A genetic algorithm-based framework for mining quantitative association rules without specifying minimum support and minimum confidence

F. Moslehi and A. Haeri*

School of Industrial Engineering, Iran University of Science & Technology, Tehran, Iran.

Abstract. Discovering association rules is a useful and common technique for data mining, in which relations and co-dependencies of datasets are shown. One of the most important challenges of data mining is to discover the rules of continuous numerical datasets. Furthermore, another restriction imposed by algorithms in this area is the need to determine the minimum threshold for the support and confidence criteria. In this paper, a multi-objective algorithm for mining quantitative association rules is proposed. The procedure is based on the genetic algorithm, and there is no need to determine the extent of the threshold for the support and confidence criteria. By proposing a multi-criteria method, useful and attractive rules and the most suitable numerical intervals are discovered, without the need to discretize numerical values and determine the minimum support threshold and minimum confidence threshold. Different criteria are considered to determine appropriate rules. In this algorithm, selected rules are extracted based on confidence, interestingness, and cosine². The results obtained from real-world datasets demonstrate the effectiveness of the proposed approach. The algorithm is used to examine three datasets, and the results show the superior performance of the proposed algorithm compared to similar algorithms.

© 2020 Sharif University of Technology. All rights reserved.

1. Introduction

Association Rule Mining (ARM) is one of the most important techniques of Data Mining (DM) [1]. It should be noted that DM consists of a set of computational techniques that are applied to discover knowledge, hidden patterns, and rules obtained from data in various sciences [2]. Among several techniques for discovering meaningful knowledge from datasets, ARM has an important role [3]. ARM is a descriptive approach that can extract interesting patterns among a large collection of data. Mining or generating useful rules from the dataset is done by ARM techniques [4]. The technique used for extracting association rules was first proposed to analyze the customers' market basket, expressing the correlation among the features.

Many algorithms have been developed to identify association rules from a database. However, most of these algorithms are designed to deal with categorical attributes. However, the data in real-world databases consist of both categorical (e.g., education) and quantitative attributes (e.g., weight). The difficulty in dealing with both quantitative and categorical attributes and the problem of working with large databases are the reasons that make it complex to extract rules form both quantitative and categorical data [5]. The majority of the proposed methods produce numerous laws along

*. Corresponding author. Tel.: +98 21 73225019;
Fax: +98 21 73225098
E-mail addresses: moslehi_fateme@ind.iust.ac.ir (F. Moslehi); ab.haeri@gmail.com (A. Haeri)

with a large number of fields, and they cannot mine the association rules. These methods apply the idea of dividing a range of values of all numerical fields into a series of distances, called discretization [6]; therefore, the information will be lost. Moreover, most of the proposed methods need to determine the threshold limit value such as minimum support and minimum confidence before the DM process begins. Many developed rule mining algorithms need to specify the minimum support threshold and the minimum confidence threshold that cause these methods to be dependent on datasets, and they must be run several times [7]; furthermore, it is difficult to define them for each database.

The appropriate rules for the Quantitative Association Rules (QAR) process are selected according to several criteria and objects. Multi-Objective Programming (MOP) has been one of the most common research areas [8]. Due to the growing importance of MOP, a wide diversity of Multi-Objective Evolutionary Algorithms (MOEAs) has been developed in different fields [9–11]. In this study, the MOEA is used to optimize the QAR problem. This approach can lead to an optimal solution to QAR problem.

This study aims to identify both quantitative and categorical association rules using a Multi-Objective Genetic Algorithm (MOGA) approach. In the proposed approach, there is no need to determine the minimum support threshold and the minimum confidence threshold, and rules are extracted without any discretization of a dataset. Three criteria including confidence, cosine^2, and interestingness are used to evaluate the generated rules. The criterion of cosine^2 in this method has been considered for selecting useful rules in the fitness function for the first time. Since this criterion is considered to evaluate the resemblance degree of antecedent and consequent and the fitness function chooses rules based on this criterion, the rules resulting from this method enjoy a higher confidence level than other proposed methods. The main objective of this paper is to develop a novel framework for Quantitative Association Rule Mining (QARM) using a Genetic Algorithm (GA), where there is no need to determine the extent of the threshold for the support and confidence criteria. To do so, the criterion cosine^2 is applied with the aim of increasing the number of extracted rules and improving the support percentage and the confidence measure.

The present research aims to answer the following questions:

- Will the clustering of records and the implementation of the QAR algorithm improve the efficiency of QAR algorithm?
- Does the use of the confidence, cosine^2, and interestingness criteria in the fitness function of an

evolutionary algorithm (EA) lead to the extraction of attractive and useful rules from a numerical dataset?

- Can the adoption of a multi-objective fitness function increase the percentage of support and confidence criteria of extracted rules?

This paper is organized as follows. Section 2 presents a brief overview of MOEAs for mining quantitative association rules. Section 3 introduces the proposed model. Then, Section 4 describes experimental results. Finally, Section 5 draws the conclusions.

1.1. Related works
In recent years, the frequent association rule has been one of the interesting research topics in the field of descriptive DM. Therefore, a large number of efficient algorithms have been proposed by researchers in this field. Kaya and Alhajj [12] developed a hybrid approach based on GA and clustering to find fuzzy sets used for mining meaningful association rules. In addition, they [13] proposed a cluster-based GA with two fitness functions related to minimum support and confidence. Alatas and Akin [14] proposed a novel EA based on rough particle swarm optimization algorithm to solve the problem of numerical ARM. Ayubi et al. [15] developed a novel method for mining general association rules from tabular data based on GA. This method was improved in terms of time complexity and memory requirements. In this method, candidate general item sets were stored in a tree structure. Nasiri et al. [16] in 2010 applied simulated annealing as a multi-objective numerical ARM. Qodmanan et al. [17] proposed a method based on GA without taking the minimum support and confidence into account. The best rules indicate that a perfect correlation between support and confidence has been drawn out from a dataset by their algorithm. A new algorithm based on Ant colony optimization was proposed by Moslehi et al. [7]. This algorithm was suggested for mining numerical rules without determining minimum support. Djenouri et al. [18] proposed an efficient hybrid algorithm based on bees swarm optimization and Tabu search to extract rules. Agbehadji et al. [19] applied a meta-heuristic algorithm to find interesting rules in numerical datasets. Wolf search algorithm technique was applied in their proposed approach. Can and Alatas [20], for the first time, applied Gravitational Search Algorithm (GSA) to search for a quantitative association rule. The results of their method have demonstrated that GSA is an effective technique for finding rules in numerical datasets.

A review of the literature reveals that there are many researchers who have used EAs, especially Gas [21], for finding QARs from datasets with quantitative values [22,23]. GENetic Association Rules

(GENAR) is an EA-based approach, as proposed by Mata et al. [24]. The fitness function of this algorithm was based on support, confidence, and cover measures. They [25] also presented an EA for extracting the frequent item sets in numerical databases. Their method was developed based on GA. They used an elastic strategy for the selection phase. Alatas and Akin [22] developed a GA to mine positive and negative association quantitative rules. The tournament selection method and uniform crossover were used in this algorithm. In their method, there is no need to determine the minimum support and the minimum confidence. QuantMiner was proposed by Salleb-Aouissi et al. [26]. They applied GA for mining quantitative association rules and defined the fitness function based on gain measure to select appropriate rules. Taboada et al. [27] developed a hybrid method based on graph theory and GA called Genetic Network Programming (GNP). The purpose of this algorithm is to extract rules among numerical datasets. Yan et al. [28] presented an evolutionary tool Evolutionary Association Rules Mining with Genetic Algorithm (EAR-MGA) for finding association rules in quantitative numerical databases. The proposed GA performs an algorithm that does not rely upon the minimum support and the minimum confidence thresholds. Martínez-Ballesteros et al. [29] developed a new algorithm called QARGA (Quantitative Association Rules by Genetic Algorithm). A roulette selection method was used in their proposed algorithm. Martín et al. [30] applied GA to maximize the interestingness, comprehensibility, and performance of quantitative association rules. Álvarez and Vazquez [5] designed a GA-based strategy (GAR-plus) for identifying association rules without the need to partition the domain of the numerical attributes. Three measures of confidence, interestingness, and comprehensibility as fitness functions were defined for finding meaningful quantitative association rules by Minaei-Bidgoli et al. [31]. The proposed algorithm (MOGAR) draws on the Michigan approach for representing rules in chromosomes. Martín et al. [32] proposed QAR-CIP-NSGA-II, a mining quantitative association rules system. This method is based on GA that extracts good intervals in association rules. They [33] also presented another algorithm to mine interesting quantitative association rules by using a Niching GA. By drawing on Greedy Particle Swarm Optimization (GPSO), a hybrid mechanism was proposed by Indira and Kanmani [34] for the ARM. GPSO improves the ARM by integrating the unique features of both GA and PSO. The GPSO methodology outperformed the GA and PSO algorithms in ARM regarding predictive accuracy and consistency when tested on five benchmark datasets. Agarwal and Nanavati [35] developed a novel hybrid ARM algorithm based on PSO and GA algorithms. The estimation

of their method for Bakery dataset has shown that this algorithm is an effective technique for finding rules in non-numerical datasets. Sarkar et al. [36] proposed a new rule mining algorithm by using GA that selects and extracts useful rules according to optimal support and confidence. Djenouri et al. [37] proposed a novel parallel for ARM with the aim of reducing computation time in finding appropriate rules process. This algorithm used GA that was implemented on Graphics Processing Units (GPUs) cluster. Kumar and Singh [38] proposed a mining quantitative association rule system. This method is based on a GA that extracts good intervals in association rules. Martínez-Ballesteros et al. [39] attempted to apply GA to improve the scalability of the methods for QARM and, thus, to process large datasets without sacrificing quality. The results of the comparative analysis mirrored the significant improvements made by the proposed methods in the computational costs of QARGA-M and previously developed GAs without reducing the quality of the results. Padilla et al. [40] developed a new genetic programming algorithm for large data ARM. The genetic operators involved in our proposed technique were particularly designed to avoid high solution complexity and any improvements in values of the fitness function. In 2018, a new generic parallel framework called MRQAR was proposed by Martín et al. [41] to unveil the QAR in large amounts of data. The MapReduce paradigm formed the basis of this approach along with Apache Spark. The incremental learning process provided by MapReduce Quantitative Association Rules (MRQAR) can apply any sequential QAR algorithm to large datasets without redesigning the algorithm.

Table 1 summarizes the applied measures for both evaluation and optimization in several works that are reviewed in this section.

As mentioned above, the researchers have carefully considered the ARM domain and that the application of EAs has been welcomed because they can search for and solve problems in this area (Table 1). Initially, Studies have focused more on the ARM in the categorical attributes' dataset. Furthermore, since many datasets in the real world are characterized by continuous numerical values, it is necessary to provide algorithms for identifying association rules from a dataset with quantitative attributes. Recently, numerous methods have been proposed in the QAR area, many of which have used the metaheuristic approach. In the QAR area, which is the subject of this paper, several metaheuristic algorithms have been used in many studies. GA is the most widely used evolutionary technique for QAR problems due to its longer historical background. GA has been highly enhanced so far, the general motivation of which points to search mechanism, representation, and fitness

Table 1. Quality measures used in the literature.

Related work	Quality measures considered						
	Support	Confidence	Recovered	Comprehen.[a] or # attributes	Amplitude	Interest	Leverage
GENAR	✓	✓	✓				
GAR	✓	✓	✓	✓	✓		
Alatas & Akin	✓	✓	✓	✓	✓		
QuantMiner	✓	✓					
GNP				✓		✓	
EARMGA		✓					
QARGA	✓	✓	✓	✓	✓		
NSGA-II-QAR	✓	✓		✓		✓	
GAR-plus	✓	✓	✓	✓	✓		
MOGAR	✓	✓		✓		✓	
QAR-CIP-NSGA-II	✓			✓		✓	
NICGAR	✓			✓		✓	
GPSO	✓	✓		✓			
Agarwal & Nanavati	✓	✓		✓		✓	
Sarkar et al.	✓						
CGPUGA	✓	✓					
Kumar & Singh	✓	✓				✓	
QARGA-M	✓	✓					✓
Padilla et al.	✓	✓					✓
MRQAR	✓						

[a]: Comprehen.: Comprehensibility

function. Different criteria have been applied to select useful rules in the works reviewed. Different combinations of these criteria have been applied in different studies. In this study, the cosine^2 measure is used for the fitness function along with two other criteria. As can be seen in Table 1, this criterion has not been used in the previous work to extract rules in QAR algorithms. A hybrid of interestingness, confidence, and cosine^2 has been adopted in the fitness function of the proposed approach. Similarly, a combination of these criteria is used in previous works including MOGAR [31]. Nevertheless, the advantage of using cosine^2 is that it examines the degree of similarity between the antecedent and consequential extractive rules and, thus, chooses rules that are more appropriate and useful. This property has made extraction rules more qualitative in this study than the previous ones with the same fitness function. The combination of criteria has increased the support and confidence of selected rules shown in the comparison results, as presented in Section 4 of this paper.

Another point that has been highlighted in previous works is the change of trend in the ARM field researches into a rule-extracting algorithm from large-scale datasets. Due to a large amount of information and the progress of data storage equipment, the data dimension is increasing. Therefore, the need to provide methods for dealing with big data is necessary more than ever. In the QAR area, researchers have not ignored this subject, and papers such as Martínez-Ballesteros et al. [39] and Martín et al. [41] have been proposed to extract quantitative association rules in the large-scale dataset.

2. Multi-objective evolutionary algorithms

ARM is considered to be one of the most common techniques of DM to discover patterns and rules [42]. Association rules that are applied to find a correlation between the existing set of items in a dataset are defined as $X \rightarrow Y$, where X and Y represent the sets of items and $X \cap Y = \phi$ [1,32]. If the interval is continuous, association rules are considered to be quantitative, and each item is a pair of distance properties [43]. The general classification of different methods of association rules is shown in Figure 1. In order to discover rules in a continuous numerical dataset, different techniques such as distribution and discretization are adopted. An exponent is used to indicate a numerical interval, which can also be used
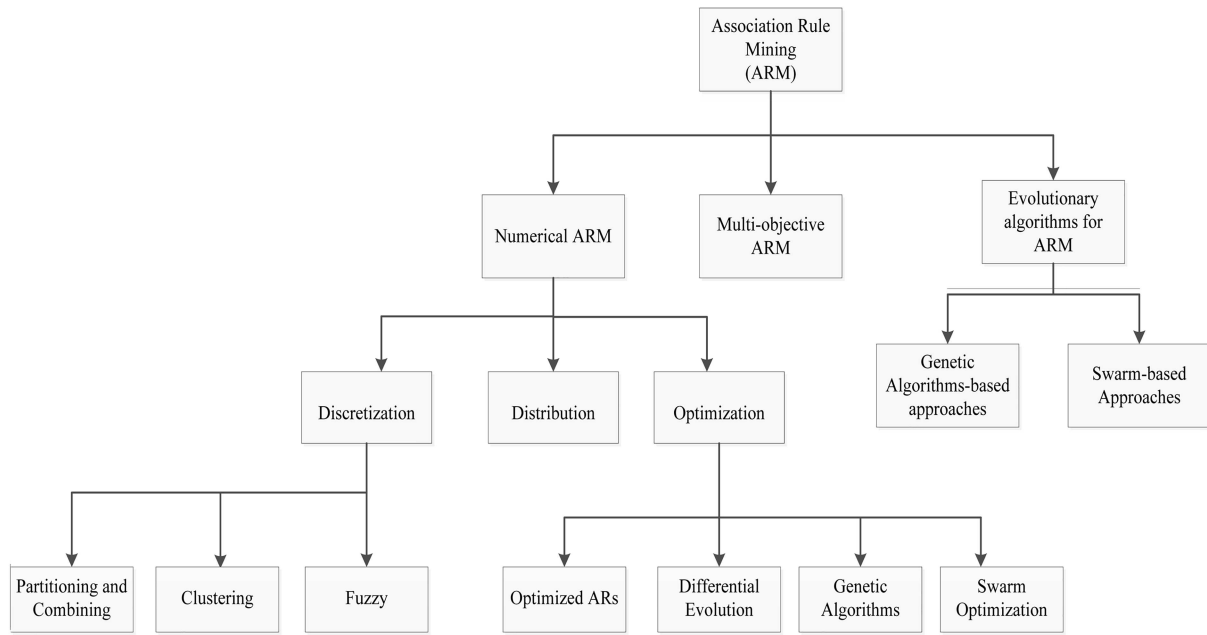
Figure 1. Numerical ARs mining approaches [46].

for the formation of rules. However, the application of such techniques and the discretization of data lead to the loss of valuable information and cause the resulting rules not to be precise enough. In these methods, the desirability of rules is only assessed based on a single criterion.

Some researchers have recently removed some restrictions of the current methods and considered association rules rather than a single-objective one [44,45]. A variety of types of objectives are considered in the process of mining association rules, resulting in more interesting and precise rules.

The investigation of association rules in DM consists of a procedure used for determining the relations between the properties or between their values in a dataset so as to facilitate the decision-making. These relations can be presented in the form of IF-THEN. The condition examined in the IF section is called Antecedent, and what the THEN section contains is called Consequent. Thus, the relation between the two can be shown as $X \rightarrow Y$. Different criteria are used to study the value and desirability of such rules, the most important and common of which are the coefficients of confidence and those of support.

- *Support:* This measure indicates the number of times a rule has appeared in transaction data. This measure is computed through Eq. (1). The number of records covered by both antecedent and consequent of the rule is determined by the support as a statistical measure, which is an indicator that presents the number of times a set of items comes along in a dataset.

$$Support\,(X \rightarrow Y) = \frac{SUP\,(XY)}{|D|}. \qquad (1)$$

- *Confidence:* According to confidence, if the first side of a rule holds, the second side will happen, as well. Thus, it indicates the validity of a rule and is calculated as expressed in Eq. (2). The strength of a rule is measured by the confidence recognized as predictive accuracy, which is a demonstration of the validity of a rule. The quality of a rule is assessed by the confidence criterion based on the occurrence frequency of an AR in the entire dataset. When a rule occurs more frequently in the dataset, it is accounted to be of better quality [46].

$$Confidence\,(X \rightarrow Y) = \frac{SUP\,(XY)}{SUP\,(X)}. \qquad (2)$$

- *Interestingness:* This measure uses the confidence of the antecedent and consequent to calculate the interestingness of the rules and is described as in Eq. (3). It consists of three sides: the first side expresses the probability of production rule according to the antecedent. The second side expresses the probability of creating a rule according to the consequent, while the third side indicates the probability that a rule may not be produced according to the whole dataset. What this criterion is that rules with a high degree of confidence are characterized by a lower degree of interestingness [31]. As a part of DM, the ARM process needs to excerpt some unseen information, that is, there should be relatively a smaller occurrence number of the extracted rules in the entire database. While

it is challenging to measure such astonishing rules, users may find themselves more interested in their extraction. For classification rules, this factor can be measured by information gain theoretic measures [47], which is not of effective interestingness for the calculation of ARs. This study uses a different approach according to the support count of both antecedent and consequent parts [48]. Interestingness $(X \rightarrow Y)$ of a rule indicates a measure of the rule's surprisingness to the user. The discovery of some unseen information from the data is the most important concept in ARM. Not only are rules with higher frequency discovered by the interestingness of a rule, but also they are relatively less frequent in the database [49].

$$Interestingness = \left[\frac{SUP\,(XY)}{SUP\,(X)} \right] \times \left[\frac{SUP\,(XY)}{SUP\,(Y)} \right]$$

$$\times \left[1 - \frac{SUP\,(XY)}{SUP\,(D)} \right]. \tag{3}$$

- $Cosine^2$: The criterion $Cosine^2$ measures the similarity between the consequent and antecedent. The quantity amplitude of such a law ranges from 0 to 1, and its square is also present in this amplitude. A $cosine^2$ value of approx. 1 suggests that the majority of transactions holding item X also contain item Y, and vice versa. Likewise, a cosine value of nearly 0 specifies that most transactions containing item X do not possess item Y, and vice versa [49]. Most of the transactions possessing item X also hold item Y, and vice versa (i.e., many of transactions having item Y contain item X, too) when $cosine^2$ value approximates to one. With the approximation of $cosine^2$ value to zero, the majority of transactions possess item X, but lack item Y, and vice versa [50]. The square of this measure is used in this paper since it is more accurate [51]. It is defined through Eq. (4):

$$Cosine^2 = \frac{SUP\,(XY)}{SUP\,(X)} \times \frac{SUP\,(XY)}{SUP\,(C)}. \tag{4}$$

2.1. Genetic Algorithm (GA)

In recent years, most researchers have used EAs [14,33,52], especially GA [21,28,32], in order to discover quantitative association rules of quantitative datasets [23]. GA is a widely renowned evolutionary approach, which randomizes the generation of a population of strings. The best string in the population will access an optimal solution by applying evolutionary operators such as crossover, mutation, and reproduction [53]. GAs are mathematical tools with a wide range of applications. They are very efficient in the optimization of the problems, especially when the respective objective functions are discontinuous and exhibit many local optima [54]. GA is a heuristic algorithm of random

search, which denotes the biological evolution. Such algorithms have been successfully used in solving hard optimization problems, those for which the steepest ancestry techniques go through local minima or remain incapable due to the complications [55]. Processing of numerical dataset for getting information is a tedious and computationally complex task because of the nature of the data. ARM for a numerical dataset is a challenging task as the huge number of expected rules will be generated. Therefore, a technique that runs for many solutions in parallel is required so that a large number of rules can be processed. The pre-existing technique that follows these criteria is the genetic algorithm that has the property of running for solution in parallel and improving their own solution in all iterations, i.e., generation in terms of GA. Given the features of the GA such as parent selection, mutation, and crossover, this algorithm passes appropriate responses to the next generation. In the QAR process, the exploration property is required for providing a number of answers and the exploitation property for their desirability. Given the steps in the GA, this algorithm allows for two exploration and exploitation needs. Therefore, the GA is selected to solve the problem of ARM. Based on the evolutionary ideas of natural selection and genetic, GAs are adaptive heuristic search algorithms. The concept of GAs is basically intended to mimic natural system processes essential for evolution. Thus, GAs manifest the intelligent utilization of a random search within a defined search space for problem solving. There is little knowledge about the fact that GAs are among the best methods for solving a problem. As a very general algorithm, it will function properly in any search space. GA is a stochastic search algorithm, modeled on the natural selection process underlining biological evolution [56]. Numerous search, optimization, and machine learning problems have successfully employed GA, working iteratively by producing new populations of strings from the older ones. Each string is an encoded binary, real, etc. version of a candidate solution. An evaluation function follows the fitness criterion for each string, suggesting its fitness with the problem [57]. Because of its longer historical background, GA is the most commonly used evolutionary procedure for solving QAR problems. GA has received many improvements, of which the general attention is given to search mechanism, representation, and fitness function.

3. Proposed method

In this paper, an MOEA is proposed for discovering quantitative association rules (MOEA-QAR), as explained in this section. This method is a combination of the MOGAs and clustering. In this method, first, the dataset is clustered; then, the records of each

cluster are determined as the input of each GA and, then, the attractive rules of each cluster are extracted. Characterized by a number of advantages and practicality, the K-means clustering algorithm is employed in the present research. The K-means algorithm is advantageous because of its simplicity and rapidity for low-dimensional data. It can also detect pure sub-clusters when defining a huge number of clusters. Besides, K-means computation is considered to be quicker than hierarchical clustering and is projected to yield tighter clusters than hierarchical clustering [58]. In the K-mean algorithm, there is a possibility to change and select the number of clusters by the user. This makes the K-mean algorithm flexible unlike hierarchical algorithms. In this algorithm, according to the requirements of the problem, the expert can test a number of different clusters and determine the proper number of clusters according to the results. The K-mean algorithm allows the user to participate and is in fact an interactive process between the algorithm and the expert. In the K-means method, the records are grouped so that the records inside a cluster can be similar (SS_W) and the records in the individual clusters be different (SS_B). This algorithm attempts to minimize SS_B and maximize SS_W. Therefore, it provides an appropriate tradeoff for forming clusters. Since the similarity of the records for rule mining algorithm can be an important advantage, the K-means method is selected for the clustering records. Because of these benefits, the criteria for evaluating the rules extracted from each cluster are calculated by using the whole dataset. In this method, there is one archive to keep the attractive rules of each generation. Finally, the rules or chromosomes remaining in the archive are introduced as the extracted rules. This idea is similar to the elitism method in the GA that was applied to store the best chromosomes of each generation. The first input generation is selected randomly, and the initial chromosome gains a value. The algorithm is run until the terminating conditions are met. These conditions may include the number of algorithm repetitions or the given number of the created rules. The running of the algorithm ends if the number of the rules existing in the archive remains unchanged. In the following, first, the method of presenting chromosomes is explained; then, the mutation, crossover, and calculation of the fitness function are expressed. The pseudocode of the proposed algorithm is described in Figure 2, and the flowchart of the algorithm is shown in Figure 3.

3.1. Genetic operations

Like the methods that rely on GAs, in the first step, the encoding or representation of the applied chromosomes must be defined. Then, how to compute the fitness value of chromosome, mutation, and crossover operations is described.

3.1.1. Presentation of chromosomes

Each chromosome shows a rule. Each feature of the dataset is presented by a gene. The ith gene expresses the ith feature. Each gene includes 2 bits and 2 numbers. The first 2 bits determine whether the feature is involved in creating the rule or not and they also determine the section of the rule where the feature is present. The bit 00 represents the presence of this feature in the rule consequence; further to this, if these two fields are equal to 11, this feature is involved in creating the rule. In addition, if these two bits are 01 or 10, then this feature is not involved in the creation of the rule. The two other numbers represent the lower and upper limits of the given feature interval. The first number is considered as the lower limit of the feature, and the second number is the upper limit.

$$Gene_i = (ac_i, lb_i, ub_i), \quad i = 1, ..., m,$$

$$C_T = Gene_1 Gene_2 .... Gene_m.$$

3.1.2. Fitness function

To achieve the projected aim, it is required to define a suitable objective function that comprises a number of parameters that are sensible enough to identify the whole plausible structural damages [59]. As previously mentioned, the proposed method is of multi-criteria type, and several criteria are used to choose the generation. The fitness of each chromosome is measured by three parameters, as shown in Eq. (5). This function attempts to extract the rules that are more valuable and attractive. In this method, the problem is solved using non-dominated solutions. If solution (B) is better than solution (A) with respect to the pre-determined criteria, then solution (A) is dominated by another solution, i.e., solution (B). In this case, solution (B) is called the non-dominated solution. The value of each rule is assessed by three parameters such as interestingness, cosine^2, and confidence coefficient. The chromosomes that can dominate other chromosomes with respect to the values of these three factors will remain in the archive. The rules are ranked based on these three factors, and the rules with greater scores will be extracted as the final rules.

$$Fitness = Confidence + Interestingness$$

$$+ Cosine^2. \tag{5}$$

The value and attractiveness of these rules are studied based on different criteria, all of which aim to excerpt beneficial rules, with each criterion used for assessing the rules from a different viewpoint. Although there are different definitions of some of these benchmarks, they share a goal, *viz.* elevation and vividness, both of which attempt to choose fascinating rules. Most often, it suffices to concentrate on the integration

```
Start:
        Clustering Total data
        data_T = Total data;
        If (data_clu_use=1)
                data_M = cluster;
        else
                data_M = Total data;

        function GENETIC-ALGPRITHM (population, FITNESS-FN, data_M) returns an individual
          inputs:  population, a set of individuals
                        FITNESS-FN, a function that measure the fitness of an individual

                Generate the initial population with N chromosomes.
                Evaluate the initial population

        repeat

                new_population <-- empty set
                 for i= = 1 to SIZE(population)  do
                x <-- RANDOM-SELECTION (population, FITNESS-FN)
                y <-- RANDOM-SELECTION (population, FITNESS-FN)
                        Child <-- REPRODUCE (x ,y)
                        if (small random probability) then child <-- MUTATE(child)
                        add child to new_population

                        Population <-- new_population

        until some individual is fit enough , or enough time has elapsed

         return the best individual in population, according to  FITNESS-FN
        function FITNESS-FN (data_T)

                Compute Support;

                Compute Confidence;

                Compute Interestingness;

                Compute Cosine²;

                Compute Fitness;
                return FITNESS-FN measure;
end.
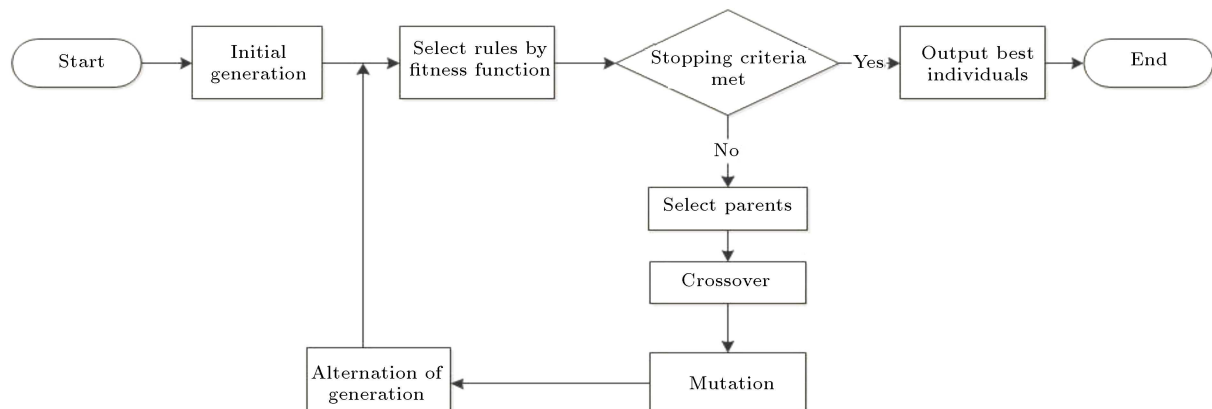```

**Figure 2.** The pseudocode of the proposed algorithm.



**Figure 3.** The flowchart of the proposed algorithm.

of support, confidence, and either elevation or influence to measure the "quality" of a rule quantitatively. However, the real value of a rule is subjective in terms of effectiveness and functionality and is strongly dependent on the specific area and business purposes [60].

### 3.1.3. Crossover and mutation operators
Regarding the methods of presenting the chromosomes and calculating the fitness function, the crossover and

mutation operators are introduced in this section. The position of the features does not change at mutation and intersection stages, i.e., if a variable represents the second gene, it will remain the second gene after the aforementioned stages. A two-point crossover method is used to perform the crossover operator, as shown in Figure 4. A high rate of crossover causes the genomes in the next generation to be more random, as there will be more genomes that represent a mix of
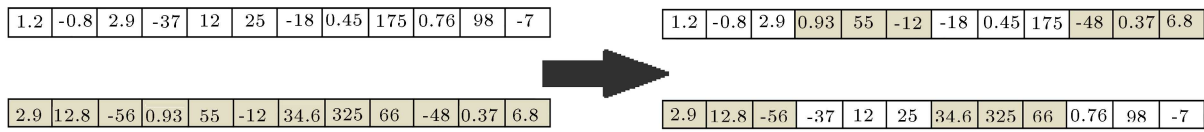
| 1.2 | -0.8 | 2.9 | -37 | 12 | 25 | -18 | 0.45 | 175 | 0.76 | 98 | -7 |

| 1.2 | -0.8 | 2.9 | 0.93 | 55 | -12 | -18 | 0.45 | 175 | -48 | 0.37 | 6.8 |

| 2.9 | 12.8 | -56 | 0.93 | 55 | -12 | 34.6 | 325 | 66 | -48 | 0.37 | 6.8 |

| 2.9 | 12.8 | -56 | -37 | 12 | 25 | 34.6 | 325 | 66 | 0.76 | 98 | -7 |

**Figure 4.** A simple example of the crossover operator.

previous generation genomes. A low rate of crossover keeps the fit genomes from the previous generation, although it decreases the chance that a very fit genome will be produced by the crossover operation. Uniform crossover will create genomes that will be very different from their parents if their parents are not similar. If its parents are similar, the offspring will be similar to its parents. Using the 1-point crossover means that offspring genomes will be less diverse; they will be quite similar to their parents. Therefore, the trade of these options is appropriate, and the two-point crossover is chosen in this paper. Both parental genotypes in a two-point crossover are cleaved at two points, making a new progeny through parts one and three from the first and the middle part from the second ancestor. A more thorough searching problem space is made possible through the two-point crossover. The schema remains safe against disruption by applying an operator with single-point and two-point crossovers; however, search space grows narrower in case of a homogeneous population [61].

The first 2 bits of the chromosome can change to do the mutation process, and the given feature may be involved in creating the rule or not. The 4 states mentioned for these two bits (00, 11, 10, and 01) are changed into one another randomly. Also, the lower and upper limits of the values can change. Of course, one should be careful that increasing or reducing the upper and lower limits must not exceed the permissible variables and variable limits; by the same token, the lower unit should not exceed the upper limit. The changes to the mutation process are made randomly.

According the output of implementing algorithm on different datasets that are shown in Figures 5 and 6, the provided solutions had exploitation requirement. These figures proved the convergence of the proposed responses.

To meet the exploration requirement, it can be expressed that all of the required changes and alterations have been considered to generate new chromosomes. Determining the crossover, initial population, and mutation provided sufficient changes and the population generation. To do so, all types of mutation are applied to chromosomes including 2 bits and the lower or upper limits of the values mutation. The first 2 bits of the chromosome can be modified to perform the mutation process, and the given feature may be involved in creating the rule. The four states mentioned for these
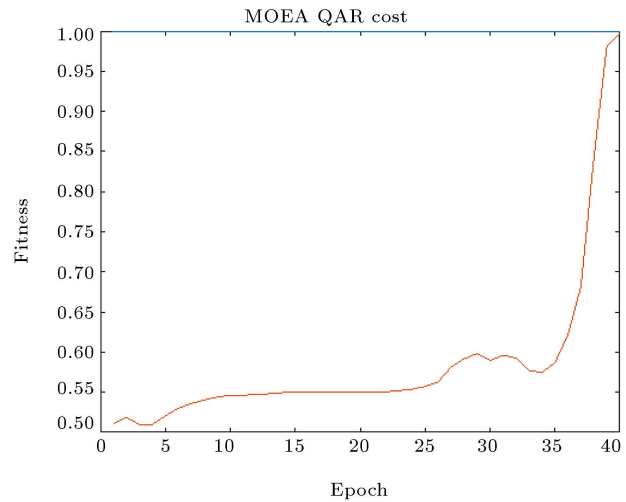


**Figure 5.** The output of the implementation of algorithm on basketball dataset.
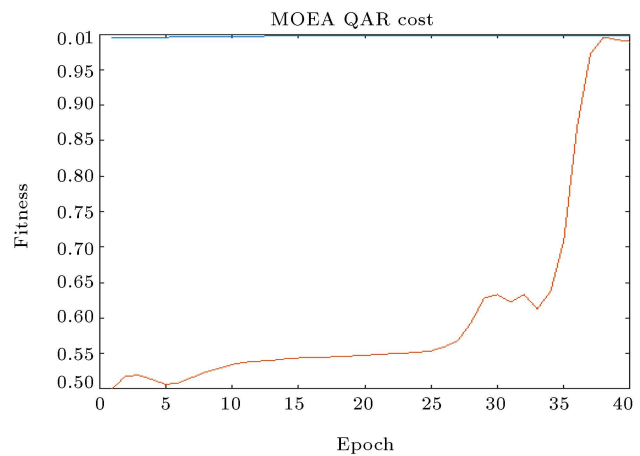


**Figure 6.** The output of the implementation of algorithm on basketball dataset.

two bits (00, 11, 01, and 10) change into one another randomly. Moreover, the lower and upper limits of the values can vary. Of course, one should exercise care when increasing or reducing the upper and lower limits no to exceed the permissible variables and variable limits, and also the lower unit should not exceed the upper limit. The changes in the mutation process are made randomly. The two-point crossover method has been used for the crossover operator. A high crossover rate causes the genomes in the next generation to be more random, as there will be more genomes that are a mix of previous generation genomes. A low crossover rate keeps the fit genomes from the previous

generation, although it decreases the chance that a very fit genome will be produced by the crossover operation. The uniform crossover will create genomes that will be very different from their parents if their parents are not similar. If its parents are similar, the offspring will be similar to its parents. The application of 1-point crossover means that offspring genomes will be less diverse; they will be quite similar to their parents. Therefore, the trade of these options is appropriate, and the two-point crossover is chosen in this paper. In the two-point crossover, both parental genotypes are split at two points, constructing a new offspring by using parts one and three from the first and the middle part from the second ancestor. The application of the two-point crossover will enable searching the problem space more thoroughly. The application of single-point and two-point crossover operators prevents schema from disruption; however, when population becomes homogeneous, the search space becomes smaller [61]. Moreover, the ability of the method to extract more rules and improve the results of the evaluation criteria (support and confidence) in comparison with the results of the previous work proves that the proposed algorithm enjoys desirable performance at the point of exploration.

## 4. Results

To evaluate the efficiency of the proposed approach, the presented algorithm was examined in three datasets, and the results were compared with those of similar models. Such datasets include properties with continuous and numerical quantities. The test data are shown in Table 2. The datasets are available at "the Bilkent University Function Approximation Repository" [62]. The adjusted parameters in the implementation of the algorithm are presented in Table 3. To achieve the accurate results of transactions, the presented algorithm was implemented on each dataset more than 10 times. The results are shown in Table 3. The frequency of the GA was considered between 100 and 400 each time so that the results might be achieved in different scenarios. A two-point crossover method was used in this algorithm.

As mentioned before, first, the dataset is clustered and, then, the useful rules of each cluster are mined by means of the GA. The number of clusters of each dataset is selected according to the Silhouette value.

**Table 3.** The parameter settings for algorithm implementation.

| Parameter | Population size | Crossover probability | Mutation probability |
|---|---|---|---|
| Value | 800 | 0.7 | 0.2 |

The number of clusters with higher Silhouette values is chosen as the optimal number of clusters. The clustering results of these three datasets are presented in Table 4.

In the following sections, the results of the proposed algorithm (MOEA-QAR) were compared with those of previous studies in similar cases such as Alatas and Akin [22], RPSO [14], GAR [25], and MOGAR [31]. Mining association rules of continuous numerical datasets using EAs is the main objective of such studies. The descriptions of the method used in the above-mentioned algorithms are provided in the first part of the introduction. In each study, a different fitness function is defined to select and mine the final rules. The results of these methods are directly derived from the papers presented by them. The developed algorithm was implemented in two ways: once on the entire dataset and once on each cluster after clustering the data. The results of both implementations are provided in the following sections. The comparison of the number of rules extracted by other algorithms and those of the proposed approach is presented in Table 5.

According to the results shown in Table 5, the proposed model can detect more useful and interesting results from datasets. The performance improvement of the proposed approach in comparison with that of other methods in terms of the number of extracted rules is also shown. Another result presented in this table is that implementing algorithm on clusters can extract more rules than implementing it on the whole dataset. The aim of applying the clustering approach before executing the QAR algorithm is to examine the difference between the results of the QAR algorithm on the total dataset records and on the records of a cluster. Since the records of a cluster are closer and more similar to the total records of the dataset, the goal is to determine whether this similarity has an effect on the results of the QAR algorithm. The results demonstrate that such similarity and proximity cause the QAR algorithm to extract a greater number of rules in a cluster than the whole dataset, and this

**Table 2.** Definition of used datasets for comparisons.

| Target feature | No. of features | No. of records | Dataset |
|---|---|---|---|
| Basketball | 96 | 5 | Points per minute |
| Bodyfat | 225 | 18 | Body height |
| Quake | 2178 | 4 | Richter |

**Table 4.** The results of three datasets clustering.

| Dataset | No. of records | | | |
|---|---|---|---|---|
| Basketball | **Cluster 1**: 16 records | **Cluster 2**: 27 records | **Cluster 3**: 20 records | **Cluster 4**: 33 records |
| Bodyfat | **Cluster 1**: 147 records | | **Cluster 2**: 105 records | |
| Quake | **Cluster 1**: 1516 records | | **Cluster 2**: 662 records | |

**Table 5.** Comparison of the number of the obtained rules.

| Dataset | Alatas & Akin [22] | RPSO [14] | MOGAR [31] | MOEA-QAR (this study) | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | Total data set: | Cluster 1: | Cluster 2: | Cluster 3: | Cluster 4: |
| Basketball | 33.8 | 34.2 | 50 | 54.71 | 61 | 57.27 | 57 | 56.37 |
| | | | | Total data set: | Cluster 1: | | Cluster 2: | |
| Bodyfat | 44.2 | 46.4 | 84 | 84.33 | 84.4 | | 84.66 | |
| | | | | Total data set: | Cluster 1: | | Cluster 2: | |
| Quake | 43.8 | 46.4 | 44.87 | 54.28 | 66.16 | | 57.4 | |

superiority is shown in Figures 7–9. As is clear in these figures, the number of extracted rules from the total dataset is shown by the blue columns, the number of extracted rules from each cluster is determined by the yellow columns, and the green columns illustrate the total number of rules from all clusters. Whereas the number of records in a cluster is less than the whole dataset, the percentage of support rules decreases.

Table 6 shows the results of confidence coefficient, and Table 7 shows the support percentage obtained by
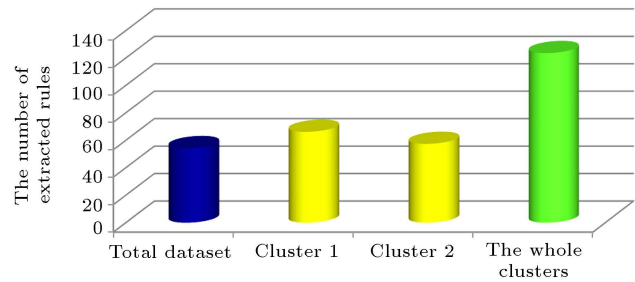


**Figure 9.** Comparison of the number of extracted rules from the total quake dataset and clusters.

implementing the proposed algorithm along with the results obtained from previous studies. The averages of these coefficients are presented in the above-mentioned tables. As Table 6 shows, this method can significantly improve the confidence of rules. The rules extracted by this method had a higher confidence coefficient than the previous methods. In the mode of implementing algorithm, support percentage would be very high for extracting the rules. In this case, the support percentage of rules of DM via this method also showed a significant difference compared to previous methods, and there was an improvement. Based on the obtained results from the two different modes of algorithm, it can be mentioned that the rules extracted from the whole dataset had higher support percentage than those of clusters. The improvement of results in terms of support and confidence indicated that the results obtained by the proposed method were more reliable and that the algorithm used to discover association rules outperformed other methods.

The criterion of interestingness in Table 8, which is used in this study, is compared with MOGAR method in the fitness function. Based on the results
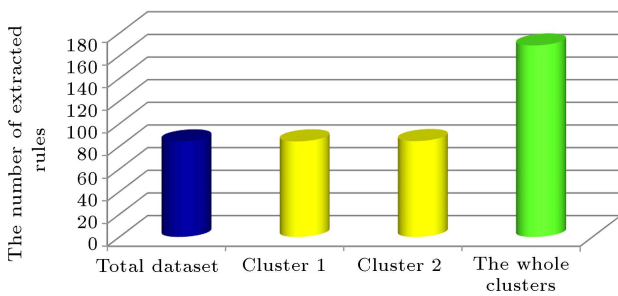


**Figure 7.** Comparison of the number of extracted rules from the total bodyfat dataset and clusters.
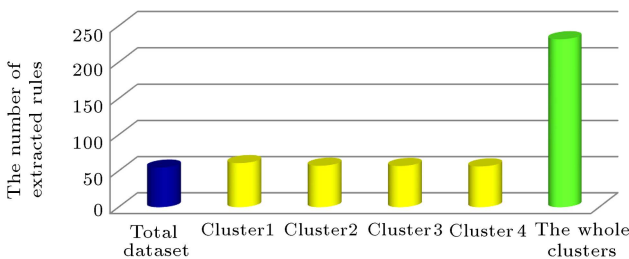


**Figure 8.** Comparison of the number of extracted rules from the total basketball dataset and clusters.

**Table 6.** Comparison of confidence (%) of the obtained rules.

| Dataset | Alatas & Akin [22] | RPSO [14] | MOGAR [31] | MOEA-QAR (this study) | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | Total data set: | Cluster 1: | Cluster 2: | Cluster 3: | Cluster 4: |
| Basketball | 60 | 6 | 83 | 98 | 97.1 | 97.7 | 96 | 97.6 |
| | | | | Total data set: | Cluster 1: | | Cluster 2: | |
| Bodyfat | 59 | 61 | 85 | 99.4 | 99.4 | | 99.4 | |
| | | | | Total data set: | Cluster 1: | | Cluster 2: | |
| Quake | 62 | 63 | 82 | 99 | 99 | | 99 | |

**Table 7.** Comparison of support (%) of the obtained rules.

| Dataset | Alatas & Akin [22] | RPSO [14] | GAR [25] | MOGAR [31] | MOEA-QAR (this study) | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Total data set: | Cluster 1: | Cluster 2: | Cluster 3: | Cluster 4: |
| Basketball | 32.21 | 36.44 | 36.69 | 50.28 | 83.11 | 34.27 | 43.34 | 27.27 | 44.69 |
| | | | | | Total data set: | Cluster 1: | Cluster 2: | |
| Bodyfat | 63.29 | 65.22 | 65.22 | 65.26 | 86.57 | 76.70 | 71.34 | |
| | | | | | Total data set: | Cluster 1: | Cluster 2: | |
| Quake | 38.74 | 38.74 | 36.96 | 30.12 | 60.11 | 59.8 | 40.12 | |

**Table 8.** Results of interestingness criteria.

| Dataset | MOGAR | MOEA-QAR (this study) | | | | |
|---|---|---|---|---|---|---|
| Basketball | 0.53 | Total data set: | Cluster 1: | Cluster 2: | Cluster 3: | Cluster 4: |
| | | 0.213 | 0.182 | 0.119 | 0.071 | 0.206 |
| Bodyfat | 0.56 | Total data set: | Cluster 1: | Cluster 2: | | |
| | | 0.209 | 0.230 | 0.173 | | |
| Quake | 0.46 | Total data set: | Cluster 1: | Cluster 2: | | |
| | | 0.173 | 0.161 | 0.151 | | |

**Table 9.** Results of cosine$^2$ of the obtained rules.

| Dataset | MOEA-QAR (this study) | | | | |
|---|---|---|---|---|---|
| Basketball | Total data set: | Cluster 1: | Cluster 2: | Cluster 3: | Cluster 4: |
| | 0.93 | 0.497 | 0.49 | 0.309 | 0.584 |
| Bodyfat | Total data set: | Cluster 1: | Cluster 2: | | |
| | 0.882 | 0.827 | 0.819 | | |
| Quake | Total data set: | Cluster 1: | Cluster 2: | | |
| | 0.61 | 0.615 | 0.485 | | |

obtained from Table 8, the rules extracted by MOGAR method are characterized by higher interestingness. As said before, regarding the definition of criterion of interestingness, the rules with higher support percentage will have lower interestingness, and since the rules examined by the proposed method in this study have higher support percentage than MOGAR method, they have a lower interestingness coefficient. The results of cosine$^2$ criterion are presented in Table 9. Since this criterion has not been used in algorithms presented so far, it has not been compared to other methods.

Regarding the results of the proposed algorithm, it could be said that the proposed method outperformed the proposed algorithm at the same tasks, and it would be an appropriate and trustworthy algorithm for finding useful and attractive rules on the sets of continuous numerical data. In comparison to other methods, this method achieved significant improve-

**Table 10.** The improvement of MOEA-QAR compared to other methods.

| Previous methods | No. of rules | Confidence | support% |
|---|---|---|---|
| Alatas & Akin [22] | 35.03% | 38.93% | 41.22% |
| RPSO [14] | 32.32% | 37.92% | 38.78% |
| MOGAR [31] | 8.77% | 15.65% | 38% |

ments with respect to the number of the rules extracted from the dataset, high confidence measure, and support percentage. Table 10 shows the average improvements of different criteria resulting from the rules extracted by the proposed method in the examined dataset in relation to the compared methods. In this table, the percentage of progression of the proposed method to the comparative methods is mentioned in terms of the criteria considered. The superiority of the proposed method is surveyed based on the following perspectives: the number of extracted rules, the support percentage, and the confidence measure. These criteria are the main and basic factors in evaluating ARM algorithms and are used to prove the efficiency of the proposed approach. These factors were used in all of the previous works, some of which are mentioned in the "related works" section. The highest value of improvement belongs to the criterion of confidence. This method can extract more rules with higher support percentage confidence coefficient in relation to other methods of sets of continuous numerical data. The criterion of cosine$^2$ has not been used in previous methods. Since this criterion evaluates the degree of resemblance of antecedent and consequent and fitness function chooses rules based on this criterion, this criterion has positive impact on the results of extracted rules in terms of confidence measure.

## 5. Discussion

In this paper, a novel QAR adopts GA algorithm. The fitness function is defined based on Confidence, cosine$^2$, and Interestingness criteria. The efficiency of the algorithms proposed is assessed and compared regarding three measures (the number of rules, confidence, and support). The proposed method is compared with those in previous works (i.e., Alatas and Akin [22], RPSO [14], GAR [25], and MOGAR [31]) that are generated using GA to evaluate the effectiveness of the proposed algorithm. As mentioned before, this algorithm was implemented in two separate ways: one way on the entire dataset and the other on the clustered dataset. In the latter case, more rules are identified from the dataset. The aim of using the clustering approach before implementing the QAR algorithm is to examine the difference between the results of the QAR algorithm on the total dataset records and those on the records of a cluster. Since the records of

a cluster are closer and more similar to the total records of the dataset, the goal is to determine whether this similarity has an effect on the results of the QAR algorithm. The results demonstrated that this similarity and proximity caused the QAR algorithm to extract a greater number of rules in a cluster than the whole dataset. Whereas the number of records in a cluster is less than the whole dataset, the percentage of support rules decreases. The average number of rules extracted from different methods is compared and shown in Table 5, indicating that the presented method is capable of extracting further rules than other approaches. Comparisons between the MOEA-QAR and the aforementioned approaches regarding the average value of support criterion and the average confidence of ARs are shown in Tables 6 and 7; these are considered to be the power of ARs. In comparison to other approaches, MOEA-QAR absolutely rivals the confidence and support of ARs. The MOEA-QAR-extracted confidence of ARs has substantially enhanced the concerning confidence and support of the rules extracted in all of the datasets.

This study characterizes the problem of mining numerical ARs as a multi-objective optimization problem. In addition, the MOEA-QAR algorithm, namely MOEA-QAR, is recommended for mining all precise and understandable ARs. A summary of the major characteristics of the proposed algorithm is presented below:

- The confidence, comprehensibility, and interestingness measures are concurrently optimized by MOEA-QAR by making no item sets;

- Unlike the conventional ARM approaches, the user does not determine the minimum support and minimum confidence, hence automating the mining process;

- In the majority of the experiments, the success of the proposed method demonstrates that MOEA-QAR is a method capable of numerical ARM featuring a simple particle design with no need for a priori discretization;

- The results of using clustering method and the QAR algorithm in a hybrid framework confirmed that the QAR algorithm enjoyed the ability to mine a higher number of rules in a cluster than the entire dataset.

The dynamic and competitive nature of today's

business environment has prompted managers and business analysts to apply Decision Support Systems (DSS) to provide quick answers to complex business questions. The organizations will not be able to meet customer expectations without a deeper awareness of customers' habitual and behavioral patterns, and this can have a significant effect on their business. The rapid growth of the volume of data and the importance of using DSS systems demonstrate the necessity of adopting DM techniques. The application of DM technology enables the companies to pay greater attention to the customers' interactions, gain proper knowledge more rapidly through behavioral and systemic information, create more effective practical attitudes, and eventually function more wisely based on such attitudes. The decision-makers will accomplish their chief useful purposes by effectively managing and analyzing the volume, velocity, and variety of the present data, which are rapidly increasing, and benefiting from proper tools and skills for a better understanding of their operations, customers, and market. In this study, a novel hybrid QAR approach is proposed for numerical datasets. DM is concerned with finding the hidden relationships present in business data in order to allow businesses to make predictions for future use [63]. ARM is one of important techniques in DM that provides valuable information for decision-makers by extracting helpful and applicable rules. The method can extract more rules with higher confidence measure and support percentage than similar and previous approaches. A larger number of rules with higher quality allow decision-makers to have more alternatives and make more appropriate executive decisions.

## 6. Conclusion

In the present study, a new approach using multi-objective Genetic Algorithms (GA) was proposed to mine the numerical association rule. One of the most important applications and interesting fields of Data Science is to extract association rules. Extracting the rules of the dataset containing continuous data is one of the challenges of this field. In this study, a multi-criteria approach based on GA was developed to extract meaningful and attractive rules from the dataset containing continuous numerical variables. Association Rule Mining (ARM) is not a single objective problem and, naturally, is a multi-objective one. In the proposed algorithm, three criteria were employed in the fitness function. The efficiency of the proposed method was examined using three continuous numerical datasets. The results of the algorithm implementation were compared with those of previous works in similar cases. The distinction between the proposed method and the similar cases can be expressed as follows: in this method, there is

no need to determine a certain limit for the minimum support and minimum confidence; however, this could not affect the number of the rules extracted, and this method improved the number of these rules compared to other approaches. Second, one of these criteria used in fitness function's definition was $cosine^2$, which was not employed in this field of study. This study found that the application of $cosine^2$ criteria in the fitness function could significantly improve the usefulness and attractiveness of the extracted rules. The proposed approach led to an improvement in the results of the evaluation of support and confidence criteria compared to other methods; both support and confidence criteria increased in the proposed method, and tradeoff was made between the values of these two criteria. Third, no limitation was considered on the number of rules in this method, causing the extracted rules to have higher desirability than other methods. In addition, support and confidence values in the rules extracted from the proposed method improved. According to the mentioned points, it can be expressed that the determined goals of this study were achieved. The results obtained through comparisons showed the improvement of the results of a couple of criteria for proper rule extraction: the support percentage and the confidence measure. The ability of this method to extract more rules with higher confidence measure and support percentage proved the applicability of the proposed approach. This algorithm was implemented in two ways: once on the entire dataset and once on each cluster after clustering the data. Implementing the algorithm on each cluster separately extracted more rules than those on the entire dataset. Considering that only the positive rules were mined in the present research, the extraction of positive and negative rules for future studies is suggested.

## References

1. Agrawal, R., Imieliński, T., and Swami, A. "Mining association rules between sets of items in large databases", In *ACM SIGMOD Record*, **22**(2), pp. 207–216 (June 1993).

2. Haeri, A. and Tavakkoli-Moghaddam, R. "Developing a hybrid data mining approach based on multi-objective particle swarm optimization for solving a traveling salesman problem", *Journal of Business Economics and Management*, **13**(5), pp. 951–967 (2012).

3. Soysal, Ö.M. "Association rule mining with mostly associated sequential patterns", *Expert Systems with Applications*, **42**(5), pp. 2582–2592 (2015).

4. Rana, M. and Mann, P.S. "Analysis of MFGA to extract interesting rules" *International Journal of Computer Applications*, **84**(3), pp. 15–21 (2013).

5. Álvarez, V.P. and Vázquez, J.M. "An evolutionary algorithm to discover quantitative association rules

from huge databases without the need for an a priori discretization", *Expert Systems with Applications*, **39**(1), pp. 585–593 (2012).

6. Miller, R.J. and Yang, Y. "Association rules over interval data", *ACM SIGMOD Record*, **26**(2), pp. 452–461 (1997).

7. Moslehi, P., Bidgoli, B.M., Nasiri, M., and Salajegheh, A. "Multi-objective numeric association rules mining via ant colony optimization for continuous domains without specifying minimum support and minimum confidence", *International Journal of Computer Science Issues (IJCSI)*, **8**(5), pp. 1–34 (2011).

8. Ehrgott, M., *Multicriteria Optimization*, **491**. Springer Science & Business Media (2005).

9. Coello, C.A.C., Lamont, G.B., and Van Veldhuizen, D.A., *Evolutionary Algorithms for Solving Multi-Objective Problems*, Springer, 800 (2002).

10. Coello, C.A.C., Lamont, G.B., and Van Veldhuizen, D.A., *Evolutionary Algorithms for Solving Multi-Objective Problems*, **5**, New York: Springer (2007).

11. Deb, K., *Multi-Objective Optimization Using Evolutionary Algorithms*, **16**, John Wiley & Sons (2001).

12. Kaya, M. and Alhajj, R. "Genetic algorithm based framework for mining fuzzy association rules", *Fuzzy Sets and Systems*, **152**(3), pp. 587–601 (2005).

13. Kaya, M. and Alhajj, R. "Utilizing genetic algorithms to optimize membership functions for fuzzy weighted association rules mining", *Applied Intelligence*, **24**(1), pp. 7–15 (2006).

14. Alatas, B. and Akin, E. "Rough particle swarm optimization and its applications in data mining", *Soft Computing*, **12**(12), pp. 1205–1218 (2008).

15. Ayubi, S., Muyeba, M.K., Baraani, A., et al. "An algorithm to mine general association rules from tabular data", *Information Sciences*, **179**(20), pp. 3520–3539 (2009).

16. Nasiri, M., Taghavi, L.S., and Minaee, B. "Multi-objective rule mining using simulated annealing algorithm", *Journal of Convergence Information Technology*, **5**(1), pp. 60–68 (2010).

17. Qodmanan, H.R., Nasiri, M., and Minaei-Bidgoli, B. "Multi objective association rule mining with genetic algorithm without specifying minimum support and minimum confidence", *Expert Systems with Applications*, **38**(1), pp. 288–298 (2011).

18. Djenouri, Y., Drias, H., and Chemchem, A. "A hybrid bees swarm optimization and tabu search algorithm for association rule mining", In *Nature and Biologically Inspired Computing (NaBIC)*, 2013 World Congress on IEEE pp. 120–125, (Aug, 2013).

19. Agbehadji, I.E., Fong, S., and Millham, R. "Wolf search algorithm for numeric association rule mining", In *Cloud Computing and Big Data Analysis (ICC-*

*CBDA )*, 2016 IEEE International Conference on IEEE pp. 146–151 (July, 2016).

20. Can, U. and Alatas, B. "Automatic mining of quantitative association rules with gravitational search algorithm", *International Journal of Software Engineering and Knowledge Engineering*, **27**(3), pp. 343–372 (2017).

21. Erickson, M., Mayer, A., and Horn, J. "The niched Pareto genetic algorithm 2 applied to the design of groundwater remediation systems", In *International Conference on Evolutionary Multi-Criterion Optimization*, (pp. 681–695), Springer, Berlin, Heidelberg (March, 2001).

22. Alataş, B. and Akin, E. "An efficient genetic algorithm for automated mining of both positive and negative quantitative association rules", *Soft Computing*, **10**(3), pp. 230–237 (2006).

23. Alcala-Fdez, J., Flugy-Pape, N., Bonarini, A., and Herrera, F. "Analysis of the effectiveness of the genetic algorithms based on extraction of association rules", *Fundamental Informaticae*, **98**(1), pp. 1–14 (2010).

24. Mata, J., Alvarez, J.L., and Riquelme, J.C. "Mining numeric association rules with genetic algorithms", In *Artificial Neural Nets and Genetic Algorithms*, pp. 264–267, Springer, Vienna (2001).

25. Mata, J., Alvarez, J.L., and Riquelme, J.C. "Discovering numeric association rules via evolutionary algorithm", *Advances in Knowledge Discovery and Data Mining*, pp. 40–51 (2002).

26. Salleb-Aouissi, A., Vrain, C., and Nortet, C. "Quant-Miner: A genetic algorithm for mining quantitative association rules", In *IJCAI*, **7**, pp. 1035–1040 (2007, January).

27. Taboada, K., Gonzales, E., Shimada, K., et al. "Association rule mining for continuous attributes using genetic network programming", *IEEJ Transactions on Electrical and Electronic Engineering*, **3**(2), pp. 199–211 (2008).

28. Yan, X., Zhang, C., and Zhang, S. "Genetic algorithm-based strategy for identifying association rules without specifying actual minimum support", *Expert Systems with Applications*, **36**(2), pp. 3066–3076 (2009).

29. Martínez-Ballesteros, M., Martínez-Álvarez, F., Troncoso, A., and Riquelme, J.C. "An evolutionary algorithm to discover quantitative association rules in multidimensional time series", *Soft Computing*, **15**(10), p. 2065 (2011).

30. Martín, D., Rosete, A., Alcalá-Fdez, J., and Herrera, F. "A multi-objective evolutionary algorithm for mining quantitative association rules", In *Intelligent Systems Design and Applications (ISDA)*, *2011 11th International Conference on*, pp. 1397–1402, IEEE (November, 2011).

31. Minaei-Bidgoli, B., Barmaki, R., and Nasiri, M. "Mining numerical association rules via multi-objective genetic algorithms", *Information Sciences*, **233**, pp. 15–24 (2013).

32. Martín, D., Rosete, A., Alcalá-Fdez, J., and Herrera, F. "QAR-CIP-NSGA-II: A new multi-objective evolutionary algorithm to mine quantitative association rules", *Information Sciences*, **258**, pp. 1–28 (2014).

33. Martín, D., Alcalá-Fdez, J., Rosete, A., and Herrera, F. "NICGAR: A niching genetic algorithm to mine a diverse set of interesting quantitative association rules", *Information Sciences*, **355**, pp. 208–228 (2016).

34. Indira, K. and Kanmani, S. "Mining association rules using hybrid genetic algorithm and particle swarm optimisation algorithm", *International Journal of Data Analysis Techniques and Strategies*, **7**(1), pp. 59–76 (2015).

35. Agarwal, A. and Nanavati, N. "Association rule mining using hybrid GA-PSO for multi-objective optimisation", In *Computational Intelligence and Computing Research (ICCIC), 2016 IEEE International Conference on*, pp. 1–7, IEEE (December, 2016).

36. Sarkar, S., Lohani, A., and Maiti, J. "Genetic algorithm-based association rule mining approach towards rule generation of occupational accidents", In *International Conference on Computational Intelligence, Communications, and Business Analytics*, pp. 517–530, Springer, Singapore (March, 2017).

37. Djenouri, Y., Belhadi, A., Fournier-Viger, P., and Fujita, H. "Mining diversified association rules in big datasets: A cluster/GPU/genetic approach", *Information Sciences*, **459**, pp. 117–134 (2018).

38. Kumar, P. and Singh, A.K. "Efficient generation of association rules from numeric data using genetic algorithm for smart cities", In *Security in Smart Cities: Models, Applications, and Challenges* pp. 323–343, Springer, Cham. (2019).

39. Martínez-Ballesteros, M., Bacardit, J., Troncoso, A., and Riquelme, J.C. "Enhancing the scalability of a genetic algorithm to discover quantitative association rules in large-scale datasets", *Integrated Computer-Aided Engineering*, **22**(1), pp. 21–39 (2015).

40. Padillo, F., Luna, J.M., Herrera, F., et al. "Mining association rules on big data through MapReduce genetic programming", *Integrated Computer-Aided Engineering*, **25**(1), pp. 31–48 (2018).

41. Martín, D., Martínez-Ballesteros, M., García-Gil, D., et al. "MRQAR: A generic MapReduce framework to discover quantitative association rules in big data problems", *Knowledge-Based Systems*, **153**, pp. 176–192 (2018).

42. Haery, A., Salmasi, N., Yazdi, M.M., and Iranmanesh, H. "Application of association rule mining in supplier selection criteria", *World Academy of Science, Engineering and Technology*, **40**(1), pp. 358–362 (2008).

43. Srikant, R. and Agrawal, R. "Mining quantitative association rules in large relational tables", In *AcmSigmod Record*, **25**(2), pp. 1–12, ACM (1996, June).

44. Shenoy, P.D., Srinivasa, K.G., Venugopal, K.R., and Patnaik, L.M. "Dynamic association rule mining using genetic algorithms", *Intelligent Data Analysis*, **9**(5), pp. 439–453 (2005).

45. Kuo, R.J., Chao, C.M., and Chiu, Y.T. "Application of particle swarm optimization to association rule mining", *Applied Soft Computing*, **11**(1), pp. 326–336 (2011).

46. Beiranvand, V., Mobasher-Kashani, M., and Bakar, A.A. "Multi-objective PSO algorithm for mining numerical association rules without a priori discretization", *Expert Systems with Applications*, **41**(9), pp. 4259–4273 (2014).

47. Freitas, A.A., *Data Mining and Knowledge Discovery with Evolutionary Algorithms*, Springer (2002).

48. Ghosh, A. and Nath, B. "Multi-objective rule mining using genetic algorithms", *Information Sciences*, **163**, pp. 123–133 (2004).

49. Prajapati, D.J., Garg, S., and Chauhan, N.C. "Interesting association rule mining with consistent and inconsistent rule detection from big sales data in distributed environment", *Future Computing and Informatics Journal*, **2**(1), pp. 19–30 (2017).

50. Merceron, A. and Yacef, K. "Interestingness measures for association rules in educational data", *EDM*, **8**, pp. 57–66 (2008).

51. Rokh, B., Mirvaziri, H., and Eftekhari, M. "Proposing an efficient combination of interesting measures for mining association rules via NSGA-II", In *Technology, Communication and Knowledge (ICTCK)*, 2014 International Congress on, pp. 1–7, IEEE (November, 2014).

52. Luna, J.M., Romero, J.R., and Ventura, S. "Grammar-based multi-objective algorithms for mining association rules", *Data & Knowledge Engineering*, **86**, pp. 19–37 (2013).

53. Hsieh, Y., Lee, P., and You, P. "Immune based evolutionary algorithm for determining the optimal sequence of multiple disinfection operations", *Scientia Iranica*, **26**(2), pp. 959–974 (2019). DOI: 10.24200/sci.2018.20324

54. Sadeghi, H., Zolfaghari, M., and Heydarizade, M. "Estimation of electricity demand in residential sector using genetic algorithm approach", *Journal of Industrial Engineering & Production Research*, **22**(1), pp. 43–50 (2011).

55. Ostadi, B., MotamediSedeh, O., Husseinzadeh Kashan, A., and Amin-Naseri, M. "An intelligent model to predict the day-ahead deregulated market clearing price: a hybrid NN, PSO and GA approach", *Scientia Iranica*, **26**(6), pp. 3846–3856 (2019). DOI: 10.24200/sci.2018.50910.1909

56. Russell, S.J. and Norvig, P., *Artificial Intelligence A Modern Approach*, Pearson Education (2008).

57. Goldberg, D.E., *Genetic Algorithms in Search Optimization and Machine Learning*, Addison Wesley, p. 41 (1989).

58. Sonagara, D. and Badheka, S. "Comparison of basic clustering algorithms", *Int. J. Comput. Sci. Mob. Comput*, **3**(10), pp. 58–61 (2014).

59. Dabbagh, H., GhodratiAmiri, G., and Shaabani, S. "Modal data-based approach to structural damage identification by means of imperialist competitive optimization algorithm", *Scientia Iranica*, **25**(3), pp. 1070–1080 (2018). DOI: 10.24200/sci.2017.4590

60. Martínez-Ballesteros, M. and Riquelme, J.C. "Analysis of measures of quantitative association rules", In *International Conference on Hybrid Artificial Intelligence Systems*, pp. 319–326, Springer, Berlin, Heidelberg (May, 2011).

61. Picek, S. and Golub, M. "Comparison of a crossover operator in binary-coded genetic algorithms", *WSEAS Transactions on Computers*, **9**, pp. 1064–1073 (2010).

62. Guvenir, H. and Uysal, I., *Bilkent University Function Approximation Repository* (2000). <http://funapp.cs.bilkent.edu.tr/>

63. Moslehi, F., Haeri, A., and Moini, A. "Analyzing and investigating the use of electronic payment tools in Iran using data mining techniques", *Journal of AI and Data Mining*, **6**(2), pp. 417–437 (2018). DOI: 10.22044/jadm.2017.5352.1643

**Biographies**

**Fateme Moslehi** received her BSc degree in Computer Engineering from University of Shahid Rajaee, Iran in 2014. She pursued MSc degree in 2015 and received it in Information Technology Engineering from Iran University of Science and Technology in 2017. Her research interests are data mining techniques such as rule mining and clustering and soft computing methods such as genetic algorithm and particle swarm optimization.

**Abdorrahman Haeri** is an Assistant Professor of Industrial Engineering at Iran University of Science and Technology, Iran. He received his PhD from the Department of Industrial Engineering, College of Engineering, and University of Tehran, Iran in 2013. He holds an MSc degree in Industrial Engineering, Sharif University of Technology, Tehran, Iran in 2008. His main areas of teaching and research interests include data envelopment analysis, network design, and data mining. He has published several papers in international conferences and academic journals including Safety Science, Scientometrics, Energy Sources, Part B: Economics, Planning, and Policy, Journal of Engineering Manufacture: Part B, Int. J. Production Research, Int. J. Advanced Manufacturing and Technology, Int. J. Productivity and Quality Management (IJPQM).