains. Data Min Knowl Discov 19(2):210–216

Han J, Pei J, Yin Y (2000) Mining frequent patterns without candidate generation. In: Proceedings of the 2000 ACM SIGMOD international conference on management of data. ACM Press, pp 1–12

Herrera F (2008) Genetic fuzzy systems: taxomony, current research trends and prospects. Evol Intell 1:27–46

Hong TP, Chen CH, Wu YL, Lee YC (2006) A GA-based fuzzy mining approach to achieve a trade-off between number of rules and suitability of membership fuctions. Soft Comput 10(11): 1091–1101

Hüllermeier E (2005) Fuzzy methods in machine learning and data mining: status and prospects. Fuzzy Sets Syst 156(3):387–406

Ishibuchi H (2007) Multiobjective genetic fuzzy systems: review and future research directions. In: IEEE international conference on fuzzy systems, pp 913–918

Jovanoski V, Lavrac N (2001) Classification rule learning with APRIORI-C. In: 10th Portuguese conference on artificial intelligence on progress in artificial intelligence, knowledge extraction, multi-agent systems, logic programming and constraint solving, vol 2258. Springer, Berlin, pp 44–51

Kavsek B, Lavrac N (2006) APRIORI-SD: adapting association rule learning to subgroup discovery. Appl Artif Intell 20:543–583

Kaya M (2006) Multi-objective genetic algorithm based approaches for mining optimized fuzzy association rules. Soft Comput 10(7):578–586

Kielan K, Kucharska-Pietura K, Warchala A, Konopka M, Pieniazek P, Hartel M (2004) "saba"—application of knowledge and state-of-the-art technologies in the field of psychiatry for development of new diagnostics prevention and therapeutic tools for schizophrenia. In: vol 57(1), pp 152–157

Kloesgen W (1996) Explora: a multipattern and multistrategy discovery assistant. In: Advances in knowledge discovery and data mining. American Association for Artificial Intelligence, pp 249–271

Kralj P, Lavrac N, Gamberger D, Krstaic A (2007) Contrast set mining through subgroup discovery applied to brain ischaemina data. In: 11th Pacific-Asia conference on knowledge discovery and data mining, vol 4426. Springer, Berlin, pp 579–586

Lavrac N, Flach PA, Zupan B (1999) Rule evaluation measures: a unifying view. In: Proceedings of the 9th international workshop on inductive logic programming, vol 1634. Springer, Berlin, pp 174–185

Lavrac N, Flach P, Kavsek B, Todorovski L (2002) Rule induction for subgroup discovery with CN2-SD. In: Proceedings of the 2nd international workshop on integration and collaboration aspects of data mining, decision support and meta-learning, pp 77–87

Lavrac N, Cestnik B, Gamberger D, Flach PA (2004a) Decision support through subgroup discovery: three case studies and the lessons learned. Mach Learn 57(1–2):115–143

Lavrac N, Kavsek B, Flach PA, Todorovski L (2004b) Subgroup Discovery with CN2-SD. J Mach Learn Res 5:153–188

López B, Barrera V, Meléndez J, Pous C, Brunet J, Sanz J (2009) Subgroup discovery for weight learning in breast cancer diagnosis. In: Proceedings of the 12th conference on artificial intelligence in medicine, vol 5651. LNAI, pp 360–364

Mantzaris D, Anastassopoulos G, Iliadis L, Adamopoulos A (2009) An evolutionary technique for medical diagnostic risk factors selection. IFIP Int Federation Inform Process 195–203

Masuda G, Sakamoto N, Yamamoto R (2002) A framework for dynamic evidence based medicine using data mining. In: Proceedings of the IEEE symposium on computer-based medical systems, pp 117–122

Michie D, Spiegelhalter DJ, Tayloy CC (1994) Machine learning. Ellis Horwood

Mueller M, Rosales R, Steck H, Krishnan S, Rao B, Kramer S (2009) Subgroup discovery for test selection: a novel approach and its application to breast cancer diagnosis. In: Proceedings of the 8th international symposium on intelligent data analysis, vol 5772. Springer, Berlin, pp 119–130

Nannings B, Bosnian RJ, Abu-Hanna A (2009) A subgroup discovery approach for scrutinizing blood glucose management guidelines by the identification of hyperglycemia determinants in ICU patients. Methods Inform Med 47(6):480–488

Papageorgiou E, Papandrianos N, Apostolopoulos D, Vassilakos P (2008) Fuzzy cognitive map based decision support system for thyroid diagnosis management. In: IEEE international conference on fuzzy systems, pp 1204–1211

Romero C, González P, Ventura S, del Jesus MJ, Herrera F (2009) Evolutionary algorithm for subgroup discovery in e-learning: a practical application using Moodle data. Expert Syst Appl 36:1632–1644

ValdTs J, Barton A, AS Haqqani A (2008) Analysis of mass spectrometry data of cerebral stroke samples: an evolutionary computation approach to resolve and quantify peptide peaks. Genet Program Evol Mach 9(3):257–274

Wrobel S (1997) An algorithm for multi-relational discovery of subgroups. In: Proceedings of the 1st European symposium on principles of data mining and knowledge discovery, vol 1263. Springer, Berlin, pp 78–87

Yardimci A (2009) Soft computing in medicine. Appl Soft Comput J 9(3):1029–1043

Yu L, Wu C, Yeh J, Jang F (2008) HAL-based evolutionary inference for pattern induction from psychiatry web resources. IEEE Trans Evol Comput 12(2):160–170

Zadeh LA (1975) The concept of a linguistic variable and its applications to approximate reasoning. Parts I, II, III. Information Science 8–9:199–249,301–357,43–80

Zelezny F, Lavrac N (2006) Propositionalization-based relational subgroup discovery with RSD. Mach Learn 62:33–63

Zitzler E, Laumanns M, Thiele L (2002) SPEA2: Improving the strength pareto evolutionary algorithm for multiobjective optimization. In: International congress on evolutionary methods for design optimization and control with applications to industrial problems, pp 95–100