MOEA-EFEP: Multi-Objective Evolutionary Algorithm for Extracting Fuzzy Emerging Patterns

Ángel Miguel García-Vico, Cristóbal José Carmona, Pedro González, and María José del Jesus

***Abstract*—Emerging pattern mining is a data mining task that belongs to the supervized descriptive rule discovery framework. Its objective is to find rules that describe emerging behavior or differentiating characteristics with respect to a property of interest. A multiobjective evolutionary algorithm for the extraction of fuzzy emerging patterns (MOEA-EFEP) is described and analyzed in this paper. MOEA-EFEP is the first multi-objective evolutionary algorithm proposed for emerging pattern mining. This approach allows us to get rules whose descriptions of the emerging phenomena are simpler than previous approaches. It is based on the well-known NSGA-II algorithm adapted for the extraction of emerging patterns. The proposal also uses fuzzy logic to deal with numeric variables in order to obtain a knowledge representation close to human reasoning. An experimental study was performed to verify the validity of the proposal. First, it presents a comparison of different rule representations and postprocessing filter strategies, in order to determine an optimal configuration of the proposal. Finally, it is compared with other algorithms for emerging pattern mining in order to determine the quality of the knowledge extracted. The results show that MOEA-EFEP obtains rules with a better description of the emerging or discriminative behavior than other algorithms of the task. The conclusions of this study are supported by the use of statistical tests.**

***Index Terms*—Emerging pattern mining (EPM), evolutionary fuzzy systems (EFSs), fuzzy rules, multi-objective evolutionary algorithms (MOEAs).**

I. Introduction

TRADITIONALLY, data mining has been approached from two perspectives: supervized and nonsupervized learning. In general, supervized methods have a predictive approach and nonsupervized methods have a descriptive behavior. Nowadays, there are several techniques that are halfway between both approaches. They attempt to describe data with respect to a property of interest. This concept is known as supervized descriptive rule discovery (SDRD) [1] and it includes techniques such as subgroup discovery [2], contrast set mining [3], and emerging pattern mining (EPM) [4], amongst others. This paper is focused on EPM, whose main aim is to describe discriminative relationships between different values of the property of interest, or emerging behavior on data. Both EPM and other techniques within SDRD have been applied in a successful way in several fields such as psychology [5], chemistry [6], bioinformatics [7]–[9], or management [10], amongst others [11]–[13].

Throughout the literature, high-precision rule models have been obtained with EPM algorithms. Nevertheless, EPM is considered as an NP-hard problem [14] and the number of emerging patterns (EPs) extracted by these methods is huge. As a consequence, the underlying phenomena of the problem is not easily understandable by the experts. The objective of this paper is to propose an algorithm for the extraction of rules with better descriptive capacity. In this way, evolutionary fuzzy systems (EFSs) [15] have been widely used in other SDRD tasks, such as subgroup discovery, in order to get rules with higher descriptive capacity than classical models [9], [16], [17]. EFSs are a soft-computing technique that combines evolutionary algorithms [18], which allows us to find a good solution in big search spaces in a reasonable time, that considers fuzzy rules [19], using a knowledge representation closer to human reasoning than other representations such as discretization.

In this paper, a new proposal based on the NSGA-II [20] algorithm for the induction of fuzzy emerging patterns (FEPs) is presented: A multi-objective evolutionary algorithm for the extraction of fuzzy emerging patterns (MOEA-EFEP). This method employs a cooperative-competitive schema where individuals cooperate by means of the maximization of the average unusualness of the population, and compete in order to obtain the optimal solution to the problem. The algorithm permits the selection of two possible rule representations, with the objective of finding the best description of the data: conjunctions of features, i.e., attribute-value pairs; or features in disjunctive normal form (DNF), which allows the presence of more than one value within a variable. These characteristics allow us to find highly descriptive rules. After that, the application of a postprocessing filter is carried out in MOEA-EFEP in order to obtain reliable and descriptive rules.

To achieve this, the paper is organized as follows: A brief description of EPM is presented in Section II. After that, the MOEA-EFEP algorithm and its main characteristics are presented in Section III. Section IV presents the experimental study carried out to show the validity of MOEA-EFEP. Finally, Section V concludes this paper.

Manuscript received November 10, 2017; revised February 15, 2018; accepted February 27, 2018. Date of publication March 9, 2018; date of current version October 4, 2018. *(Corresponding author: Ángel Miguel Garcéa-Vico.)*

The authors are with the Department of Computer Science, University of Jaén, Jaén 23071, Spain (e-mail: agvico@ujaen.es; ccarmona@ujaen.es; pglez@ujaen.es; mjjesus@ujaen.es).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

This paper has supplementary downloadable multimedia material available at http://ieeexplore.ieee.org provided by the authors. This includes results tables for the whole experimentation carried out throughout this paper and source code of the MOEA-EFEP algorithm. This material is 114 MB in size.

Digital Object Identifier 10.1109/TFUZZ.2018.2814577

II. Emerging Pattern Mining

The EPM task was defined by Dong and Li [4], [21] as the search for patterns whose supports increase significantly from one dataset (D_1) to another (D_2). In this way, a pattern is considered as emerging if and only if its growth rate (GR) is greater than a threshold ρ. The GR is defined as follows:

$$\mathrm{GR}(x) = \begin{cases} 0, & \text{IF } \mathrm{Sup}_{D_1}(x) = \mathrm{Sup}_{D_2}(x) = 0 \\ \infty, & \text{IF } \mathrm{Sup}_{D_1}(x) \neq 0 \wedge \mathrm{Sup}_{D_2}(x) = 0 \\ \frac{\mathrm{Sup}_{D_1}(x)}{\mathrm{Sup}_{D_2}(x)}, & \text{another case} \end{cases} \tag{1}$$

where $\mathrm{Sup}_{D_i}(x)$ is the support of the pattern x on dataset i (D_i).

The main objectives of EPM are the discovery of differentiating characteristics between classes, emerging trends in temporal datasets or the detection of differences among variables. Nevertheless, the former objective is the most used throughout the literature in order to build powerful classifiers. This study is focused on this first objective, describing these characteristics in order to make them easily understandable by the experts.

Generally, these patterns can be represented as rules in the form [22]

$$R : \text{Cond} \rightarrow \text{Class} \tag{2}$$

where Cond are features that form the antecedent part of the rule and Class is the property of interest or class, being the consequent part of the rule. In the following, the terms pattern or rule are used interchangeably.

There are several algorithms for EPM developed so far, which can be classified according to the approach used to mine EPs. A complete review has been presented in [23].

Along the literature, the majority of the approaches presented have been used to build powerful classifiers. Although EPM can be used to classify new instances of data, the objectives of EPM go further than classification. EPM tries to describe other kinds of phenomena that are out of the scope of classification tasks, such as the description of emerging trends. Therefore, Section II-A presents the main differences between EPM and classification tasks. Next, Section II-B presents the main quality measures used to quantify the quality of a descriptive rule. Finally, Section II-C presents the advantages of using EFSs to obtain rules with high descriptive capacity.

A. EPM Versus Classification

Within data mining, there are two perspectives.

1) Predictive induction, whose main objective is to find knowledge in order to predict or classify new incoming instances. Classification [24], regression [24], and time series forecasting [25] are tasks that use this approach.
2) Descriptive induction, whose main objective is the extraction of knowledge from data in order to find interesting or surprising relationships in data. Association rules [26], summarisation [27], or EPM [4], amongst others, use this approach.

In Fig. 1, the main differences between both kinds of inductions are presented. In Fig. 1(a), a complex and precise classifier, based on supervized induction, is shown. In this case, the whole search space is split into two regions, depending on the type of objects in the set with respect to the target variable. In Fig. 1(b), an unsupervized model is shown, in which four groups of elements with common characteristics are described. In these models, a target variable is not available. These groups are based on other criteria such as support and confidence. Therefore, it can be seen that the model on the left, which corresponds to supervized induction, has a quite different objective with respect to the model on the right, which corresponds to unsupervized induction.

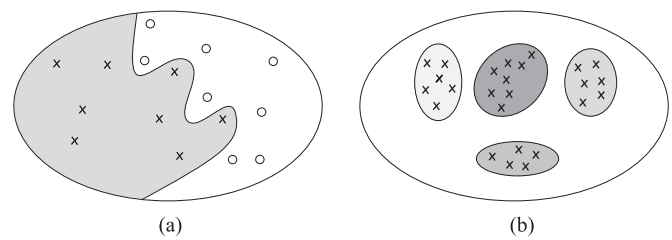

Fig. 1. Examples of supervized (a) and unsupervized (b) data mining models.

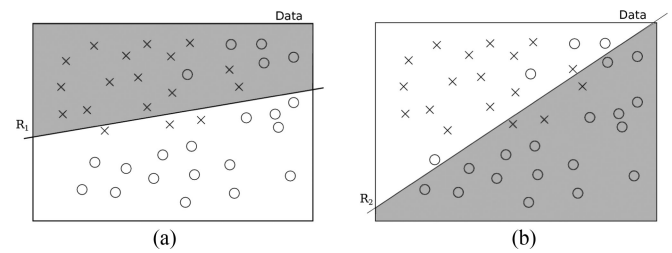

Fig. 2. Representation of a dataset in which two EPs, R_1 (a) and R_2 (b) are defined. R_1 defines emerging behavior for variable "x" and R_2 for variable "o."

However, EPM is a technique for the extraction of patterns with respect to a property of interest or target variable in the data. In this way, it is halfway between both supervized and unsupervized inductions. Its main goal is to generate relations in order to describe emerging behavior between independent variables of two timestamped datasets with respect to a value of the target variable.

In supervized learning, in order to classify new unseen objects, the rules present dependencies with respect to each other. Within the prediction process of an example class, all rules that describe information about the zone of the example take part in the prediction process. This fact means that the underlying phenomena, which defines the behavior of the dataset is not clearly defined. In EPM, the description of the emerging behavior is related to the definition of some underlying phenomena and the rules define independent pieces of knowledge. Therefore, the knowledge extracted in EPM should be as simple as possible in order to facilitate analysis of this kind of phenomena to the expert.

As an example, a dataset is represented in Fig. 2. In this dataset, two emerging rules are defined. Fig. 2(a) represents the rule R_1, which defines a discriminative behavior for examples that belong to the "x" class. On the other hand, Fig. 2(b) represents the rule R_2, which defines a discriminative behavior for examples that belong to the "o" class. As can be observed, the rules correctly define an emerging or discriminative behav-

TABLE I
CONFUSION MATRIX FOR A RULE

	Predicted condition	
True condition	Positive	Negative
Positive	$p = tp$	$\overline{p} = fn$
Negative	$n = fp$	$\overline{n} = tn$

ior. The objective of EPM is the extraction of knowledge that precisely describes the emerging phenomena in an easy and understandable way to the expert. In this way, it is important to highlight that the descriptions obtained should be simple, with high coverage of positive examples and low error rate. In the example, these descriptions are graphically defined as straight lines, which are easy to understand. These rules also have a high coverage of positive examples and low error rate. It is important to remark that it is not necessary to find rules with zero error rate; rules with a good one but simpler are desirable. In addition, these rules can be used independently in order to understand the underlying phenomena in data.

B. Quality Measures for EPM

A quality measure is defined to quantify the interest of a rule, but there is no clear consensus about how to determine it within the SDRD context. In EPM, the measures used are focused on reliability, novelty, generality, and interpretability, in most cases [23].

EPM was conceived for the analysis between two datasets, and by extension, of a single problem with two classes. With a higher number of classes, it is necessary to perform a One-vs-All [28] decomposition of the problem, where the positive class is a value of the target variable, and the negative the remaining values of the variable. Therefore, it is necessary to show the information of the confusion matrix of a rule or the set of rules in order to determine its quality.

Table I presents the confusion matrix of a rule where p is the number of examples correctly covered, n is the number of incorrectly covered examples, $\overline{p}$ is the number of incorrectly noncovered examples, and $\overline{n}$ is the number of correctly noncovered examples.

The most widely used descriptive quality measures for EPM, and for this reason considered in this paper, are [23].

1) Number of rules (n_r). It computes the number of rules induced.
2) Number of variables (n_v). It measures the number of variables in the antecedent part of the rule. For a set of rules, this result is the average of variables for each rule of the set.
3) Confidence (Conf). This measure is defined as the ratio of the predictive capacity of the rule for the positive class with respect to the examples it covers [29].

$$\text{Conf}\,(R) = \frac{p}{p+n}. \tag{3}$$

4) Weighted relative accuracy (WRAcc). Also known as unusualness, this descriptive measure quantifies the tradeoff between generality and accuracy gain of the rule [22]

$$\text{WRAcc}(R) = \frac{p+n}{P+N}\left(\frac{p}{p+n} - \frac{P}{P+N}\right) \tag{4}$$

where $P = p + \overline{p}$ and $N = n + \overline{n}$. The domain of this measure is dependent on the percentage of examples for the positive class. Therefore, it is necessary to perform a normalization of the measure in order to make comparisons.

5) True positive rate (TPR). It measures the percentage of examples correctly covered with respect to the total number of positive examples [30]

$$\text{TPR}(R) = \frac{p}{P}. \tag{5}$$

6) False positive rate (FPR). It determines the percentage of examples incorrectly covered with respect to the total amount of negative examples [31]. This value must be minimized in order to obtain reliable rules

$$\text{FPR}(R) = \frac{n}{N}. \tag{6}$$

C. EFSs for EPM

An EFS [15] is a fuzzy system [32] augmented with a learning process based on an EA [33]. Usually fuzzy systems consider the model structure in the form of rules, which are called fuzzy rule based systems. These kinds of systems have been widely used throughout the literature with applications in finance, control, engineering, and medicine [9], [34]–[37], amongst others. These kinds of systems provide a comprehensible representation of the knowledge extracted and a good approach to handling continuous variables.

The use of fuzzy LLs [32] to represent numeric variables is easier to understand than a discretized representation [19]. In addition, it avoids the lose of information that produces the discretization process. On the other hand, the inclusion of an evolutionary learning process allows an efficient global search, which allows a good solution, or even the optimal one, in a reasonable time. Additionally, it can use the same quality measures used to quantify the quality of a rule to guide the search process.

Within the learning procedure of an EFS, two strategies could be used for encoding the individuals of the population [38]: "Chromosome = set of rules" also known as Pittsburgh [39] and "Chromosome = rule," where an individual represents a single rule such as "cooperative-competitive" approach [40]. Nowadays, the EvAEP algorithm [36] is the only EFS developed focused on the extraction of EPs within the literature. This method is based on a mono-objective EA, which follows the "chromosome = rule" approach together with an iterative rule learning (IRL).

III. MOEA-EFEP: MULTI-OBJECTIVE EVOLUTIONARY ALGORITHM FOR EXTRACTING FUZZY EMERGING PATTERNS

The proposed algorithm is an EFS whose objective is to extract emerging fuzzy and/or crisp rules, depending on the type of variables the problem contains, with a good tradeoff between reliability and descriptive capacity.

The induction of EPs with high descriptive capacity can be considered as a multi-objective problem. The search of this kind of rule implies obtaining general rules with high TPR. However, these rules should be as accurate as possible. These objectives are conflicting. Thus, the use of a multi-objective evolutionary

algorithm (MOEA) is a well-suited approach for EPM because they can optimize several quality measures simultaneously and efficiently search for an optimal solution in huge search spaces.

There are several MOEAs developed so far throughout the literature, where the MOEA/D [41], IBEA [42], SPEA2 [43], and NSGA-II [20] algorithms are highlighted due to its success. In particular, MOEA-EFEP is a MOEA based on the well-known NSGA-II [20] algorithm. We have decided the use of the NSGA-II approach to deal with multi-objective optimization problems in EPM because of its success in similar tasks such as subgroup discovery [16], and its performance with conflicting objectives [44]. NSGA-II is a computationally fast MOEA based on a nondominated sorting approach. MOEA-EFEP is oriented toward EPM and contains specific operators in order to find simple, general, and trustworthy rules. The selection of two different kinds of rules representations in order to obtain good results on different kinds of problems is also possible. These representations will be explained in detail in Section III-A.

The algorithm uses a "chromosome = rule" approach including both the antecedent and consequent of the rule in order to extract knowledge for all the classes in a single execution. It uses a cooperative-competitive approach where individuals cooperate and compete in order to achieve the optimal solution. The diversity of the individuals of the population is improved by the use of a guided reinitialization procedure and the niching technique based on crowding distance. The generalization of the individuals is favored by the use of a general rule-based initialization and an oriented mutation operator. Finally, the use of an elite population and a postprocessing phase based on a filter allows us to obtain rules with good confidence. The postprocessing filter removes those rules extracted, which do not achieve some minimal properties that favors the reliability of the knowledge extracted.

It is important to remark that in the elite population, individuals are forced to compete among themselves by the use of a token competition procedure [45], sorting the population by WRAcc. Such a method keeps only those rules with a good tradeoff between confidence and generality. They cooperate in replacing the elite population if their average WRAcc is higher than the average WRAcc of the elite population. This procedure promotes the obtaining of the optimal solution.

The following sections describe the key elements of the proposal: First, the structure and representation of the rules of MOEA-EFEP are shown in Section III-A. Next, the quality measures considered as objectives in the algorithm are presented in Section III-B. Finally, the operational scheme of the algorithm and its components are outlined in Section III-C.

A. Rule Representation in MOEA-EFEP

Rules obtained by MOEA-EFEP use fuzzy logic for the representation of continuous variables by means of linguistic labels (LLs). The use of LLs avoids the use of a previous discretization phase, which produces a loss of information. In addition, the knowledge obtained is more interpretable due to LLs being closer to human reasoning than other representations [19]. The continuous variables are considered as a set of LLs.

$$\begin{array}{c} \textit{Genotype} \\ \left| \begin{array}{c} X_1 \\ 1 \end{array} \right| \begin{array}{c} X_2 \\ \emptyset \end{array} \left| \begin{array}{c} X_3 \\ 3 \end{array} \right| \begin{array}{c} X_4 \\ \emptyset \end{array} \left| \begin{array}{c} Class \\ 1 \end{array} \right| \\ \Downarrow \\ \textit{Phenotype} \\ IF\ (X_1 = Low) \wedge (X_3 = Sports)\ THEN\ (Class = Positive) \end{array}$$

Fig. 3. Representation of a fuzzy rule with continuous and categorical variables in MOEA-EFEP for the canonical form.

$$\begin{array}{c} \textit{Genotype} \\ \left| \begin{array}{c} X_1 \\ 1\ \emptyset\ 1 \end{array} \right\| \begin{array}{c} X_2 \\ 1\ 1\ 1 \end{array} \left\| \begin{array}{c} X_3 \\ 1\ \emptyset\ \emptyset\ \emptyset \end{array} \right\| \begin{array}{c} X_4 \\ \emptyset\ \emptyset\ \emptyset \end{array} \left\| \begin{array}{c} Class \\ 2 \end{array} \right| \\ \Downarrow \\ \textit{Phenotype} \\ IF(X_1 = (Low \vee High)) \wedge (X_3 = Arts)\ THEN\ (Class = Negative) \end{array}$$

Fig. 4. Representation of a fuzzy rule with continuous and categorical variables in MOEA-EFEP for DNF.

The fuzzy sets that correspond to each LL can be specified by the user or defined by means of a uniform partition with triangular membership functions, if expert knowledge is not available.

MOEA-EFEP is able to obtain rules following two representations, according to the expert's necessities.

1) Canonical form (CAN). The antecedent part of the rule is formed by conjunctions of features. Fig. 3 shows an example of the representation of this kind of rule. The chromosome that represents the antecedent part of the rule is an integer vector whose length is equal to the number of variables of the problem. These integer values represent the LL for numeric variables, such as in X_1 or the ith categorical value for categorical variables, such as in X_3. Note that a value of zero means no participation of the variable in the rule.
2) DNF. The antecedent part of the rule is formed by features in DNF. This representation means that a rule can have several values for a variable, joined by disjunctions. Fig. 4 shows the representation of a chromosome. It uses a bit vector with a length equal to the number of possible features. The number of possible features for categorical variables is indicated by the problem; for numerical variables it is the number of LLs considered. A feature participates in the rule if it contains a one. Note that a variable does not participate in the rule if all its values are zero or one.

One of the most important characteristics of MOEA-EFEP is that the consequent part of rules is also represented in chromosomes of both representations. This is represented with an integer indicating the categorical value of the class. This allows rules for all the classes to be obtained in a single execution. To the best of our knowledge, MOEA-EFEP is the only method for EPM that follows this approach. The other algorithms must be executed once per class in order to get rules for all classes.

The evaluation of a chromosome in MOEA-EFEP is performed by means of the confusion matrix of the rule, where each quality measure can be computed. For the understanding of the construction of the confusion matrix presented in Table I,

it is necessary to define when an example is covered by a rule. Considering:

1) $\{X_m/m = 1, \ldots, n_v\}$ is a set of variables that can be categorical or numeric, and n_v represents the number of variables.
2) $\{\text{Class}_j/j = 1, \ldots, n_c\}$ is the set of values for the target variable, and n_c is the number of values.
3) $\{E^k = \left(e_1^k, e_2^k, \ldots, e_{n_v}^k, \text{Class}_j^k\right)/k = 1, \ldots, n_{ex}\}$ is a set of examples, where Class_j^k is the value of the target variable for the example E^k and n_{ex} is the number of examples of the problem.

An example E^k is considered as correctly covered by a rule R_i if and only if

$$\text{APC}\left(E^k, R_i\right) > 0 \wedge \text{Class}_j^k = \text{Class}_j. \tag{7}$$

The antecedent part compatibility (APC) value is the degree of compatibility of the example with respect to the antecedent part of the rule. Thus, an example is correctly covered by a rule if the example has a degree of membership higher than zero in the fuzzy subspace delimited by the antecedent of the rule, and the class of the example is equal to the consequent of the rule.

The APC value is calculated as follows:

$$\text{APC}\left(E^k, R\right) = T\left(\text{TC}\left(\mu_{\text{LL}_1^1}\left(e_1^k\right), \ldots, \mu_{\text{LL}_1^{l_1}}\left(e_1^k\right)\right), \ldots, \text{TC}\left(\mu_{\text{LL}_{n_v}^1}\left(e_{n_v}^k\right), \ldots, \mu_{\text{LL}_{n_v}^{l_{n_v}}}\left(e_{n_v}^k\right)\right)\right) \tag{8}$$

where

- $\text{LL}_{n_v}^{l_{n_v}}$ LL number l_{n_v} for the variable n_v;
- $\mu_{\text{LL}_{n_v}^{l_{n_v}}}\left(e_{n_v}^k\right)$ degree of membership of the example e^k in the LL number l_{n_v} for the variable n_v. For categorical variables, this value is 1 if $e_i^k = X_i$ or zero otherwise;
- T selected t-norm to represent the fuzzy AND operator, i.e., the fuzzy intersection. In this case, the minimum t-norm is selected;
- TC selected t-conorm to represent the fuzzy OR operator, i.e., the fuzzy union. In this case, the maximum t-conorm is selected.

B. Quality Measures Considered as Objectives

The quality measures related to the supervized descriptive capacity of the rule are those that can quantify concepts such as interest, interpretability, generality, and reliability [23]. In this way, several objectives should be improved in order to get an optimal solution for the problem. However, some objectives are conflicting. For example, the extraction of more general rules degrades the reliability of the knowledge extracted and vice-versa [46].

A multi-objective optimization algorithm attempts to find decision vectors, which correspond to objective vectors, that cannot be improved in a dimension without degrading another. This situation is called the Pareto front [47]. Therefore, the aim of MOEA-EFEP is the extraction of those rules that belong to the Pareto front. MOEA-EFEP is based on the NSGA-II approach, widely used throughout the literature. However, a study presented in [44] demonstrates that the performance of this algorithm decreases when the number of objectives increases. So, it is necessary to select some quality measures that can bind these objectives together without degrading the performance of the NSGA-II approach.

According to this, MOEA-EFEP tries to improve two main objectives that are suitable for a Pareto front: generality and reliability. As mentioned previously, generality and reliability are, in general, conflicting objectives because rules with high generality contain a higher error rate than rules with less generality. So, it is necessary the extraction of a Pareto front in order to find the optimal solution.

The objectives used in the optimization are represented by the following the quality measures.

1) Geometric mean TPR-TNR (g-mean). It represents the reliability objective. This measure quantifies the tradeoff between the accuracy of a rule with respect to positive and negative instances as the product of the TPR and the true negative rate (TNR). The measure allows rules to be obtained, which correctly cover examples in both the positive and negative classes. The maximization of this measure forces FPR to be minimized, so more reliable rules are obtained. As a side effect, the confidence and WRAcc of the rule are also improved. This measure has been widely used on problems with imbalanced data as a substitute for the accuracy measure. It is computed as [48]

$$\text{g-mean}\left(R\right) = \sqrt{\frac{p}{P} \cdot \frac{\overline{n}}{N}}. \tag{9}$$

2) Support difference (SuppDiff). It represents the generality objective. In [22], it is shown that this quality measure has a good behavior within the SDRD framework. It measures the support difference between the positive and negative classes, so the measure returns discriminative rules with high support in the positive class. The measure attempts to maximize GR, improving TPR, so the generality of the rules is high. It is calculated as [3]

$$\text{SuppDiff}\left(R\right) = \frac{p}{P} - \frac{n}{N}. \tag{10}$$

C. Evolutionary Model

The MOEA-EFEP algorithm is a MOEA based on NSGA-II [20] for the extraction of FEPs. Specifically, the following changes are introduced with respect to the original NSGA-II approach.

1) Diversity is improved by means of a general rule-based initialization and oriented mutation operators.
2) A guided reinitialization procedure is employed in order to prevent falling into local maxima.
3) The use of a cooperative-competitive approach on an elite population. The individuals of the Pareto front compete following a token competition procedure, which allows a reduced set of rules to be obtained, with minimum overlapping. Also, they cooperate in order to obtain the population with the best average WRAcc value. This elite population is returned to the expert.

The operation scheme of the algorithm is shown in Fig. 5. As can be observed, the algorithm starts with an initial population

```
while MaxEvaluations not reached do
    Q_t ← GeneticOperators(P_t)
    R_t ← P_t ∪ Q_t
    Generate, for each class j, all non-dominated fronts F^j = (F_1^j, F_2^j, ..., F_i^j)
    if not HasEvolved(F_1^j ∀j = 1, ..., n_c) then
        P_aux ← Elit ∪ F_1^j ∀j = 1, ..., n_c
        P_aux ← Sort P_aux by dominance and extract the Pareto front F_1^j
        P_aux ← TokenCompetition(P_aux)
        if AvgWRAcc(P_aux) > AvgWRAcc(Elit) then
            Elit ← P_aux
        end if
        P_t ← GuidedReinitialisation()
    else
        for j = 1 to n_c do
            i ← 1
            N ← NumIndividuals(P_t, j)
            while N ≥ NumIndividuals(F_i^j) do
                Introduce F_i^j in P_t+1
            end while
            Introduce N individuals of F_i^j in P_t+1 by crowding distance
        end for
    end if
    t ← t + 1
end while
P_aux ← Elit ∪ F_1^j ∀j = 1, ..., n_c
P_aux ← Sort P_aux by dominance and extract the Pareto front F_1^j
P_aux ← TokenCompetition(P_aux)
if AvgWRAcc(P_aux) > AvgWRAcc(Elit) then
    Elit ← P_aux
end if
return ApplyFilter(Elit)
```

Fig. 5. Operation scheme of the MOEA-EFEP algorithm.

(P_t) of a predetermined size including individuals from all the classes in order to promote obtaining rules for all the classes. These are generated by the general rule-based initialization procedure. This procedure tries to generate individuals with high generality, generating 75% of the individuals of each class with only a maximum of 25% of the variables initialized. The rest of these individuals (25%) are generated completely at random.

After that, an offspring population (Q_t) of the same size as P_t is generated by means of the application of genetic operators. The operators used are the classical tournament selection [49] of two individuals, a multipoint crossover operator [50] and an oriented mutation operator. The oriented mutation operator removes a variable of the rule or randomly changes a gene. These procedures are applied with the same probability. It is important to remark that the genetic operators are applied between individuals of the same class. These populations are joined together on (R_t), where the fast nondominated sorting algorithm is applied and forms different fronts in the following way: the first front (F_1) contains the nondominated individuals, which is the Pareto front. The second front (F_2) contains individuals, which are dominated by only one individual; the third front (F_3) is composed of individuals dominated by two individuals, and so on. This procedure is applied in order to obtain the Pareto front for each class (F_1^j).

The next step of the algorithm is to determine the population of the next generation (P_{t+1}). This population is generated following a conditional approach: if any F_1^j does not evolve, i.e., it does not change for at least a 5% of the total evaluations, two procedures are performed.

1) Join all F_1^j with the elite population, sort them by dominance in order to get the Pareto front and perform a token competition procedure. The resulting population will overwrite the elite one if and only if their average unusualness is greater than the elite.
2) P_{t+1} is completely generated by means of the guided reinitialization procedure. This procedure generates the remaining individuals of the population by means of covering an example not covered up to the moment with a maximum of 90% of the variables of the rule initialized. With this procedure, diversity is improved and it avoids falling into local maxima.

If the population evolves, P_{t+1} is generated by introducing in order the first complete fronts of R_t until there are the same number of individuals for each class. When the number of individuals in a front F_i^j exceeds the size of the remaining individuals to be introduced for a class j, F_i^j is sorted by crowding distance in descending order, and the first individuals are introduced on P_{t+1} following the NSGA-II procedure.

At the end of the evolutionary process, the last Pareto fronts F_1^j and the elite population are joined. After that, rules are sorted by dominance and the token competition is applied over the resulting Pareto front. The resulting population is returned as the set of optimal rules if the average WRAcc is greater than the elite. Finally, a postprocessing filter is applied. The objective is to keep those descriptive rules that are also reliable. Therefore, in Section IV-C a study of different filters is proposed where the best is chosen as the postprocessing filter of MOEA-EFEP.

IV. Experimental Study

The objective of this experimental study is to compare the performance of MOEA-EFEP against the most relevant EPM algorithms with respect to their descriptive properties. First, the determination of the best knowledge representation for the algorithm is necessary in order to achieve this objective. Additionally, it is necessary to determine the best postprocessing filter strategy, in order to keep those rules with the best tradeoff between reliability and generality.

For this purpose, the study has been divided into the following sections.

1) Section IV-A presents the experimental framework, where the configuration used on the algorithms and the statistical tests used in the study are introduced.
2) Section IV-B presents a comparative of canonical and DNF rule representations for the MOEA-EFEP algorithm in order to select the best one.
3) Section IV-C presents a study of the use of different postprocessing filter strategies in order to determine the most appropriate one for MOEA-EFEP.
4) Section IV-D presents the comparison of MOEA-EFEP, using the best knowledge representation and filter, against iEPMiner [51], FEPM [52], and EvAEP [36]. These algorithm were demonstrated in [23] as the most relevant with respect to their descriptive properties, and allow us to ease the analysis.

A. Experimental Framework

In this section, the datasets used in the experiments and their main characteristics are presented. Next, the algorithms used in

TABLE II
PROPERTIES OF THE DATASETS USED IN THE EXPERIMENTS

Name	$N_v/C_v/n_e$	n_c	Name	$N_v/C_v/n_e$	n_c
appendicitis	7/0/106	2	monk-2	6/0/432	2
australian	8/6/690	2	mushroom	0/22/5644	2
automobile	15/10/150	6	newthyroid	5/0/215	3
balance	4/0/625	3	nursery	0/8/12690	5
bands	19/0/365	2	page-blocks	10/0/5472	5
breast	0/9/277	2	phoneme	5/0/5404	2
bupa	6/0/365	2	pima	8/0/768	2
car	0/6/1728	4	ring	20/0/7400	2
chess	0/36/3196	2	saheart	8/1/462	2
coil2000	85/0/9822	2	satimage	36/0/6435	7
contraceptive	9/0/1473	3	segment	19/0/2310	7
crx	6/9/653	2	shuttle	9/0/58000	7
dermatology	34/0/358	6	sonar	60/0/208	2
flare	0/11/1066	6	spectfheart	44/0/267	2
german	7/13/1000	2	splice	0/60/3190	3
glass	9/0/214	7	texture	40/0/5500	11
haberman	3/0/306	2	thyroid	21/0/7200	3
hayes-roth	4/0/160	3	tic-tac-toe	0/9/958	2
heart	13/0/270	2	titanic	3/0/2201	2
hepatitis	19/0/80	2	twonorm	20/0/4700	2
housevotes	0/16/232	2	vehicle	18/0/846	4
ionosphere	33/0/351	2	wdbc	30/0/569	2
iris	4/0/150	3	wisconsin	9/0/683	2
magic	10/0/19020	2	yeast	8/0/1484	10
mammographic	5/0/830	2	zoo	0/16/101	7

TABLE III
PARAMETERS USED BY THE ALGORITHMS

Algorithm	Parameters
iEPMiner	Minimum support = 0.02, minimum growth rate = 1, minimum chi-squared = 3.84
FEPM	max of items = 5, growth rate = 10, subset relation = superset, maximum depth = 10, tree count = 100, number of labels = 4, hedges = very, somewhat, extremely, little, slightly, positively, generally.
EvAEP	number of labels = 3, number of evaluation = 10000, population length = 100, crossover probability = 0.6, mutation probability = 0.01
MOEA-EFEP	number of labels = 3, number of evaluation = 10000, population length = 51, crossover probability = 0.6, mutation probability = 0.1

the experimentation together with their parameters configuration are shown. Finally, a brief description of the statistical tests used throughout the experimentation is explained.

1) Datasets. The experimental study was carried out using a set of 50 well-known real datasets from the UCI repository [53]. The properties of these datasets are presented in Table II with the number of numerical variables (N_v), categorical variables (C_v), number of examples (n_e), and number of classes (n_c) are presented.
2) Algorithms. The experiments were carried out with the parameters shown in Table III. The values used were those recommended by the experts.
   For the iEPMiner algorithm, it is necessary to perform a previous discretization of the numerical variables of datasets presented in Table II. The discretization process used is the Fayyad discretize [54], which has been widely used throughout the literature.
3) Quality measures analyzed. The quality measures shown in the results tables are the average of the rule set: number of rules (n_r) and number of variables (n_v) for measuring conciseness; WRAcc for measuring novelty and reliability; GR, CONF, and FPR for measuring reliability and TPR for measuring generality. It is important to remark that, due to the domain $[0, \infty]$ of the GR measure, it is impossible to average the result. So the values shown for GR are the percentage of rules that are EP on test data.
   The results presented are the average of different executions following a five-fold distribution optimally balanced stratified cross-validation procedure [55] over each dataset, i.e., the average of 15 executions for the evolutionary algorithms (three executions for each fold with different seeds, due to the nondeterministic behavior of these methods), and the average of five executions (one execution per fold) for the other algorithms.
   Due to space constraints, only the results of the statistical tests are presented in this paper. The complete set of result tables is available at http://simidat.ujaen.es/papers/EFEP_MOEA, which also contains the partitions used in this study in order to be used by other researchers to extend the comparison.
4) Statistical tests. Following the recommendations of [56] a set of nonparametric tests, which provide a simple, robust, and safe method to perform the comparative analysis were used. The methods employed were the Wilcoxon signed rank [57], Iman-Davenport [58], and Holm [59].
   In all the experiments, the significance level used is $\alpha = 0.05$.

TABLE IV
RESULTS OF WILCOXON TEST FOR THE ANALYSIS OF KNOWLEDGE REPRESENTATION FOR MOEA-EFEP

Measure	Comparative	R+	R-	p-value
WRAcc	CAN vs DNF	298	**830**	**0.0005**
CONF	CAN vs DNF	**623**	553	0.7235
GR	CAN vs DNF	39	**66**	0.4144
TPR	CAN vs DNF	250	**878**	**0.0009**
FPR	CAN vs DNF	**727**	308	**0.0183**

In R+ and R− columns bold means the winner on each row. In p-value column a bold value means a values less than 0.05

TABLE V
NUMBER OF RULES AND VARIABLES OBTAINED BY MOEA-EFEP WITH DIFFERENT KNOWLEDGE REPRESENTATIONS

Representation	Rules	Variables
CAN	4.2346	**4.5967**
DNF	**4.1587**	4.6195

The bold means the best value for each column.

B. Analysis of the Knowledge Representation

The representation of knowledge is one of the key factors in obtaining a descriptive model. For EPM, a canonical representation has been widely used throughout the literature. This representation, together with the use of fuzzy logic, is a natural and interpretable way for humans to understand the underlying phenomena of the problem. However, there is also another interesting rule representation by means of DNF rules, known in the EPM literature as disjunctive emerging patterns (DEPs) [23]. To the best of our knowledge, there are no current studies on the determination of the best rule representation for an EPM algorithm. Therefore, the objective of this study is to determine

TABLE VI
RESULTS OF THE FRIEDMAN AND HOLM TEST

	CONF				FPR					
Filter	Ranking	$+$	$\pm$	Filter	Ranking	$+$	$\pm$	Filter	n_r	n_v
Confidence	1.5700	3	3	Confidence	1.7200	1	3	Confidence	2.1307	4.2059
Minimals	2.7000	0	2	Minimals	2.0700	1	3	Maximals	4.1187	4.6225
Maximals	2.7400	0	2	Maximals	2.2700	1	3	Minimals	4.1493	4.6215
Unfiltered	2.9900	0	2	Unfiltered	3.9400	0	0	Unfiltered	4.1587	4.6195

the best representation for use on the comparison with the rest of the EPM algorithms. The statistical tests were performed independently for each quality measure analyzed.

Table IV shows the results of the Wilcoxon signed rank test and Table V shows the average number of rules and variables obtained for each representation. The number of rules obtained for DNF is lower than canonical with a slight difference in the number of variables. However, although the number of variables on DNF rules is higher they obtain on average the best results on the majority of measures studied, with significant differences on WRAcc and TPR, and without significant differences on CONF and GR. Nevertheless, there is a significant difference on FPR in favor of CAN rules. According to these results, the use of DEPs is very well suited to descriptive induction on EPM. The possibility of choosing more than one value for a variable makes a rule more general, and allows it to cover a higher number of examples. This high coverage, together with an evolutionary process, promotes a good tradeoff between reliability and generality, as can be observed with a great improvement of TPR, which counteracts the effect of higher FPR values.

In the following sections, the knowledge representation used in MOEA-EFEP will be the DNF representation, and it will be noted as MOEA-EFEP^{DNF}.

C. Filter Analysis

The huge amount of rules obtained by EPM algorithms makes it necessary to perform some filtering strategy in order to keep only high-quality rules. In [23], an experimental analysis was carried out in order to determine if the application of a postprocessing filter was a good strategy for achieving this objective. Following this study, in this section three filters are applied with respect to the original set of rules and then compared with respect to each other, in order to select the best one for MOEA-EFEP^{DNF}. These filters are: minimal EPs, maximal EPs, and EPs with confidence higher than 60%.

The application of the Iman-Davenport test presents significant differences between filters on CONF and FPR, and there are no significant differences in the remaining measures. According to these results, it is necessary to apply the Holm test in those measures with significant differences in order to find that filter has significant differences with respect to the others.

Table VI summarises the results of the Friedman and Holm tests with a complete pairwise comparison of filters. For each measure, the Friedman rank, followed by the number of methods for which it is statistically better ($+$) or equal or better ($\pm$) is shown.

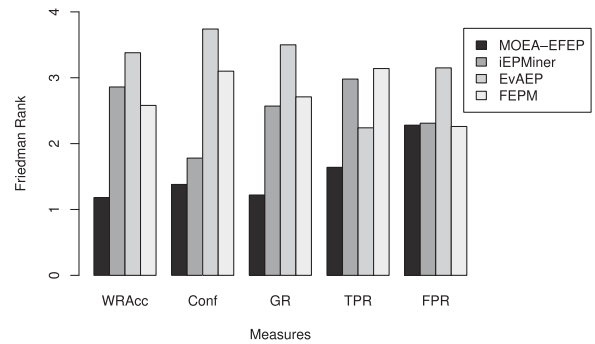

Fig. 6. Friedman rankings of the algorithms analyzed on the studied quality measures.

It can be observed that the filters applied have similar behavior but the confidence filter is slightly better. This filter gets the best results with significant differences on CONF and it is the best without significant differences on FPR with respect to the maximals and minimals filters. In addition, the number of rules and variables obtained by the confidence filter is the lowest. Therefore, the best option is the confidence filter, because it obtains better results than the other filters analyzed with less number of rules and variables. In the following sections, MOEA-EFEP with a confidence postprocessing filter and DNF rule representation is used. It will be noted as MOEA-EFEP$^{\text{DNF}}_{\text{Conf}}$.

D. Comparison of MOEA-EFEP and the Most Relevant Algorithms for EPM

In this section, the main objective is to analyze the descriptive behavior of MOEA-EFEP$^{\text{DNF}}_{\text{Conf}}$ with respect to the most relevant algorithms in the literature according to its descriptive properties [23]: EvAEP, iEPMiner, and FEPM. Following the results of the previous analysis, the knowledge representation used in MOEA-EFEP$^{\text{DNF}}_{\text{Conf}}$ was the DNF representation with a postprocessing filter based on keeping those rules whose confidence is greater than 60% on training data.

Once more, the Iman-Davenport test was carried out. Results of this test show that, for all quality measures analyzed, it exists at least one pair of algorithms with significant differences. Thus, it is necessary to apply the Holm test in order to find the methods with these significant differences. The determination of a control algorithm for each quality measure is necessary in order to ease the analysis. This control algorithm is the method with the best Friedman ranking for each quality measure.

Fig. 6 presents the Friedman rank for each algorithm on each quality measure. The minimum value is considered the best.

TABLE VII
RESULTS OF THE HOLM TEST FOR THE QUALITY MEASURES ANALYZED

Measure	Alg_{Ctrl}	Algorithm	$p-value$
WRAcc	MOEA-EFEP$^{DNF}_{Conf}$	EvAEP	**2.20 · 10^{-16}**
		iEPMiner	**1.53 · 10^{-10}**
		FEPM	**5.88 · 10^{-8}**
CONF	MOEA-EFEP$^{DNF}_{Conf}$	EvAEP	**2.20 · 10^{-16}**
		FEPM	**5.42 · 10^{-11}**
		iEPMiner	0.1213
GR	MOEA-EFEP$^{DNF}_{Conf}$	EvAEP	**2.20 · 10^{-16}**
		FEPM	**1.57 · 10^{-8}**
		iEPMiner	**1.70 · 10^{-7}**
TPR	MOEA-EFEP$^{DNF}_{Conf}$	FEPM	**1.88 · 10^{-8}**
		iEPMiner	**4.21 · 10^{-7}**
		EvAEP	**0.0201**
FPR	FEPM	EvAEP	**0.0017**
		MOEA-EFEP$^{DNF}_{Conf}$	1.0000
		iEPMiner	1.0000

The bold means the best value for each column.

TABLE VIII
NUMBER OF RULES AND VARIABLES FOR EACH ALGORITHM

	n_r	n_v
MOEA-EFEP$^{DNF}_{Conf}$	**2.1307**	4.2059
iEPMiner	1829.4400	**2.0166**
EvAEP	11.0320	6.6758
FEPM	376.1160	2.9842

Therefore, MOEA-EFEP$^{DNF}_{Conf}$ will be the control for WRAcc, CONF, GR, and TPR; and FEPM for FPR.

Once the control algorithms for each measure are determined, the Holm test can be applied. Table VII presents the results of the test on each quality measure and Table VIII presents the average number of rules and variables obtained. The analysis for each quality measure is presented below.

1) WRAcc. The null hypothesis is rejected in all cases in favor of the MOEA-EFEP$^{DNF}_{Conf}$ algorithm. This may be due to the use of an elite population with the cooperative-competitive approach based on WRAcc, which keeps those rules with the highest values.
2) CONF. The null hypothesis is rejected in favor of the MOEA-EFEP$^{DNF}_{Conf}$ algorithm with respect to EvAEP and FEPM. Nevertheless there are no significant differences with respect to iEPMiner. These good results of MOEA-EFEP$^{DNF}_{Conf}$ are due to the use of a filtering strategy based on confidence, which keeps only those rules that are trustworthy.
3) GR. Again, the null hypothesis is rejected for all the algorithms analyzed in favor of MOEA-EFEP$^{DNF}_{Conf}$. In this case, the use of the SupDiff and g-mean as optimization objectives in the evolutionary algorithm obtains rules that are easily generalizable, with a great discriminative power.
4) TPR. For this measure, the null hypothesis is rejected in favor of MOEA-EFEP$^{DNF}_{Conf}$ in all cases. These good results obtained in MOEA-EFEP$^{DNF}_{Conf}$ may be due to the use of the diversity mechanisms proposed, which promotes the use of rules with a low number of variables.

TABLE IX
AVERAGE EXECUTION TIMES (IN SECONDS) FOR THE ALGORITHMS STUDIED IN DIFFERENT DATASETS

MOEA-EFEP$^{DNF}_{Conf}$	EvAEP	iEPMiner	FEPM
20.72 ± 42.17	115.68 ± 379.15	28.57 ± 120.26	339.21 ± 809.93

The bold value means it is the best one.

5) FPR. The null hypothesis is rejected in favor of FEPM with respect to EvAEP. iEPMiner and MOEA-EFEP$^{DNF}_{Conf}$ have no significant differences with respect to FEPM. In this way, the use of SuppDiff and g-mean as objectives of MOEA-EFEP$^{DNF}_{Conf}$ allow general rules to be obtained without degrading reliability.

In EPM context, it is important to find simple rules that describe as precisely as possible the discriminating characteristics or the emerging behavior in data. In this way, MOEA-EFEP$^{DNF}_{Conf}$ obtains the best results on WRAcc, CONF, GR, and TPR. This means that MOEA-EFEP$^{DNF}_{Conf}$ extracts rules, which cover a high number of positive instances with a good tradeoff with respect to their confidence. The high GR means that the underlying phenomena is well described. In addition, although the average number of variables obtained is slightly high, the average number of rules obtained is the best. This allows an easy comprehension of the problem by the expert. For CONF and FPR, which are measures related to reliability, MOEA-EFEP$^{DNF}_{Conf}$ obtains better results on CONF without significant differences with respect to iEPMiner. FEPM also obtains the best results on FPR without significant differences. Rules extracted by iEPMiner and FEPM cover a lower number of examples with respect to MOEA-EFEP$^{DNF}_{Conf}$ and EvAEP, due to the high Friedman rank in TPR obtained. Therefore, these models should extract a big number of rules in order to obtain a good definition of data, which have a negative impact in the interpretability of the model, so they are more oriented to supervized induction. However, with respect to these measures, the results of MOEA-EFEP$^{DNF}_{Conf}$ are not significantly worse than those algorithms. Therefore, MOEA-EFEP can be considered a promising alternative for the extraction of a reduced set of rules with simple and trustworthy descriptions about the discriminating characteristics of classes or the emerging behavior on timestamped data.

E. Execution Time Analysis

Another important aspect for the comparison of the algorithms is their cost in time. Therefore, an empirical study has been carried out in order to measure the cost of the training time of the algorithms.

Table IX presents the average time in seconds for the experiments performed on each algorithm. These experiments were carried out using a computer with an Intel Xeon processor at 2.66 GHz and 64 GB of RAM.

The results show that MOEA-EFEP$^{DNF}_{Conf}$ obtains the best average execution time. This is because the execution time of MOEA-EFEP$^{DNF}_{Conf}$ is related to the number of variables and instances, while for FEPM and iEPMiner, it is related to the number of variables and items, i.e., the different variable-value pairs of the problem, which are normally much bigger than the num-

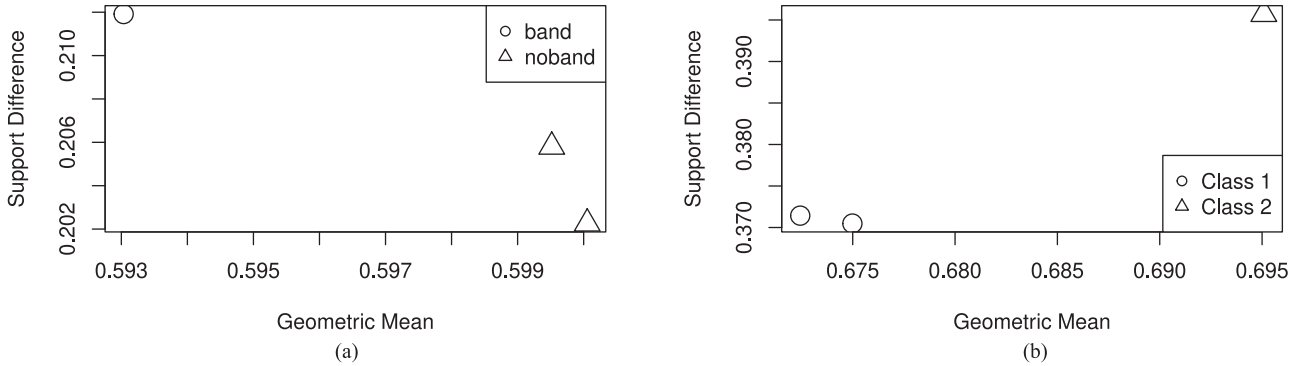

Fig. 7. Examples of the Pareto front extracted from MOEA-EFEP in different datasets. Each class represents an individual Pareto front. (a) Bands. (b) German.

TABLE X
EXAMPLES OF RULES AND THE RESULTS FOR THE QUALITY MEASURES FOR MOEA-EFEP IN BANDS AND GERMAN DATASETS

bands		*training*		*test*				
Rule		G-mean	SuppDiff	WRAcc	CONF	GR	TPR	FPR
R_1:	If (Proof_cut = LL_2) AND (Ink_temperature = LL_2 OR LL_3) AND (Solvent_pct = LL_2) AND (Esa_voltage = LL_1) AND (ESA_amperage = LL_1) AND (Density = LL_2) AND (Anode_ratio = LL_1) THEN band	0.593	0.212	0.6059	0.5118	1.7861	0.4814	0.2695
R_2:	If (Ink_temperature = LL_2) AND (Roughness = LL_2) AND (Press_speed = LL_3) AND (Ink_pct = LL_1) AND (Wax = LL_2) AND (Density = LL_2) AND (Anode_ratio = LL_3) THEN noband	0.599	0.206	0.6028	0.7337	1.6173	0.5391	0.3333
R_3:	If (Ink_temperature = LL_1 OR LL_2) AND (Blade_pressure = LL_2) AND (Varnish_pct = LL_2) AND (Press_speed = LL_3) AND (Solvent_pct = LL_2) AND (Wax = LL_2) AND (Hardener = LL_2) AND (Anode_ratio = LL_2) AND (Chrome_content = LL_3) THEN noband	0.600	0.202	0.6011	0.7262	1.5572	0.5652	0.3629
german		*training*		*test*				
Rule		G-mean	SuppDiff	WRAcc	CONF	GR	TPR	FPR
R_1:	If (StatusAccount = A14 OR A13) AND (CreditAmount = LL_1) AND (Age = LL_2) THEN 1	0.672	0.371	0.6857	0.8772	3.0634	0.5514	0.1800
R_2:	If (StatusAccount = A14 OR A13) AND (Age = LL_2) THEN 1	0.675	0.370	0.6852	0.8706	2.8837	0.5671	0.1967
R_3:	If (StatusAccount = A11 OR A12) AND (Guarantors = A101 OR A102) AND (Age = LL_1 OR LL_2) AND (ForeignWorker = A201) THEN 2	0.695	0.396	0.6978	0.4720	2.0862	0.7600	0.3642

ber of variables of the problem. The execution time of EvAEP comes from the use of an IRL approach, because it can be executed several times. Nevertheless, it is important to remark that iEPMiner is the fastest algorithm on simple problems.

F. Analysis of the Rules Obtained by MOEA-EFEP

One of the key points in EPM is the interpretability of the results. A simple model allows an easy interpretation of the underlying phenomena or the comprehension of the changes in data throughout time. However, some concepts used to measure the interpretability of the models are conflicting. As an example, a more general rule is normally less reliable than a more specific one. For this reason, a MOEA has been proposed in this paper in order to find rules with the best tradeoff between these objectives, i.e., the Pareto front.

Fig. 7 shows different Pareto fronts obtained by MOEA-EFEP$^{\text{DNF}}_{\text{Conf}}$ on each class of the *bands* and *german* datasets. The Paretos of the different classes of each dataset are represented in the same figure. These Pareto fronts show that the MOEA approach and the objectives proposed within it are well-suited for the extraction of descriptive knowledge with a good tradeoff between the generality and reliability of the results.

In order to demonstrate the interpretability of the results extracted from MOEA-EFEP$^{\text{DNF}}_{\text{Conf}}$, Table X presents the whole rules sets extracted by MOEA-EFEP$^{\text{DNF}}_{\text{Conf}}$ on the datasets presented in Fig. 7. The rules obtained by MOEA-EFEP$^{\text{DNF}}_{\text{Conf}}$ use three LLs for the continuous variables, so $\text{LL}_1 = \text{Low}, \text{LL}_2 = \text{Medium}$, and $\text{LL}_3 = \text{High}$. Table X also presents the quality measures obtained in test data for each rule.

V. CONCLUSION

In this paper, a MOEA for extracting FEPs for EPM has been proposed. The main objective of this algorithm is to find simple and reliable EPs in order to describe emerging behavior in timestamped data or discriminating characteristics among classes.

MOEA-EFEP is a MOEA based on the NSGA-II approach, but oriented toward the extraction of high-quality EPs, using specific operators for this purpose. In particular, it uses a guided reinitialization and the niching technique based on the crowding distance, which promotes diversity. It uses a general rule-based initialization and oriented mutation operators, as well as the quality measures proposed as objectives to promote the generality of the rules obtained. Finally, MOEA-EFEP uses an elite population where a cooperative-competitive approach based on token competition and replacing based on global measures is performed in order to promote the reliability of results. All together with the possibility of using a knowledge representation based in canonical form or in DNF.

The suitability of using a DNF representation in order to obtain high-quality rules that can easily describe the underlying phenomena has been demonstrated in this paper. This allows us to develop future proposals with this kind of knowledge representation in order to improve the results obtained. Specifically, the reliability has been improved by means of a confidence threshold for MOEA-EFEP. Finally, the proposed algorithm re-

turns a set of rules, which is simpler, more easy to understand and as reliable as the most relevant algorithms in the EPM literature. Therefore, the use of MOEAs or other kind of metaheuristics based on fuzzy logic for EPM is a promising future research line for new developments.

REFERENCES

[1] P. Kralj-Novak, N. Lavrac, and G. I. Webb, "Supervised descriptive rule discovery: A unifying survey of constrast set, emerging patern and subgroup mining," *J. Mach. Learn. Res.*, vol. 10, pp. 377–403, 2009.

[2] N. Lavrac, B. Kavsek, P. A. Flach, and L. Todorovski, "Subgroup discovery with CN2-SD," *J. Mach. Learn. Res.*, vol. 5, pp. 153–188, 2004.

[3] S. D. Bay and M. J. Pazzani, "Detecting group differences: Mining contrast sets," *Data Mining Knowl. Discovery*, vol. 5, no. 3, pp. 213–246, 2001.

[4] G. Z. Dong and J. Y. Li, "Efficient mining of emerging patterns: Discovering trends and differences," in *Proc. 5th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 1999, pp. 43–52.

[5] Y. Yu, K. Yan, X. Zhu, and G. Wang, "Detecting of PIU behaviors based on discovered generators and emerging patterns from computer-mediated interaction events," in *Proc. 15th Int. Conf. Web-Age Inf. Manage.*, 2014, pp. 277–293.

[6] J.-P. Métivier *et al.*, "Discovering structural alerts for mutagenicity using stable emerging molecular patterns," *J. Chem. Inf. Model.*, vol. 55, no. 5, pp. 925–940, 2015.

[7] C. J. Carmona, C. Chrysostomou, H. Seker, and M. J. del Jesus, "Fuzzy rules for describing subgroups from influenza a virus using a multi-objective evolutionary algorithm," *Appl. Soft Comput.*, vol. 13, no. 8, pp. 3439–3448, 2013.

[8] Y. Nie, H. Wang, X. Lu, and Y. Qin, "Parallel emerging patterns in microarray," in *Proc. 6th Int. Conf. Intell. Hum.-Mach. Syst. Cybern.*, 2014, vol. 1, pp. 82–85.

[9] C. J. Carmona *et al.*, "A fuzzy genetic programming-based algorithm for subgroup discovery and the application to one problem of pathogenesis of acute sore throat conditions in humans," *Inf. Sci.*, vol. 298, pp. 180–197, 2015.

[10] G. Li, R. Law, H. Q. Vu, J. Rong, and X. R. Zhao, "Identifying emerging hotel preferences using emerging pattern mining technique," *Tourism Manage.*, vol. 46, pp. 311–321, 2015.

[11] M. J. del Jesus, P. González, F. Herrera, and M. Mesonero, "Evolutionary fuzzy rule induction process for subgroup discovery: A case study in marketing," *IEEE Trans. Fuzzy Syst.*, vol. 15, no. 4, pp. 578–592, Aug. 2007.

[12] C. J. Carmona, S. Ramírez-Gallego, F. Torres, E. Bernal, M. J. del Jesus, and S. García, "Web usage mining to improve the design of an e-commerce website: OrOliveSur.com," *Expert Syst. Appl.*, vol. 39, pp. 11 243–11 249, 2012.

[13] C. J. Carmona, P. González, B. García-Domingo, M. J. del Jesus, and J. Aguilera, "MEFES: An evolutionary proposal for the detection of exceptions in subgroup discovery. An application to concentrating photovoltaic technology," *Knowl.-Based Syst.*, vol. 54, pp. 73–85, 2013.

[14] L. Wang, H. Zhao, G. Dong, and J. Li, "On the complexity of finding emerging patterns," in *Proc. 28th Annu. Int. Comput. Softw. Appl. Conf.*, 2004, vol. 2, pp. 126–129.

[15] F. Herrera, "Genetic fuzzy systems: Taxomony, current research trends and prospects," *Evol. Intell.*, vol. 1, pp. 27–46, 2008.

[16] C. J. Carmona, P. González, M. J. del Jesus, and F. Herrera, "NMEEF-SD: Non-dominated multi-objective evolutionary algorithm for extracting fuzzy rules in subgroup discovery," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 5, pp. 958–970, Oct. 2010.

[17] C. J. Carmona, P. González, M. J. del Jesus, and F. Herrera, "Overview on evolutionary subgroup discovery: Analysis of the suitability and potential of the search performed by evolutionary algorithms," *WIREs Data Mining Knowl. Discovery*, vol. 4, no. 2, pp. 87–103, 2014.

[18] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading, MA, USA: Addison-Wesley, 1989.

[19] E. Hüllermeier, "Fuzzy methods in machine learning and data mining: Status and prospects," *Fuzzy Sets Syst.*, vol. 156, no. 3, pp. 387–406, 2005.

[20] K. Deb, A. Pratap, S. Agrawal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.

[21] G. Z. Dong and J. Y. Li, "Mining border descriptions of emerging patterns from dataset pairs," *Knowl. Inf. Syst.*, vol. 8, no. 2, pp. 178–202, 2005.

[22] C. J. Carmona, M. J. del Jesus, and F. Herrera, "A unifying analysis for the supervised descriptive rule discovery via the weighted relative accuracy," *Knowl.-Based Syst.*, vol. 139, pp. 89–100, 2018.

[23] A. M. García-Vico, C. J. Carmona, D. Martín, M. García-Borroto, and M. J. del Jesus, "An overview of emerging pattern mining in supervised descriptive rule discovery: Taxonomy, empirical study, trends and prospects," *WIREs, Data Mining Knowl. Discovery*, vol. 8, no. 1, 2018.

[24] V. Cherkassky and F. Mulier, *Learning From Data. Concepts, Theory and Methods*, 2nd ed. Piscataway, NJ, USA: IEEE Press, 2007.

[25] G. Box, G. Jenkins, and G. Reinsel, *Time Series Analysis: Forecasting and Control*, 4th ed. Hoboken, NJ, USA: Wiley, 2008.

[26] R. Agrawal, T. Imieliski, and A. Swami, "Mining association rules between sets of items in large databases," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 1993, pp. 207–216.

[27] R. Zembowicz and J. M. Zytkow, "From contingency tables to various forms of knowledge in databases," in *Advances in Knowledge Discovery and Data Mining*. Cambridge, MA, USA: MIT Press, 1996, pp. 329–349.

[28] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, "An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes," *Pattern Recognit.*, vol. 44, no. 8, pp. 1761–1776, 2011.

[29] U. M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery: an overview," in *Advances in Knowledge Discovery and Data Mining*. Cambridge, MA, USA: MIT Press, 1996, pp. 1–34.

[30] W. Kloesgen, "Explora: A multipattern and multistrategy discovery assistant," in *Advances in Knowledge Discovery and Data Mining*. Palo Alto, CA, USA: American Association for Artificial Intelligence, 1996, pp. 249–271.

[31] D. Gamberger and N. Lavrac, "Expert-guided subgroup discovery: Methodology and application," *J. Artif. Intell. Res.*, vol. 17, pp. 501–527, 2002.

[32] L. A. Zadeh, "The concept of a linguistic variable and its applications to approximate reasoning. Parts I, II, III," *Inf. Sci.*, vol. 8, no. 3, pp. 199–249 (part I), 1975; vol. 8, no. 4, pp. 301–357 (part II), 1975; vol. 9, no. 1, pp. 43–80 (part III), 1975.

[33] E. Hüllermeier, "Fuzzy sets in machine learning and data mining," *Appl. Soft Comput.*, vol. 11, no. 2, pp. 1493–1505, 2011.

[34] E. Onieva, U. Hernandez-Jayo, E. Osaba, A. Perallos, and X. Zhang, "A multi-objective evolutionary algorithm for the tuning of fuzzy rule bases for uncoordinated intersections in autonomous driving," *Inf. Sci.*, vol. 321, pp. 14–30, 2015.

[35] J. A. Sanz, D. Bernardo, F. Herrera, H. Bustince, and H. Hagras, "A compact evolutionary interval-valued fuzzy rule-based classification system for the modeling and prediction of real-world financial applications with imbalanced data," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 4, pp. 973–990, Aug. 2015.

[36] A. M. García-Vico, J. Montes, J. Aguilera, C. J. Carmona, and M. J. del Jesus, "Analysing concentrating photovoltaics technology through the use of emerging pattern mining," in *Proc. 11th Int. Conf. Soft Comput. Models Ind. Environ. Appl.*, 2016, pp. 1–8.

[37] M. Antonelli, D. Bernardo, H. Hagras, and F. Marcelloni, "Multiobjective evolutionary optimization of type-2 fuzzy rule-based systems for financial data classification," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 2, pp. 249–264, Apr. 2017.

[38] O. Cordón, F. Herrera, F. Hoffmann, and L. Magdalena, *Genetic Fuzzy Systems: Evolutionary Tuning and Learning of Fuzzy Knowledge Bases*. Singapore: World Scientific, 2001.

[39] K. DeJong, W. Spears, and D. F. Gordon, "Using genetic algorithms for concept learning," *Mach. Learn.*, vol. 13, no. 2, pp. 161–188, 1997.

[40] M. L. Wong and K. S. Leung, *Data Mining Using Grammar Based Genetic Programming and Applications*. Norwell, MA, USA: Kluwer, 2000.

[41] H. Li and Q. Zhang, "Multiobjective optimization problems with complicated Pareto sets, MOEA/D and NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 13, no. 2, pp. 284–302, Apr. 2009.

[42] E. Zitzler and S. Künzli, "Indicator-based selection in multiobjective search," in *Proc. Int. Conf. Parallel Problem Solving Nature*, 2004, pp. 832–842.

[43] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength Pareto evolutionary algorithm for multiobjective optimization," in *Proc. Int. Congr. Evol. Methods Des. Optim. Control Appl. Ind. Probl.*, 2002, pp. 95–100.

[44] H. Ishibuchi, N. Tsukamoto, Y. Hitotsuyanagi, and Y. Nojima, "Effectiveness of scalability improvement attempts on the performance of nsga-ii for many-objective problems," in *Proc. 10th Annu. Conf. Genetic Evol. Comput.*, 2008, pp. 649–656.
[45] K. S. Leung, Y. Leung, L. So, and K. F. Yam, "Rule learning in expert systems using genetic algorithm: 1, concepts," in *Proc. 2nd Int. Conf. Fuzzy Logic Neural Netw.*, 1992, pp. 201–204.
[46] M. Buckland and F. Gey, "The relationship between recall and precision," *J. Amer. Soc. Inf. Sci.*, vol. 45, no. 1, pp. 12–19, 1994.
[47] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*. Hoboken, NJ, USA: Willey, 2001.
[48] M. Kubat and S. Matwin, "Addressing the curse of imbalanced training sets: One-sided selection," in *Proc. 14th Int. Conf. Mach. Learn.*, 1997, vol. 97, pp. 179–186.
[49] B. L. Miller and D. E. Goldberg, "Genetic algorithms, tournament selection, and the effects of noise," *Complex Syst.*, vol. 9, pp. 193–212, 1995.
[50] J. H. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI, USA: University of Michigan Press, 1975.
[51] H. Fan and K. Ramamohanarao, "Efficiently mining interesting emerging patterns," in *Proc. 4th Int. Conf. Web-Age Inf. Manage.*, 2003, pp. 189–201.
[52] M. García-Borroto, J. Martínez-Trinidad, and J. Carrasco-Ochoa, "Fuzzy emerging patterns for classifying hard domains," *Knowl. Inf. Syst.*, vol. 28, no. 2, pp. 473–489, 2011.
[53] A. Asuncion and D. J. Newman, "UCI Machine Learning Repository," 2007. [Online]. Available: http://www.ics.uci.edu/ mlearn/ MLRepository.html
[54] U. M. Fayyad and K. B. Irani, "Multi-interval discretization of continuous-valued attributes for classification learning," in *Proc. 13th Int. Joint Conf. Artif. Intell.*, 1993, pp. 1022–1029.
[55] J. G. Moreno-Torres, J. A. Sáez, and F. Herrera, "Study on the impact of partition-induced dataset shift on-fold cross-validation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 8, pp. 1304–1312, Aug. 2012.
[56] S. García, A. Fernández, J. Luengo, and F. Herrera, "Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power," *Inf. Sci.*, vol. 180, pp. 2044–2064, 2010.
[57] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics*, vol. 1, pp. 80–83, 1945.
[58] R. L. Iman and J. M. Davenport, "Approximations of the critical region of the friedman statistic," *Commun. Statist. A, Theory Methods*, vol. 9, pp. 571–595, 1980.
[59] S. Holm, "A simple sequentially rejective multiple test procedure," *Scand. J. Statist.*, vol. 6, pp. 65–70, 1979.

Ángel Miguel García-Vico received the B.Sc. degree in computer science from the University of Jaén, Jaén, Spain, in 2015 and the M.Sc. degree in data science and computer engineering from the University of Granada, Granada, Spain, in 2016. He is currently working toward the Ph.D. degree in computer science from the University of Jaén.

His research interests include emerging patterns, subgroups discovery, supervised descriptive rule discovery, evolutionary fuzzy systems, big data, and data stream mining.

Cristóbal José Carmona received the M.Sc. and Ph.D. degrees in computer science from the University of Jaén, Jaén, Spain, in 2007 and 2011, respectively.

He is currently an Assistant Professor with the Department of Computer Science, University of Jaén, and Honorary Senior Research Fellow with the De Montfort University, Leicester, U.K. He is the author of more than 20 of original articles and is involved in teaching artificial intelligence for students of computer sciences. His research interests include subgroup discovery, emerging patterns, contrast mining, supervised descriptive rule discovery, web usage mining, evolutionary algorithms, evolutionary fuzzy systems, data mining, big data, and real applications through soft computing techniques.

Pedro González received the M.Sc. and Ph.D. degrees in computer science from the University of Granada, Spain, in 1991 and 2008, respectively.

He is an Assistant Professor with the Department of Computer Science at University of Jaén, Jaén, Spain. He has published more than 20 journal papers and received 800 citations (Scholar Google, H-index 13). His research interest include subgroup discovery, supervised descriptive rule discovery, knowledge extraction based on evolutionary algorithms, fuzzy rule-based systems, soft computing, data streaming, data science and big data.

María José del Jesus received the M.Sc. and Ph.D. degrees in computer science from the University of Granada, Granada, Spain, in 1994 and 1999, respectively.

She is currently a Professor with the Department of Computer Science, University of Jaén, Jaén, Spain. She is the Supervisor of seven Ph.D. students. She has published more than 65 journal papers that have received more than 4500 citations (Scholar Google, H-index 31). She is a coauthor of the book *Multilabel Classification-Problem Analysis, Metrics and Techniques* (Springer, 2016). Her current research interests include fuzzy rule-based systems, subgroup discovery, emerging pattern mining, multilabel classification, data preparation, radial basis neural networks, knowledge extraction based on evolutionary algorithms, and data science and big data.