Subgroup Discovery: Real-World Applications

C. J. Carmona
Languages and Computer Technology Systems, Department of Civil Engineering, University of Burgos, 09006, Burgos (Spain)
D. Elizondo
School of Computer Science and Informatics, De Montfort University, The Gateway, Leicester, LE1 9BH, (United Kingdom)

Abstract

Subgroup discovery is a data mining technique which extracts interesting rules with respect to a target variable. An important characteristic of this task is the combination of predictive and descriptive induction. In this paper, an overview about subgroup discovery is performed. In addition, different real-world applications solved through evolutionary algorithms where the suitability and potential of this type of algorithms for the development of subgroup discovery algorithms are presented.

1. Introduction

This papers present a review of subgroup discovery by analysing the main properties, models and quality measures employed by subgroup discovery approaches.

Subgroup discovery (Kloesgen, 1996, Wrobel, 1997) is a broadly applicable data mining technique aimed at discovering interesting relationships between different objects in a set with respect to a specific property which is of interest to the user the target variable. The patterns extracted are normally represented in the form of rules and called subgroups (Siebes, 1995).

There is an important difference between subgroup discovery and classification because subgroup discovery attempts to describe knowledge for the data while a classifier attempts to predict it. Furthermore, the model obtained by a subgroup discovery algorithm is usually simple and interpretable, while that obtained by a classifier is complex and precise.

In this way, subgroup discovery is an interesting descriptive data mining task in order to analyse problem and in order to obtain interesting and unusual knowledge with respect to an interest property or class. This work presents the most important real-world applications solved through subgroup discovery algorithms based on evolutionary fuzzy systems.

The paper is organised as follows. Section 2 introduces subgroup discovery, its positioning in data mining and its main elements; Section 3 classifies the most important quality measures used in subgroup discovery; Section 4 presents a historical revision of the subgroup discovery algorithms; finally, Section 5 shows the most relevant real-world problems using algorithms of subgroup discovery based on evolutionary fuzzy systems.

2. Subgroup Discovery

In the following subsections the formal definition of the subgroup discovery task, the relation with other data mining tasks and the main elements of a subgroup discovery algorithm are depicted.

2.1. Definition of subgroup discovery

The concept of subgroup discovery was initially introduced by Kloesgen (Kloesgen, 1996) and Wrobel (Wrobel, 1997), and more formally defined by Siebes (Siebes, 1995) but using the name Data Surveying for the discovery of interesting subgroups. It can be defined as (Wrobel, 2001):

> *"In subgroup discovery, we assume we are given a so-called population of individuals (objects, customer, ...) and a property of those individuals we are interested in. The task of subgroup discovery is then to discover the subgroups of the population that are statistically "most interesting", i.e., are as large as possible and have the most unusual statistical (distributional) characteristics with respect to the property of interest."*

Subgroup discovery attempts to search relations between different properties or variables of a set with respect to a target variable. Due to the fact that subgroup discovery is focused in the extraction of relations with interesting characteristics, it is not necessary to obtain complete but partial relations. These relations are described in the form of individual rules.

Then, a rule (R), which consists of an induced subgroup description, can be formally defined as (Gamberger and Lavrac, 2002, Lavrac et al., 2004a):

$$R : Cond \rightarrow Target_{value}$$

where $Target_{value}$ is a value for the variable of interest (target variable) for the subgroup discovery task (which also appears as $Class$ in the literature), and $Cond$ is commonly a conjunction of features (attribute-value pairs) which is able to describe an unusual statistical distribution with respect to the $Target_{value}$.

As an example, let D be a dataset with three variables $Age = \{Less\ than\ 25,\ 25\ to\ 60,\ More\ than\ 60\}$, $Sex = \{M,\ F\}$ and $Country = \{Spain,\ USA,\ France,\ German\}$, and a variable of interest target variable $Money = \{Poor,\ Normal,\ Rich\}$. Some possible rules containing subgroup descriptions are:

$$R_1 : (Age = Less\ than\ 25\ AND\ Country = German) \rightarrow Money = Rich$$
$$R_2 : (Age = More\ than\ 60\ AND\ Sex = F) \rightarrow Money = Normal$$

where rule R_1 represents a subgroup of German people with less than 25 years old for which the probability of being rich is unusually high with respect to the rest of the population, and rule R_2 represents that women with more than 60 years old are more likely to have a normal economy than the rest of the population.

2.2. Subgroup discovery versus classification

Data mining is a stage of the Knowledge Discovery in Databases defined as “the non-trivial extraction of implicit, unknown, and potentially useful information from data” (Fayyad et al., 1996). Description of the ten most used data mining algorithms can be found in (Wu et al., 2009). Data mining techniques can be applied from two different perspectives:

- Predictive induction, whose objective is the discovery of knowledge for classification of prediction. Among its features, we can find classification (Cherkassky and Mulier, 2007), regression (Cherkassky and Mulier, 2007), or temporal series (Box et al., 2008).

- Descriptive induction, whose main objective is the extraction of interesting knowledge from the data. Its features include association rules (Agrawal et al., 1993), summarisation (Zembowicz and Zytkow, 1996) or subgroup discovery (Kloesgen, 1996, Wrobel, 1997) can be mentioned.

In Fig. 1 the main difference between descriptive and predictive induction can be found. Fig. 1.a represents a precise and complex model (classifier) for predictive induction which divides perfectly the space in two determined regions with respect to the type of objects in the set. This model is based on precision and interpretability. However, Fig. 1.b represents a model for descriptive induction which describes groups of elements (clusters) in the set, without a target variable, and based on support and confidence of the objects. As can be observed the model on the left (predictive induction) has a different goal with respect to the model on the right (descriptive induction). Therefore, different heuristics and evaluation criteria in both types of learning are employed.

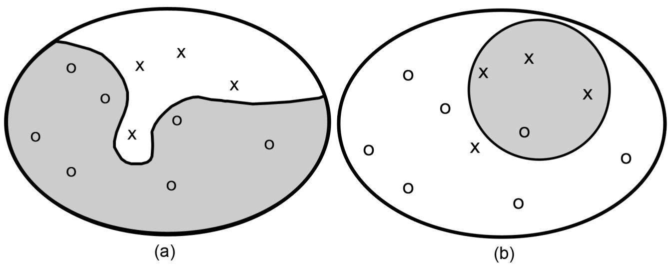

Figure 1: Models of different techniques of knowledge discovery of databases

Subgroup discovery (Kloesgen, 1996) is a technique for the extraction of patterns, with respect to a property of interest in the data, or target variable. This technique is somewhere halfway between predictive and descriptive induction, and its goal is to generate in a single and interpretable way subgroups to describe relations between independent variables and a certain value of the target variable. The algorithms for this task must generate subgroups for each value of the target variable. Therefore, an execution for each value of the variable must be performed.

A rule for subgroup discovery is represented in Fig. 2, where two values for the target variable can be found ($Target_{value} = x$ and $Target_{value} = o$). In this representation a subgroup for the first value of the target variable can be observed, where the rule attempts to cover a high number of objects with a single function: a circle. As can be observed the subgroup does not cover all the examples for the target value x even the examples covered are not positive in all the cases, but the form of this function is uniform and very interpretable with respect others. In this way the algorithm achieves a reduction of the complexity. Furthermore, the true positive rate for the value of the target variable is high, with a value of 75%.

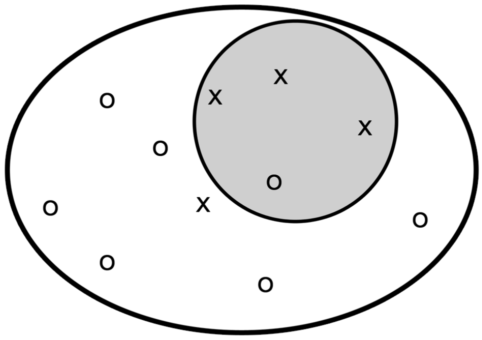

Figure 2: Representation of a subgroup discovery rule with respect to a value (x) of the target variable

The subgroup discovery task is differentiated from classification techniques basically because subgroup discovery attempts to describe knowledge for the data while a classifier attempts to predict it. Furthermore, the model obtained by a subgroup discovery algorithm is usually simple and interpretable, while that obtained by a classifier is complex and precise.

Currently, several techniques lie halfway between descriptive and predictive data mining. "Supervised Descriptive Rule Induction" (Kralj-Novak et al., 2009) is a new recently proposed paradigm which includes techniques combining the features of both types of induction, and its main objective is to extract descriptive knowledge from the data of a property of interest. These techniques use supervised learning to solve descriptive tasks. Within this new paradigm, the following data mining techniques are included:

- Subgroup Discovery (Kloesgen, 1996, Wrobel, 1997), defined as the extraction of interesting subgroups for a target value.
- Contrast Set Mining (Bay and Pazzani, 2001), defined as "a conjunction of attribute-value pairs defined on groups with no attribute occurring more than once".
- Emerging Pattern Mining (Dong and Li, 1999) defined as "patterns whose frequencies in two classes differ by a large ratio".

The main difference between these techniques is that while subgroup discovery algorithms attempt to describe unusual distributions in the search space with respect a value of the target variable, contrast set and emerging pattern algorithms seek relationships of the data with respect to the possible values of the target variable. Contrast set algorithms attempts to obtain high differences of support between the possible values and emerging pattern algorithms search patterns with different frequencies in two classes of the target variable. These last two techniques are based on measures of coverage and accuracy and subgroup discovery is also focused on novelty and unusualness measures as can be observed in the following sections.

2.3. *Main elements in a subgroup discovery algorithm*

Different elements can be considered the most important when a subgroup discovery approach must be applied. These elements are defined below (Atzmueller et al., 2004):

- *Type of the target variable*. Different types for the variable can be found: binary, nominal or numeric. For each one different analyses can be applied considering the target variable as a dimension of the reality to study.
  - Binary analysis. The variables have only two values (True or False), and the task is focused on providing interesting subgroups for each of the possible values.
  - Nominal analysis. The target variable can take an undetermined number of values, but the philosophy for the analysis is similar to the binary, to find subgroups for each value.
  - Numeric analysis. This type is the most complex because the variable can be studied different ways such as dividing the variable in two ranges with respect to the average, discretisising the target variable in a determined number of intervals (Moreland and Truemper, 2009), or searching for significant deviations of the mean among others.
- *Description language*. The representation of the subgroups must be suitable for obtaining interesting rules. These rules must be simple and therefore are represented as attribute-value pairs in conjunctive or disjunctive normal form in general. Furthermore, the values of the variables can be represented as positive and/or negative, through fuzzy logic, or through the use of inequality or equality and so on.
- *Quality measures*. These are a key factor for the extraction of knowledge because the interest obtained depends directly on them. Furthermore, quality measures provide the expert with the importance and interest of the subgroups obtained. Different quality measures have been presented in the specialised bibliography (Gamberger and Lavrac, 2003, Kloesgen, 1996, Kloesgen and Zytkow, 2002, Lavrac et al., 2004a), but there is no consensus about which are the most suitable for use in subgroup discovery.
- *Search strategy*. This is very important, since the dimension of the search space has an exponential relation to the number of features and values considered. Different strategies have been used up to the moment, for example beam search, evolutionary algorithms, search in multi-relational spaces, etc.

3. Quality measures

A wide number of quality measures have been presented in the subgroup discovery literature both to guide the search process in order to find the best subgroup discovery rules and to measure the quality of the subgroup discovery rule set finally obtained (Kloesgen, 1996, Lavrac et al., 2004b). The most common quality measures used in subgroup discovery can be classified by their main objective such as:

- Complexity measures, related to the interpretability of the subgroups, i.e. to the simplicity of the knowledge extracted.

- Generality measures, used to quantify the quality of individual rules according to the individual patterns of interest covered.
- Precision measures, showing the precision of the subgroups.
- Interest measures, intended for selecting and ranking patterns according to their potential interest to the user.
- Hybrid, that attempt to obtain a good trade-off between different objectives.

Table 1 summarises the *Quality measures* most used in subgroup discovery (Herrera et al., 2011) and their main characteristics.

Table 1: Classification of the quality measures used in subgroup discovery

Quality measure	*C*	*G*	*P*	*I*
Number of rules	X			
Number of variables	X			
Coverage (Lavrac et al., 2004b)		X		
Support (Lavrac et al., 2004b)		X		
Confidence (Agrawal et al., 1996)			X	
Precision measure Q_c (Gamberger and Lavrac, 2002)			X	
Precision measure Q_g (Kloesgen, 1996)			X	
Q_g-Weight (Gamberger and Lavrac, 2002)			X	
Interest (Noda et al., 1999)				X
Novelty (Wrobel, 1997)				X
Lift (Brin et al., 1997)				X
Significance (Kloesgen, 1996)				X
Sensitivity (Kloesgen, 1996)		X	X	
False Alarm (Gamberger and Lavrac, 2002)		X	X	
Specificity (Kloesgen, 1996)		X	X	
Unusualness (Lavrac et al., 1999)		X	X	X
Piatetstky-Shapiro (Grosskreutz et al., 2008)		X	X	X

C=Complexity, G=Generality, P=Precision and I=Interest

According to the subgroup discovery concept the obtaining of interesting, simple and interpretable subgroups, covering the majority of the examples of the interest property (target variable) is desirable. Considering this definition and the analysis of the different quality measures used in the literature, we propose three guidelines in order to establish the type of measure more suitable, to guide the search process and to analyse the quality of the subgroups obtained by any subgroup discovery algorithm:

- *Interpretability*. A subgroup discovery proposal must obtain few rules containing a low number of variables in the antecedent part in order to help to the experts to understand and use the extracted knowledge, i.e. simple and interpretable subgroups are preferred in subgroup discovery task.
- *Relation sensitivity-confidence*. A subgroup discovery algorithm must obtain results with a good precision, where the majority examples covered belong to the target variable, i.e.

it must achieve the best possible relation between sensitivity and confidence. Both quality measures are primordial in order to provide subgroups to the experts that cover the higher number of described correctly examples. It is difficult for the algorithms to obtain this compromise due to the loss suffered by a measure when trying to increase the other.

- *Novelty*. A subgroup discovery model must contribute novel knowledge, providing the experts with information in order to describe unusual and interesting behaviour within the data. This objective could be measured with a wide number of quality measures as novelty, interest or significance, among others. Nevertheless, it is important to highlight the utility of the unusualness to measure this objective because it contributes with generality and confidence to the problem. Moreover, this quality measure is widely used in the specialised bibliography.

It can be considered that the main purpose of a subgroup discovery algorithm is to find a good trade-off between these three guidelines because this lead to the obtaining of good results in a wide number of quality measures and not only in the ones used in the search process.

4. Algorithms

In this section, the subgroup discovery approaches developed so far are described. There are several proposals of algorithms for subgroup discovery. To classify these algorithms it can be distinguished between extensions of classification algorithms, extensions of association algorithms and evolutionary fuzzy systems. Table 2 shows the main algorithms for subgroup discovery developed so far under this classification.

Table 2: Classification of the algorithms for subgroup discovery developed so far

Extensions of classification algorithms
EXPLORA (Kloesgen, 1996)
MIDOS (Wrobel, 1997)
SubgroupMiner (Kloesgen and May, 2002)
SD (Gamberger and Lavrac, 2002)
CN2-SD (Lavrac et al., 2004b)
RSD (Lavrac et al., 2003, Zelezny and Lavrac, 2006)
Extensions of association algorithms
APRIORI-SD (Kavsek and Lavrac, 2006, Kavsek et al., 2003)
SD4TS (Mueller et al., 2009)
SD-MAP (Atzmueller and Puppe, 2006)
DpSubgroup (Grosskreutz et al., 2008)
Merge-SD (Grosskreutz and Rueping, 2009)
IMR (Boley and Grosskreutz, 2009)
Evolutionary algorithms
SDIGA (del Jesus et al., 2007c)
MESDIF (Berlanga et al., 2006, del Jesus et al., 2007b)
NMEEF-SD (Carmona et al., 2009b, 2010)

4.1. *Extensions of classification algorithms*

Among the algorithms for subgroup discovery developed as extensions of classification algorithms it can be distinguished between the pioneering algorithms and those based on classification algorithms. Both are described below.

4.1.1. *The pioneering algorithms*

The first algorithms developed for subgroup discovery –EXPLORA and MIDOS– are extensions of classification algorithms and use decision trees. They can employ two different strategies for the search, exhaustive and heuristic search, and several quality measures to evaluate the quality of the subgroups.

- EXPLORA (Kloesgen, 1996) was the first approach developed for subgroup discovery. It uses decision trees for the extraction of rules. The rules are specified by defining a descriptive schema and implementing a statistical verification method. The interest of the rules is measured using statistical measures (Kloesgen, 1996) such as evidence, generality, redundancy or simplicity, among others. EXPLORA can apply both exhaustive and heuristic subgroup discovery strategies without pruning.

- MIDOS (Wrobel, 1997) applies the EXPLORA approach to multi-relational databases. The goal is to discover subgroups of the target variable (defined as first order conjunctions) that have an unusual statistical distribution with respect to the complete population. MIDOS uses optimistic estimation and searches the space of rules exhaustively, except for safe pruning (using a minimum support pruning) for binary target variables. In order to find precise and significant subgroups the size of the subgroups and the distributional unusualness are considered. This algorithm can also use sampling in the example space to reduce the search space and speed up the search process.

These algorithms can use exhaustive or heuristic search. Exhaustive evaluation of the candidate rules allows the best subgroups to be found, but if the search space becomes too large this is not affordable. Then a heuristic search can be used to reduce the number of potential subgroups to consider (if the rules are ordered by generality, parts of the search space can be pruned). EXPLORA performs the subgroup discovery task from data in a single relation, while MIDOS can handle multiple relational tables.

4.1.2. *Algorithms based on classification*

Several algorithms developed for the subgroup discovery task have been developed by means of adaptations of classification rule learners. Classification rule learning algorithms have the objective of generating models consisting of a set of rules inducing properties of all the classes of the target variable, while in subgroup discovery the objective is to discover individual rules of interest. Moreover, classification rule learning algorithms do not appropriately address the subgroup discovery task as they use the covering algorithm for rule set construction. So in order to use a classification rule learning algorithm for subgroup discovery some modifications must be implemented. The algorithms described here attempt to overcome the inappropriate bias of the standard covering algorithm (only the first induced rules may be of interest as subgroup descriptions). They then use a modified weighted covering algorithm and introduce example weights to modify the search heuristic. They are briefly detailed below:

- SubgroupMiner (Kloesgen and May, 2002) is an extension of EXPLORA and MIDOS. It is an advanced subgroup discovery system that uses decision rules and interactive search in the space of the solutions, allowing the use of very large databases by means of the efficient integration of databases, multi-relational hypotheses, visualisation based on interaction options, and the discovery of structures of causal subgroups. This algorithm can use several quality measures to verify if the statistical distribution of the target is significantly different in the extracted subgroup, but the most usual is the binomial test (Kloesgen, 1996). This can handle both numeric and nominal target attributes, but for numeric variables a previous discretisation is performed.

- SD (Gamberger and Lavrac, 2002) is a rule induction system based on a variation of the beam search algorithms and guided by expert knowledge: instead of defining an optimal measure to discover and automatically select the subgroups, the objective is to help the expert in performing flexible and effective searches on a wide range of optimal solutions. Discovered subgroups must satisfy the minimal support criteria and must also be relevant. The algorithm keeps the best subgroups descriptions in a fixed width beam and in each iteration a conjunction is added to every subgroup description in the beam, replacing the worst subgroup in the beam by the new subgroup if it is better. To evaluate the quality of the subgroups Q_g measure is used. High quality subgroups must cover as many $Target_{value}$ examples and as few non-$Target_{value}$ examples as possible. The algorithm SD uses a visualisation method (Gamberger et al., 2002) to provide the experts with an easy tool to test the subgroups. In (Gamberger and Lavrac, 2003) methods to avoid noise and outliers values are presented. Different constraints (filtering of subgroups) are described for the algorithm SD in (Lavrac, 2005): Maximum length (a threshold defined by the user) and minimum support for the rules. This algorithm is implemented in a module of the software tool Orange(Demsar et al., 2004).

- CN2-SD (Lavrac et al., 2004a) is a subgroup discovery algorithm obtained by adapting a standard classification rule-learning approach CN2 (Clark and Niblett, 1989, Clark and Boswell, 1991) to subgroup discovery. It induces subgroups in the form of rules using a modified unusualness as the quality measure for rule selection. This approach performs subgroup discovery through the following modifications of CN2: a) replacing the accuracy-based search heuristic with a new unusualness heuristic that combines the generality and accuracy of the rule; b) incorporating example weights into the covering algorithm; c) incorporating example weights into the unusualness search heuristic; d) using probabilistic classification based on the class distribution of examples covered by individual rules. An extension of this algorithm named CN2-MSD (Abudawood and Flach, 2009) has been developed to manage multi-class target variables. This algorithm is implemented in a module of the software tool Orange(Demsar et al., 2004).

- RSD (Relational subgroup discovery) (Lavrac et al., 2003, Zelezny and Lavrac, 2006) has the objective of obtaining population subgroups which are as large as possible, with a statistical distribution as unusual as possible with respect to the property of interest, and different enough to cover most of the target population. It is an upgrade of the CN2-SD algorithm which enables relational subgroup discovery.

4.2. Extensions of association algorithms

An association rule algorithm attempts to obtain relations between the variables of the data set. In this case, several variables can appear both in the antecedent and consequent of the rule. In contrast, in subgroup discovery the consequent of the rule, consisting of the property of interest is prefixed. The characteristics of the association rule algorithms make it feasible to adapt these algorithms for the subgroup discovery task. The algorithms based on association rule learners are briefly described below:

- APRIORI-SD (Kavsek and Lavrac, 2006, Kavsek et al., 2003) 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). APRIORI-SD uses a postprocessing mechanism, unusualness, as the quality measure for the induced rules and probabilistic classification of the examples. For the evaluation of the set of rules the area under the ROC curve is used, in conjunction with the support and significance of each individual rule, and the size and accuracy of the set of rules. This algorithm is implemented in a module of the software tool Orange (Demsar et al., 2004).

- SD4TS (Mueller et al., 2009) is an algorithm based on APRIORI-SD but using the quality of the subgroup to prune the search space even more. The quality measure used is specified by the problem analysed. The algorithm does not need a covering heuristic.

- SD-Map (Atzmueller and Puppe, 2006) is an exhaustive subgroup discovery algorithm that uses the well-known FP-growth method (Han et al., 2000) for mining association rules with adaptations for the subgroup discovery task. It uses an implicit depth-first search step for evaluating the subgroup hypotheses, included in the divide-and-conquer frequent pattern growth (that is, by the reordering/sorting optimization). SD-Map uses a modified FP-growth step that can compute the subgroup quality directly without referring to other intermediate results. It can use several quality functions like Piatetsky-Shaphiro (Kloesgen, 1996), unusualness, or the binomial test (Kloesgen, 1996), among others. The adaptations of the algorithms based on APRIORI for subgroup discovery are also valid for the FP-growth method. This algorithm is implemented in the software tool VIKAMINE (Atzmueller and Puppe, 2005). An extension named SD-Map⋆ (Atzmueller and Lemmerich, 2009) has been developed which is applicable for binary, categorical, and continuous target variables.

- DpSubgroup (Grosskreutz et al., 2008) is a subgroup discovery algorithm that uses a frequent pattern tree to obtain the subgroups efficiently. It incorporates tight optimistic estimate pruning and focuses on binary and categorical target variables. DpSubgroup uses an explicit depth-first search step for evaluating the subgroup hypotheses. It makes use of the FpTree-based data representations introduced by SD-Map algorithm and focuses on binary and categorical target concepts.

- Merge-SD (Grosskreutz and Rueping, 2009) is a subgroup discovery algorithm that prunes large parts of the search space by exploiting bounds between related numerical subgroup descriptions. In this way the algorithm can manage data sets with numeric attributes. The algorithm uses a new pruning scheme which exploits the constraints among the quality of subgroups ranging over overlapping intervals. Merge-SD performs a depth-first search in the space of subgroup descriptions.

- IMR (Boley and Grosskreutz, 2009) is an alternative algorithmic approach for the discovery of non-redundant subgroups based on a breadth-first strategy. It searches for equivalence classes of descriptions with respect to their extension in the database rather than individual descriptions. So the algorithm searches in the space of subgroup extensions and has a potentially reduced search space returning at most one description of each extension to the user. It can use several quality measures, but in the experimentations the one used is the binomial test. To manage continuous attributes, they must be previously discretised (the authors use the minimal entropy discretisation).

4.3. Evolutionary algorithms for extracting subgroups

Subgroup discovery is a task which can be approached and solved as optimisation and search problems. Evolutionary algorithms (Bäck et al., 1997) imitate the principles of natural evolution in order to form searching processes. One of the most widely used types of evolutionary algorithms are genetic algorithms, inspired by natural evolution processes and initially defined by Holland (Goldberg, 1989, Holland, 1975). The heuristic used by this type of algorithm is defined by a fitness function, which determines which individuals (rules in this case) will be selected to form part of the new population in the competition process. This makes genetic algorithms very useful for the subgroup discovery task. The evolutionary algorithms proposed for extracting subgroups are explained below:

- SDIGA (del Jesus et al., 2007c) is an evolutionary fuzzy rule induction system. It uses as quality measures for the subgroup discovery task adaptations of the measurements used in the association rules induction algorithms, confidence and support, and can also use other measures such as interest, significance, sensitivity or unusualness. The algorithm evaluates the quality of the rules by means of a weighted average of the measures selected. An analysis of different combinations of quality measures can be observed in (Carmona et al., 2009a). SDIGA uses linguistic rules (del Jesus et al., 2007a) as description language to specify the subgroups. This algorithm is implemented in the software tool KEEL(Alcalá-Fdez et al., 2009, Alcalá-Fdez et al., 2011).

- MESDIF (Berlanga et al., 2006, del Jesus et al., 2007b) is a multi-objective genetic algorithm for the extraction of fuzzy rules which describe subgroups. The algorithm extracts a variable number of different rules expressing information on a single value of the target variable. The search is based on the multi-objective SPEA2 (Zitzler et al., 2002) approach, and so applies the concepts of elitism in the rule selection (using a secondary or elite population) and the search for optimal solutions in the Pareto front. It can use several quality measures at a time to evaluate the rules obtained such as confidence, suport, significance or unusualness. This algorithm is implemented in the software tool KEEL(Alcalá-Fdez et al., 2009, Alcalá-Fdez et al., 2011).

- NMEEF-SD (Carmona et al., 2009b, 2010) is an evolutionary fuzzy system whose objective is to extract descriptive fuzzy and/or crisp rules for the subgroup discovery task, depending on the type of variables present in the problem. NMEEF-SD has a multi-objective approach whose search strategy is based on NSGA-II (Deb et al., 2002), which is based on a non-dominated sorting approach, and on the use of elitism. This algorithm uses specific operators to promote the extraction of simple, interpretable and high quality rules. It allows a number of quality measures to be used both for the selection and the evaluation of rules within the evolutionary process, including confidence, support, sensitivity, significance and unusualness. This algorithm is implemented in the software tool KEEL(Alcalá-Fdez et al., 2009, Alcalá-Fdez et al., 2011).

The evolutionary algorithms proposed so far for the subgroup discovery task are based on the hybridisation between fuzzy logic and evolutionary algorithms, known as evolutionary fuzzy systems (Herrera, 2008). They provide novel and useful tools for pattern analysis and for extracting new kinds of useful information.

As claimed in (Dubois et al., 2005), "the use of fuzzy sets to describe associations between data 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 the attribute domains". It is especially useful in domains where the boundaries of a piece of information used may not be clearly defined. The evolutionary algorithms allow the inclusion of quality measures in the process in order to obtain rules with suitable values not only in the selected quality measures but also in the others. The best way to obtain solutions with a good compromise between the quality measures for subgroup discovery is through a multi-objective evolutionary algorithm approach.

5. Real-world applications analysed through evolutionary algorithms for extracting interesting subgroups

A wide range of contributions in the specialised literature related to different fields can be found, where descriptive knowledge associated with a specific target value has a special interest.

It is important to note the importance of the evolutionary fuzzy system within the subgroup discovery task. The following sections present the most important real-world applications solved through this type of algorithms within subgroup discovery.

5.1. Marketing

In the area of marketing, and specifically in the planning of trade fairs, it is important to extract conclusions from the information on previous trade fairs to determine the relationship between the trade fair planning variables and the success of the stand. For this problem a subgroup discovery rule induction algorithm is well suited.

From a review of the literature and by asking the exhibitors, a questionnaire was designed to reflect the variables that allow a better explanation of trade fair success, later contrasted by experts. This questionnaire contains 104 variables, seven of which are numerical and the rest are categorical features (obtained by the experts by means of discretization). The questionnaire contains questions relating to the prior planning of the fair (which must be answered before the celebration of the fair) to the valuations on the participation in the fair as well as the actions to develop by the company after the fair (which will be answered once the fair has finished), and other questions to be answered during the fair.

In this way, once the data for each exhibitor have been gathered, the stands global efficiency is rated as high, medium or low, in terms of the level of achievement of objectives set for the trade fair, based on various marketing criteria.

The data contained in this data set were collected in the Machinery and Tools biennial fair held in Bilbao, Spain, in March 2002 and contain information on 228 exhibitors. With the data collected for each exhibitor, the stands were characterized according to their level of achievement of objectives, obtaining the class distribution (low, medium, or high efficiency).

The use of SDIGA for this problem is well suited because in a subgroup discovery task, the objective is not to generate a set of rules that cover all the data set examples but individual rules that, given a property of interest of the data, describe the most interesting subgroups for the user. This is the type of knowledge we want to obtain.

Table 3 represents the most relevant rules obtained by SDIGA in this dataset.

Table 3: Rules obtained by SDIGA in the trade fairs problem

Rule	*Description*
R_1	IF Zone=(N OR S) AND GratitudeLetter=NO AND StandDifHeights=NO THEN Low
R_2	IF Zone=N AND ImprovImagen=(Medium OR High) AND Stewardess=Yes THEN Medium
R_3	IF Zone=C AND GratitudeReports=NO THEN High

The knowledge discovered for each one of the target variable values is understandable by the user due to the use of DNF fuzzy logic and the low number of rules and conditions in the rule antecedents (below 10% of the 104 variables). Moreover, the rules obtained with the SDIGA algorithm are very simple, due to the application of a hill-climbing algorithm, which optimizes each extracted rule and increases their simplicity.

5.2. e-Learning

This section presents a case study based on Moodle platform. Specifically, we have used the students usage data of the Moodle system, which is one of the most widely used e-learning systems. Moodle has a large and diverse users community with over 75 languages in over 160 countries

The objective in using subgroup discovery in an e-learning system is to analyse what relation the usage of complementary activities of a course can have to final mark obtained. We have used the final mark as the variable to characterise.

The Moodle system contains a great deal of detailed information on course content, users, usage, etc., stored in a relational data base. We have applied a pre-processing step to the information, obtaining a new summary table (see Table 4) with the most important information related to our objective.

Table 4: Description of the attributes employed for each student

Name	**Description**	**Type**
course	Identification of the course	Discrete
nAssigment	Number of assignments completed	Continuous
nAssigmentP	Number of assignments passed	Continuous
nAssigmentF	Number of assignments failed	Continuous
nQuizz	Number of quizzes completed	Continuous
nQuizzP	Number of quizzes passed	Continuous
nQuizzF	Number of quizzes failed	Continuous
nMessagesC	Number of messages sent to the chat	Continuous
nMessagesT	Number of messages sent to the teacher	Continuous
nMessagesF	Number of messages sent to the forum	Continuous
nRead	Number of forum messages read	Continuous

Table 4 contains a summary of the activities completed and the mark obtained by each student in an e-learning course. We have discretised the marks into classes (fail, pass, good and excellent) in order to codify them as the values of the rule consequent. The information is corresponding to 192 different courses of the University of Cordoba. Among all these courses it is available in the dataset only the 5 courses (with a total of 293 students) with the highest usage of the activities and resources available in Moodle.

Some of the most relevant rules obtained by SDIGA are summarised below:

- *IF course = C110 AND nAssignment = High AND nPosts = High THEN mark = Good (Accuracy: 0.9285, Significance: 6.5348, Coverage:0.1575).*

  This rule shows that in the ProjectManagement (C110) course, the students who have completed a high number of assignments and sent a lot of messages to the forum, have also obtained good marks. The teacher must continue to promote these types of activity in this course because of their effectiveness for the students in the final mark obtained.

- *IF course = C29 AND nMessagesT = Very low THEN mark = Fail (Accuracy: 0.8560, Significance: 59.1774, Coverage: 0.2520).*

  In the AppliedComputerScienceBasis (C29) course, most of the students who have sent a very low number of messages to the teacher have failed. Using this information, the teacher can direct more attention to these students because they have a higher probability of failing.

- *IF course = C110 OR C88 AND nPosts = High OR Very High AND nQuizP = Medium OR High OR Very High THEN mark = Good (Accuracy: 0.7382, Significance: 43.4771, Coverage: 0.2431).*

  This rule shows that if the students of the course Project- Management (C110) or ComputerScienceBasis (C88) have sent a high or very high number of messages to the forum, and they have also obtained a medium, high or very high score in the quizzes, then they obtain good marks.

- *IF course = C29 OR C110 OR C111 AND nAssignmentF = Very High OR High OR Medium AND nQuizF = Very High OR High OR Medium AND nMessagesT = Very low OR Low THEN mark = Fail (Accuracy: 0.8667, Significance: 61.8034, Coverage: 0.4726).*

  This rule shows that if the students of the course ProgrammingForEngineers (C29) or ProjectManagement (C110) or ComputerScienceBasis (C88) have failed in a very high, high or medium number of assignments, have failed in a very high, high or medium number of quizzes, and have sent a very low or low number of messages to the teacher, then they have obtained a fail in their final marks.

5.3. Medical domain

This study analyses 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.

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, in order 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.

5.3.1. 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 5 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 3 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 3 continuous variables:

- *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.

5.3.2. Results and analysis

Table 6 shows the average of the results obtained by MESDIF, SDIGA and CN2-SD for the test partitions, where: *Label* 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, SIG is the significance, $WRAcc$ is the unusualness, SUP is the sensitivity and CNF is the confidence.

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.

In order to analyse the results shown in Table 6, it is important to consider that an algorithm shows good behaviour for subgroup discovery 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. So, 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:

Table 5: Description of the variables

N	*Name*	*Description*	*Values*
0	Derivation	Derivation	0,1,. . .,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,. . .,8
5	Coexist	Number of people living with the patient	0,1,. . .,5
6	Reason for consultation	Reason why the patient comes for consultation	0,1,. . .,20
7	Medical history	Medical history	0,1,. . .,20
8	Psychiatric history	Psychiatric history	0,1,. . .,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,. . .,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,. . .,6
37	Previous consultation	Previous consultation	0,1,2,3,4
38	Consultation time	Consultation time	0,1,. . .,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
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,. . .,8
66	Intervention	Intervention	1,2,. . .,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,. . .,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

- 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

Table 6: Results obtained by the algorithms from the data set

Algorithm	*Labels*	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

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). In average, the rule sets obtained by MESDIF cover more than 50% of the examples.
- 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 account 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. Some rules obtained by MESDIF can be observed below in Table 7.

Table 7: Rules obtained for MESDIF in the complete experimentation

♯	*Rule*	SIG	$WRAcc$	SUP_c	CNF
R_1	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_2	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_3	IF (Developmental disorder=0 AND Autolytic behaviour=0) THEN Admission Time=Evening	1.471	0.009	0.884	0.427
R_4	IF (Litium=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

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. Experts have highlighted that the results obtained by MESDIF not only match the previous reported, 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.

Acknowledgment

This work was supported by the Spanish Ministry of Economy and Competitiveness under projects TIN2012-33856, and by the Andalusian Research Plan under project TIC-3928.

This paper was written in the Centre for Computational Intelligence, De Montfort University Leicester, during the research visit of C.J. Carmona. Thanks to Prof. R. John and Prof. F. Chiclana for the invitation.

References

Abudawood, T., Flach, P., 2009. Evaluation Measures for Multi-class Subgroup Discovery, in: Proc. of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Springer. pp. 35–50.

Agrawal, R., Imieliski, T., Swami, A., 1993. Mining association rules between sets of items in large databases, in: Proc. of the 1993 ACM SIGMOD International Conference on Management of Data, ACM Press. pp. 207–216.

Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., Verkamo, A., 1996. Fast discovery of association rules, in: Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. (Eds.), Advances in Knowledge Discovery and data mining. AAAI Press, pp. 307–328.

Alcalá-Fdez, J., Fernández, A., Luengo, J., Derrac, J., García, S., Sánchez, L., Herrera, F., 2011. KEEL Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework. Journal of Multiple-Valued Logic and Soft Computing 17, 255–287.

Alcalá-Fdez, J., Sánchez, L., García, S., del Jesus, M., Ventura, S., Garrell, J., Otero, J., Romero, C., Bacardit, J., Rivas, V., Fernández, J., Herrera, F., 2009. KEEL: A Software Tool to Assess Evolutionary Algorithms for Data Mining Problems. Soft Computing 13, 307–318.

Atzmueller, M., Lemmerich, F., 2009. Fast Subgroup Discovery for Continuous Target Concepts, in: Proc. of the 18th International Symposium on Methodologies for Intelligent Systems, Springer. pp. 35–44.

Atzmueller, M., Puppe, F., 2005. Semi-Automatic Visual Subgroup Mining using VIKAMINE. Journal of Universal Computer Science 11, 1752–1765.

Atzmueller, M., Puppe, F., 2006. SD-Map - A Fast Algorithm for Exhaustive Subgroup Discovery, in: Proc. of the 17th European Conference on Machine Learning and 10th European Conference on Principles and Practice of Knowledge Discovery in Databases, Springer. pp. 6–17.

Atzmueller, M., Puppe, F., Buscher, H.P., 2004. Towards Knowledge-Intensive Subgroup Discovery, in: Proc. of the Lernen - Wissensentdeckung - Adaptivität - Fachgruppe Maschinelles Lernen, pp. 111–117.

Bäck, T., Fogel, D., Michalewicz, Z., 1997. Handbook of evolutionary computation. Oxford University Press.

Bay, S.D., Pazzani, M.J., 2001. Detecting group differences: Mining contrast sets. Data Mining and Knowledge Discovery 5, 213–246.

Berlanga, F.J., del Jesus, M.J., González, P., Herrera, F., Mesonero, M., 2006. Multiobjective Evolutionary Induction of Subgroup Discovery Fuzzy Rules: A Case Study in Marketing, in: Proc. of the 6th Industrial Conference on Data Mining, Springer. pp. 337–349.

Boley, M., Grosskreutz, H., 2009. Non-redundant Subgroup Discovery Using a Closure System, in: Proc. of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Springer. pp. 179–194.

Box, G., Jenkins, G., Reinsel, G., 2008. Time series analysis: forecasting and control. Wiley. 4th edition.

Brin, S., Motwani, R., Ullman, J.D., Tsur, S., 1997. Dynamic Itemset Counting and Implication Rules for Market Basket Data, in: Proc. of the 1997 ACM SIGMOD International Conference on Management of Data, ACM Press. pp. 255–264.

Carmona, C.J., González, P., del Jesus, M.J., Herrera, F., 2009a. An Analysis of Evolutionary Algorithms with Different Types of Fuzzy Rules in Subgroup Discovery, in: Proc. of the IEEE International Conference on Fuzzy Systems, pp. 1706–1711.

Carmona, C.J., González, P., del Jesus, M.J., Herrera, F., 2009b. Non-dominated Multi-objective Evolutionary algorithm based on Fuzzy rules extraction for Subgroup Discovery, in: Proc. of the 4th International Conference on Hybrid Artificial Intelligence Systems, Springer. pp. 573–580.

Carmona, C.J., González, P., del Jesus, M.J., Herrera, F., 2010. NMEEF-SD: Non-dominated Multi-objective Evolutionary algorithm for Extracting Fuzzy rules in Subgroup Discovery. IEEE Transactions on Fuzzy Systems 18, 958–970.

Cherkassky, V., Mulier, F., 2007. Learning from Data. Concepts, Theory and Methods. IEEE Press. 2nd edition.

Clark, P., Boswell, R., 1991. Rule Induction with CN2: Some Recent Improvements, in: Proc. of the 5th European Conference on Machine Learning, Springer. pp. 151–163.

Clark, P., Niblett, T., 1989. The CN2 Induction Algorithm. Machine Learning 3, 261–283.

Deb, K., Pratap, A., Agrawal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions Evolutionary Computation 6, 182–197.

Demsar, J., Zupan, B., Leban, G., 2004. White Paper (www.ailab.si/orange) .

Dong, G.Z., Li, J.Y., 1999. Efficient Mining of Emerging Patterns: Discovering Trends and Differences, in: Proc. of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press. pp. 43–52.

Dubois, D., Prade, H., Sudkamp, T., 2005. On the representation, measurement, and discovery of fuzzy associations. IEEE Transactions on Fuzzy Systems 13, 250–262.

Fayyad, U.M., 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.

Gamberger, D., Lavrac, N., 2002. Expert-Guided Subgroup Discovery: Methodology and Application. Journal Artificial Intelligence Research 17, 501–527.

Gamberger, D., Lavrac, N., 2003. Active subgroup mining: a case study in coronary heart disease risk group detection. Artificial Intelligence in Medicine 28, 27–57.

Gamberger, D., Lavrac, N., Wettschereck, D., 2002. Subgroup visualization: A method and application to population screening, in: Proc. of the 2nd International Workshop on Integration and Collaboration Aspects of Data Mining, Decision Support and Meta-Learning, pp. 35–40.

Goldberg, D.E., 1989. Genetic Algorithms in search, optimization and machine learning. Addison-Wesley Longman Publishing Co., Inc.

Grosskreutz, H., Rueping, S., 2009. On Subgroup Discovery in Numerical Domains. Data Mining and Knowledge Discovery 19, 210–216.

Grosskreutz, H., Rueping, S., Wrobel, S., 2008. Tight optimistic estimates for fast subgroup discovery, in: European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, pp. 440–456.

Han, J., Pei, J., Yin, Y., 2000. Mining Frequent Patterns Without Candidate Generation, in: Proc. 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. Evolutionary Intelligence 1, 27–46.

Herrera, F., Carmona, C.J., González, P., del Jesus, M.J., 2011. An overview on Subgroup Discovery: Foundations and Applications. Knowledge and Information Systems 29, 495–525.

Holland, J.H., 1975. Adaptation in natural and artificial systems. University of Michigan Press .

del Jesus, M.J., González, P., Herrera, F., 2007a. Fuzzy Sets and Their Extensions: Representation, Aggregation and Models. Springer. volume 220. chapter Subgroup Discovery with Linguistic Rules. pp. 411–430.

del Jesus, M.J., González, P., Herrera, F., 2007b. Multiobjective Genetic Algorithm for Extracting Subgroup Discovery Fuzzy Rules, in: Proc. of the IEEE Symposium on Computational Intelligence in Multicriteria Decision Making, IEEE Press. pp. 50–57.

del Jesus, M.J., González, P., Herrera, F., Mesonero, M., 2007c. Evolutionary Fuzzy Rule Induction Process for Subgroup Discovery: A case study in marketing. IEEE Transactions on Fuzzy Systems 15, 578–592.

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, Springer. pp. 44–51.

Kavsek, B., Lavrac, N., 2006. APRIORI-SD: Adapting association rule learning to subgroup discovery. Applied Artificial Intelligence 20, 543–583.

Kavsek, B., Lavrac, N., Jovanoski, V., 2003. APRIORI-SD: Adapting Association Rule Learning to Subgroup Discovery, in: Proc. of the 5th International Symposium on Intelligent Data Analysis, Springer. pp. 230–241.

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.

Kloesgen, W., May, M., 2002. Census Data Mining - An application, in: Proc. of the 6th European Conference on principles of data mining and knowledge discovery, pp. 65–79.

Kloesgen, W., Zytkow, J., 2002. Handbook of Data Mining and Knowledge Discovery. Oxford.

Kralj-Novak, P., Lavrac, N., Webb, G.I., 2009. Supervised Descriptive Rule Discovery: A Unifying Survey of Constrast Set, Emerging Pateern and Subgroup Mining. Journal of Machine Learning Research 10, 377–403.

Lavrac, N., 2005. Subgroup discovery techniques and applications, in: Proc. of the 9th Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer. pp. 2–14.

Lavrac, N., Cestnik, B., Gamberger, D., Flach, P.A., 2004a. Decision Support Through Subgroup Discovery: Three Case Studies and the Lessons Learned. Machine Learning 57, 115–143.

Lavrac, N., Flach, P.A., Zupan, B., 1999. Rule Evaluation Measures: A Unifying View, in: Proc. of the 9th International Workshop on Inductive Logic Programming, Springer. pp. 174–185.

Lavrac, N., Kavsek, B., Flach, P.A., Todorovski, L., 2004b. Subgroup Discovery with CN2-SD. Journal of Machine Learning Research 5, 153–188.

Lavrac, N., Zelezny, F., Flach, P.A., 2003. RSD: Relational Subgroup Discovery through First-Order Feature Construction, in: Proc. of the 12th International Conference Inductive Logic Programming, Springer. pp. 149–165.

Moreland, K., Truemper, K., 2009. Discretization of Target Attributes for Subgroup Discovery, in: Proc. of the 6th International Conference Machine Learning and Data Mining in Pattern Recognition, Springer. pp. 44–52.

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: Proc. of the 8th International Symposium on Intelligent Data Analysis, Springer. pp. 119–130.

Noda, E., Freitas, A.A., Lopes, H.S., 1999. Discovering interesting prediction rules wih a genetic algorithm. IEEE Congress on Evolutionary Computation 2, 1322–1329.

Siebes, A., 1995. Data Surveying: Foundations of an Inductive Query Language, in: Proc. of the 1st International Conference on Knowledge Discovery and Data Mining, AAAI Press. pp. 269–274.

Wrobel, S., 1997. An Algorithm for Multi-relational Discovery of Subgroups, in: Proc. of the 1st European Symposium on Principles of Data Mining and Knowledge Discovery, Springer. pp. 78–87.

Wrobel, S., 2001. Inductive logic programming for knowledge discovery in databases. Springer. chapter Relational Data Mining. pp. 74–101.

Wu, X., Kumar, V., Ross-Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.H., Steinbach, M., Hand, D.J., Steinberg, D., 2009. Top 10 algorithms in data mining. Knowledge and Information Systems 14, 1–37.

Zelezny, F., Lavrac, N., 2006. Propositionalization-based relational subgroup discovery with RSD. Machine Learning 62, 33–63.

Zembowicz, R., Zytkow, J.M., 1996. From Contingency Tables to Various Forms of Knowledge in Databases, in: Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press, pp. 329–349.

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.