

Obtaining optimal quality measures for quantitative association rules

M. Martínez-Ballesteros, A. Troncoso, F. Martínez-Álvarez, J.C. Riquelme

A B S T R A C T

There exist several works in the literature in which fitness functions based on a combination of weighted measures for the discovery of association rules have been proposed. Nevertheless, some differences in the measures used to assess the quality of association rules could be obtained according to the values of the weights of the measures included in the fitness function. Therefore, user's decision is very important in order to specify the weights of the measures involved in the optimization process. This paper presents a study of well-known quality measures with regard to the weights of the measures that appear in a fitness function. In particular, the fitness function of an existing evolutionary algorithm called QARGA has been considered with the purpose of suggesting the values that should be assigned to the weights, depending on the set of measures to be optimized. As initial step, several experiments have been carried out from 35 public datasets in order to show how the weights for confidence, support, amplitude and number of attributes measures included in the fitness function have an influence on different quality measures according to several minimum support thresholds. Second, statistical tests have been conducted for evaluating when the differences in measures of the rules obtained by QARGA are significative, and thus, to provide the best weights to be considered depending on the group of measures to be optimized. Finally, the results obtained when using the recommended weights for two real-world applications related to ozone and earthquakes are reported.

Keywords:

Quality measures
Quantitative association rules
Fitness function
Evolutionary algorithms

1. Introduction

The discovery of association rules is an effective computational technique focused on the extraction and representation of meaningful relationships among different variables. It was first defined by Agrawal et al. in [1] but the authors only considered the use of discrete variables. Nonetheless, when the domain of the variables involved in the rule extraction process is continuous, the rules obtained are called quantitative. In this case, they are called quantitative association rules and will hereinafter be referred to QAR. One of the first and most used algorithms is Apriori, which was also proposed by Agrawal et al. [2] one year later.

The mining of QAR is typically associated with the non-supervised learning, in which datasets lack a priori information about the internal structure of data or about how attributes are interrelated. Given this kind of input, the challenging task faced by QAR is to find groups of attributes that exhibit similar behavior. These groups must be formulated as comprehensive rules so that the relationships existing among the attributes can be easily interpreted.

There exist several measures to assess the quality of the QAR. All these measures are conceived to separately evaluate different properties of the rules. In this sense, it is quite common to model the rule extraction process by means of a multi-objective (MO) fitness function. This function aims at jointly maximizing a set of measures and its election depends on the type of rules searched for.

Different strategies can be found in the literature to solve MO problems. On one hand, there exist Pareto-based MO algorithms, which attempt at discovering the best trade-off among conflicting objectives. On the other hand, many fitness functions seeking for the optimization of a single objective can also be found in the literature. In them, some input parameters are required to weight, and therefore favor or penalize, the relevance of the measures. The improvement of certain measures is easily achieved by using these functions, insofar as such measures do not assess conflicting properties.

A vast majority of works focused on QAR mining only include the measures support and confidence in the fitness function. Nevertheless, there are many other measures, as gathered and reported in [17], that are ignored in this configuration. The negative effect that might be caused in other properties is simply not taken into account. All the measures considered in this work as well as their mathematical formulation are summarized in Table 1, where $n(X)$ is the number of occurrences of the itemset X in the dataset and N is the total number of instances in the dataset. ND

Table 1
Measures to assess the quality of QAR.

Measures	Equation	Description	Range
$Sup(X)$	$n(X)/N$	Coverage of X	[0, 1]
$Sup(X \Rightarrow Y)$	$n(X \cap Y)/N$	Generality of the rule	[0, 1]
$Conf(X \Rightarrow Y)$	$sup(X \Rightarrow Y)/sup(X)$	Reliability of the rule	[0, 1]
$Lift(X \Rightarrow Y)$	$sup(X \Rightarrow Y)/(sup(X) \cdot sup(Y))$	Interest of the rule <ul style="list-style-type: none"> • Value < 1 : X and Y (ND) • Value = 1 : X and Y (I) • Value > 1 : X and Y (PD) 	[0, $+\infty$)
$Gain(X \Rightarrow Y)$	$conf(X \Rightarrow Y) - sup(Y)$	Implication of the rule	[-0.5, 1]
$Certainty\ Factor(X \Rightarrow Y)$	$(conf(X \Rightarrow Y) - sup(Y))/(1 - sup(Y))$ <ul style="list-style-type: none"> • If $conf(X \Rightarrow Y) > sup(Y)$: • If $conf(X \Rightarrow Y) \leq sup(Y)$: $(conf(X \Rightarrow Y) - sup(Y))/sup(Y)$	Gain normalized <ul style="list-style-type: none"> • Value < 0 : X and Y (ND) • Value = 0 : X and Y (I) • Value > 0 : X and Y (PD) 	[-1, 1]
$Leverage(X \Rightarrow Y)$	$sup(X \Rightarrow Y) - sup(X)sup(Y)$	Strength of the rule <ul style="list-style-type: none"> • Value < 0 : X and Y (ND) • Value = 0 : X and Y (I) • Value > 0 : X and Y (PD) 	[-0.25, 0.25]
$Accuracy(X \Rightarrow Y)$	$sup(X \Rightarrow Y) + sup(\neg X \Rightarrow \neg Y)$	Veracity of the rule	[0, 1]

stands for negatively dependent, *PD* for positively dependent and *I* for independent.

A preliminary study about how the weights, in a fitness function based on a sum of weighted measures, have an influence on the quality measures was provided in [20]. However, this study was reduced as it was applied only for the weights associated to the support and confidence measures.

The main goal of this work is to conduct an extensive study to evaluate the effect of varying different weights for different ranges in a fitness function. Another significant goal is to provide the researcher with several guidelines to set the weights of any fitness function, according to the preset objectives that are wanted to be maximized. Additionally, multiple relationships between the weight variations and the quality measures are inferred in this work. The algorithm selected to perform such tasks, to which different experimental setups have been applied, is QARGA [16].

The remainder of the paper is as follows. Section 2 overviews the most relevant works recently published with techniques using a weighted sum-based fitness function. The QARGA algorithm as well as the design of the experimental setup is described in Section 3. Section 4 provides a description of the datasets. It also includes the setup of the parameters involved in the process. Moreover the analysis itself is reported as well as some statistical tests. Finally, Section 5 summarizes the conclusions drawn from the analysis conducted.

2. Related work

This section is devoted to examine the latest and most relevant works recently published. In particular, those focused on the extraction of QAR by means of fitness functions constructed as a combination of weighted objectives.

Although the optimization of only support and confidence is a usual strategy for defining fitness functions, some works use fitness functions even simpler. Such is the case of EARMGA, an evolutionary algorithm introduced in [28] that only considered the confidence as quality measure to be maximized. It is worth mentioning that the authors did not compute the actual minimum support.

An approach called QuantMiner was introduced in [26]. It used confidence and support as metrics to evaluate its performance. This genetic algorithm was defined in order to provide satisfactory intervals for numeric attributes. The algorithm was compared to GAR [22,23], another algorithm with similar features, to assess the accuracy.

An approach based on an evolutionary algorithm providing an antecedent with a variable number of attributes was published in 2001 [21]. This algorithm, called GENAR, included support, confidence and number of recovered instances in the fitness function. A similar fitness function, but adding the comprehensibility and the amplitude of the intervals to the aforementioned set of attributes, was used in Pachón et al. [24] naming the algorithm GAR-Plus. A comparative analysis of the quality of EARMGA, GAR and EARMGA was presented in [5], where all the three algorithms were applied to two different datasets than those of the original papers. The coverage and their efficiency were reported.

Another genetic algorithm was proposed in 2011 by Soto and Olaya-Benavides [27]. This algorithm included four weighted measures in the fitness function: support, confidence, comprehensibility and interest. A weighted support based on the individual weight of the items according to their importance in the dataset was calculated in [25].

The extraction of QAR can also be found in bioinformatics. For instance, the analysis of microarray gene-expression data by means of QAR based on half-spaces was introduced in [10]. The measures selected to form the fitness function were interestingness, support, confidence and coverage.

All the algorithms mentioned so far only explored positive dependencies. However, it is well-known that negative dependencies can also provide meaningful information. In this sense, a genetic algorithm to extract QAR with positive and negative dependencies was proposed in [3]. To achieve such task, the authors formed a fitness function composed of the five quality measures: support, confidence, number of attributes, recovered records and amplitude. A multiobjective version was introduced two years later by the same authors using, this time, a Pareto-based evolutionary algorithm [4].

Recently, the study of the positive and negative dependencies by means of a multi-objective evolutionary algorithm has been proposed [19]. MOPNAR's fitness function consists of three measures to be optimized: the interestingness, the comprehensibility and what they named performance, a modification of the coverage. The performance and efficiency of MOPNAR was evaluated by comparing the algorithm to up to nine different approaches. One of the main benefits of this algorithm claimed by the authors is its low computational cost. Note that this algorithm extends MOEA/D [14] by introducing two new components into the evolutionary model: an external population and a restarting process. MOEA/D was designed to deal with especially complicated Pareto sets.

The extraction of fuzzy QAR has also been addressed. In particular, the use of Pareto-optimal fuzzy rules as candidate rules

Table 2
Summary of measures used in fitness functions reviewed in Section 2.

Algorithms	Quality measures considered						
	Support	Confidence	Re-covered	# Attributes	Amplitude	Interest	Coverage
GENAR, 2001 [21]	✓	✓	✓				
Georgii et al., 2005 [10]	✓	✓				✓	✓
Alatas and Akin, 2006 [3]	✓	✓		✓	✓		
QuantMiner, 2007 [26]	✓	✓					
EARMGA, 2009 [28]		✓					
MOEA/D, 2009 [14]					✓	✓	
Soto et al., 2011 [27]	✓	✓		✓		✓	
GAR-Plus, 2012 [24]	✓	✓	✓	✓	✓		
Pears et al., 2013 [25]	✓						
MOPNAR, 2014 [19]					✓	✓	

in classification problems by means of a genetic algorithm was proposed in [12]. The standard single objective genetic algorithm (SOGA) maximizes a weighted sum-based fitness function composed of the number of correctly classified training patterns, the number of fuzzy rules of such set and the total number of attributes in the antecedent of the rule. Almost the same authors examined the effect of using weighted sum-based fitness functions for parent selection and generation update [13]. That effect was tested on the performance of NSGA-II [7] for a high-dimensional space of a multi-objective problem. An interesting review of QAR learning was presented in [8]. It explores fuzzy (and also crisp) rule extraction processes by means of evolutionary algorithms.

Table 2 summarizes the measures used in the works reviewed in this section. It can be noted that support and confidence present a generalized use. However, the optimization of only these measures is not usually enough since: (a) if only the support is maximized, very general QAR could be obtained since the interval amplitudes might reach the whole domain of each attribute and (b) if only the confidence is maximized, rules with few instances might be generated and negative dependencies among attributes might be missed. For this reason, other existing measures to assess the quality of QAR were introduced in Section 1 and will be jointly used for the rest of this work.

From the analysis of this section, it can be easily concluded that there exist many fitness function proposals based on a combination of weighted objectives in a single equation. However, their performance is very sensitive to the choice of the weights of the measures included within the fitness function and there is no even consensus on which measures and what weights have to be used. Therefore, there exists a need to provide some hints to create and initialize fitness functions to mine QAR. To fill such a gap is, precisely, the main goal of the work proposed here.

3. Methodology

This section presents a general description of the QARGA algorithm used to obtain QAR. The experimentation framework carried out in order to analyze how QAR are influenced regarding the weights of the fitness function is also described.

3.1. Description of QARGA

The QARGA algorithm is used to study the QAR obtained, according to the weights which provide more or less relevance to the measures included in the fitness function. A detailed description of the algorithm can be found in [18].

QARGA is a real-coded genetic algorithm designed to discover existing relationships, specifically QAR, among several variables. Namely, each individual of the population constitutes a rule. These

Table 3
Public datasets.

Dataset	Records	Attributes
Ailerons (AI)	7154	41
Baseball (BA)	337	17
Basketball (BK)	96	5
Bodyfat (FA)	252	18
Bolts (BL)	40	8
Buying (BU)	100	40
Computer activity (CA)	8192	22
Country (CN)	122	21
College (CO)	236	21
Education (ED)	1500	44
Elevators (EV)	16 599	19
Fried (FR)	40 768	11
House_16H (HH)	22 784	17
Kinematics (KI)	8192	9
Longley (LO)	16	7
Mortgage (MO)	1049	17
Normal Body Temperature (NT)	130	3
Plastic (PL)	1650	3
2Dplanes (PN)	40 768	11
Pw Linear (PW)	200	11
Pollution (PO)	60	16
Pole Telecomm (PT)	9065	49
Pyrimidines (PY)	74	28
Quake (QU)	2178	4
Read (RE)	681	26
School (SC)	62	20
Sleep (SL)	57	8
Stock price (SP)	950	10
Televisions (TV)	40	5
Treasury (TR)	1049	17
Triazines (TZ)	186	61
Usnews College (US)	1269	32
Vineyard (VY)	52	4
Weather Ankara (WA)	1609	11
Weather Izmir (WI)	1641	11

rules are then subjected to an evolutionary process, in which the mutation and crossover operators are applied and, at the end of the process, the individual that presents the best fitness is designated as the best rule.

The individuals are codified using two arrays, each of them with a number of elements equal to the number of attributes of the dataset. The first array is composed of upper and lower limits of intervals to which the attributes belong to, and the second one is composed of values 1 and 2, depending on if the attribute belongs to antecedent or consequent, respectively, or value 0 if the attribute is not in the rule.

The individuals of the initial population are randomly generated. That is, the number of attributes appearing in the rule, the values that indicate if the attribute belongs to antecedent or consequent and the intervals for each attribute are randomly generated. Several

constraints are considered to assure that the individuals represent sound rules when the genes are generated.

The genetic operators are briefly described as follows.

1. *Selection*: An elitist strategy is used to replicate the individual with the best fitness. By contrast, a roulette selection method is used for the remaining individuals rewarding the best individuals according to the fitness.
2. *Crossover*: Two parent individuals, chosen by means of the roulette selection, are combined to generate a new individual. First, all the attributes associated to each parent are analyzed. The attribute would belong to the antecedent in the new individual if the attribute in both parents belongs to the antecedent, and to the consequent if the attribute belongs to the consequent in both parents. Moreover, in these two cases, the interval is obtained generating two random numbers among the limits of the intervals of both parents. Thus, the lower interval is generated by means of a random number that belongs to the interval formed by both lower intervals of the parents; the upper interval is analogously calculated. Otherwise, if the attribute belongs to the antecedent in the father individual and to the consequent in the mother individual or vice versa, then the attribute would be randomly chosen between both parents. In this case, the interval of such attribute is the interval of the ancestor selected.
3. *Mutation*: It consists in varying one gene of the individuals. The mutation can be focused on the modification of the attribute from antecedent to consequent, from consequent to antecedent or from antecedent or consequent to null, or on the intervals, in which three different cases are possible: equiprobable mutation of the upper limit, of the lower limit or of both limits of the interval. For this aim, a random value between 0 and the maximum amplitude is generated and will be added or subtracted to the limit of the interval which is randomly selected.

The fitness of each individual in the evolutionary process allows determining which are the best candidates to remain in subsequent generations. In order to make this decision, its calculation involves several measures that provide information about the rules. The fitness function has been designed to maximize a combination of different measures of QAR. The fitness function proposed in [16] to be maximized by QARGA was

$$f(\text{rule}) = w_s \cdot \text{sup} + w_c \cdot \text{conf} + w_n \cdot n\text{Attrib} - w_a \cdot \text{ampl} - w_r \cdot \text{recov} \quad (1)$$

where *sup* is the support of the rule, *conf* is the confidence of the rule, *recov* is the ratio of instances which had already been covered, *nAttrib* is the number of attributes appearing in the rule and *ampl* is the average size of intervals of the attributes belonging to the rule.

It can be appreciated that the fitness function presents a set of weights, w_s, w_c, w_n, w_a and w_r , to drive the process of search of rules and will vary depending on the required rules. Therefore, the user should be aware of the importance of each measure in order to specify the weights considering that significant differences in the QAR quality measures could be obtained.

3.2. Experimental design

It is well known that one of the shortcomings of a fitness function defined as a sum of weighted measures is the parametrization of such weights. A detailed study of the relative influence of each weight on the rules obtained by QARGA is provided in Section 4.

With that purpose, several experiments were designed to highlight the main differences in measures performance when the

weights of the fitness function are modified according to several minimum support thresholds, namely 0, 0.05 and 0.1. Thus, QARGA is forced to learn QAR with an established minimum support as it penalizes the fitness function of the individuals of the population which do not satisfy the minimum support threshold.

First of all, the experimentation is based on the analysis with regard to the weights for the confidence and support measures, and second, with regard to the number of attributes and the amplitude of the intervals. In particular, two experiments, one for each pair of measures, have been carried out.

The first experiment consists in varying the weights for confidence and support measures by using a given minimum support threshold. Specifically, the weight values for support and confidence measures, w_s and w_c , vary from 0 to 1 with increments of 0.1 (11 different values for both weights). Hence, QARGA was run 363 ($3 \times 11 \times 11$) times in total for each dataset. The results are summarized in a table for each minimum support threshold as follows: the weight of confidence is fixed and the weight of support varies from 0 to 1 (or vice versa) in order to analyze how the weight of the confidence measure (or the support measure) affects on the remaining measures, independently of the weights of last ones. Note that the rest of the weights included in QARGA's fitness function have been set to 1 in order to avoid their influence in the results and to ensure that the remaining measures are present in the fitness function.

The same experiment for the weights of the amplitude and number of attributes measures is made with a minimum support threshold of 0.05. In this case, the weights considered for the remaining measures have been set to 1 for w_s and 0.5 for w_c . Thus, QARGA was run 121 (11×11) times for each dataset.

Second, statistical tests have been conducted for all measures to evaluate if the differences in measures of the rules obtained by QARGA, reported from two previous experiments, are significative. For that, a minimum support threshold of 0.05 has been considered, for each dataset a weight, w_n, w_a, w_s or w_c , is set and the mean of each measure for the 11 different values of the weight, w_a, w_n, w_c or w_s , respectively, is computed. For example, w_n is fixed and the mean of each measure for the 11 values of w_a is calculated.

Finally, QARGA was executed 5 times and 20 rules were obtained in each execution to compare the measures when using a set of default weights and the recommended weights provided for the results of statistical tests for two real-world applications related to ozone and earthquakes.

4. Experimental study

4.1. Datasets description

This section presents the main features of the datasets used in the study carried out in this work. Several datasets have been retrieved from the public BUFA repository [11]. In particular, the thirty-five public datasets from BUFA repository used in [17]. Note that Buying, Country, College, Education, Read and Usnews College have been preprocessed using K-means Imputation method proposed in [9] (available in the KEEL tool [6]) in order to deal with missing values. Table 3 describes the number of attributes and instances for the datasets from BUFA repository.

Likewise, two different real-world datasets have been analyzed to validate the conclusions achieved after the study performed. Specifically, datasets from a meteorological and seismological context have been used.

First, the tropospheric ozone dataset is composed of climatological time series such as temperature, humidity, direction and speed of the wind, several variables such as the hour of the day and the day of the week and, finally, the tropospheric ozone. These variables have

influence on the ozone concentration in the atmosphere, which is the target agent. All variables have been retrieved from the meteorological station of the city of Seville in Spain for the months from July to August during the years 2003 and 2004, generating a dataset with 7 quantitative attributes and 1488 instances. These months have been selected due to the highest concentration of ozone reported.

The earthquake dataset [15] has been retrieved from the Spanish's National Geographical Institute. This dataset consists of 4 quantitative attributes and 873 instances related to the location and the magnitude of Spanish earthquakes collected from 1981 to 2008. Table 4 summarizes the number of attributes and instances for the ozone and earthquakes datasets.

4.2. Parameters setup

As for the values for the main parameters of QARGA, it is noteworthy that these values have been used for all executions independently of the different values for the weights.

The main parameters of QARGA are 100 for the size of the population, 100 for the number of generations, 0.1 for the mutation probability p_{mut} of the individuals and 0.2 for the mutation probability p_{mutgen} of each gene in the individual. The maximum number of attributes which could be included in both the antecedent and consequent are 10 and 5, respectively. Note that antecedent and consequent must contain one attribute at least. QARGA has obtained 100 QAR for each public dataset and for each configuration of the weights of the fitness function.

In contrast, only 20 QAR have been found for the two real-world datasets.

4.3. Analysis of the quality measures

A detailed analysis of the QAR quality measures according to variations in the weights of the measures included in the fitness

Table 4
Ozone and earthquakes datasets.

Dataset	Records	Attributes
Ozone	1488	7
Earthquakes	873	4

Table 5
Average values obtained by QARGA with minimum support threshold of 0 according to w_s and w_c .

ID	w_s	w_c	Covered instances (%)	Rule support (%)	Lift	Confidence (%)	Accuracy (%)	Amplitude (%)	#Attributes	Leverage	Certainty factor	Gain
0	0.0	[0-1]	43.8	0.82	2354.4	99.2	93.3	0.59	13.9	0.0076	0.99	0.92
1	0.1	[0-1]	46.4	0.85	2347.1	99.2	93.3	0.62	13.7	0.0076	0.98	0.92
2	0.2	[0-1]	48.0	0.88	2331.9	99.2	93.3	0.64	13.6	0.0077	0.99	0.92
3	0.3	[0-1]	50.3	0.92	2342.7	99.2	93.4	0.63	13.5	0.0078	0.99	0.92
4	0.4	[0-1]	53.1	0.95	2365.6	99.2	93.2	0.64	13.4	0.0078	0.99	0.92
5	0.5	[0-1]	55.9	0.99	2335.4	99.1	93.3	0.66	13.4	0.0079	0.98	0.92
6	0.6	[0-1]	60.0	1.04	2322.7	99.2	93.4	0.67	13.4	0.0079	0.98	0.92
7	0.7	[0-1]	63.1	1.09	2319.7	99.1	93.3	0.71	13.4	0.0080	0.98	0.91
8	0.8	[0-1]	67.3	1.16	2343.9	99.2	93.4	0.71	13.4	0.0081	0.98	0.91
9	0.9	[0-1]	70.0	1.22	2345.0	99.1	93.5	0.73	13.4	0.0081	0.98	0.91
10	1.0	[0-1]	74.6	1.30	2365.0	99.2	93.5	0.77	13.5	0.0081	0.98	0.91
11	[0-1]	0.0	57.1	1.00	2224.7	92.3	95.4	0.79	13.4	0.0078	0.91	0.87
12	[0-1]	0.1	57.8	1.02	2256.5	99.2	93.2	0.75	13.6	0.0078	0.98	0.91
13	[0-1]	0.2	58.1	1.03	2328.7	99.8	93.0	0.68	13.5	0.0079	0.99	0.92
14	[0-1]	0.3	57.8	1.03	2316.9	99.9	93.1	0.68	13.5	0.0079	0.99	0.92
15	[0-1]	0.4	57.9	1.03	2347.5	100.0	93.2	0.65	13.5	0.0079	0.99	0.92
16	[0-1]	0.5	57.0	1.01	2369.5	100.0	93.2	0.65	13.6	0.0079	0.99	0.92
17	[0-1]	0.6	58.2	1.03	2367.5	100.0	93.1	0.66	13.6	0.0079	0.99	0.92
18	[0-1]	0.7	57.9	1.03	2387.5	100.0	93.2	0.65	13.5	0.0079	0.99	0.92
19	[0-1]	0.8	56.8	1.02	2382.0	100.0	93.2	0.63	13.5	0.0079	0.99	0.92
20	[0-1]	0.9	57.1	1.01	2386.2	100.0	93.2	0.62	13.4	0.0078	0.99	0.92
21	[0-1]	1.0	57.0	1.01	2406.3	100.0	93.3	0.61	13.5	0.0079	0.99	0.92

function to be optimized by QARGA is presented in this section. To this aim, QARGA has been applied following two phases.

In the first phase, the performance of several quality measures is studied according to different minimum support thresholds and variations of w_s and w_c . In the second phase, the behavior of the same set of quality measures is analyzed regarding variations of w_n and w_g . The results obtained by QARGA when optimizing the fitness function with variations in the weights of the measures are discussed. Afterwards, a non-parametric statistical analysis will be conducted to detect significant differences in the quality measures of the rules obtained by QARGA. Finally, the conclusions achieved will be applied to real-world datasets to show if better results are obtained when the best sets of weights are applied instead of using non-weighted objectives in the fitness function.

As described in Section 3.2, QARGA has been executed 363 times for each dataset, that is, a total of 12 705 executions in the first phase. In addition, QARGA has been run 121 times for each dataset, that is, 4235 executions in the second phase. In order to perform the study, the average results for the 35 datasets using the same values of the weights as much in the first as the second phase have been retrieved. Several interestingness measures have been calculated to assess the quality of the QAR obtained by QARGA for each execution. In particular, support, confidence, lift, gain, leverage, accuracy, number of attributes, amplitude of the attributes, number of the rules obtained and percentage of covered records have been computed. The equation of these measures can be found in Table 1 and are thoroughly detailed in [16].

4.3.1. Analyzing the influence of w_s and w_c

Regarding the first phase, the average results of the 12 705 runs for each minimum support threshold are reported in Tables 5, 7 and 9, respectively.

Tables 6, 8 and 10 show the studied quality measures grouped by similar behavior according to the w_s and w_c variations and the minimum support threshold 0, 0.05 and 0.1, respectively.

Tables 5, 7 and 9 are described as follows:

- In the first rows from 0 to 10, each row represents the average values of the considered quality measures for the QAR obtained in each dataset from executions that have a fixed w_s value and values for w_c (11 values) that vary from 0 to 1 summing up

11 × 35 runs for each row. Thus, the average results of the row identified by 0 correspond to the runs that use $w_s=0$ and $w_c \in [0, 1]$ in the fitness function for each dataset. The rest of the rows corresponds to each fixed value for w_s , respectively.

- According to the rows from 11 to 21, each row indicates the average values of the considered quality measures of the QAR found in each dataset from the runs that have a fixed w_c value and w_s varies from 0 to 1 (11 values for $w_s \times 35$ executions for each row).

Table 5 shows the average values in terms of covered instances, rule support, lift, confidence, accuracy, number of attributes, leverage, certainty factor and gain of the QAR obtained by QARGA with a minimum support threshold set to 0. It can be observed that the average confidence, accuracy, certainty factor and gain reach values close to 100% and 1. However, the average support obtained is around 1%, the rules found have extremely narrow intervals and a high number of attributes is obtained. In general terms, all the measures present a similar behavior regardless of w_s and w_c values. However, the percentage of covered instances is highly increased when increasing the w_s value. This measure achieves the maximum value when w_s is 1 but no dataset is completely covered due to the low support of the rules.

Table 6 shows the ten studied quality measures split into six groups according to the results reported in Table 5. It can be noted

Table 6
Performance of quality measures with minimum support equal to 0 according to w_s and w_c .

Weights	Quality measures grouped by similar behavior with minimum support 0					
	Support Leverage Covered instances	Gain Confidence Certainty factor	Accuracy	Lift	#Attributes	Amplitude
$w_s \uparrow$	\uparrow	=	=	=	=	\uparrow
$w_c \uparrow$	=	\uparrow if $w_c \leq 0.1$ = if $w_c > 0.1$	\downarrow if $w_c \leq 0.1$ = if $w_c > 0.1$	\uparrow	=	\downarrow

Table 7
Average values obtained by QARGA with minimum support threshold of 0.05 according to w_s and w_c .

ID	w_s	w_c	Covered instances (%)	Rules support (%)	Lift	Confidence (%)	Accuracy (%)	Amplitude (%)	#Attributes	Leverage	Certainty factor	Gain
1	0.0	[0-1]	96.1	4.4	519.8	86.0	68.0	11.3	7.3	0.0222	0.76	0.52
2	0.1	[0-1]	96.3	4.4	534.4	86.0	68.2	11.3	7.3	0.0223	0.76	0.52
3	0.2	[0-1]	96.6	4.5	521.5	85.8	68.3	11.3	7.2	0.0223	0.76	0.52
4	0.3	[0-1]	96.8	4.5	518.9	85.8	68.5	11.3	7.3	0.0223	0.76	0.52
5	0.4	[0-1]	97.1	4.5	520.9	85.8	68.7	11.2	7.2	0.0224	0.76	0.52
6	0.5	[0-1]	97.4	4.6	515.1	85.5	68.7	11.1	7.2	0.0221	0.75	0.52
7	0.6	[0-1]	97.5	4.6	531.2	85.5	68.8	11.2	7.2	0.0220	0.75	0.52
8	0.7	[0-1]	98.0	4.7	557.6	85.2	69.0	11.1	7.2	0.0221	0.75	0.52
9	0.8	[0-1]	98.3	4.8	583.8	85.4	69.3	11.0	7.2	0.0220	0.75	0.52
10	0.9	[0-1]	98.5	4.8	617.0	85.3	69.7	10.9	7.2	0.0220	0.75	0.53
11	1.0	[0-1]	98.7	4.8	610.4	85.4	69.5	10.9	7.2	0.0219	0.75	0.52
12	[0-1]	0.0	91.3	3.1	968.6	70.2	83.2	6.6	7.7	0.0172	0.63	0.56
13	[0-1]	0.1	93.6	3.4	951.4	75.5	80.8	7.2	7.7	0.0183	0.68	0.57
14	[0-1]	0.2	95.7	3.7	844.5	78.7	78.0	8.0	7.6	0.0194	0.70	0.56
15	[0-1]	0.3	97.2	4.0	782.7	82.2	75.1	8.9	7.5	0.0205	0.73	0.56
16	[0-1]	0.4	98.2	4.3	681.0	84.7	71.7	10.0	7.4	0.0216	0.75	0.54
17	[0-1]	0.5	98.8	4.7	583.2	87.3	68.5	11.0	7.2	0.0225	0.77	0.53
18	[0-1]	0.6	99.1	5.0	444.5	89.2	65.2	12.2	7.1	0.0234	0.78	0.51
19	[0-1]	0.7	99.3	5.3	334.4	91.3	62.2	13.3	7.0	0.0243	0.80	0.49
20	[0-1]	0.8	99.3	5.5	233.1	92.9	59.6	14.2	6.8	0.0250	0.81	0.48
21	[0-1]	0.9	99.4	5.7	145.9	94.3	57.2	15.1	6.7	0.0255	0.82	0.46
22	[0-1]	1.0	99.5	5.9	61.2	95.5	55.4	15.9	6.7	0.0259	0.83	0.45

that w_s is positively correlated with support, leverage, covered instances and amplitude, whereas the remaining quality measures are not affected by the variations of w_s . With respect to w_c , some differences can be observed from this table. For instance, although support, leverage and covered instances measures are dependent of w_s , these measures are not influenced by w_c . Alternatively, the lift is positively correlated with w_c but the amplitude is negatively correlated. However, confidence, gain and certainty factor are only increased when w_c achieves values between 0 and 0.1, contrary to the accuracy that presents an opposite behavior. A w_c greater than 0.1 does not cause alterations in the performance of these last three measures.

Figs. 1–3 summarize the values obtained for each group of measures when the minimum support threshold is 0. Note that only the three most representative measures are displayed due to the similar performance among the measures of the same group. Fig. 1 represents the support, leverage and covered instances measures. It can be observed that their values form an increasing inclined plane according to w_s . Fig. 2 visualizes the values obtained for the confidence, certainty factor and gain measures. These measures present an awning model reaching the highest values when w_c is greater than 0.1. Finally, the accuracy measure is shown in Fig. 3. It achieves the highest value when w_c is 0. The performance of this measure can be considered as a valley model.

Table 7 summarizes the average values in terms of covered instances, rule support, lift, confidence, accuracy, number of attributes, leverage, certainty factor and gain of the QAR found by QARGA when considering a minimum support threshold of 0.05.

Even though confidence, accuracy, certainty factor and gain have lesser values than those obtained for a minimum support threshold of 0, the percentage of covered records reaches values close to 100% and the support of the rules achieves values around 5%. On the other hand, the rules present a stronger dependency between the antecedent and consequent (leverage measure) when a minimum support threshold of 0.05 is considered. Furthermore, QAR have a less number of attributes and broader intervals, then the rules provided are more comprehensible for the user.

Table 8 displays the ten measures under study clustered into only three groups when a minimum support threshold of 0.05 is applied. It can be appreciated that the performance of these measures is completely different when no minimum support threshold is applied.

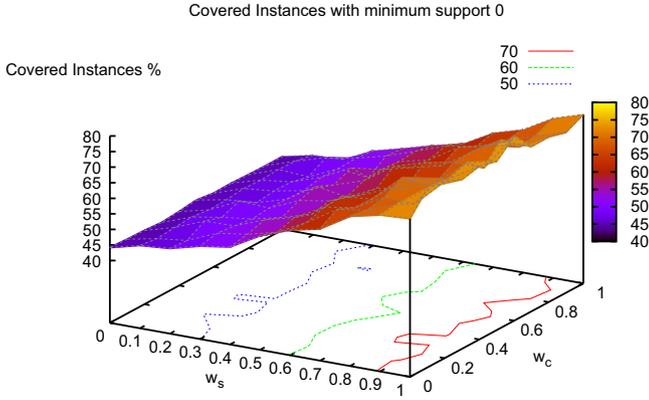


Fig. 1. Covered instances with minimum support 0 according to w_s and w_c .

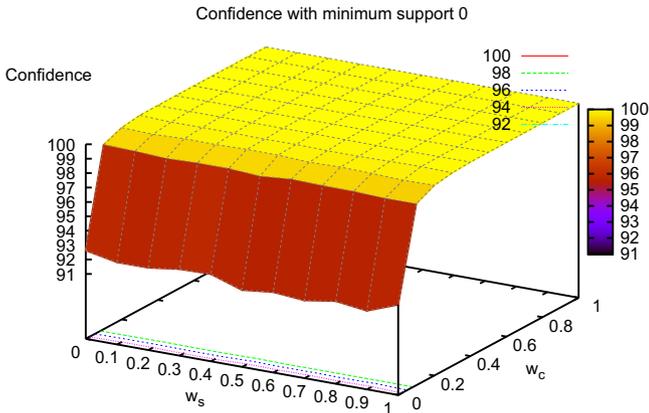


Fig. 2. Confidence with minimum support 0 according to w_s and w_c .

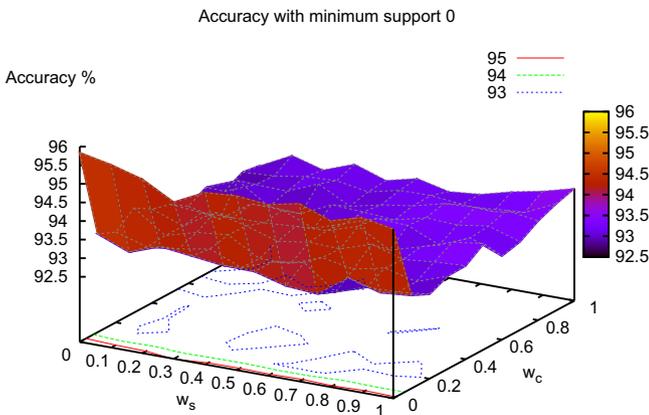


Fig. 3. Accuracy with minimum support 0 according to w_s and w_c .

For instance, the group composed of support, confidence, certainty factor, leverage and amplitude and the group formed by lift, accuracy, gain and number of attributes are only affected by w_c . These groups are positively and negatively correlated respectively according to w_c . Regarding the percentage of covered instances, it is positively correlated with w_s and w_c .

Figs. 4 and 5 illustrate the values obtained for two groups of measures when the minimum support threshold is 0.05. Fig. 4 represents the performance of confidence, support, leverage, amplitude and certainty factor measures. These measures reach the highest values when w_c is 1. Note that the confidence behaves as an increasing inclined plane with respect to w_c instead of presenting an awning model as in Fig. 2. Fig. 5 shows the values obtained for the accuracy measure and summarizes the behavior

Table 8

Performance of quality measures with minimum support 0.05 according to w_s and w_c .

Weights			
Quality measures grouped by similar behavior with minimum support 0.05			
Support	Lift	Gain	Covered instances
Leverage			
Amplitude			
Confidence			
Certainty factor			
$w_s \uparrow$	=	=	\uparrow
$w_c \uparrow$	\uparrow	\downarrow	\uparrow

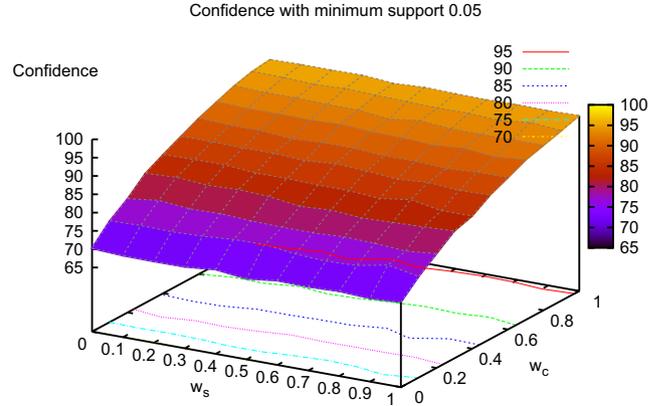


Fig. 4. Confidence with minimum support 0.05 according to w_s and w_c .

of lift, gain and number of attributes. In this case, these measures get the highest values when w_c is 0 and perform as a decreasing inclined plane relative to w_c . The percentage of covered records does not exhibit the same behavior and it is positively correlated according to w_s and w_c .

Table 9 presents the average values in terms of covered instances, rule support, lift, confidence, accuracy, number of attributes, leverage, certainty factor and gain of the QAR discovered by QARGA with a minimum support threshold of 0.1. It can be observed that support of the rules is above 10% and leverage measure is better than that of other minimum support thresholds, in particular, 0 and 0.05. Nevertheless, confidence, accuracy, certainty factor and gain measures present the worst values compared to the values achieved when a minimum support threshold of 0 or 0.05 is used. Specifically, confidence and accuracy are below 70% and certainty factor never reaches values above 50%. From this observation, if a minimum support threshold of 0.1 is applied, QARGA discovers rules less accurate, less interesting and reaches lesser gain of information on the rules regarding the consequent when the antecedent is also present.

Table 10 summarizes the ten measures under study grouped into only two groups when a minimum support threshold of 0.1 is considered. It can be noticed that all the measures behave equally regardless of the values of w_s and w_c . Only the confidence and certainty factor are slightly increased when increasing w_c .

Fig. 6 shows the average values obtained by the group of the confidence and certainty factor. These measures reach the highest values when w_c is 1. Similar to Fig. 4, the confidence behaves as an increasing inclined plane with regard to w_c .

From the first phase of the experimentation, we provide the following final remarks. First, the QAR obtained when the minimum support threshold is 0 are more specific. Therefore, the support and instances covered values are lesser and the number of attributes and accuracy are greater compared to the values obtained when the

minimum support threshold is 0.05 or 0.1. Second, although the confidence, gain, accuracy, and lift are better when the minimum support threshold is 0, it is desirable to apply a minimum support threshold in order to avoid support values below 1%. Moreover, it can be observed that better results are obtained if a minimum support threshold of 0.05 is applied instead of using 0.1. Taking into account this fact, the w_s setting is not important in the final results. And third, it has been observed that w_c is the most influential weight. Thus, values of w_c around 0.5 are desirable because not all measures are increased according to w_c .

Thereafter, it would be interesting to study the influence on the QAR quality measures of the rest of weights included in the fitness function of QARGA. Hence, a second phase has been carried out to analyze the influence of w_n and w_a over the quality measures under study.

4.3.2. Analyzing the influence of w_n and w_a

In the light of the results obtained in the first phase, a minimum support threshold of 0.05 is only considered and the remaining measures have been set to 1 for w_r and w_s , respectively, and 0.5 for w_c .

The average results of the 4235 runs are in Table 11 in a similar way to Tables 5, 7 and 9. Table 11 reports the average values in terms of covered instances, rule support, lift, confidence, accuracy, number of attributes, leverage, certainty factor and gain of the

Table 10

Performance of quality measures with minimum support 0.1 according to w_s and w_c .

Weights	Quality measures grouped by similar behavior with minimum support 0.1	
Lift	Gain	Confidence
Support	Leverage	Certainty factor
Accuracy	Amplitude	
#Attributes	Covered instances	
$w_s \uparrow$	=	=
$w_c \uparrow$	=	\uparrow

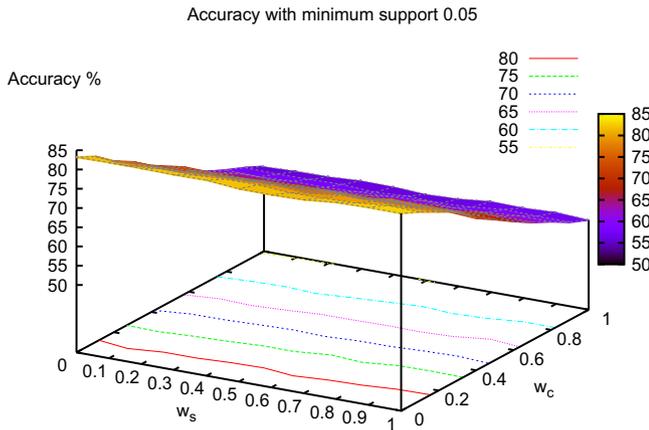


Fig. 5. Accuracy instances with minimum support 0.05 according to w_s and w_c .

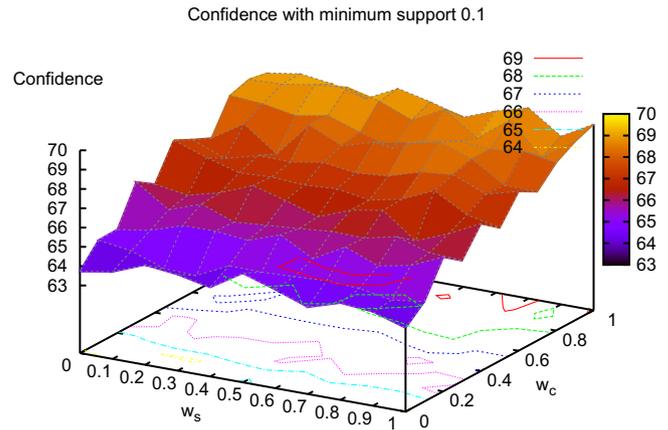


Fig. 6. Confidence with minimum support 0.1 according to w_s and w_c .

Table 9

Average values obtained by QARGA with minimum support threshold of 0.1 according to w_s and w_c .

ID	w_s	w_c	Covered instances (%)	Rule support (%)	Lift	Confidence (%)	Accuracy (%)	Amplitude (%)	#Attributes	Leverage	Certainty factor	Gain
1	0.0	[0-1]	99.9	12.7	702.0	66.8	68.2	12.1	4.5	0.0404	0.47	0.35
2	0.1	[0-1]	99.9	12.8	687.6	66.7	68.0	12.2	4.5	0.0402	0.47	0.35
3	0.2	[0-1]	99.9	12.8	694.8	66.7	68.0	12.2	4.5	0.0402	0.47	0.35
4	0.3	[0-1]	99.9	12.9	745.8	66.9	68.1	12.2	4.5	0.0403	0.47	0.35
5	0.4	[0-1]	99.9	12.9	708.5	66.8	68.0	12.2	4.5	0.0400	0.47	0.35
6	0.5	[0-1]	100.0	13.2	194.3	66.8	67.7	12.2	4.4	0.0414	0.47	0.35
7	0.6	[0-1]	99.9	12.9	677.2	66.8	68.2	12.3	4.5	0.0403	0.47	0.35
8	0.7	[0-1]	99.9	12.9	745.3	67.0	68.1	12.2	4.5	0.0402	0.47	0.35
9	0.8	[0-1]	99.9	13.0	736.2	67.1	68.3	12.2	4.5	0.0403	0.47	0.35
10	0.9	[0-1]	99.9	13.0	712.7	66.6	68.2	12.2	4.5	0.0401	0.47	0.35
11	1.0	[0-1]	99.8	12.8	715.8	66.8	67.7	12.2	4.5	0.0393	0.47	0.35
12	[0-1]	0.0	99.9	12.8	683.3	64.2	68.6	12.2	4.5	0.0399	0.45	0.35
13	[0-1]	0.1	99.9	12.9	650.9	64.5	68.4	12.2	4.5	0.0401	0.45	0.34
14	[0-1]	0.2	99.9	12.9	680.5	65.6	68.4	12.2	4.5	0.0402	0.46	0.35
15	[0-1]	0.3	99.9	12.9	652.3	66.0	68.2	12.2	4.5	0.0405	0.46	0.35
16	[0-1]	0.4	99.9	12.8	640.2	66.4	68.3	12.3	4.5	0.0405	0.47	0.35
17	[0-1]	0.5	99.9	12.9	653.7	67.0	68.0	12.2	4.5	0.0401	0.47	0.35
18	[0-1]	0.6	99.9	12.9	685.6	67.4	68.0	12.2	4.5	0.0402	0.48	0.35
19	[0-1]	0.7	99.9	12.9	675.9	67.9	68.0	12.2	4.5	0.0401	0.48	0.35
20	[0-1]	0.8	99.9	12.9	678.7	68.3	67.8	12.2	4.5	0.0406	0.48	0.35
21	[0-1]	0.9	99.9	13.0	675.7	68.8	67.6	12.2	4.5	0.0402	0.49	0.36
22	[0-1]	1.0	99.9	13.0	658.4	68.8	67.3	12.2	4.5	0.0402	0.49	0.35

Table 11Average values obtained by QARGA with minimum support threshold of 0.05 according to w_n and w_a .

ID	w_n	w_a	Covered instances (%)	Rule support (%)	Lift	Confidence (%)	Accuracy (%)	Amplitude (%)	#Attributes	Leverage	Certainty factor	Gain
1	0.00	[0-1]	100.0	7.1	14.61	95.5	65.2	17.9	4.92	0.0361	0.86	0.54
2	0.10	[0-1]	100.0	7.2	9.86	95.2	68.2	18.0	5.04	0.0383	0.86	0.57
3	0.20	[0-1]	100.0	7.3	7.53	94.7	70.8	18.0	5.12	0.0401	0.86	0.59
4	0.30	[0-1]	100.0	7.3	8.25	94.1	73.2	18.0	5.24	0.0416	0.86	0.61
5	0.40	[0-1]	100.0	7.4	6.85	93.4	75.3	17.9	5.33	0.0431	0.86	0.62
6	0.50	[0-1]	100.0	7.4	8.73	92.4	77.4	17.9	5.38	0.0442	0.85	0.63
7	0.60	[0-1]	100.0	7.4	7.28	91.3	79.4	17.9	5.46	0.0453	0.84	0.64
8	0.70	[0-1]	100.0	7.4	8.68	89.9	81.0	17.8	5.47	0.0460	0.83	0.65
9	0.80	[0-1]	100.0	7.5	8.00	89.1	82.2	17.7	5.51	0.0465	0.83	0.65
10	0.90	[0-1]	100.0	7.5	9.02	88.2	83.2	17.7	5.54	0.0472	0.82	0.65
11	1.00	[0-1]	100.0	7.5	8.13	87.6	84.1	17.7	5.56	0.0479	0.81	0.66
12	[0-1]	0.00	100.0	7.4	7.93	91.2	71.1	17.7	5.10	0.0379	0.81	0.56
13	[0-1]	0.10	100.0	7.3	9.69	91.6	73.1	17.8	5.17	0.0396	0.82	0.58
14	[0-1]	0.20	100.0	7.3	8.92	91.8	74.6	17.7	5.26	0.0413	0.83	0.60
15	[0-1]	0.30	100.0	7.3	10.63	91.8	75.6	17.8	5.28	0.0421	0.84	0.61
16	[0-1]	0.40	100.0	7.3	7.75	91.9	76.3	17.9	5.33	0.0431	0.84	0.62
17	[0-1]	0.50	100.0	7.4	8.47	91.9	77.0	17.8	5.34	0.0438	0.84	0.62
18	[0-1]	0.60	100.0	7.4	11.32	92.2	77.6	17.8	5.37	0.0446	0.85	0.63
19	[0-1]	0.70	100.0	7.4	7.01	92.2	78.0	17.9	5.39	0.0451	0.85	0.64
20	[0-1]	0.80	100.0	7.4	8.51	92.2	78.6	18.0	5.42	0.0457	0.85	0.64
21	[0-1]	0.90	100.0	7.3	8.56	92.3	78.8	18.0	5.45	0.0463	0.86	0.65
22	[0-1]	1.00	100.0	7.3	8.18	92.4	79.3	18.0	5.46	0.0466	0.86	0.65

QAR obtained by QARGA with a minimum support threshold of 0.05 and variations in w_n and w_a .

It can be observed that the percentage of covered instances is 100% and the support of the rules is around 7% regardless of w_n and w_a values. The lift measure achieves the maximum value when w_n is fixed to 0. Regarding the confidence, accuracy and certainty factor, these measures reach values above 90%, 70% and 80%, respectively, for almost all the variations of weights. In addition, the maximum value is achieved when w_n is lesser than 0.2 for the confidence, w_n is 1 for the accuracy and w_n is lesser than 0.5 and w_a is greater than 0.8 for the certainty factor. The leverage and gain measures obtain the maximum values for high values of w_n and w_a . It can be appreciated that confidence and accuracy are the most influenced measures by the variations of w_n and w_a showing an opposite behavior when w_n is increased.

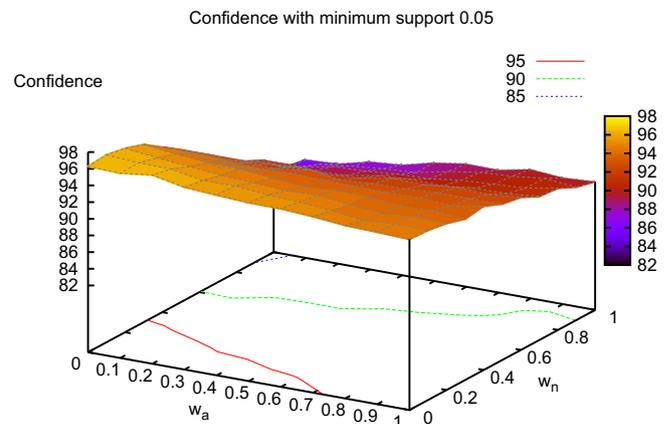
Table 12 summarizes the studied quality measures grouped by similar behavior according to the w_a and w_n variations and a minimum support threshold of 0.05. Four groups have been formed from the results reported in Table 11. It can be noted that w_n is positively correlated with gain, leverage, accuracy and number of attributes, whereas the confidence, certainty factor and the lift when w_n is less than or equal to 0.1 are negatively correlated. The remaining quality measures are not affected by the variations of w_n . Similar conclusions can be extended to w_a . However, confidence and certainty factor are positively correlated with respect to w_a as the group of the accuracy measure. Alternatively, the lift measure presents an irregular behavior in reference to w_a (represented by the symbol \sim). Figs. 7 and 8 summarize the average values obtained for the groups represented by the confidence and accuracy, respectively.

4.4. Statistical tests analysis

Finally, a non-parametric statistical analysis [6] has been carried out for all the measures to prove if significant differences are presented in the measures of QAR obtained by QARGA from the experimentation described in Section 4.3. The statistical analysis has been performed using the results retrieved when a minimum support threshold of 0.05 is considered. For each dataset, a weight w_s , w_c , w_n or w_a is set and the average of each quality measure for the 11 values of w_s , w_c , w_n or w_a , respectively, is calculated.

Table 12Performance of quality measures with minimum support threshold of 0.05 according to w_n and w_a .

Weights	Quality measures grouped by similar behavior with minimum support threshold of 0.05			
	Support Amplitude Covered Instances	Gain Leverage Accuracy #Attributes	Confidence Certainty factor	Lift
$w_n \uparrow$	=	\uparrow	\downarrow	\downarrow if $w_n \leq 0.1$ = if $w_n > 0.1$
$w_a \uparrow$	=	\uparrow	\uparrow	\sim

**Fig. 7.** Confidence with minimum support 0.05 according to w_a and w_n .

First, the Friedman and Iman-Davenport tests for a level of significance equals to 0.05 have been applied for the support, confidence, leverage, accuracy, gain, certainty factor and lift measures to find the control configuration for each weight fixed and for each measure. Afterwards, a post hoc statistical analysis has been carried out since p -values lower than the level of significance considered have been obtained. Namely, Holm has been the post hoc statistical test applied to detect significant

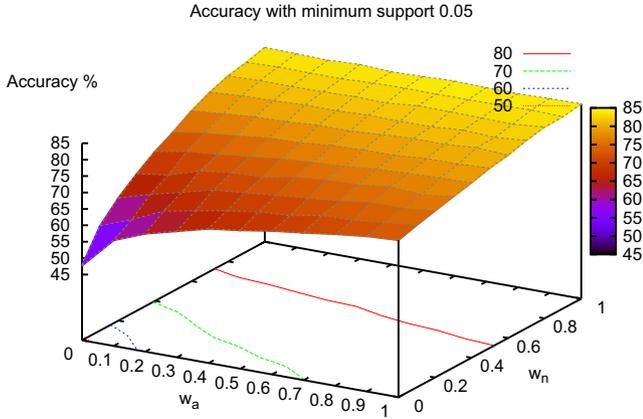


Fig. 8. Accuracy with minimum support 0.05 according to w_a and w_n .

differences between the control configuration for each weight and the remaining values.

Table 13 reports the conclusions reached after applying the Friedman and Iman–Davenport tests in addition to the Holm test. Note that only the results obtained for the support, confidence and accuracy measures are reported due to similar conclusions have been obtained for the remaining measures. Each column shows the control configuration for each phase of the experimentation and for each measure. To analyze the first phase, w_s has been fixed and w_c has been ranged from 0 to 1 and vice versa. In the case of the second phase, w_n has been fixed and w_a has been ranged from 0 to 1 and vice versa. Symbol \checkmark represents if the null hypothesis is rejected, that is, if significant differences exist between the control configuration and the remaining weights.

Table 14 summarizes the best configurations for each measure to be optimized and the measures grouped by similar behavior according to variations in the weights. The first column shows the groups identified, second column reports the best values found for w_s , w_c , w_n and w_a for each measure and third column details the recommended weights to maximize each group of measures. The three groups are composed of accuracy, gain and lift for the first group, confidence and certainty factor for the second group and finally support and leverage for the third group. It can be observed that, high values for w_n , w_a and w_s and low values for w_c are recommended to maximize the accuracy, gain and lift measures, henceforth named *Accuracy group*. Alternatively, low values for w_n and w_s , and high values for w_a and w_c are desirable to maximize the confidence and certainty factor measures, from now considered as *Confidence group*. Finally, values close to 1 for w_n , w_a , w_s and w_c are needed to maximize the support and leverage measures.

4.5. Application to real-world datasets

QARGA has been applied to the two real-world datasets related to ozone and earthquakes described in Section 4.1. This study aims at analyzing if better results are obtained when the recommended weights provided in Table 14 are used instead of set w_n , w_a , w_s and w_c with the value 1, hereinafter considered as *Default group*. Note that no group has been considered to maximize the support and leverage measures due to values close to 1 are needed, hence, these measures can be optimized with the *Default group*.

Tables 15 and 16 summarize the results obtained after executing QARGA 5 times for the ozone and earthquake datasets, respectively. Note that the best results have been highlighted in bold.

In Table 15, it can be observed that the lift, confidence, certainty factor and gain measures are improved with regard to the *Default*

Table 13

Summary of statistical tests according to the w_s , w_c , w_n and w_a with minimum support threshold 0.05.

Accuracy Control algorithm $w_s = 0.9, w_c \in [0, 1]$		Confidence Control algorithm $w_s = 0.0, w_c \in [0, 1]$		Support Control algorithm $w_s = 1.0, w_c \in [0, 1]$	
Weights w_s	H_0 rejected	Weights w_s	H_0 rejected	Weights w_s	H_0 rejected
0.0	\checkmark	0.7	\checkmark	0.0	\checkmark
0.1	\checkmark	0.9	\checkmark	0.1	\checkmark
0.2	\checkmark	1.0	\checkmark	0.2	\checkmark
0.3	\checkmark	0.8	\checkmark	0.3	\checkmark
0.4	\checkmark	0.6	\checkmark	0.4	\checkmark
0.6	\checkmark	0.5	\checkmark	0.5	\checkmark
0.5	\checkmark	0.3	\checkmark	0.6	\checkmark
0.7	\checkmark	0.4	\checkmark	0.7	\checkmark
0.8	\checkmark	0.2	\checkmark	0.8	\checkmark
1.0	\checkmark	0.1	\checkmark	0.9	\checkmark
Control algorithm $w_c = 0.0, w_s \in [0, 1]$		Control algorithm $w_c = 1.0, w_s \in [0, 1]$		Control algorithm $w_c = 1.0, w_s \in [0, 1]$	
Weights w_c	H_0 rejected	Weights w_c	H_0 rejected	Weights w_c	H_0 rejected
1.0	\checkmark	0.0	\checkmark	0.0	\checkmark
0.9	\checkmark	0.1	\checkmark	0.1	\checkmark
0.8	\checkmark	0.2	\checkmark	0.2	\checkmark
0.7	\checkmark	0.3	\checkmark	0.3	\checkmark
0.4	\checkmark	0.6	\checkmark	0.4	\checkmark
0.5	\checkmark	0.5	\checkmark	0.5	\checkmark
0.4	\checkmark	0.6	\checkmark	0.6	\checkmark
0.3	\checkmark	0.7	\checkmark	0.7	\checkmark
0.2	\checkmark	0.8	\checkmark	0.8	\checkmark
0.1	\checkmark	0.9	\checkmark	0.9	\checkmark
Control algorithm $w_n = 1.0, w_a \in [0, 1]$		Control algorithm $w_n = 0.0, w_a \in [0, 1]$		Control algorithm $w_n = 1.0, w_a \in [0, 1]$	
Weights w_n	H_0 rejected	Weights w_n	H_0 rejected	Weights w_n	H_0 rejected
0.0	\checkmark	1.0	\checkmark	0.0	\checkmark
0.1	\checkmark	0.9	\checkmark	0.1	\checkmark
0.2	\checkmark	0.8	\checkmark	0.2	\checkmark
0.3	\checkmark	0.7	\checkmark	0.3	\checkmark
0.4	\checkmark	0.6	\checkmark	0.4	\checkmark
0.5	\checkmark	0.5	\checkmark	0.5	\checkmark
0.6	\checkmark	0.4	\checkmark	0.7	\checkmark
0.7	\checkmark	0.3	\checkmark	0.6	\checkmark
0.8	\checkmark	0.2	\checkmark	0.8	\checkmark
0.9	\checkmark	0.1	\checkmark	0.9	\checkmark
Control algorithm $w_a = 1.0, w_n \in [0, 1]$		Control algorithm $w_a = 1.0, w_n \in [0, 1]$		Control algorithm $w_a = 0, w_n \in [0, 1]$	
Weights w_a	H_0 rejected	Weights w_a	H_0 rejected	Weights w_a	H_0 rejected
0.0	\checkmark	0.0	\checkmark	1.0	\checkmark
0.1	\checkmark	0.1	\checkmark	0.9	\checkmark
0.2	\checkmark	0.2	\checkmark	0.4	\checkmark
0.3	\checkmark	0.3	\checkmark	0.3	\checkmark
0.4	\checkmark	0.5	\checkmark	0.2	\checkmark
0.5	\checkmark	0.4	\checkmark	0.1	\checkmark
0.6	\checkmark	0.7	\checkmark	0.6	\checkmark
0.7	\checkmark	0.6	\checkmark	0.8	\checkmark
0.8	\checkmark	0.9	\checkmark	0.7	\checkmark
0.9	\checkmark	0.8	\checkmark	0.5	\checkmark

group when the *Confidence group* of weights is used. Finally, if QARGA is executed using the weight set from the *Accuracy group*, the lift, accuracy, leverage and gain are improved in contrast to the *Default group*.

Similar conclusions can be extended in the results obtained for the earthquake dataset. In Table 16, it can be appreciated that the same set of measures is optimized using the two groups of weights except for the lift measure for the *Confidence group* and gain for

Table 14

Recommended weights to optimize the quality measures grouped by similar behavior.

Measures	Weights				Recommended weights			
	w_n	w_a	w_s	w_c	w_n	w_a	w_s	w_c
Accuracy	[0.9,1.0]	[0.8,1.0]	[0.8,1.0]	[0.0,0.1]				
Gain	[0.7,1.0]	[0.8,1.0]	[0.0,1.0]	[0.0,0.6]	[0.9,1.0]	[0.8,1.0]	[0.8,1.0]	[0.0,0.1]
Lift	[0.7,1.0]	[0.8,1.0]	[0.5,1.0]	[0.0,0.2]				
Confidence	[0.0,0.3]	[0.6,1.0]	[0.0,0.6]	[0.9,1.0]	[0.0,0.3]	[0.8,1.0]	[0.0,0.6]	[0.9,1.0]
Certainty factor	[0.0,0.4]	[0.8,1.0]	[0.0,0.6]	[0.9,1.0]				
Support	[0.8,1.0]	[0.0,0.8]	[0.8,1.0]	[0.8,1.0]	[0.8,1.0]	0.8	[0.8, 0.9]	[0.8,1.0]
Leverage	[0.8,1.0]	[0.8,1.0]	[0.0,0.9]	[0.7,1.0]				

Table 15

Results obtained by the recommended weights for the ozone dataset.

Weight groups	Covered record (%)	Support (%)	Lift	Confidence (%)	Accuracy	Amplitude	# Attributes	Leverage	Certainty factor	Gain
Default group	95.83	11.62	3.33	61.41	67.80	28.62	3.21	0.0383	0.43	0.27
Confidence group	86.21	9.31	20.48	69.15	62.43	28.12	3.21	0.0309	0.50	0.28
Accuracy group	93.73	9.78	30.90	53.63	73.81	25.39	3.20	0.0397	0.39	0.28

Table 16

Results obtained by the recommended weights for the Earthquake dataset.

Weight groups	Covered record (%)	Support (%)	Lift	Confidence (%)	Accuracy	Amplitude	# Attributes	Leverage	Certainty factor	Gain
Default group	99.34	16.03	1.35	74.91	51.33	23.62	2.79	0.0204	0.41	0.16
Confidence group	89.69	8.76	1.29	91.02	35.73	18.26	3.10	0.0189	0.67	0.19
Accuracy group	98.30	13.29	1.45	47.01	63.73	20.81	2.81	0.0218	0.19	0.11

the *Accuracy group*. The number of attributes and amplitude of intervals are quite similar to that of the *Default group*.

From this observation, it can be concluded that QARGA discovers rules more accurate, more interesting and reaches higher gain of information on the rules when the set of recommended weights is used.

Some other interesting conclusions can be drawn from these results. Once the minimum support is reached according to the fixed threshold of minimum support, the weight of the support does not have an influence on the support of the rule. If all the weights are set to 1, the maximum support allowed for each dataset is reached. Although the support is never improved, recommended weights can be used to improve the confidence and accuracy in addition to the certainty factor, gain, lift and leverage.

5. Conclusions

An analysis of the quality measures based on the variations of the weights included in the fitness function of the QARGA algorithm has been carried out in this paper. Specifically, QARGA has been applied to several public datasets with the aim of studying how its performance is affected according to the choice of the weights. First, QARGA has been applied to 35 datasets in order to determine the values of the weights that lead to optimal QAR. Later, these assumptions have been tested on two datasets of different nature, related to ozone and earthquakes, thus confirming that the new set of weights generates better rules. Significant differences have been observed within the results of several quality measures calculated from the QAR obtained by QARGA when w_s , w_c , w_n and w_a were ranged from 0 to 1. However, w_c and w_n have been more influential than w_s and w_a over the set of quality measures studied. Furthermore, several groups of measures have been identified according to their behavior against the weight variations. The use of recommended weights outperforms the results of QARGA in real-world

applications to those obtained when non-weighted objectives are optimized in the fitness function.

Acknowledgments

The authors would like to thank Spanish Ministry of Science and Technology, Junta de Andalucía and Pablo de Olavide University for the support under projects TIN2011-28956-C02, P12-TIC-1728 and APPB813097, respectively.

References

- [1] R. Agrawal, T. Imielinski, A. Swami, Mining association rules between sets of items in large databases, in: Proceedings of the ACM SIGMOD International Conference on Management of Data, 1993, pp. 207–216.
- [2] R. Agrawal, R. Srikant, Fast algorithms for mining association rules in large databases, in: Proceedings of the International Conference on Very Large Databases, 1994, pp. 478–499.
- [3] B. Alatas, E. Akin, An efficient genetic algorithm for automated mining of both positive and negative quantitative association rules, *Soft Comput.* 10 (3) (2006) 230–237.
- [4] B. Alatas, E. Akin, A. Karci, MODENAR: multi-objective differential evolution algorithm for mining numeric association rules, *Appl. Soft Comput.* 8 (1) (2008) 646–656.
- [5] J. Alcalá-Fdez, N. Flügge-Pape, A. Bonarini, F. Herrera, Analysis of the effectiveness of the genetic algorithms based on extraction of association rules, *Fund. Inform.* 98 (1) (2010) 1001–1014.
- [6] J. Alcalá-Fdez, L. Sánchez, S. García, M.J. del Jesus, S. Ventura, J.M. Garrell, J. Otero, C. Romero, J. Bacardit, V.M. Rivas, J.C. Fernández, F. Herrera, Keel: a software tool to assess evolutionary algorithms for data mining problems, *Soft Comput.* 13 (3) (2009) 307–318.
- [7] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. Evol. Comput.* 6 (2) (2002) 182–197.
- [8] M.J. del Jesús, J.A. Gámez, P. González, J.M. Puerta, On the discovery of association rules by means of evolutionary algorithms, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1 (5), 2011, pp. 397–415.
- [9] Deogun, W. Spaulding, B. Shuart, D. Li, Towards Missing Data Imputation: A Study of Fuzzy K-means Clustering Method, *Lecture Notes on Computer Science*, vol. 3066, 2004, pp. 573–579.

- [10] E. Georgii, L. Richter, U. Rckert, S. Kramer, Analyzing microarray data using quantitative association rules, *BMC Bioinform.* 21 (2) (2005) 123–129.
- [11] H.A. Guvenir, I. Uysal, Bilkent university function approximation repository, (<http://funapp.cs.bilkent.edu.tr>), 2000.
- [12] H. Ishibuchi, Y. Nojima, Optimization of Scalarizing Functions Through Evolutionary Multiobjective Optimization, *Lecture Notes in Computer Science*, vol. 4403, 2006, pp. 51–65.
- [13] H. Ishibuchi, N. Tsukamoto, Yusuke Nojima. Empirical analysis of using weighted sum fitness functions in NSGA-II for many-objective 0/1 knapsack problems, in: *Proceedings of the International Conference on Computer Modelling and Simulation*, 2009, pp. 71–76.
- [14] H. Li, Q. Zhang, Multiobjective optimization problems with complicated Pareto sets, MOEA/D and NSGA-II, *IEEE Trans. Evol. Comput.* 13 (2) (2009) 284–302.
- [15] F. Martínez-Álvarez, A. Troncoso, A. Morales-Esteban, J.C. Riquelme, Computational Intelligence Techniques for Predicting Earthquakes, *Lecture Notes in Artificial Intelligence*, vol. 6679 (2), 2011, pp. 287–294.
- [16] M. Martínez-Ballesteros, F. Martínez-Álvarez, A. Troncoso, J.C. Riquelme, An evolutionary algorithm to discover quantitative association rules in multi-dimensional time series, *Soft Comput.* 15 (10) (2011) 2065–2084.
- [17] M. Martínez-Ballesteros, F. Martínez-Álvarez, A. Troncoso, J.C. Riquelme, Selecting the best measures to discover quantitative association rules, *Neurocomputing* 126 (2014) 3–14.
- [18] M. Martínez-Ballesteros, J.C. Riquelme, Analysis of Measures of Quantitative Association Rules, *Lecture Notes in Computer Science*, vol. 6679 (2), 2011, pp. 319–326.
- [19] D. Martín, A. Rosete, J. Alcalá-Fdez, F. Herrera, A new multiobjective evolutionary algorithm for mining a reduced set of interesting positive and negative quantitative association rules, *IEEE Trans. Evol. Comput.* 18 (1) (2014) 54–69.
- [20] M. Martínez-Ballesteros, F. Martínez-Álvarez, A. Troncoso, J.C. Riquelme, A Sensitivity Analysis for Quality Measures of Quantitative Association Rules, *Lecture Notes in Artificial Intelligence*, vol. 8073, 2013, pp. 578–587.
- [21] J. Mata, J. Álvarez, J.C. Riquelme, Mining numeric association rules with genetic algorithms, in: *Proceedings of the International Conference on Adaptive and Natural Computing Algorithms*, 2001, pp. 264–267.
- [22] J. Mata, J.L. Alvarez, J.C. Riquelme, QuantMiner: A genetic algorithm for mining quantitative association rules, in: *Proceedings of the ACM Symposium on Applied computing*, 2002, pp. 590–594.
- [23] J. Mata, J.L. Álvarez, J.C. Riquelme, Discovering Numeric Association Rules via Evolutionary Algorithm, *Lecture Notes in Artificial Intelligence*, vol. 2336, 2002, pp. 40–51.
- [24] V. Pachón Álvarez, J. Mata Vázquez, An evolutionary algorithm to discover quantitative association rules from huge databases without the need for an a priori discretization, *Expert Syst. Appl.* 39 (1) (2012) 585–593.
- [25] R. Pears, Y.S. Koh, G. Dobbie, W. Yeap, Weighted association rule mining via a graph based connectivity model, *Inf. Sci.* 218 (2013) 61–84.
- [26] A. Salleb-Aouissi, C. Vrain, C. Nortet, QuantMiner: A genetic algorithm for mining quantitative association rules, in: *Proceedings of the International Joint Conferences on Artificial Intelligence*, 2007, pp. 1035–1040.
- [27] W. Soto, A. Olaya-Benavides, A genetic algorithm for discovery of association rules, in: *Proceedings of the International Conference of the Chilean Computer Science Society*, 2011, pp. 289–293.
- [28] X. Yan, C. Zhang, S. Zhang, Genetic algorithm-based strategy for identifying association rules without specifying actual minimum support, *Expert Syst. Appl.* 36 (2) (2009) 3066–3076.