Knowledge and Information Systems manuscript No.
(will be inserted by the editor)

Mining Exceptional Relationships with Grammar-Guided Genetic Programming

J. M. Luna · M. Pechenizkiy · S. Ventura

Received: Aug 08, 2014 / Accepted: Jun 13, 2015

Abstract Given a database of records, it might be possible to identify small subsets of data which distribution is exceptionally different from the distribution in the complete set of data records. Finding such interesting relationships, which we call exceptional relationships, in an automated way would allow discovering unusual or exceptional hidden behavior. In this paper we formulate the problem of mining exceptional relationships as a special case of exceptional model mining, and propose a grammar guided genetic programming algorithm (MERG3P) that enables the discovery of any exceptional relationships. In particular, MERG3P can work directly not only with categorical, but also with numerical data. In the experimental evaluation, we conduct a case study on mining exceptional relations between well-known and widely used quality measures of association rules, which exceptional behavior would be of interest to pattern mining experts. For this purpose, we constructed a dataset comprising a wide range of values for each considered association rule quality measure, such that possible exceptional relations between measures could be discovered. Thus, besides the actual validation of MERG3P, we found that the support and leverage measure in fact are negatively correlated under certain conditions, while in general experts in the field expect these measures to be positively correlated.

Keywords Association Rules · Exceptional Subgroups · Genetic Programming

J. M. Luna
Department of Computer Science and Numerical Analysis, University of Cordoba, Rabanales Campus, 14071 Cordoba, Spain
E-mail: jmluna@uco.es

M. Pechenizkiy
Department of Computer Science, Eindhoven University of Technology, W&I, TU/e, P.O. Box 513, 5600 MB Eindhoven, the Netherlands
E-mail: m.pechenizkiy@tue.nl

S. Ventura
Department of Computer Science and Numerical Analysis, University of Cordoba, Rabanales Campus, 14071 Cordoba, Spain. Dr. Ventura also belongs to Department of Computer Science, King Abdulaziz University, Saudi Arabia Kingdom
E-mail: sventura@uco.es

1 Introduction

Currently, more and more companies gather raw data about their customers. By and large, this set of data lacks of interest as a whole, so the extraction, description and analysis of hidden knowledge in data is a requirement to make right decisions for the company.

The extraction and description of knowledge in data is a broadly studied task and many techniques are available to achieve this aim. One of the most widely known techniques for this end is association rule mining (ARM) [1,16]. Since ARM reveals relations among patterns, this task has been broadly used in the business field, where the discovery of strong and interesting relations helps in effective decision making. However, business is not the only area where this challenging task has been applied to, including medical diagnosis, census data, web fraud detection, credit car business, etc. In many of these fields the searching for exceptional models — groups of patterns whose distribution is exceptionally different from that of the entire data — could be a very interesting goal. To this end, the concept of exceptional model mining (EMM) was introduced by *Leman et al.* [11].

EMM was proposed as an interesting task, ascertaining subgroups where a model fitted to the subgroup is substantially different from that same model fitted to the entire database. These models are defined as sets of items or attributes that guarantees the relative frequency of the subgroups in the database. Nevertheless, similarly to frequent pattern mining, these subgroups are considered as a whole and do not describe relations between the attributes. At this point, ARM could provide interesting features that together with the EMM formulation may give rise to exceptional relationship mining (ERM). This new task is considered as a special case of both EMM and ARM, lying in the intersection of the aforementioned two tasks. ERM enables the description of reliable, abnormal and exceptional relations between attributes to be discovered so its the computational cost in the mining process is much higher than the one of EMM, requiring the mining of both exceptional subgroups and association rules within each exceptional subgroup.

ERM could be approached from different perspectives with respect to traversing the search space; exhaustive search or by means of specific heuristics. While in frequent itemset mining the task is typically formulated as to find *all* frequent itemsets satisfying certain constraints and therefore the use of exhaustive search with pruning is a logical choice, in EMM and ERM this formulation is less relevant, and hence, the looking for exceptional models is a task that has been usually addressed by specific meta-heuristics. For instance, an iterative model in the EMM field was proposed by *Leeuwen* [9] where the final goal was to find maximally exceptional models comprising the minimal number of features. Recently, *Leeuwen et al.* [10] proposed a beam search methodology for finding diverse subgroup sets.

Despite ERM could be addressed in many ways, we have suggested the use of a specific heuristic by means of a grammar-guided genetic programming (G3P) model [5], which achieved excellent results in unsupervised learning tasks [12]. We have not considered the use of an exhaustive search pattern mining because datasets that are too large or too complex could become intractable. For instance, exhaustive search approaches cannot deal with numeric attributes, requiring them to be discretized first. It has been shown that applying discretization to the numeric data and using pattern mining techniques could give rise to a loose of information of interest [6,18].

Furthermore, the use of a G3P methodology allows us to use a grammar to represent the association rules on any domain and enabling solutions where the shape, size and structural complexity are constrained by the grammar. Indeed, the proposed grammar could be easily adapted to consider the original EMM problem, where the induced subgroups are given by predefined target models. Additionally, the aim of the proposed approach is the discovery of exceptional rules related to any non-predefined model or context, so the use of an evolutionary algorithm is justified since the number of contexts and rules associated to that contexts could be so high to be computationally intractable for an exhaustive search methodology, specially in numerical domains.

We have conducted a series of experiments demonstrating the usefulness of the ERM and the feasibility of the proposed G3P approach. In this sense, we tackle a real application by the analysis of quality measures in the ARM field, which is an interesting and widely studied problem by expert users in the field [2], looking for exceptional behaviours in measures that, at first glance, seem rather homogeneous. For instance, Support and Leverage quality measures seem to be positively correlated. Nevertheless, the experimental stage reveals a completely hidden knowledge, stating that this correlation is quite different under some restrictions.

The main contribution of this paper is the proposal of a new task, ERM, which is based on EMM. This new task allows to describe exceptional relationships in subgroups that, at first glance, seems rather homogeneous. ERM could be considered from different perspectives, from exhaustive search to evolutionary methodologies. We proposed a way to achieve this with MERG3P. Its applicability and utility has been demonstrated, describing exceptional relationships between different quality measures in the ARM field, which provides unknown knowledge for further studies about how the measures are correlated and how the exceptionalness appears in this correlation.

The rest of the paper is organized as follows: Section 2 presents the most relevant definitions, related work and explains the novelty and contributions of our work correspondingly; Section 3 describes the proposed MERG3P algorithm; Section 4 presents the datasets used in the experiments, the experimental set-up and the results obtained; finally, some concluding remarks are outlined in Section 5.

2 Preliminaries and problem formulation

In this section, we formally define the proposed concept of exceptional relationship mining (ERM). To do so, we previously explain, in a formal way, both the ARM and EMM problems, discuss the intuition behind and justification for our approach for ERM and highlight the novelty of our work.

2.1 Association Rule Mining

ARM is considered as an important descriptive task in the data mining field, and it has received enormous attention since its introduction by *Agrawal et al.* [1] in the early 90s. ARM seeks for frequent, interesting and strong relationships among patterns that are usually hidden in data. These relationships are known as associ-

ation rules (AR), which are implications of the form $Antecedent \rightarrow Consequent$, both $Antecedent$ (A) and $Consequent$ (C) being sets with no items in common.

Definition 1 (Association rule). Let $\mathcal{I} = \{i_1, i_2, ..., i_n\}$ be a set of items, and let A and C be item-sets, that is, $A = \{i_1, ..., i_j\} \subset \mathcal{I}$ and $C = \{i_1, ..., i_k\} \subset \mathcal{I}$. An association rule (AR) is an implication of the type $A \rightarrow C$ where $A \subset \mathcal{I}$, $C \subset \mathcal{I}$, and $A \cap C = \emptyset$, where the frequency of occurrence or support of $A \rightarrow C$ is higher than a previously fixed value, i.e., $support(A \cup C) \geq minimum\_value$.

Definition 2 (Rare association rule). Let $\mathcal{I} = \{i_1, i_2, ..., i_n\}$ be a set of items, and let A and C be item-sets, that is, $A = \{i_1, ..., i_j\} \subset \mathcal{I}$ and $C = \{i_1, ..., i_k\} \subset \mathcal{I}$. A rare association rule (RAR) is an implication of the type $A \rightarrow C$ where $A \subset \mathcal{I}$, $C \subset \mathcal{I}$, and $A \cap C = \emptyset$, where the frequency of occurrence or support of $A \rightarrow C$ is less than a previously fixed value, i.e., $support(A \cup C) < maximum\_value$.

In general, the meaning of an association rule is that if the antecedent A is satisfied, then it is highly probable that the consequent C is also satisfied. Thus, the mining of ARs was originally designed for the market basket analysis to obtain relations between products like $diapers \rightarrow beer$ that describes the high probability of someone buying $diapers$ also buying $beer$. It would allow shop-keepers to exploit this relationship by moving the products closer together on the shelves.

Definition 3 (Association rule mining). Let $\mathcal{D}$ be a dataset comprising a set of transactions $\mathcal{T}$, and each transaction $t_i \subset \mathcal{T}$ having a set of items $\mathcal{I}$ or features. The association rule mining (ARM) is a task that finds all AR/RAR that satisfy specific user thresholds (according to the set of transactions satisfied), i.e., $ARM = \{\forall\ AR \mid quality(AR) \geq threshold\}$.

Generally, ARM suffers from four main problems. First, the discovery of interesting items from large datasets is a hard process since many items should be analysed. Second, some of the items simply appear by chance so they could be considered as spurious, and therefore, useless for the user. Third, the process of extracting association rules from the set of available items is computationally expensive and requires a large amount of memory. Finally, the presence of numerical attributes in a dataset is a drawback and many algorithms, specially exhaustive search algorithms, require a preprocessing step to discretize them. For a better understanding of the computational and memory requirements, a dataset containing k items could generate $2^k - 1$ item-sets and $3^k - 2^{k+1} + 1$ rules, so the use of huge datasets implies a prohibitively hard process of mining.

Many authors have dealt with these drawbacks [4,12]. An important research on this sense was carried out by *Luna et al.* [12], who proposed a grammar guided genetic programming (G3P) algorithm to solve all the aforementioned problems. This algorithm, called G3PARM, makes use of a grammar to represents association rules in any domain and not requiring a previous mining of the interesting items. Nevertheless, not only the extraction of frequent association rules is interesting, but also the discovery of abnormal or unusual behaviour hidden in data. In this regard, the extraction of reliable and rare relations [14,7] is considered as a subset of ARM and also an area of interest for further exploration.

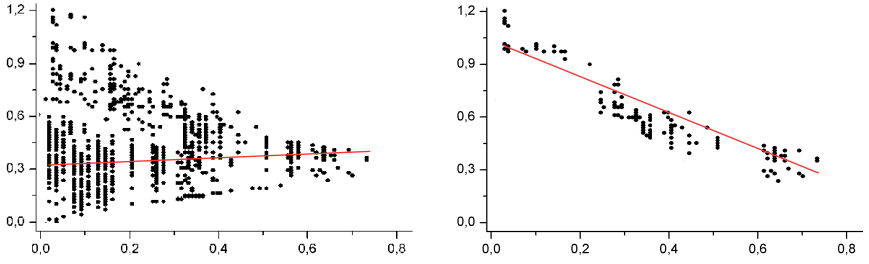

Fig. 1 Example of the scatter plot and distribution of a sample subgroup (left) and its complement (right)

2.2 Exceptional Model Mining

A pattern p in a domain D is defined as a function $p : D \rightarrow \{0, 1\}$ so that a pattern p satisfies a transaction t_i from a set of transactions $\mathcal{T}$ if and only if $p(t_i) = 1$. This definition could be moved as a set of patterns instead of an isolate pattern p.

Definition 1 (Subgroup). A subgroup that fits to a relation r is a set of transactions $t \subseteq \mathcal{T}$ that are satisfied by the relation r, i.e., $t = \{t_n \in \mathcal{T} | r(t_n) = 1\}$.

Definition 2 (Complement). The complement of a subgroup $\overline{G}$ is the set of transactions $\overline{t} \subseteq \mathcal{T}$ that are not covered by the subgroup G, i.e., $\overline{t} = \mathcal{T} \setminus \{t_n \in \mathcal{T} | G(t_n) = 1\}$.

Definition 3 (Context of a subgroup). The context c of a subgroup G is the set of features where G presents a specific behaviour. Given a set of transactions $\mathcal{T}$, the context of G is formally defined as $c^M = \{t \subseteq \mathcal{T}\}$.

EMM is defined as the discovery of features where the model fitted to a specific subgroup is significantly different to its complement (see Figure 1). EMM problem was presented by *Leman et al.* [11] and it was defined (*Problem 1*) as the discovery of subgroups where the model fitted to a subgroup is significantly different to its complement.

Problem 1 (Exceptional model mining). Suppose we are given a dataset comprising a set of transactions $\mathcal{T}$, and a measure ϕ that quantifies the exceptionalness according to a minimum threshold ϵ. The task is to find all subgroups $\mathcal{G}$ with corresponding minimal subgroup description, such that each $G \in \mathcal{G}$ implies an exceptional model M, i.e., $\phi^{\mathcal{T}^M}(G^M) \geq \epsilon$.

2.3 Exceptional Relationship Mining

We propose a task that lies in the intersection of ARM and EMM. This task, known as ERM (*Problem 2*), is defined as the discovery of accurate relations among

patterns where the subgroup fitted to a specific relation is significantly different to its complement. Similarly to ARM, ERM is a learning task where any attribute could be considered to be included in both the antecedent and consequent of a relation. Also, each relation is analysed within a context based on the patterns satisfied.

Problem 2 (Exceptional relationship mining). Suppose we are given a dataset comprising a set of transactions $\mathcal{T}$, and all the relationships defined in a model M. The task is to find a set of association rules $\mathcal{R}$ with corresponding minimal relationship description, such that each $R \in \mathcal{R}$ describes an exceptional subgroup G, i.e., R *describes* $(G \mid \phi^{\mathcal{T}^M}(G^M) \geq \epsilon)$.

The main difference between ERM and EMM lies on the fact that EMM obtains exceptional subgroups comprising a conjunction of items, and ERM obtains exceptional subgroups described by rules that describe the relationships between the conjunction of items within each subgroup. Additionally, it should be noted that the computational cost of ERM is higher than the one of EMM since it is required the mining of both exceptional subgroups and association rules within each exceptional subgroup.

The applicability of ERM is its ability to discover exceptional behaviour in a group that, at first glance, seems rather homogeneous. For example, let us consider a dataset with information about people of different countries around the world. In general, the people's features trend to present a similar distribution. Nevertheless, there exists some exceptional features whose distribution is completely different from an undeveloped country than from the rest.

2.4 Proposed Approach

As mentioned above, EMM was defined as a task to discover exceptional subgroups in data. However, not only it is interesting in data analytics the discovery of exceptional behaviour in a group that, at first glance, seems rather homogeneous, but it is also necessary to describe its behaviour by means of relationships between patterns. ERM can be approached in many different ways, for instance, the use of an exhaustive search methodology to extract exceptional relationships is a possibility, similarly to Apriori was proposed for mining ARs. Nevertheless, the use of numerical features in the defined contexts makes non-viable the use of an exhaustive search approach, requiring a discretization step. In this regard, the use of an evolutionary methodology could bring about interesting advantages for dealing with numerical domains.

In this paper, we propose a novel approach for mining what we consider as exceptional relationships or exceptional association rules, i.e., reliable relations that describe an exceptional behaviour when they are described within a context. Unlike the original EMM problem, these relationships enable the description of exceptional behaviour on any attribute, denoting the accuracy of the relations and how infrequent these exceptional relations are.

Note that, since we are considering a task that lies in the intersection between EMM and ARM, the resulting approach inherits the drawbacks considered by any association rule mining algorithm. In this regard, the use of G3P enables the

overcoming of these drawbacks as stated by *Luna et al.* [12]. Additionally, the use of a grammar to encode the solutions enables the restriction of the search space and the guiding of solutions to the users' aim. Thus, despite the fact that any evolutionary approach does not guarantee the extraction of the complete set of solutions, the grammar used in the proposed approach enables both the execution on any domain and the finding of specific rules instead of the complete set of rules discovered by any exhaustive search methodology, not requiring a post-processing step. We describe MERG3P in detail in the following section.

3 MERG3P: A Grammar-Guided Genetic Programming Approach

In this section, unusual relations between items are considered, since not only the extraction of exception subgroups is a dare, but also their description. The main interest is to carry out this description on numerical domains, so the use of exhaustive search methodologies like Apriori is dismissed. In this sense, we propose the use of an evolutionary methodology to describe exceptional subgroups in data. More precisely, we focus on the use of G3P that brings about both a pruning of the search space and the discovery of features over different domains without a preprocessing step. G3P defines solutions by using a context-free grammar, so its applicability to huge search spaces is not a problem since this methodology allows to constraint the solutions and, therefore, reduce the search space. Like any evolutionary methodology, solutions are ranked with a fitness function that, in this specific approach, is based on the absolute difference between the correlation coefficient of the rule in a context and the correlation coefficient of the complement of the rule. Additionally, descriptions about different models in different contexts could be obtained by running the proposed approach only once. Thus, solutions are grouped into different niches (one per context) and the best contexts, according to the fitness function, are described once a maximum number of generations is reached.

3.1 Encoding Criterion

G3P [5] appeared in the mid 90s as an extension of genetic programming (GP) [8]. This methodology was proposed as a way of using grammars for formalizing constraints in GP, since grammar formalisms is considered as a major representation structure in computer science. In G3P, each solution to the problem under study is represented by a genotype and a phenotype. Hence the genotype is a derivation syntax tree based on the language defined by a grammar G (Figure 2 depicts the context-free grammar defined to the proposed G3P algorithm) to enforce syntactic constraints on the tree. The term phenotype refers to the meaning of the tree structure, i.e., the phenotype represents the rule that is directly evaluated.

For a better understanding of the concept of context-free grammar, it is defined as a four-tuple $(\Sigma_N, \Sigma_T, P, S)$ where Σ_T represents the alphabet of terminal symbols and Σ_N the alphabet of non-terminal symbols. Notice that they have no common elements, i.e., $\Sigma_N \cap \Sigma_T = \emptyset$. In order to encode an individual using a G3P approach, a number of production rules from the set P are applied starting from the start symbol S. A production rule is defined as $\alpha \rightarrow \beta$ where $\alpha \in \Sigma_N$,

G = (Σ_N, Σ_T, P, S) with:

S = Rule
Σ_N = {Rule, Conditions, Consequent, Condition, Condition_Nominal, Condition_Numerical }
Σ_T = {'AND', 'Attribute', 'Consequent_value', '=', 'IN','Min_value','Max_value', 'Consequent_attribute' }
P = {Rule = Conditions, Consequent ;
Conditions = 'AND', Conditions, Condition | Condition ;
Condition = Condition_Nominal | Condition_Numerical ;
Condition_Nominal = 'Attribute', '=', 'value' ;
Condition_Numerical = 'Attribute', 'IN', 'Min_value', 'Max_value' ;
Consequent = Condition_Nominal | Condition_Numerical ; }

Fig. 2 Context-free grammar to represent exceptional rare association rules expressed in extended BNF notation

and $\beta \in \{\Sigma_T \cup \Sigma_N\}^*$. It should be noted that in any context-free grammar there may appear the production rule $\alpha \rightarrow \varepsilon$, i.e., the empty symbol ε is directly derived from α. To obtain individuals, a number of production rules is applied from the set P, obtaining a derivation syntax tree for each individual, where internal nodes contain only non-terminal symbols, and leaves contain only terminals. This process begins from the start symbol `Rule`, which always has a child node representing the antecedent of the rule, i.e., the conjunction of conditions, and the consequent. Considering the grammar defined in this approach, the following language is obtained $L(G) = \{\ (AND\ Condition)^n\ Condition\ \rightarrow\ Consequent : n \geq 0\}$.

In order to clear up the individual representation, a sample individual generated through the application of a series of production rules is illustrated in Figure 3. This sample individual was generated from the set of production rules P of the proposed grammar and the sample meta-data depicted in Table 1, where the leafs represent terminal symbols according to the meta-data. As mentioned above, the phenotype of an individual represents the meaning of the syntax-tree, i.e., the rule obtained by eliminating non-terminal genotype symbols. Thus, focusing on the sample tree depicted, the phenotype represents the following rule:

IF $Size\ IN\ [160, 350]\ AND\ Weight\ IN\ [150, 220]$
THEN $PriceIN [300, 550]$

Additionally, not only the grammar to represent individuals is required, but also the context where the individual is considered. In this regard, each individual is represented within a context comprising two attributes from the set of available attributes and not included in the tree. Therefore, each individual includes a

Table 1 Sample meta-data including four attributes for a sample dataset

Attributes	Values
Size	[100, 500]
Price	[200, 1000]
Weight	[10, 230]
Elements	[1, 15]

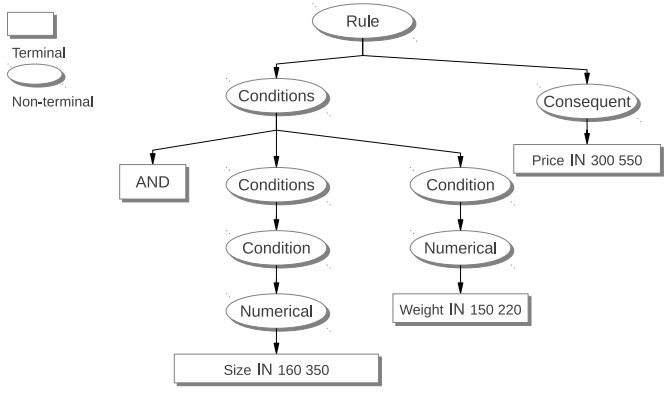

Fig. 3 Sample derivation syntax-tree conformant to the grammar defined in the proposed algorithm

derivation syntax tree and a bit-set where the i-th bit represents the i-th attribute, and only two bits could be set at once. For a better understanding, consider the sample meta-data depicted in Table 1 that comprises four attributes. Six different contexts could be mined: 0011, 0101, 1001, 0110, 1010, and 1100. Thus, a sample individual encoded by the grammar and the contextual bit-set is depicted in Figure 4, where its space of solutions is given by the contextual attributes and its phenotype provides a solution of the given space of solutions.

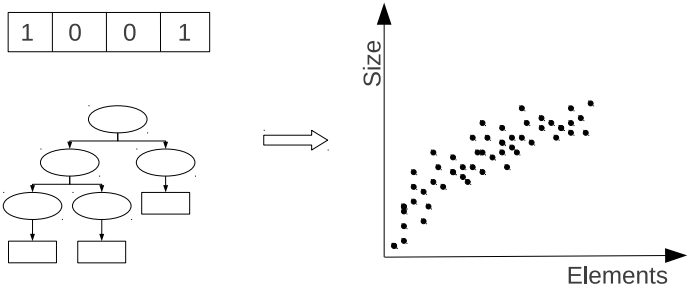

Fig. 4 Context representation of a sample individual including the syntax tree and the bit-set from the meta-data shown in Table 1

Finally, it should be noted that the search space could be huge in this problem, since the goal of this algorithm is not only the discovery of rare association rules, but it is also required to consider the context where the rule is exceptional. In this problem, the higher the number of attributes, the higher the number of contexts where exceptional rare association rules should be mined. Thus, the use of an evolutionary algorithm is highly recommended since the number of contexts and rules associated to that contexts could be so high to be considered computationally intractable.

3.2 Evaluation Process

One of the most important processes in any evolutionary algorithm is the evaluation of the solutions, determining how close a given solution is from achieving the aim. As far as the ARM problem is concern, many researchers focused their studies on the choice of good quality measures [2,15]. Two are the most widely-used measures in this field: Support and Confidence. The Support measure is used to determine whether a certain association rule is a frequent occurrence. In this regard, given a set of all transactions $\mathcal{T} = \{t_1, t_2, t_3, ..., t_n\}$ in a dataset, the Support of an association rule R is defined as the number of transactions satisfied by this rule, i.e., $\{Support(R) = |\mathcal{S}|,\ \mathcal{S} \subseteq \mathcal{T}\}$, $|\mathcal{S}|$ being the number of transactions satisfied by the rule. An association rule is established as frequent if its Support is higher than a Support threshold previously fixed.

The Support measure is the most widely used measure in ARM and some exhaustive search algorithms [1,4] used this measure to reduce the computational cost, since those items with a Support lower than a threshold are not considered. Nevertheless, some authors [13] studied the problem of mining rare or infrequent association rules, where the minimum Support threshold is considered as maximum and only rules having a Support value lower than a maximum threshold are considered. Regarding the Confidence measure, it calculates the reliability of an association rule, i.e., the probability of finding the right-hand side of the rule in the dataset under the condition that the left-hand side is also satisfied. This quality measure determines how much a given item depends on another. In a formal way, the Confidence measure is defined as $Confidence(R) = Confidence(X \rightarrow Y) = Support(R)/Support(X)$.

In the algorithm proposed in this paper, the two aforementioned quality measures are considered and the Support quality measure is not used to determine frequent rules but infrequent ones. In this sense, two thresholds are considered (a minimum and a maximum) in order to identify infrequent association rules and discard those that appear in such a low frequency that could be considered as noise, e.g., rules that cover only 1% of the transactions. As for the Confidence measure, it is used to mine only rules that are more reliable than a threshold. Thus, only Support and Confidence are considered to determine the quality of the association rules extracted. Nevertheless, these quality measures do not determine the exceptionalness of a rule so new quality measures are required in this sense. Thus, we propose the use of the Pearson's correlation coefficient ρ (see Equation 1) to determine the exceptionalness of a rule with respect to its context.

$$\rho_{X,Y} = \frac{COV(X,Y)}{\sigma_x \sigma_Y} \tag{1}$$

The Pearson's coefficient measures the strength of linear dependence between two variables, and is defined as the division between the covariance of the two variables and the product of their standard deviations. This coefficient ranges from 1 to -1, indicating a positive or negative correlation between the variables, respectively. A value of 0 indicates no linear correlation between the variables.

Once these three quality measures were described, we propose a fitness function to determine how promising each individual is. To do so, we consider the use of the three measures at once. Support and Confidence measures are used to determine whether a rule should be analized, since some quality thresholds are predefined by

the user. If a certain association rule does not satisfy the Support and Confidence thresholds, then a 0 fitness function value is assigned to this solution. On the contrary, the fitness value is determined by the absolute difference between the correlation coefficient of the rule in a context $\rho_{Context}(Rule)$ and the correlation coefficient of the complement of the rule in the same context $\rho_{Context}(\overline{Rule})$. Therefore, starting from an association rule R, the fitness function F used in this algorithm is defined in Equation 2.

$$F(R) = \begin{cases} |\rho_{Context}(R) - \rho_{Context}(\overline{R})| & \text{if } Support_{min} \leq Support(R) \ \wedge \\ & Support(R) \leq Support_{max} \ \wedge \\ & Confidence(R) \geq Confidence_{min} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

3.3 Genetic Operators

Similarly to any evolutionary algorithm, the proposed algorithm uses two genetic operators to generate new individuals from the existing ones in each generation of the evolutionary process. The first genetic operator follows a methodology widely used by many G3P algorithms [12, 19]. On the contrary, the second genetic operator proposed in this paper was designed for the problem under study, not working on the syntax-tree but on the *context* where the syntax-tree is described.

Operator to change conditions. The first genetic operator creates new individuals by generating new conditions from a set of parents randomly selected. In this regard, the genetic operator randomly chooses a condition from the set of conditions of a given parent, creating a completely new condition that replaces the old-one. The pseudo-code of this genetic operator is shown in Algorithm 1. Using this genetic operator, only the syntax-tree is modified so the context is not affected by this genetic operator.

Algorithm 1 Genetic operator that changes conditions randomly selected

Require: *parents*
Ensure: *offsprings*
1: $offsprings \leftarrow \emptyset$
2: **for all** individuals in *parents* **do**
3: $ind \leftarrow$ getIndividual(*parents*)
4: **if** random() < mutationProbability **then**
5: $condition \leftarrow$ getRandomCondition(*ind*)
6: $newCondition \leftarrow$ newCondition(*ind*)
7: $newInd \leftarrow$ exchange(*ind*,*Condition*, *newCondition*)
8: $offsprings \leftarrow offsprings \cup newInd$
9: **end if**
10: **end for**
11: **return** *offsprings*

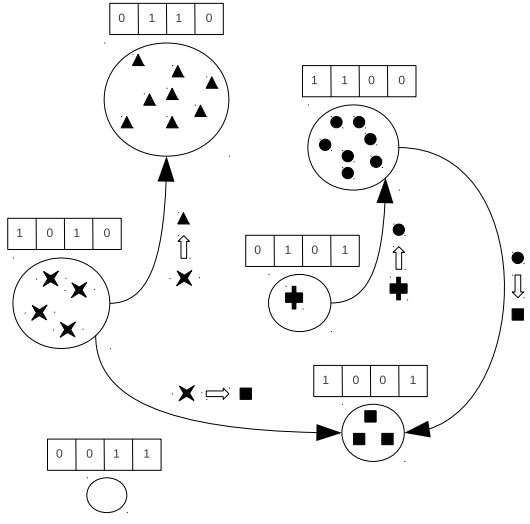

Fig. 5 Complete set of populations existing from the sample meta-data

Operator to change context. The aim of this second genetic operator is to change the context within a rule is described. As mentioned in Section 3.1, each individual is represented through a syntax-tree and a bit-set that represents the contextual attributes, i.e., those attributes where the exceptionalness of the rule is considered. Let us take into account the complete set of populations available from the sample meta-data depicted in Table 1, establishing that each pair of contextual attributes is a population including any individual described in this context. The genetic operator proposed here (Figure 5 graphically depicts the process) selects the worse individual from each population (those individuals having the lower fitness value) and move it to another population, so the contextual features are changed and the syntax-tree remains the same. Over generations, the more promising a combination of contextual attributes is, the higher is the number of association rules included in its population. Finally, it should be noted that there is no restriction about the population, so any selected rule could be moved to any population set.

3.4 G3P Algorithm

In this paper, a G3P algorithm (depicted in Figure 6) for mining exceptional association rules is proposed. The algorithm makes use of a G3P methodology to represent solutions to the problem in a flexible way through the use of a grammar, restricting the search space and working on any domain. This evolutionary algorithm considers that each association rule is obtained by means of a syntax-tree, and each individual comprises both a syntax-tree and a bit-set that represents the context where the association rule is defined.

The algorithm begins the creation of a population of size n thanks to the defined context-free grammar. Once this process is performed, the evolutionary process starts and a series of processes is carried out over generations. Thus, in

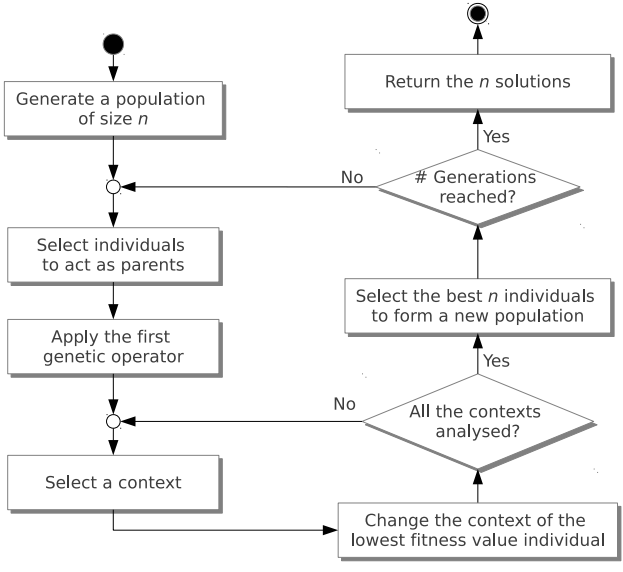

Fig. 6 Overview of the G3P algorithm for mining exceptional association rules

each generation, the algorithm randomly selects a subset of individuals from the population set by means of a tournament selector, i.e., two randomly selected individuals are compared an the one with the highest fitness value is selected as parent. From this set of parents, new individuals are obtained through the application of the genetic operators. In the first place, the first genetic operator is carried out and new individuals are obtained by replacing a random condition from each parent with a new condition randomly obtained. Then, the second genetic operator comes into play and the worse individual of each context is modified. Hence, this second genetic operator select a context, and then, from the set of individuals of this context, the operator selects the one with the worse fitness value. If more than one individual have this fitness value, all of them are taken and their context is modified. The process continues till all the contexts are analysed.

Finally, using the new generated individuals and the old-ones, those n individuals having a higher fitness function value are taken to form the new population. Noted that this final procedure is implemented to avoid an uncontrolled increase of the population size. In order to avoid repeated individuals, if two individuals are the same (considering both the syntax-tree and the context bit-set), then only one is considered. Only and only if the population of size n is not completed, repeated individuals could be used.

This set of previously described procedures is carried out over the generations. Once the algorithm reaches the maximum number of generations previously established by the data-miner, it returns the population set, comprising the best solutions discovered along the evolutionary process.

3.5 Computational complexity of MERG3P

In order to analyse the efficiency of the proposed model, we have carried out a computational complexity analysis. In this sense, we have analysed each of the main procedures of MERG3P separately: individual generator, evaluator, parent selector and genetic operator. Finally, we have determined the computational complexity of the whole MERG3P algorithm.

Firstly, the computational complexity of the generator procedure depends on the maximum derivation tree ($\mathcal{X}_{max\_der}$) and the number of individuals to be created ($\mathcal{X}_{num\_ind}$). In this regard, both variables determine the complexity of this procedure, represented as $\mathcal{O}(\mathcal{X}_{max\_der} \times \mathcal{X}_{num\_ind})$, its running time increases linearly with the maximum tree size and the number of individuals. As for the evaluator procedure, it depends on the number of individuals ($\mathcal{X}_{num\_ind}$), instances ($\mathcal{X}_{num\_ins}$) and attributes ($\mathcal{X}_{num\_att}$). Mathematically, this value is defined as $\mathcal{O}(\mathcal{X}_{num\_ins} \times \mathcal{X}_{num\_att} \times \mathcal{X}_{num\_ind})$. Additionally, the complexity order of the parent selector procedure is constant since it only requires a random value, i.e., $\mathcal{O}(1)$. Finally, the computational complexity of the genetic operator procedure depends on both the derivation tree size ($\mathcal{X}_{max\_der}$) and number of individuals ($\mathcal{X}_{num\_ind}$). In consequence, its complexity order is defined as $\mathcal{O}(\mathcal{X}_{max\_der} \times \mathcal{X}_{num\_ind})$.

Analysing the computing requirements for each procedure, it is stated that both $\mathcal{X}_{max\_der}$ and $\mathcal{X}_{num\_ind}$ are previously fixed, so they are considered as constants and their complexity order is $\mathcal{O}(1)$. Additionally, all the procedures are repeated as many times as the predefined number of generations, which is also a constant value predefined. Therefore, bearing in mind all these issues, the resultant computational complexity of MERG3P is stated as $\mathcal{O}(\mathcal{X}_{num\_ins} \times \mathcal{X}_{num\_att})$. Thus, the complexity of the proposed approach is linear with regard to the number of instances and the number of attributes.

4 Experimental analysis

The aim of this section is not only to demonstrate the behaviour of the proposed model, but also to use it in an interesting real application field. Thus, we propose the discovery and analysis of hidden exceptional relations between quality measures in the ARM field, an interesting issue in problems where the optimization of some quality measures is a dare. Thereby, a dataset comprising all the available values for each measure is used so that any exceptional relation between measures could be discovered.

In this section, we first describe the dataset used and how it was constructed. Then, we describe the parameter tuning. Finally, we describe the result of the experimental study including the performance testing on numerical, categorical and mixed datasets and demonstrating the flexibility of the proposed approach to be adapted to different problems.

4.1 Real application dataset

For the sake of mining and analysing feasible exceptional relationships between quality measures in ARM, we constructed a benchmark dataset for which we know

that it comprises any available value for seven quality measures. Thereby, it is of great interest to the ARM community since it enables any relationship between quality measures to be obtained. Existing datasets could comprise unexplored values for different quality measures, depending on the patterns distribution.

This benchmark dataset [1], was constructed by considering 100 sample transactions, i.e., $\mathcal{T} = \{t_1, t_2, t_3, ..., t_{100}\}$, which are accordingly distributed along the whole search space. From this sample dataset, the Support of any sample k-itemset could take a value from the set $\mathcal{P} = \{0/100, 1/100, ..., 100/100\}$ for any $k \geq 1$. Thus, considering a sample association rule R of the type $X \rightarrow Y$, and knowing the Support of the antecedent ($Support(X)$), consequent ($Support(Y)$) and entire rule ($Support(R)$), then it is possible to obtain the value of any quality measure, as it is described below.

In this dataset, seven quality measures are considered: Support, Confidence, Lift, Conviction, Leverage, Certainty Factor and IS. Beginning with the Support measure, it is formally defined in Equation 3 as the proportion of the number of transactions $t \subseteq \mathcal{T}$ that include X and Y in a dataset.

$$Support(R) = \frac{|\{X \cup Y \subseteq \mathcal{T}\}|}{|\mathcal{T}|} \tag{3}$$

On the other hand, the Confidence of an association rule is defined in Equation 4 as the proportion of the number of transactions that include X and Y among all the transactions that comprise X.

$$Confidence(R) = \frac{Support(R)}{Support(X)} \tag{4}$$

The Lift measure (see Equation 5) establishes how many times X and Y occur together more often than would be expected if they were statistically independent.

$$lift(R) = \frac{Support(R)}{Support(X) * Support(Y)} \tag{5}$$

The Conviction measure (see Equation 6) represents the ratio of the expected frequency that X occurs without Y, considering X and Y statistically independent sets, divided by the observed frequency of incorrect predictions.

$$conviction(R) = \frac{Support(X) - Support(X) * Support(Y)}{Support(X) - Support(R)} \tag{6}$$

In a similar way to the Lift measure, Leverage (see Equation 7) calculates the proportion of additional cases covered by both X and Y above those expected if X and Y were independent of each other.

$$leverage(R) = Confidence(R) - Support(X) * Support(Y) \tag{7}$$

The Certainty Factor (see Equation 8), also known CF, is interpreted as a measure of variation of the probability that Y is in a transaction when we consider only those transactions where X is. More specifically, a positive certainty factor

[1] The dataset and the data generator can be reached at `http://www.uco.es/grupos/kdis/kdiswiki/index.php/Exceptional_ARM`

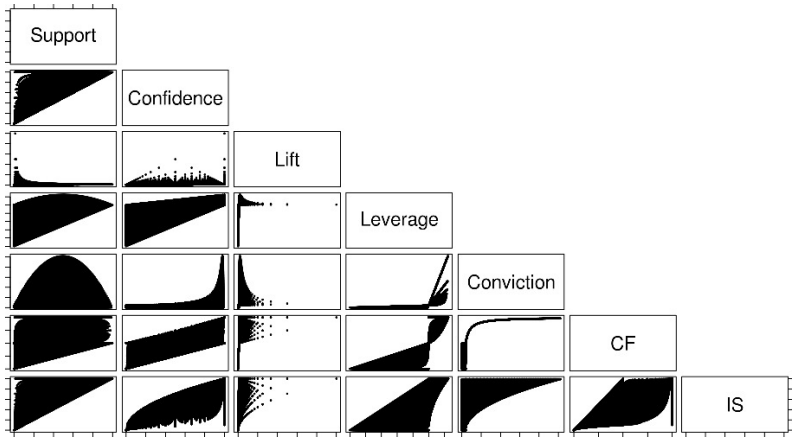

Fig. 7 Distribution of the database represented in scatter plots

measures the decrease of the probability that Y is not in a transaction, given that X is.

$$CF(R) = \begin{cases} \dfrac{Confidence(R) - Support(Y)}{1 - Support(Y)} & \text{if } Confidence(R) > Support(Y) \\ \dfrac{Confidence(R) - Support(Y)}{Support(Y)} & \text{if } Confidence(R) < Support(Y) \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

Finally, the IS measure (see Equation 9) was obtained as the product of Support and interest factor, representing the ratio between the joint probability of X and Y with respect to their expected probability under the independence assumption.

$$IS(R) = \frac{Support(R)}{\sqrt{Support(X) * Support(Y)}} \tag{9}$$

Using the seven quality measures aforementioned and 100 sample transactions, then we form a new dataset where each attribute is one of the seven quality measures and each transaction is an available value for each measure. Thus, we consider all the value combinations for $Support(X)$, $Support(Y)$ and $Support(R)$ considering that any value from the set $\mathcal{P} = \{0/100,\ 1/100,\ ...,\ 100/100\}$ could be taken. Additionally, it is necessary to consider that $Support(R) \leq Support(X)$ and $Support(R) \leq Support(Y)$. In this regard, taking into consideration a Support value of a rule $Support(R) = 23/100$, then $Support(X) = \{23/100, ..., 100/100\}$ and $Support(C) = \{23/100, ..., 100/100\}$.

To sum up, the constructed dataset includes 7 attributes (one per quality measure) and 348,511 instances, i.e., all the available combinations of values for a set of 100 transactions. Hence, using one of these quality measures, it is impossible to obtain a value that does not fit an instance of this dataset. For a better understanding, Figure 7 depicts the distribution of the instances of this dataset in a scatter plot, showing the distribution per pair of measures.

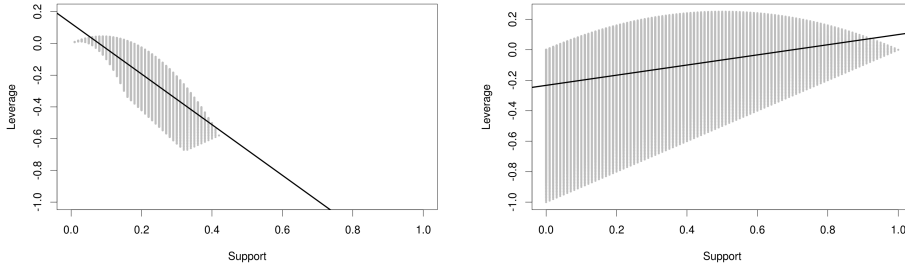

Fig. 8 Example of the scatter plot and distribution in the context of Support and Leverage of the rule **IF** IS IN $[0.2, 0.4]$ **THEN** CF IN $[-0.7, 0.5]$ (left) and its complement (right)

4.2 Parameter tuning

As any evolutionary approach, the proposed model should be configured with a set of adjustable parameters. All these parameters require previous study to determine those considered optimal [2], i.e., those that allow us to obtain the best global results. It is worth mentioning that no single combination of parameter values performs better for any datasets, and sometimes, it depends on the problem under study. In this regard, the best results for our approach are obtained with a population size of 25 individuals obtained using a context-free grammar with a maximum derivation size of 24. As for the number of generations, solutions are hardly improved after 150 generations so the evolutionary process should be carried out using this value. In this process, new individuals are obtained with a probability of 0.95 for the first genetic operator, since the second proposed genetic operator works on the worse individual of each population, not requiring a probability. The remaining values of the quality thresholds are set as follows: 0.8 for the Confidence, 0.05 and 0.4 for the minimum and maximum Support thresholds, respectively.

All the experiments were performed on an Intel Core i7 machine with 12GB main memory and running CentOS 5.4. Finally, the proposal presented in this paper was coded using JCLEC[3] [17], a Java library for evolutionary computation.

4.3 Experimental results

In this section, we describe some exceptional rare association rules discovered by means of the proposed algorithm. Thus, this algorithm was run by using the proposed dataset that includes quality measures. The best exceptional rare rule is defined in the context of the Support and Leverage measures and represents the rule **IF** IS IN $[0.2, 0.4]$ **THEN** CF IN $[-0.7, 0.5]$. This association rule (see Figure 8) appears in 22.99% of the instances, indicating that, if the IS measure has

[2] A sensitivity analysis was carried out. The results and statistical analysis could be reached at `http://www.uco.es/grupos/kdis/kdiswiki/index.php/Exceptional_ARM`

[3] JCLEC is available for download (`http://jclec.sourceforge.net`).

a value between 0.2 and 0.4, then it is highly probable that the CF measure defines a value between -0.7 and 0.5. More specifically, this association rule provides an accuracy of 92.88%, i.e., it is extremely highly probable that if the antecedent is satisfied, then the consequent is also satisfied. As shown Figure 8, the rule describes a set of instances which distribution is correlated with $\rho = 0.3954$. On the contrary, the distribution of its complement is correlated with $\rho = -0.7890$, i.e., the association rule discovered is exceptional with respect to the context where it is defined. Thus, despite the fact that Support and Leverage seems to be positively correlated, the algorithm discover a completely hidden knowledge, stating that this correlation is completely different if we consider that the IS measures is in $[0.2, 0.4]$ and the Certainty Factor is in $[-0.7, 0.5]$.

Now, a more frequent and reliable association rule is analysed. This rule describes that if the IS measure provides a value between 0.5 and 0.9, then it is highly probable that the Lift value is between 0.1 and 41.0. This rule is defined in the context of Support and CF, where the instances are positively correlated as depicted in Figure 9. However, if we consider the association rule IS IN $[0.5, 0.9]$ **THEN** $Lift$ IN $[0.1, 41.0]$, this correlation is negative and completely different. This rule identifies the 31.1198% of the instances in the dataset and provides an accuracy of 99.99%. In fact, the context of Support and CF appears as a good context descriptor, including a high number of exceptional rare association rules so, as mentioned in Section 3.3, this context represents a high quality population.

Rule **IF** $Confidence$ IN $[0.7, 0.8]$ **THEN** IS IN $[0.4, 0.8]$ (see Figure 10) is other exceptional rare rule discovered in the context of Support and CF . This rule describes instances with a correlation completely opposite to its complement, i.e., $\rho = -0.5690$. The depicted rule covers 10.25% of the instances and provides a reliability of 85.78%. Thus, this infrequent rule –covering 35,733 instances out of 348,511– describes a negative correlation when the Confidence value is between 0.7 and 0.8 and the IS measure calculates a value that is between 0.4 and 0.8. Finally, notice that this is a highly reliable rule, describing that, in the 85.78% of the instances that the antecedent is satisfied, then the consequent is also satisfied.

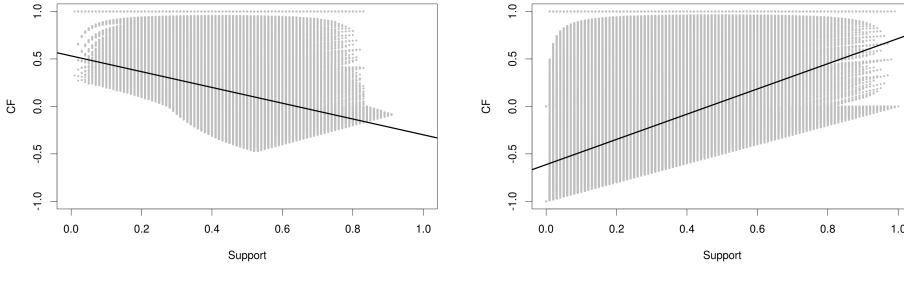

Fig. 9 Example of the scatter plot and distribution in the context of Support and Certainty Factor of the rule **IF** IS IN $[0.5, 0.9]$ **THEN** $Lift$ IN $[0.1, 41.0]$ (left) and its complement (right)

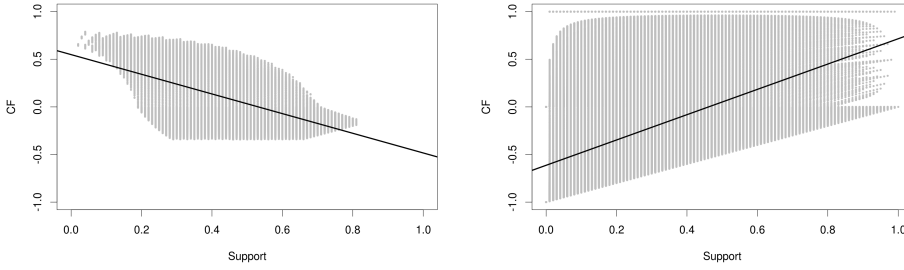

Fig. 10 Example of the scatter plot and distribution in the context of Support and Certainty Factor of the rule **IF** $Confidence$ IN $[0.7, 0.8]$ **THEN** IS IN $[0.4, 0.8]$ (left) and its complement (right)

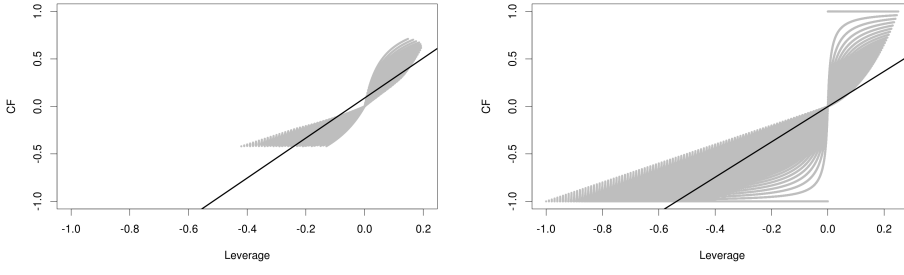

Fig. 11 Example of the scatter plot and distribution in the context of Leverage and Certainty Factor of the rule **IF** $Confidence$ IN $[0.6, 0.8]$ **THEN** $Support$ IN $[0.2, 0.9]$ (left) and its complement (right)

At this point, let us consider a rule having a low fitness value to demonstrate the difference with respect to the aforementioned rules. The rule **IF** $Confidence$ IN $[0.6, 0.8]$ **THEN** $Support$ IN $[0.2, 0.9]$, described in the context of Leverage and CF, appears as a rule with a Support of 14.06% and a Confidence of 89.26%. As it is graphically depicted in Figure 11, the correlation of this association rule is similar to the correlation of its complement. In fact, the rule calculates a correlation coefficient $\rho = 0.9249$, whereas the context obtains $\rho = 0.7687$. Thus, we demonstrate the difference between high quality exceptional rare association rules and poor quality rules. Notice that both of them cover a small set of instances and obtain high confidence values, but the difference is the distribution of the covered instances.

Finally, we show the possibility of discovering sets of instances that does not represent an association rule though their distribution is completely opposite to their context. For example, it is possible to configure the algorithm to be able to discover rules of the type **IF** $Confidence$ IN $[0.6, 0.8]$ **THEN** $Confidence$ IN

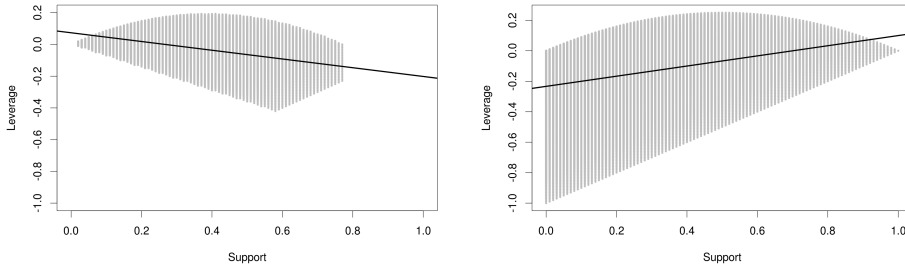

Fig. 12 Example of the scatter plot and distribution in the context of Support and Leverage of the set of instances satisfied by $Confidence$ IN $[0.6, 0.8]$ (left) and its complement (right)

$[0.6, 0.8]$. In this regard, the obtained rule is not an association rule, since the antecedent and consequent are not disjoint sets. However, this set of instances could be interesting to be mined since its distribution ($\rho = -0.4057$) is different to the distribution of its complement ($\rho = 0.3197$) as shown Figure 12. This set of instances covers 15.74% of the instances and indicates that the distribution of the Support and Leverage, which is positively correlated, behaves negatively correlated if the Confidence value is between 0.6 and 0.8.

Let us now consider a set of datasets[4] including different types of domains and different number of both attributes and instances. Thus, we study and demonstrate the capability of the proposed model to be used on a variety of problems. Table 2 depicts the set of datasets used in this experiment. Since the goal is to show the strength of the proposed model to be used in different domains, only the best rule discovered for each dataset is depicted.

Table 2 Datasets and their main characteristics

Dataset	**#Instances**	**#Attributes**	**Type of Attributes**
Automobile	392	8	Numerical
German Credit	1000	21	Categorical, Numerical
Soybean	683	36	Categorical

The best exceptional rare rule obtained from the *Automobile* dataset is defined in the context of the *milespergallon* and *horsepower*, and represents the rule **IF** *displacement* IN $[153.3, 206.8]$ **THEN** *weight* IN $[2129.5, 3784.4]$, obtaining a Confidence value of 96.66% and covering 7.39% of the instances. The correlation of this association rule is $\rho = 0.1373$, whereas the context obtains $\rho = -0.7931$. Thus, at first glance, the increment of the *horsepower* implies a decrement of the *milespergallon*. As for the *GermanCredit* dataset, the best rule is **IF** $InstallmentPlans = A141$ **THEN** $DurationMonth$ IN $[6.9, 57.7]$, which is

[4] All the datasets are publicly available for download from the UCI machine learning repository (`http://archive.ics.uci.edu/ml/datasets/`).

defined in the context $CreditAmount$ and $ForeignWorker$. This rule describes that if the antecedent is satisfied, then the consequent is satisfied in 90.64% of the situations, and also covering 12.60% of the instances. The fitness function value obtained with this solution is 0.2761, i.e., the difference between its coefficient $\rho = 0.1927$ and its complement $\rho = -0.0834$. Finally, a completely categorical dataset is used (*soybean* dataset). The best rule obtained is of high interest, since its coefficient $\rho = 0.9532$ is completely different to the coefficent of its complement $\rho = -0.1793$. The rule **IF** $seed = brown - w/blk - specks$ **THEN** $precip = gt - norm$ is defined in the context of *leaves* and *mycelium*, covering 7.61% of the instances and providing a reliability of 91.22%.

Finally, to conclude this experimental analysis, we adapt the proposed approach to produce only specific association rules. Thus, it is possible to use a grammar that enables users to find exceptional behaviour by using a specific target attribute in the consequent, e.g., $hail = yes$ in the *soybean* dataset. In this sense, the production rule $P = \{$Consequent = 'Consequent_attribute', '=', 'Consequent_value'; $\}$ is replaced by the production rule $P = \{$Consequent = 'hail', '=', 'yes'; $\}$, so any rule encoded by the grammar includes both the specific attribute and value in its consequent. Table 3 shows the results obtained when running the modified algorithm. Notice that, since we are focusing on a specific attribute, the aforementioned parameters were briefly modified, so a 0.8 Confidence threshold is now used. Finally, Figure 13 shows a bloxplot including the average values obtained from the rules obtained. It depicts that the set of rules are highly reliable, having a low deviation with regard to the mean. Furthermore, the correlation coefficients are quite different, denoting a high degree of exceptionalness between a subgroup and its opposite.

Table 3 Set of rules discovered by the modified algorithm on the *soybean* dataset

Context	Rule	Support	Confidence	Coefficients
leafspots-halo & fruiting-bodies	**IF** mycelium = firm-and-dry **THEN** hail = yes	16.25%	82.22%	$\rho_{rule} = 0.7710$ $\rho_{\overline{rule}} = -0.5379$
leafspots-halo & mycelium	**IF** fruiting-bodies = brown **THEN** hail = yes	10.10%	83.13%	$\rho_{rule} = 0.8399$ $\rho_{\overline{rule}} = 0.0736$
seed-discolor & roots	**IF** leafspot-size = gt-1/8 **THEN** hail = yes	39.38%	82.26%	$\rho_{rule} = -0.0126$ $\rho_{\overline{rule}} = 0.6802$
leafspots-halo & external-decay	**IF** severity = severe **THEN** hail = yes	5.27%	80.00%	$\rho_{rule} = -0.3707$ $\rho_{\overline{rule}} = 0.2433$
plant-growth & mycelium	**IF** seed = colored **THEN** hail = yes	9.66%	88.00%	$\rho_{rule} = -0.3339$ $\rho_{\overline{rule}} = 0.2349$
external-decay & fruit-spots	**IF** seed-tmt = none **THEN** hail = yes	35.87%	80.32%	$\rho_{rule} = -0.0085$ $\rho_{\overline{rule}} = -0.4257$

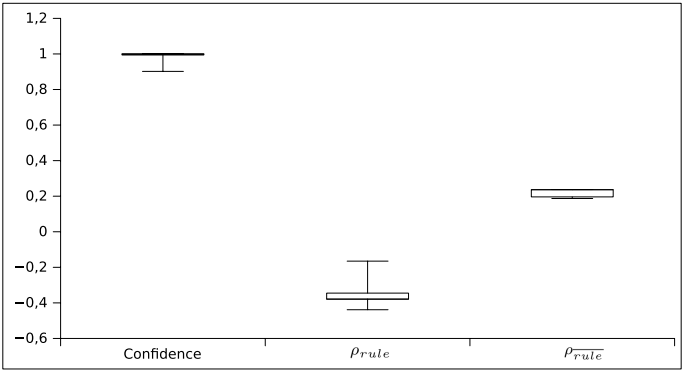

Fig. 13 Boxplot for the average values of different metrics of Distribution of Table 3

Additionally, it is also possible to restrict the subgroups where the user wants to obtain an exceptional description. Thus, the algorithm could be also adapted to previously fix the context, e.g., *precip* and *hail*, and then discovering association rules within these contexts (see Table 4). Similarly, Figure 14 shows a bloxplot including the average values obtained from the rules obtained, which shows that the set of rules are highly reliable, having a low deviation with regard to the mean. Finally, the correlation coefficient values are accordingly distributed along

Table 4 Set of rules discovered on the context *precip* and *hail* and using the *soybean* dataset

Context	Rule	Support	Confidence	Coefficients
precip & hail	**IF** leafspot-size = lt-1/8 **THEN** leaf-mild = absent	6.73%	90.19%	$\rho_{rule} = -0.4385$ $\rho_{\overline{rule}} = 0.2375$
precip & hail	**IF** leafspot-size = lt-1/8 **THEN** sclerotia = absent	7.46%	100.00%	$\rho_{rule} = -0.3789$ $\rho_{\overline{rule}} = 0.2365$
precip & hail	**IF** leafspot-size = lt-1/8 **THEN** mycelium = absent	7.46%	100.00%	$\rho_{rule} = -0.3789$ $\rho_{\overline{rule}} = 0.2365$
precip & hail	**IF** leafspot-size = lt-1/8 **THEN** leaves = abnorm	7.46%	100.00%	$\rho_{rule} = -0.3789$ $\rho_{\overline{rule}} = 0.2365$
precip & hail	**IF** leaf-mild = present **THEN** leaves = abnorm	6.58%	100.00%	$\rho_{rule} = -0.3739$ $\rho_{\overline{rule}} = 0.1924$
precip & hail	**IF** roots = rotted **THEN** leaves = abnorm	12.44%	98.83%	$\rho_{rule} = -0.3162$ $\rho_{\overline{rule}} = 0.1870$
precip & hail	**IF** fruit-spots = brown-w/blk-specks **THEN** stem = abnorm	8.34%	100.00%	$\rho_{rule} = -0.1652$ $\rho_{\overline{rule}} = 0.1998$

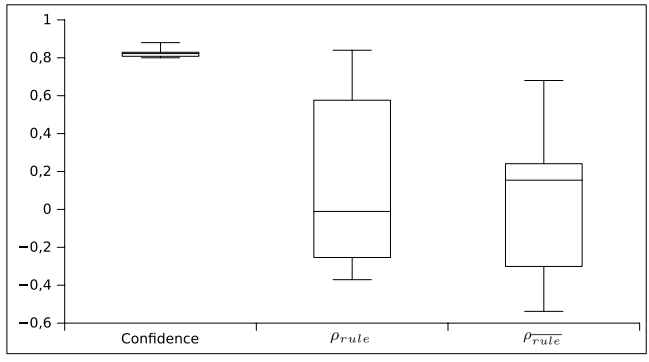

Fig. 14 Boxplot for the average values of different metrics of Distribution of Table 4

Table 5 Set of rules discovered for the two attributes of interest, i.e., lot size and sales price, for the benchmark dataset proposed by *Leman et al.*

Context	Rule	Support	Confidence	Coefficients
lot size & sales price	**IF** recroom = yes **THEN** driveway = yes	14.84%	91.83%	$\rho_{rule} = 0.3042$ $\rho_{\overline{rule}} = 0.5476$
lot size & sales price	**IF** bathrms IN [1,3 2,1] **THEN** recroom = yes	6.41%	26.31%	$\rho_{rule} = 0.0548$ $\rho_{\overline{rule}} = 0.5476$
lot size & sales price	**IF** driveway = no **THEN** bathrms IN [1,7 3,1]	2.93%	20.77%	$\rho_{rule} = 0.3143$ $\rho_{\overline{rule}} = 0.5476$

the whole space of values, so the proposed model is able to discover both positive and negative correlation.

Finally, we have considered the benchmark dataset proposed by *Leman et al.* [11], which contains information on 546 houses that were sold in Windsor, Canada in the summer of 1987. The information for each house includes the two attributes of interest, lot size and sales price. An additional 10 attributes are available to define candidate subgroups, including the number of bedrooms and bathrooms and whether the house is located at a desirable location. Table 5 shows the exceptional subgroups discovered by MERG3P, which are similar to those discovered by *Leman et al.*. It should be noted that, whereas the proposal proposed by *Leman et al.* denotes patterns in the form of $bathrms \geq 2$, our proposal specifies a range of values, i.e., $bathrms\ IN\ [1, 73, 1]$. Analysing the subgroups, the group formed by **IF** $recroom = yes$ **THEN** $driveway =$ yes was also discovered by *Leman et al.* [11] in the form of $recroom = yes \wedge driveway = yes$.

As an additional feature, our model is able to discover exceptional subgroups without requiring any attribute of interest, i.e., the two attributes that determine the context where the model is defined. Thus, MERG3P looks for exceptional subgroups in the whole search space, discovering the rules depicted in Table 6. This is a interesting feature that enables to discover any exceptional subgroup without a previous knowledge of the data under study. As shown, MERG3P discovers

Table 6 Set of rules discovered by MERG3P without considering any attribute of interest for the benchmark dataset proposed by *Leman et al.*

Context	Rule	Support	Confidence	Coefficients
price & bedrooms	**IF** gashw = yes **THEN** fullbase = no	3.11%	65.38%	$\rho_{rule} = -0.5248$ $\rho_{\overline{rule}} = 0.3603$
lot size & bedrooms	**IF** bathrms IN [2,3 3,6] **THEN** prefarea = no	1.46%	72.72%	$\rho_{rule} = 0.7059$ $\rho_{\overline{rule}} = -0.1041$
lot size & recroom	**IF** bathrms IN [2,3 3,6] **THEN** fullbase = no	1.09%	54.54%	$\rho_{rule} = -0.6388$ $\rho_{\overline{rule}} = 0.1583$
gashw & garagepl	**IF** price IN [117345,7 160297,2] **THEN** airco = no	1.46%	28.57%	$\rho_{rule} = 0.7071$ $\rho_{\overline{rule}} = -0.0548$

very interesting subgroups. For instance, considering the price and the number of bedrooms as attributes of interest, it discovers the rule **IF** $gashw = yes$ **THEN** $fullbase = no$, which produces a variation on the correlation from $\rho = 0.3603$ to $\rho = -0.5248$, which is a high deviation.

The results obtained in this experimental stage demonstrate the strength of the proposed model for mining exceptional association rules. This model enables the discovery of any rule in any context (attributes of interest). The results demonstrate the ability of this model for mining and describing hidden relations that behave completely different than they seem to do. Thus, we discover exceptional rare relations between quality measures which distribution is completely different to the context where they are defined. Considering the proposed benchmark dataset in the ARM field, Support and Leverage seem to be positively correlated, but in fact, they are negatively correlated if IS is in [0.2, 0.4] and the Certainty Factor is in [-0.7, 0.5]. This exceptional behaviour is of interest in fields such as multi-objective optimization, where it is necessary to attain an optimal trade-off between conflicting objectives. Thus, CF values increase when the Support values also increase, since they seem to be positively correlated. In this regard, these two measures could not be considered as conflicting objectives and searching for a set of rules that satisfied these two measures is not an option. Notice that the fact of maximizing the Support measure, implies a maximization of the CF measure. However, in this experimental analysis, we have demonstrated that this study could be of high interest if we consider the instances with a Confidence value in the interval [0.7, 0.8] and an IS value in the interval [0.4, 0.8]. For this set of instances, the fact of maximizing the Support does not imply a maximization of the CF, and a multi-objective optimization could be of interest. Finally, the capability of the proposed model to be used on a variety of problems was demonstrated, showing how the algorithm could be easily modified to extract only a specific type of rules. For fairness, we should state that the proposed model may leave unexplored a section of solutions. It is possible because of the evolutionary process and its predefined parameter values. Therefore, it is quite important to tune these parameters in a right way to achieve as good results as possible.

5 Concluding remarks

The problem of mining exceptional subgroups in data is giving rise to a increasing interest in the discovery of subgroups where the model fitted to a specific subgroup is significantly different to its complement. This information could be of great interest in many fields to identify specific groups of features that form subgroups whose distribution is completely different from the rest. Starting from this problem, we have proposed the concept of exceptional relationship mining, a special case of EMM and ARM. This task for mining exceptional relationships enables relationships between patterns within exceptional subgroups to be obtained.

In this paper, we have formally presented the new concept of exceptional relationship mining, which could be approached from either an exhaustive search or a specific heuristic. For the sake of applying this new concept, we have considered the analysis of exceptional relations between quality measures in the field of association rules. Since this quality measures are defined in a continuous domain, and exhaustive search approaches are not appropriated for this type of domains, we consider the use of an evolutionary methodology based on G3P, named MERG3P.

Discovering exceptional relationships between quality measures in the ARM field revealed that some quality measures that seemed to be positively correlated appear as negatively correlated under some restrictions. Thus, the fact of maximizing a quality measure does not imply a maximization of other positively correlated measure, and a deep analysis is required to extract hidden exceptional behaviour. In relation to this it might be interesting to use ERM and our approach for benchmark construction to study relationship between many other existing measures of interest [3] for pattern mining or study the behaviour of other metrics [5].

Finally, we have considered a benchmark dataset in the EMM field. Results have revealed that our proposal is able to discover similar subgroups as the existing approaches. As an additional feature, MERG3P is able to discover exceptional subgroups without requiring to predefine attribute of interest where the subgroups are mined. Instead, our algorithm is capable of discovering any exceptional subgroup on any context.

MERG3P is just one approach for ERM. It would be interesting to see how far we can go with other heuristic search and exhaustive search approaches for pattern mining in case of the ERM problem formulation.

Acknowledgements This research was supported by the Spanish Ministry of Economy and Competitiveness, project TIN-2014-55252-P, and by FEDER funds. This research was partly supported by STW CAPA project. Finally, this research was also supported by the Spanish Ministry of Education under FPU grant AP2010-0041.

References

1. R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. In *Proceedings of 20th International Conference on Very Large Data Bases*, VLDB'94, pages 487–499, Santiago de Chile, Chile, 1994. Morgan Kaufmann.
2. F. Berzal, I. Blanco, D. Sánchez, and M. A. Vila. Measuring the accuracy and interest of association rules: A new framework. *Intelligent Data Analysis*, 6(3):221–235, 2002.

[5] The dataset and the data generator can be reached at `http://www.uco.es/grupos/kdis/kdiswiki/index.php/Exceptional_ARM`

3. Liqiang Geng and Howard J. Hamilton. Interestingness measures for data mining: A survey. *ACM Comput. Surv.*, 38(3), 2006.
4. J. Han, J. Pei, Y. Yin, and R. Mao. Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data Mining and Knowledge Discovery*, 8:53–87, 2004.
5. R. I. Hoai, N. X. Whigham, P. A. Shan, Y. O'neill, and M. McKay. Grammar-based genetic programming: A survey. *Genetic Programming and Evolvable Machines*, 11(3), 2010.
6. Szymon Jaroszewicz. Minimum variance associations — discovering relationships in numerical data. In *The Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, pages 172–183, Osaka, Japan, 2008.
7. Y. S. Koh and N. Rountree. *Rare Association Rule Mining and Knowledge Discovery: Technologies for Infrequent and Critical Event Detection*. Information Science Reference, Hershey, New York, 2010.
8. J. R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. The MIT Press, December 1992.
9. M. Leeuwen. Maximal exceptions with minimal descriptions. *Data Mining Knowledge Discovery*, 21(2):259–276, 2010.
10. Matthijs Leeuwen and Arno Knobbe. Diverse subgroup set discovery. *Data Mining Knowledge Discovery*, 25(2):208–242, 2012.
11. D. Leman, A. Feelders, and A. J. Knobbe. Exceptional model mining. In *Proceedings of the European Conference in Machine Learning and Knowledge Discovery in Databases*, volume 5212 of *ECML/PKDD 2008*, pages 1–16, Antwerp, Belgium, 2008. Springer.
12. J. M. Luna, J. R. Romero, and S. Ventura. Design and behavior study of a grammar-guided genetic programming algorithm for mining association rules. *Knowledge and Information Systems*, 32(1):53–76, 2012.
13. J. M. Luna, J. R. Romero, and S. Ventura. On the adaptability of G3PARM to the extraction of rare association rules. *Knowledge and Information Systems*, 38(2):391–418, 2014.
14. C. Romero, J. M. Luna, J. R. Romero, and S. Ventura. Mining Rare Association Rules from e-Learning Data. In *Proceedings of the 3rd International Conference on Educational Data Mining, EDM 2010*, pages 171–180, 2010.
15. C. Romero, J. M. Luna, J. R. Romero, and S. Ventura. RM-Tool: A framework for discovering and evaluating association rules. *Advances in Engineering Software*, 42(8):566–576, 2011.
16. A. Salam and M. Khayal. Mining top-k frequent patterns without minimum support threshold. *Knowledge and Information Systems*, 30:57–86, 2012.
17. S. Ventura, C. Romero, A. Zafra, J. A. Delgado, and C. Hervás. JCLEC: A java framework for evolutionary computation. *Soft Computing*, 12(4):381–392, 2008.
18. Geoffrey I. Webb. Discovering associations with numeric variables. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '01, pages 383–388, New York, NY, USA, 2001. ACM.
19. A. Zafra, M. Pechenizkiy, and S. Ventura. ReliefF-MI: An extension of ReliefF to multiple instance learning. *Neurocomputing*, 75(1):210–218, 2012.