A Sensitivity Analysis for Quality Measures of Quantitative Association Rules

María Martínez-Ballesteros[1], Francisco Martínez-Álvarez[2], Alicia Troncoso[2], and José C. Riquelme[1]

[1] Department of Computer Science, University of Seville, Spain
{mariamartinez,riquelme}@us.es
[2] Department of Computer Science, Pablo de Olavide University of Seville, Spain
{fmaralv,ali}@upo.es

Abstract. There exist several fitness function proposals based on a combination of weighted objectives to optimize the discovery of association rules. Nevertheless, some differences in the measures used to assess the quality of association rules could be obtained according to the values of such weights. Therefore, in such proposals it is very important the user's decision in order to specify the weights or coefficients of the optimized objectives. Thus, this work presents an analysis on the sensitivity of several quality measures when the weights included in the fitness function of the existing QARGA algorithm are modified. Finally, a comparative analysis of the results obtained according to the weights setup is provided.

Keywords: Data mining, sensitivity analysis, quantitative association rules, quality measures.

1 Introduction

The use of efficient computational techniques has become a task of the utmost importance due to the high volume of data that can be stored nowadays. In this context, the discovery of association rules (AR) –and particularly of quantitative association rules (QAR) in this work– is a popular and well-known methodology to discover significant and apparently hidden relations among variables that form databases [2]. This discovery of knowledge is based on statistical techniques such as correlation analysis and variance. One of the most used and cited algorithms is Apriori [1].

When the domain is continuous, the AR are known as QAR. In this context, let $F = \{F_1, ..., F_n\}$ be a set of features, with values in $\mathbb{R}$. Let A and C be two disjunct subsets of F, that is, $A \subset F$, $C \subset F$, and $A \cap C = \emptyset$. A QAR is a rule $X \Rightarrow Y$, in which features in A belong to the antecedent X, and features in C belong to the consequent Y, such that X and Y are formed by a conjunction of multiple boolean expressions of the form $F_i \in [v_1,\ v_2]$. The consequent Y is usually a single expression.

The AR extraction process is a non-supervised learning technique to explore data properties. The main goal pursuit is, then, to find groups of attributes appearing frequently together in a dataset, so to provide comprehensive rules able to explain the existing relations among them. The mining process of AR is usually modeled as a multi-objective problem in which several quality measures of AR are the objectives to be optimized since there not exist an unique measure to determine the AR quality. There are several approaches to solve multi-objective problems. The most common approaches are focused in Pareto-based multi-objective algorithms which try to find the best trade-off between two or more conflicting objectives. However, many others are based in weighted sum fitness functions which formulate the problem as a single-objective optimization problem using parameters of scalarization. Such weighted sum fitness functions allow to find solutions according to the user preferences and emphasize some objectives over others. Most of the existing techniques to discover AR are typically focused in using the support and confidence measures as objectives to be optimized by a weighted sum fitness function. Therefore, the main goal of this work is to conduct an analysis on the sensitivity of such quality measures when the weights in the fitness function vary. Nonetheless, there also exist other measures widely used for both evaluation and optimization of AR [9]. Some of such quality measures are described in Table 1. Note that $n(X)$ is the number of occurrences of the itemset X in the dataset and N is the total number of instances in the dataset. ND stands for negatively dependent, PD for positively dependent and I for indepedent.

Table 1. Quality measures for quantitative association rules

Measures	Equation	Description	Range
$Sup(X)$	$n(X)/N$	Coverage of X	[0, 1]
$Sup(X \Longrightarrow Y)$	$n(X \cap Y)/N$	Generality of the rule	[0, 1]
$Conf(X \Longrightarrow Y)$	$sup(X \Longrightarrow Y)/sup(X)$	Reliability of the rule	[0, 1]
$Lift(X \Longrightarrow Y)$	$sup(X \Longrightarrow Y)/(sup(X) \cdot sup(Y))$	Interest of the rule • Value < 1: X and Y (ND) • Value = 1: X and Y (I) • Value > 1: X and Y (PD)	$[0, +\infty)$
$Gain(X \Longrightarrow Y)$	$conf(X \Longrightarrow Y) - sup(Y)$	Implication of the rule	[-0.5, 1]
$Certainty\ Factor(X \Longrightarrow Y)$	• If $conf(X \Longrightarrow Y) > sup(Y)$: $(conf(X \Longrightarrow Y) - sup(Y))/(1 - sup(Y))$ • If $conf(X \Longrightarrow Y) <= sup(Y)$: $(conf(X \Longrightarrow Y) - sup(Y))/sup(Y)$	Gain normalized • Value < 0: X and Y (ND) • Value = 0: X and Y (I) • Value > 0: X and Y (PD)	[-1, 1]
$Leverage(X \Longrightarrow Y)$	$sup(X \Longrightarrow Y) - sup(X)sup(Y)$	Strength of the rule • Value < 0: X and Y (ND) • Value = 0: X and Y (I) • Value > 0: X and Y (PD)	[-0.25, 0.25]
$Accuracy(X \Longrightarrow Y)$	$sup(X \Longrightarrow Y) + sup(\neg X \Longrightarrow \neg Y)$	Veracity of the rule	[0, 1]

Thus, we aim to provide guidelines to set the weights of the fitness function according to the objectives to be satisfied by the rules. On the other hand, we intend to establish multiple relationships between the quality measures and variations in the weights of the fitness function by means of the results obtained by the QARGA algorithm [8].

The remainder of the paper is as follows. Section 2 describes some techniques which included a weighted sum fitness function. The QARGA algorithm used in the study performed and the experimental setup is detailed in Section 3. Section 4 presents and discusses the results obtained by QARGA using different weights in the fitness function. Finally, Section 5 summarizes the conclusions drawn from the analysis conducted.

2 Related Work

There exist several fitness functions proposals based on a combination of weighted objectives in a single equation. Hence, their performance is very sensitive to the choice of the weights of the measures included within the fitness function. Actually, many algorithms to discover AR can be found in the literature. Most of them are based on the methods proposed by Agrawal et al. [1] but such methods require high computational cost and memory. Genetic algorithms, colony algorithms, evolutionary algorithms (EA) and particle swarm algorithms are usually used to overcome such drawbacks. Techniques based on EA have been extensively used for the optimization and adjustment of models in data mining. Evolutionary computation is usually used to discover AR in both EA and genetic programming since they offer a set of advantages for knowledge extraction and specifically for rule induction processes.

A wide range of methods have been proposed to address the discovery and optimization of AR by a weighted sum fitness function. This kind of fitness function has been applied into several optimization problems. For instance, the authors in [7] examined the effect of using weighted sum fitness functions for parent selection and generation update. Such an effect was tested on the performance of NSGA-II for a high-dimensional space of a multi-objective problem.

The authors in [11] proposed an EA-based approach capable of obtaining an undetermined number of quantitative attributes in the antecedent of the rule. Their approach, called GENAR, optimized a weighted fitness function based on the support and confidence measures and the number of recovered instances. The same quality measures plus the comprehensibility and the amplitude of the intervals forming the rule were included in the weighted sum fitness function of the GAR-plus algorithm [12].

In [2] a GA is proposed as a search strategy for both positive and negative QAR mining within databases. The discovery of QAR was optimized by a weighted sum fitness function composed of support, confidence, number of attributes and amplitude. Later, the same authors proposed a multi-objective Pareto-based EA called MODENAR in [3]. Those measures and the recovered records were included within the fitness function to be optimized in such works.

A genetic algorithm was proposed in [14] which optimized support, confidence, comprehensibility and interest of AR included in a weighted fitness function. A weighted support based on the individual weight of the items according to their importance in the dataset was calculated in [13].

3 Methodology

3.1 Description of QARGA

This section describes the main features of the QARGA algorithm, which is used to perform the fitness function sensitivity study according to the weighting of the measures. QARGA is a real-coded genetic algorithm designed to discover existing relationships, specifically QAR, among several variables. A detailed description of the algorithm can be found in [10].

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 AR.

The fitness function proposed in [8] to be maximized by QARGA was:

$$\begin{aligned} f(rule) = w_s \cdot sup + w_c \cdot conf + w_n \cdot nAttrib \\ -w_a \cdot ampl - w_r \cdot recov \end{aligned} \tag{1}$$

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. Moreover, the fitness function was provided with 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.

However, the user should be aware of the importance of each measure, in order to specify the weights or coefficient because significant differences in the AR quality measures could be obtained. Next section describes the experimentation framework carried out to assess how the weights included in QARGA's fitness function may influence, in order to provide a guide for the user's decision.

3.2 Experimental Design

It is well known that one of the shortcomings of the weighted sum fitness function is the parametrization of such weights. A fitness function-based sensitivity analysis and a detailed study of some weights are discussed in Section 4 to ascertain the relative influence of each weight on the final results obtained by QARGA. The aim of this study is to analyze the behavior of QARGA to achieve optimal solutions.

Therefore, three sets of experiments have been performed by varying the weights for support and confidence measures included in QARGA's fitness function. The first set of experiments has used a minimum support threshold equal to 0 to obtain all the QAR found by QARGA. The second and third set of experiments have used a minimum support threshold equal to 0.05 and 0.1 respectively to penalize the fitness function of those individuals of the population which do not satisfy the minimum support thresholds, respectively. We aim to force QARGA to learn QAR with the established minimum support.

Different configurations of QARGA have been executed by modifying the weights of the support and confidence measures optimized in the fitness function for each set of experiments. Specifically, the weight values for support and confidence measures, henceforth called w_s and w_c respectively, have been varied from 0 to 1 with increments of 0.1 (11 different values for both). Hence, QARGA was run 363 (3 x 11 x 11) times in total. To be precise, 363 executions have carried out for each dataset, using them as training data. Note that the rest of the weights included in the fitness function of QARGA have been set to 1 in order to avoid the influence of such weight in the results and to ensure that the remaining measures are present in the fitness function.

These set of 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.

4 Experimental Study

4.1 Datasets Description and Parameters Setup

This section presents the main features of the datasets used in the sensitivity study carried out. Several datasets have been tested from the public BUFA repository [6]. In particular, the thirty-five public datasets from BUFA repository used in [9]. Note that Buying, Country, College, Education, Read and Usnews Colleged have been preprocessed using K-means Imputation method proposed in [5] (available in the KEEL tool [4]) in order to deal with missing values.

As for the values for the main parameters of QARGA, it is noteworthy that these values have been used for each execution to assess the performance of QARGA according to the different values of the weights included in the fitness function.

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 include both the antecedent and consequent are 10 and 5, respectively. Note that both the antecedent and consequent must contain one attribute at least. QARGA has obtained 100 QAR for each dataset and each setting of the fitness function weights.

4.2 Sensitivity Analysis of the Quality Measures

In this section, the results obtained by QARGA when optimizing the fitness function with variations in the weights of the measures are discussed. Specifically, the results obtained by QARGA, first, when a minimum support threshold is applied and second, when w_s and w_c are modified in the fitness function are compared.

As described in Section 3.2, QARGA has been executed 363 times for each dataset, that is, a total of 12705 executions. In order to perform the parametric sensitivity study, the average results for the 35 datasets using the same configuration has been calculated. Several interestingness measures have been calculated to assess the quality of the AR obtained by QARGA for each run. 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. A detailed explanation of these measures can be found in [8].

Tables 2 and 3 summarize the behavior of the quality measures depending on the minimum support threshold used by QARGA. Note that similar results have been obtained when the minimum support threshold is 0.05 and 0.1, hence, only the results obtained by QARGA with a minimum support threshold equal to 0.05 are shown in Table 3.

Each table presents the studied quality measures grouped by their performance when the weights associated with the support and confidence measures in the fitness function are increased or decreased.

Table 2. Performance of quality measures according to the support and confidence weights with minimum support equal to 0

Weight	Quality measures grouped by similar behavior					
$minsup = 0$	Support Leverage Covered instances	Confidence Gain #Rules	Accuracy	Lift	#Attributes	Amplitude
Support ↑	↑	=	=	=	↓	↑
Confidence ↑	=	>0.1 = ≤0.1 ↑	>0.1 ↑ ≤0.1↓	↑	↑	↓

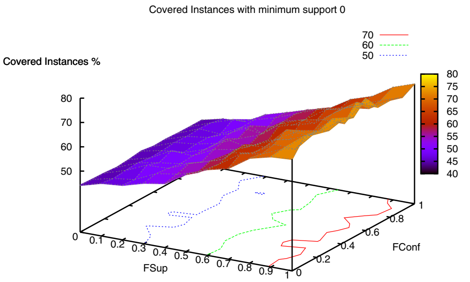

Fig. 1. Covered instances with minimum support 0

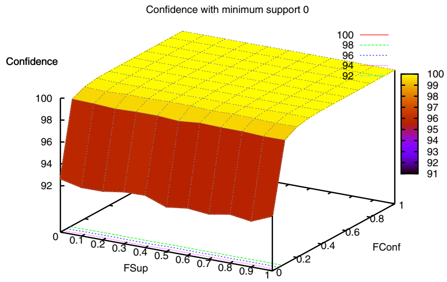

Fig. 2. Confidence with min. sup. 0

Table 2 shows the ten studied quality measures arranged into six groups. It can be noted that w_s is positively correlated with support, leverage, covered instances and amplitude whereas is negatively correlated with the number of attributes if no minimum support threshold is applied. The performance of the other measures under study is not affected by the variations of w_s. With respect to w_c, some differences can be observed. For instance, although support, leverage and covered instances are dependent of w_s, such measures are not influenced by

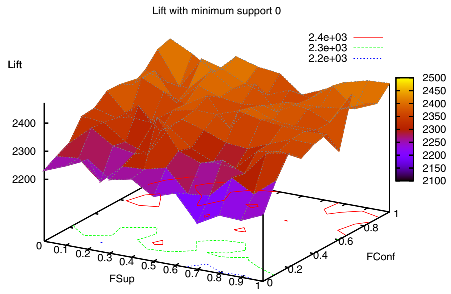

Fig. 3. Lift with minimum support 0

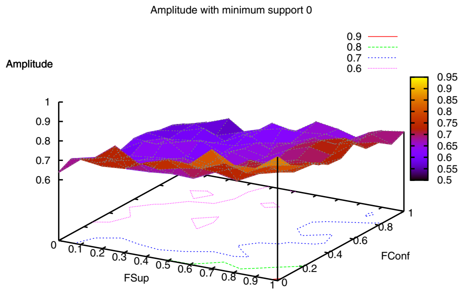

Fig. 4. Amplitude with minimum support 0

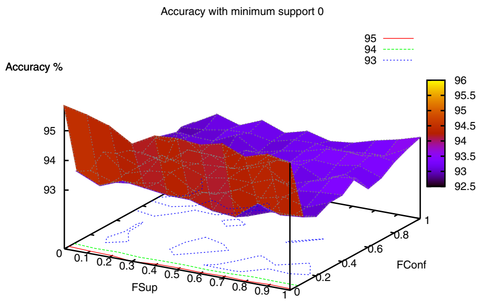

Fig. 5. Accuracy with minimum support 0

w_c. Lift measure and number of attributes are positively correlated with w_c and amplitude is negatively correlated. However, confidence, gain and the number of rules are only increased when w_c achieves values of 0 and 0.1. A w_c greater than 0.1 does not cause alterations in the performance of such measures. It can be observed an opposite behavior in the accuracy since it is negatively correlated with the confidence if such weight is 0 or 0.1 and positively correlated if w_c is greater than 0.1.

Figures 1, 2, 3, 4 and 5 summarize the values obtained for each group of measures when the minimum support threshold is 0. Note that only one measure of each group is displayed due to the similar performance among the measures of each group and space limitations. Figure 1 represents the support, confidence and covered instances measures. It can be observed that their values form an increasing inclined plane relative to w_s. Figure 2 visualizes the values obtained for the confidence measure and its behavior can be extended to the gain measure and number of rules. These measures present an awning model reaching their highest values when w_c is greater than 0.1. Figures 3 and 4 show the lift and amplitude values respectively when the minimum support threshold is 0. These measures do not follow any specific behavior pattern and can be considered as rough models. Finally, the accuracy measure is displayed in Figure 5. It achieves the highest value when w_c is 0. The performance of this measure can be considered as a valley model.

Table 3. Performance of quality measures according to the support and confidence weights with minimum support equal to 0.05

Weight	Quality measures grouped by similar behavior		
$minsup = 0.05$	Support Confidence Leverage Amplitude	Lift Accuracy Gain #Attributes	Covered instances #Rules
Support ↑	=	=	<0.8 ↑ ≥0.8 =
Confidence ↑	↑	↓	≤0.3 ↑ >0.3 =

Table 3 displays the ten measures under study grouped into only three groups when a minimum support threshold is applied. It can be appreciated that the performance of these measures are completely different when a minimum support threshold is not applied. For instance, the group composed of support, confidence, 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 with w_c. Regarding the third group, that is, covered instances and number of rules are only affected when w_s is less than 0.8 and w_c is less or equal to 0.3, both positively correlated. Weights above these values do not cause performance variations on such measures.

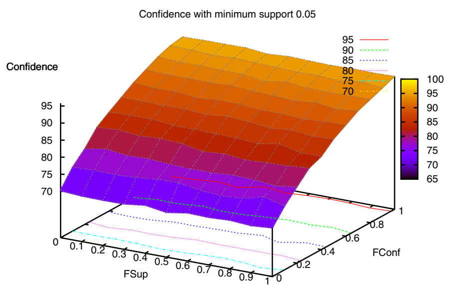

Fig. 6. Confidence with minimum support 0.05

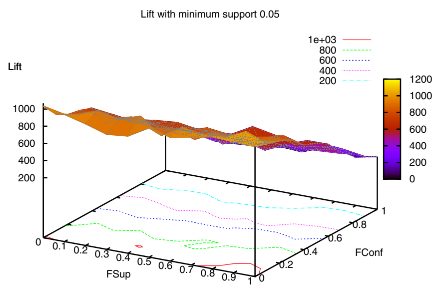

Fig. 7. Lift with minimum support 0.05

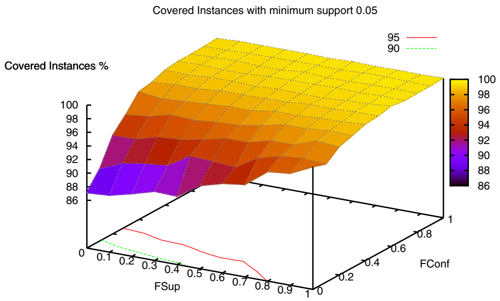

Fig. 8. Covered instances with minimum support 0.05

Figures 6, 7 and 8 illustrate the values obtained for each group of measures when the minimum support threshold is 0.05.

Note that a similar behavior has been obtained when the minimum support threshold is 0.1. Figure 6 represents the performance of confidence, support, leverage and amplitude measures. These measures reach their highest values when w_c is 1. In this case, the confidence behaves as an increasing inclined plane with respect to w_c instead of presenting an awning model as Figure 2. Figure 7 shows the values obtained for the lift measure and summarizes the behavior of accuracy, gain and number of attributes in addition to lift. In this case, these measures get their highest values when w_c is 0 and perform as a decreasing inclined plane relative to w_c. Finally, Figure 8 shows the values obtained for the covered instances. This measure exhibits its maximum value when w_s is 1 and w_c is above 0.3. The covered instances present a behavior similar to an awning model.

As final remarks, we provide the following use recommendations. First, the obtained AR are more specific when the minimum support threshold is 0. 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. 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%. Taking into account such decision, 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.

Finally, we note that it would be interesting to study the rest of weights included in the fitness function of QARGA to analyze their influence on the AR quality measures.

5 Conclusions

An analysis based on the sensitivity 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. Significant differences have been observed in the results of ten AR quality measures calculated from the AR obtained by QARGA when w_s and w_c were ranged from 0 to 1. However, w_c has been more influential than w_s over the set of AR quality measures studied. Furthermore, several groups of measures have been identified according to their behavior against the weight variations.

Acknowledgments. The financial support from the Spanish Ministry of Science and Technology, project TIN2011-28956-C02-00, and from the Junta de Andalucía, project P11-TIC-7528, is acknowledged.

References

1. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large databases. In: Proceedings of the International Conference on Very Large Databases, pp. 478–499 (1994)
2. Alatas, B., Akin, E.: An efficient genetic algorithm for automated mining of both positive and negative quantitative association rules. Soft Computing 10(3), 230–237 (2006)
3. Alatas, B., Akin, E., Karci, A.: MODENAR: Multi-objective differential evolution algorithm for mining numeric association rules. Applied Soft Computing 8(1), 646–656 (2008)
4. Alcalá-Fdez, J., Sánchez, L., García, S., del Jesus, M.J., Ventura, S., Garrell, J.M., Otero, J., Romero, C., Bacardit, J., Rivas, V.M., Fernández, J.C., Herrera, F.: Keel: A software tool to assess evolutionary algorithms for data mining problems. Soft Computing 13(3), 307–318 (2009)
5. Li, D., Deogun, J., Spaulding, W., Shuart, B.: Towards missing data imputation: A study of fuzzy K-means clustering method. In: Tsumoto, S., Słowiński, R., Komorowski, J., Grzymała-Busse, J.W. (eds.) RSCTC 2004. LNCS (LNAI), vol. 3066, pp. 573–579. Springer, Heidelberg (2004)
6. Guvenir, H.A., Uysal, I.: Bilkent university function approximation repository (2000), `http://funapp.cs.bilkent.edu.tr`
7. Ishibuchi, H., Tsukamoto, N., Nojima, Y.: 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, pp. 71–76 (2009)
8. Martínez-Ballesteros, M., Martínez-Álvarez, F., Troncoso, A., Riquelme, J.C.: An evolutionary algorithm to discover quantitative association rules in multidimensional time series. Soft Computing 15(10), 2065–2084 (2011)
9. Martínez-Ballesteros, M., Martínez-Álvarez, F., Troncoso, A., Riquelme, J.C.: Selecting the best measures to discover quantitative association rules. Neurocomputing (in press, 2013), doi: `http://dx.doi.org/10.1016/j.neucom.2013.01.056`
10. Martínez-Ballesteros, M., Riquelme, J.C.: Analysis of measures of quantitative association rules. In: Corchado, E., Kurzyński, M., Woźniak, M. (eds.) HAIS 2011, Part II. LNCS, vol. 6679, pp. 319–326. Springer, Heidelberg (2011)
11. Mata, J., Álvarez, J., Riquelme, J.C.: Mining numeric association rules with genetic algorithms. In: Proceedings of the International Conference on Adaptive and Natural Computing Algorithms, pp. 264–267 (2001)
12. Pachón Álvarez, V., Mata Vázquez, J.: 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), 585–593 (2012)
13. Pears, R., Koh, Y.S., Dobbie, G., Yeap, W.: Weighted association rule mining via a graph based connectivity model. Information Sciences 218, 61–84 (2013)
14. Soto, W., Olaya-Benavides, A.: A genetic algorithm for discovery of association rules. In: Proceedings of the International Conference of the Chilean Computer Science Society, pp. 289–293 (2011)