ADV MATH SCI JOURNAL
Advances in Mathematics: Scientific Journal **9** (2020), no.11, 9489–9508
ISSN: 1857-8365 (printed); 1857-8438 (electronic)
https://doi.org/10.37418/amsj.9.11.52

SELECTING, SORTING AND RANKING ASSOCIATION RULES WITH MULTIPLE CRITERIA USING DOMINANCE RELATION

AZZEDDINE DAHBI[1], SIHAM JABRI, YOUSSEF BALOUKI, AND TAOUFIQ GADI

ABSTRACT. Datamining is the process of extracting interesting knowledge and information of patterns from large databases. Using association rules in datamaining is one of the most relevant tasks in modern society, which aim to discover interesting relationship and correlation among sets of items in large transactional databases. One of the main problems related to the discovery of these association (that a decision maker can face) is the huge number of association rules extracted. Hence, the knowledge post-processing phase becomes very challenging to rank and select the most interesting AR, Various interestingness measures have been proposed as a post processing phase. However, the abundance of these measures caused a new problem because there is no optimal measure and there is no measure which is better than others. To overcome this challenge we propose a new algorithm based on dominance relation aiming to find a good compromise without favoring or excluding any measures. Numerical experiments and comparison with other approaches are made on benchmark datasets and confirm a significant performance of the proposed approach.

[1]*corresponding author*

2020 *Mathematics Subject Classification.* 68P99, 68W30.
Key words and phrases. Data mining, Association rules, Interestingness measures, Multiple criteria, dominance relation.

1. INTRODUCTION

Data mining (DM) is defined as the set of techniques and methods for exploring and analyzing big datasets in order to find (among these data) certain unknown or hidden rules, associations or tendencies. It is the art of extracting information from data to gain insight into the data as a core element of the whole knowledge discovery process. It is widely used in business (insurance, retail, banking, credit card fraud detection system), science research (medicine, astronomy, biological data analysis), and government security (detection of criminals and terrorists). Association rules mining is one of the most important topics in data mining research and development, and it is the subject matter of this paper. It aims to extract interesting correlations, frequent patterns, associations among sets of items in the transaction databases or other data repositories. Association rules are widely used in various areas such as telecommunication networks, market and risk management, inventory control, etc. Several algorithms have been developed on the basis of threshold fixing [1,2] or the use of different measures other than Support and Confidence, or else on the basis of other criteria [3]. However, existing association rules algorithms produce an important number of rules. Hence, the decision maker is unable to determine the most interesting ones and is consequently unable to make decisions. In order to overcome this problem, an efficient post-treatment phase has become a compelling need. Many interestingness measures were proposed in literature to determine the interestingness of the rules [4] and [5]. In the last decade, interestingness measure field had an important activity. A high number of measures was proposed in the literature. Indeed, the results of evaluations vary from a measure to another and can even be contradictory since the measures evaluate differently the envisaged rules. That is why some rules may be relevant according to a measure and not relevant according to another. Many works in the literature proposed to find a solution and help the users in the choice of the measure to be the most adequate to the decision scope. Our paper lies within this scope: We propose an approach which permits selecting, classifying and ranking the association rules without favoring, or excluding any measures. The paper is structured as follows. In Section 2, we display an overview of related works. In Section 3, we present the necessary scientific background and an overview association rules mining, interestingness measures and dominance relation. Section 4 presents our approach based on dominance relation. In Section 5, we will discuss the experimental results and its analysis. Conclusion and scope for future work is given in the last section.

2. State of art

After the process of association rules mining (ARM), the post-treatment comes by using the interestingness measures (IM) to evaluate and find the most interesting rules. The huge number of the proposed IM caused a new problem of the selection of the measure to be the most adequate to the decision. Some researches [6] compared the evaluation results by a selected interestingness measures for the discovered and that yield by the human experts, and choose the measure that gives the nearest one to the expert evaluation result. However, it is not always possible to get the experts evaluation and their results can not be taken as a general conclusion. Other approaches [7] proposed an evolutionary approach using a genetic algorithm to test the effectiveness of different interestingness and choose the appropriate IM according to each dataset in application domain. Julian Blanchard et al. [8] and M.Gavrilov [9] present a new approach to study the similarity between the interestingness measures by classifying them according to some criteria such as the subject, the scope and the nature of the measure to help the user choose the adequate measure to be applied for a given application domain. Tan et al. [10] proposes and describes several properties to design a good IM in order to select the right measure for a given application domain, but those approaches do not guarantee the selection of the adequate and the best measure for the simple reason that not verify the used properties. Lenca [11] proposes a new methodology to select the best rules by describing the properties oh interestingness measures then applying a multi-criteria decision aid process. Our previous work [12] used an approach based on k-means to partition the generated AR into k disjoint clusters and then classify the obtained clusters from the best to the worst. Chen et al [13] and Toloo et al [14] propose a new methodology to estimate and rank the efficiency of AR resulted from the DM process by applying a decision analysis method using a new integrated DEA model. Some authors discovered significant and relevant rules without favoring or excluding any measures by adopting the notion of dominance between rules and the sky pattern[15] and [16].

3. Backround

3.1. **Association Rules Mining.**

Association rule mining is one of the major techniques od DM aims to discover interesting relationship, correlation, frequent patterns or causal structure among a sets of binary variable in transaction database and this type of analysis is called "market basket analysis" Let $I = \{i_1, i_2, ..., i_n\}$. be a set of all items and D designe a transactional database of N transactions. Assocation rules are generated over a large set of transaction denoted by T with $T = \{t_1, t_2, ..., t_m\}$. , and every transaction t_i is an itemset and meet $t_i \subseteq I$. Given a non empty set I, an AR is a statement of the form $X \to Y$, where $X, Y \subseteq I$ such $X \cap Y = \oslash$ .It indicates that the presence of items in the antecedent of rule (X) implies the presence of items in the consequent of rule(Y). An association rule can be considered interesting if the items involved occur together often and there are suggestions that one of the sets might in some sense lead to the presence of the other set. These.association are not based on the characteristics of a domain (as in functional dependency) but on the co-occurrence of data items in the dataset, it is totally data driven technique. The strength of an association rule can be measured by mathematical notions called: 'support', and 'confidence'. The notation sup(X)=P(X): support of an itemset X is the fraction of transaction that contain X,is used to represent the proportion of times that the set X appears in the transaction T.

The support of the rule $X \to Y$ in D is the percentage of transaction in a database D that contain $X \cup Y$ and is represented as:

$$Support(X \to Y) = P(X, Y) = \frac{n(X, Y)}{n}.$$

The confidence of a rule $X \to Y$ is computed as the percentage of transactions containing X which also contain Y and is represented as:

$$Confidence(X \to Y) = \frac{P(X, Y)}{P(X)} = \frac{n(X, Y)}{n(X)}.$$

Here: n(X) is the number of transaction containing X , n(X,Y) is the number of transactions that contain items $X \cup Y$, and n is the total number of transactions in D. The problem of ARM is decomposed into two phases : the first phase is the discovery of frequents itemsets and the second phase is the generation of rules from the frequent itemsets.Numerous algorithms for ARM existing in the literature differ in their approaches to generate all AR from given database that have support and confidence greater than user specified minimum of support denoted as minsup, and minimum of confidence denoted as minconf.

3.2. **Interestingness Measures.**

One of the main issue of ARM is the huge amount of generated patterns and rules where most of them are redundant and not interesting to the user. Then various IM have been proposed, developed and applied to evaluate and select the useful ones. They can be classified into two categories: subjectives [17] and objectives [12]. Subjective measures e.g. unexpectedness and actionability [18] (data user) focus on finding interesting patterns by matching against a given set of user beliefs and knowledge,while objectives ones [11] (datadriven) are numerical indexes and measure the interestingness in terms of their probabilities and statistics. Support, confidence, and lift are the classic objective measures, now they are many other available to the analyst based on these three measures such as conviction, J-measures, Gini index, Laplace [10]. Some authors [19] discover interesting rules by using a new methodology for combining data-driven (objective) and user-driven (subjective) evaluation measures. Their methodology is that the objective measures are first used to filter the rule set and then subjective measures are used to assist the user in analyzing the rules according to his knowledge and goals. Razan Paul [20] and [21] use a semantic interestingness measures for discovering association rules. Semantic interestingness measures take into account how data attributes are semantically related. It makes use of the structure of the ontology that hosts the corresponding items (e.g. generalization, specialization, etc.). Owing to the large number of interesting measures existing in the literature, how to select suitable measures becomes a major challenge. To overcome that problem, several approaches and techniques were presented ,by proposing intuitive formal criteria that a good measure should verify to evaluate the degree of interest of rule [22]. Hiep Xuan et al [23] try to solve this problem by ranking objective IM with sensitivity value and help the user to have an insight view on the behaviors of IM and as a final purpose Tan et al [10] discuss the properties of twenty-one measures and concludes that there is no measure better than others in all application domain. Some objective measures shown in Table I and used to evaluate the performance or interestingness of rules.

TABLE 1. SOME INTERESTINGNESS MEASURES

measure	formula
Lift	$lift(X \rightarrow Y) = \frac{P(X,Y)}{P(X)P(Y)}$
Information Gain	$GI(X \rightarrow Y) = log\frac{P(X,Y)}{P(X)P(Y)}$
Example Counter Example Rate	$ECR(X \rightarrow Y) = 2 - \frac{1)}{conf(X \rightarrow Y)}$
Jaccard	$JRD(X \rightarrow Y) = \frac{P(X,Y)}{P(X\ negY)+P(Y)}$
Cosinus	$COS(X \rightarrow Y) = \frac{P(X,Y)}{\sqrt{P(X)P(Y)}}$
Pearl	$PRL(X \rightarrow Y) = \frac{P(X)}{P(\frac{X}{Y})-P(Y)}$
Loevinger	$LVG(X \rightarrow Y) = \frac{P(\frac{X}{Y})-P(Y)}{1-P(Y)}$
Conviction	$CNV(X \rightarrow Y) = \frac{P(X)\neg P(Y)}{P(XY)}$
Zhang	$ZHN(X \rightarrow Y) = \frac{P(X,Y)-P(X)P(Y)}{max\{P(XY)P(\neg Y),P(Y)P(X\neg Y)\}}$
Piatetsky Shapiro	$PS(X \rightarrow Y) = P(XY) - P(X)P(Y)$
Sebag-Schoenauer	$SBG(X \rightarrow Y) = \frac{P(X,Y)}{P(X\neg Y)}$

3.3. **Dominance Relation.**

In a generic decision-making context, a decision Pareto dominates another if it is strictly favored in at least one aspect of the decision and at least as good as the other in all other aspects [24] The notion of Pareto optimality started in social welfare and economic theory, and the Pareto dominance relation is usually related in that area and many other related decision areas, such as collective decision and voting theory, decision making under uncertainty, and multi-criteria decision making and optimization [25]. Let (P) be an optimization problem in the presence of multiple criteria that may be connecting. The criteria to be optimized are explicit functions of decision variables. To compare candidate solutions in multi objective optimization problems, the concept of Pareto dominance is used. A decision vector x is said to dominate another y when it is as good as y regarding each objective, and there is at least one objective with respect to which x is better than y. In this case, the solution x is called the non-dominated solution. For a maximization problem, a solution vector x is said to dominate the solution vector y when:

$$\forall i \in \{1, 2, \ldots, I\} f_i(x) \geq f_i(y)$$

TABLE 2. Example of dominance

Rules/Measures	M1	M2	M3	M4
R1	0.9	0.8	0.8	0.6
R2	0.8	0.7	0.3	0.4
R3	0.8	0.9	0.4	0.8
R4	0.3	0.3	0.1	0.3
R5	0.5	0.5	0.8	0.8
R6	0.4	0.4	0.5	0.7

$and\ \exists i \in \{1, 2, \ldots, I\} f_i(x) > f_i(y).$
If there is no feasible solution y that dominates x, x is said to be an effective solution or a Pareto optimal solution of the multi−objective optimization problem. The solutions that are non−dominated within the entire search space are denoted as Pareto optimal and constitute the Pareto optimal set.

4. PROPOSED APPROACH

In order to help the user's in the procedure of taking the right decision against the large quantities of data, a new method proposed based on the use of the MCDA in the association rules. As we know, Roy [26] proposed four different categories as problematiques in MCDA: the selection problem, the sorting problem, ranking problem and description problem. In the first time, we will treat the selection problem, by using Pareto dominance for choosing the most interesting association rules evaluated with a set of interestingness measures and not only one.

4.1. **Selecting AR.**

Dominance of rules.

Let R and R' be two association rules, we say that an association rule R dominates another association rule R' if and only if R is more relevant than R' for all measurements, and it is denoted as $R \succ R'$.

$R_1 \succ R_2 \succ R_4$ If the ruleR_1 dominates R_2, then R_1 is equivalent or better than R_2 for all the selected measures. The rules dominated by other (at least) are not relevant and are eliminated, and they remain as the set of rules that are not dominated by any other, following all the measures M. We eliminate a rule R of the end result, not because it is not interesting for one measure but because it is not relevant in a combination of a set of measures. In this selection phase, it is not guaranteed that the rules suppressed are not interesting, because it may discard some dominated rules that contain valuable information. Another disadvantage of the selection problem is the number of the selected AR (may not be compatible with the number desired by the users ie: the user requests a set of relevant rules larger than the set of the selected rules (non dominated)). In this case, a sorting problem becomes very necessary.

4.2. **The sorting AR.**

The sorting AR aims to arrange and classify the AR into a few groups in preference order, so that the user can manage them more effectively. First, we will call select AR and apply it to the set of all AR to select the non dominated rules, and we put them in the first group. Then, we apply the select AR to the set of the dominated rules and select the non dominated rules of this groups and put them in the second class and so on until no dominated rules existed.

Algorithm class AR.

$Input : W = (R, M)$

$Output : the\ ordered\ classes\ of\ association\ rules$

1.$Begin$

2.| $n \leftarrow 0$

3.| $WhileR \neq 0do$

4.| $\quad |n \leftarrow n + 1$

5.| $\quad |Cn \leftarrow selectar(\Omega)$

6.| $\quad |R \leftarrow R \backslash Cn$

7.| $\quad \lfloor \Omega \leftarrow (R, M)$

8.| $return(C1, ..., Cn)$

9.End

This classification or sorting phase can answer some point but there is still another gap to differentiate between the rules in the same cluster. Hence, we propose our major contribution which permit to rank all the AR.

TABLE 3. RESULTS OF THE EXAMPLE

The rules	Ndom	Ndomb
R1	2	0
R2	1	2
R3	2	0
R4	0	5
R5	2	0
R6	1	1

4.3. **Ranking AR.**

In this part we use a notion of dominance relation to find a good compromise without excluding or favoring any measure, which permit to rank association rules according to a real value, the rank of AR by our method depends on the number of rules which dominate the tested rule and which are dominated by this later, because the rule, which dominates 20, is better than the rule dominates 10, and the rule, which dominated by 20, is worse than the rule dominated by 10. Our method proposes a way in which our tested rule is ranked not only by the number of rules it dominates (Ndom) ; but is also examined by which rules are more significant and in turn dominate the tested rule (Ndomb). This cross-examination method allows us to find the most relevant rule among very large datasets. We say that the rule is better than another rule according to a combination of a set of measures and not only one measure as previously proposed. Considering example shown in the table 2, using the data set D and supposing that $M = \{M1, M2, M3, M4\}$. The first rule "R1" strictly dominates the second rule"R2" because R1(M1) = 0.9, R1(M2) = 0.8 and R1(M3) = 0.8 and R1(M4) = 0.6 which are all (pair by pair) bigger than R2(M1) = 0.7, R2(M2)=0.7, R2 (M3)= 0.3,R2(M4)=0.4.Similarly, we suppose have R2 dominates R4 .
So we get: $R_1 \succ R_2 \succ R_4$.

Firstly, our approach is to propose two-ways to rank association rules based on dominance relation: The first takes as score $Ndom(R_s, \{R_k\})$ the number of how many rules has the selected rule R_s dominates. $\{R_k\})$ is the set of all association rules without R_s.

On the other hand, the second approach takes as score $Ndomb(\{R_k\}, R_s)$ the number of rules whose dominate the selected rule.

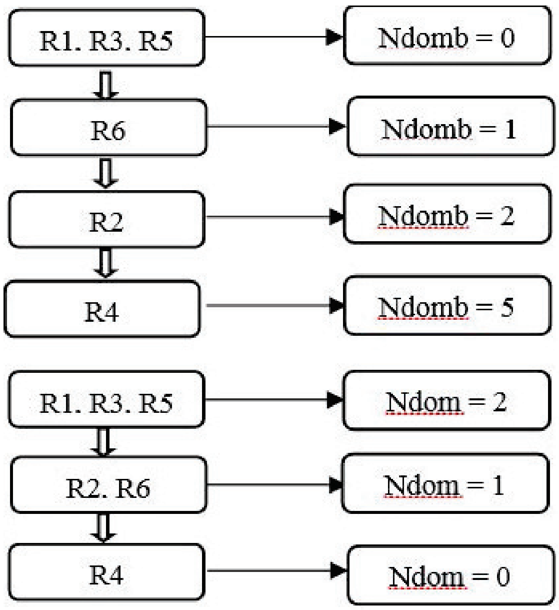

FIGURE 1. Schema of example result

The best rule is the one that is more dominant than it is dominated, and the higher is the Ndom score the more relevant is the rule and the higher is the Ndomb the more irrelevant is the rule. In the example the rule R1 dominate R2 and R4, and it is not dominated by any other rules.

$$Ndom(R_1, \{R_2, R_3, R_4, R_5, R_6\}) = 2$$

$$Ndomb(\{R_2, R_3, R_4, R_5, R_6\}, R_1) = 0.$$

The result of the example is given in the table 3.

We deduce that the rule R4 is the least significant because it has a minimum Ndom and maximum Ndomb. On the other hand, the most relevant rules are R1,R3,and R5. Fig. 1 shows a diagram to illustrate the example results. We choose one of the two approaches to rank the association rules. As we can see the rules R2 and R6 have the same Ndom but the R6 has a lower Ndomb than R2, then we can deduce that the R6 is has a higher significance than R2. Another contribution is to combine ndom and ndomb into a new score: scordom obtained using the DEA method by maximizing ndom and minimizing ndomb since the best rule is the one that is more dominant than it is dominated. DEA (data envelopment analysis) is a linear programming methodology to measure the efficiency of multiple decision making units(DMU) when the production process presents a structure of multiple inputs and outputs. DEA provides an ordinal ranking of relative efficiency compared to the pareto-efficient frontier; We tried to determine the optimal weight, using linear programming so as to maximize the ratio=virtual output/virtual input. Suppose each DMU consume m inputs $(Xi : i = \dots m)$ to produce s outputs $(yr : r = 1 \dots s)$. the chosen dea model evaluates the efficiency of DMU_o, DMU under consideration, by solving the following linear program:

$$Score_i = Maximize \sum_{r=1}^{s} u_r y_{rj}$$

$$Subject\ to : \sum_{i=1}^{m} W_i x_{i0} \ = 1, \qquad i = 1, 2, \dots, m$$

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} w_i x_{ij} \ \leq 0, \qquad j = 1, 2, \dots, n$$

$$w_i \geq \epsilon \qquad i = 1, 2, \dots, m \qquad u_r \geq \epsilon \qquad r = 1, 2, \dots, s$$

Here: x_{ij} and y_{rj} are the inputs and the outputs of the DMU_j, w_i and u_r are the inputs and the outputs weights,x_{i0} and y_{r0} are the inputs and the outputs of the DMU_0.

4.4. **Approach structure.**

In this section, we are presenting the structure of the proposed algorithm. Firstly, we import a transactional dataset which contain the information recorded from transaction. Then, we apply one of the algorithms of the mining association rules process, we choose the simplest one and the best known Apriori algorithm to find the frequent itemsets and generate all association rules. The next step is to evaluate the generated association rules by calculating a set of the interestingness measures and not only one, we can use the objectives measures or the subjectives measures or the both. The last step is to ask the users to give their needs and objectives and choose one of the approaches among the proposed approaches: selection of association rules (select AR), classification and sorting association rules (sort AR), or ranking association rules (rank AR) permit of ranking AR according to a value chosen from three calculated values (ndom, ndomb, scoredom).

The flow chart of the proposed algorithm is shown in Fig 2.

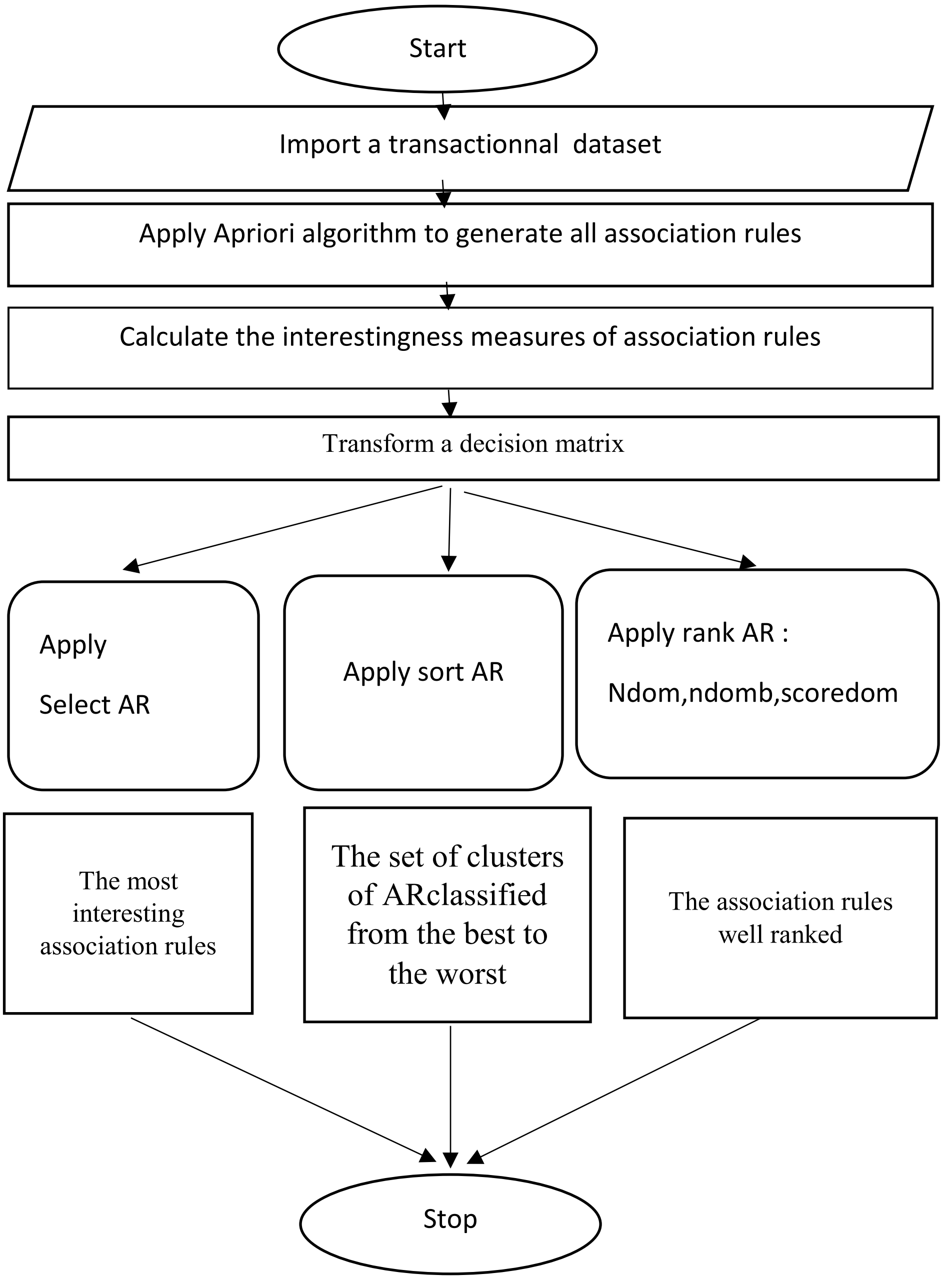

FIGURE 2. Schema of example result

5. EXPERIMENT STUDY

In this section we will illustrate the benefits of the proposed approach. In the first time, we use apriori algorithm to generate the associacion rules from a set

TABLE 4. CHARACTERISTICS OF THE USED DATASETS

Data set	items	transactions
Mushroom	22	8124
Flare1	32	323
Flare2	32	1066
Monks1	19	432
Monks2	19	432
Monks3	19	432
zoo	28	101

TABLE 5. NUMBER OF AR GENERATED FOR EACH DATASET

Datasets	minsup	Number of rules generated
Mushroom	40	2654
Flare1	10	16748
Flare2	10	17174
Monks1	1	38062
Monks2	1	39478
Monks3	1	36470
Zoo	10	11484

of dataset (mushroom, flare1, flare2,monks1, monks2, monks3, Zoo) which are available from the UCI Machine Learning Repository [1]. Table 4 summarizes the characteristics of the used datasets.Table 5 shows the minimum support taken for each dataset chosen and the number of rules extracted from the different datasets. After the generation of association rules, we use the interestingness measures explained in the previous section to evaluate these rules. The measures chosen for the performed test are: : Support (SUP),Confidence(CONF), Lift, Information Gain(IG), Example Counter Example Rate(ECR), Piatetsky Shapiro (PS), Cosinus(COS) and Jacard(JRD). These measures are calculated using the formulas cited in the table 1.

[1]http://archive.ics.uci.edu/ml/

TABLE 6. NUMBER OF AR GENERATED FOR EACH DATASET

Datasets	Mushroom	flare1	flare2	Monks1	Monks2	Monks3	Zoo
AR	2654	16748	17174	38062	39478	36470	11484
Select AR	658	613	3377	509	1788	387	3329

TABLE 7. the first five clusters for each datasets

Datasets	Mushroom	flare1	flare2	Monks1	Monks2	Monks3	Zoo
Cluster1	658:0.98	613:0.87	3377:0.93	509:043	1788:0.34	387:0.48	3329:0.82
Cluster2	155:0.94	254:0.87	454:0.78	2408:033	4111:0.26	1633:0.37	75:0.73
Cluster3	226:0.91	292:0.86	314:0.79	1692:0.30	3304:0.24	1099:0.40	388:0.56
Cluster4	149:0.81	162:0.85	1870:078	2730:0.25	2938:0.20	4154:041	704:0.54
Cluster5	55:0.78	186:083	306:076	2646:0.26	5220:0.20	8422:0.28	140:0.48
Number of cluster	33	55	58	12	14	24	34

5.1. **Selection problem.**

In this subsection, we apply select AR to choose and select the most interesting AR by keeping only the non dominated AR and delet the dominated AR. We show through experiment that select AR can significantly reduce the huge number of rules generated from the datasets. For all measurements, Table 6 compare the size of non-dominated rules of select AR with all the AR. The reduction of AR helps the users and make the interpretation easy and see the most interesting ones;

5.2. **Sorting problem.**

In this subsection, we apply sort AR to cluster the generated AR. Table 7 shows the obtained results for the first five clusters for each datasets. The result of the sort AR show that it can easily classify the clusters of rules which makes it easy for the user to choose the best AR and the worst AR and all that is done without discarding any rule.

5.3. **Ranking problem.**

Now, we apply the proposed algorithm to calculate the values, ndom, ndomb for each rule and then apply DEA to maximize ndom and minimize ndomb and obtained the new score scordom for each rule. The Tables 8, 9 and 10 show the obtained results from a sample of 10 rules for the rules generated by apriori for each datasets (zoo, muchroom,monks).

TABLE 8. RESULT OF A SAMPLE OF 10 RULES GENERATED FROM ZOO DATASET

Association rules	SUP	CONF	LIFT	GI	ECR	PS	JRD	COS	NDom	NDomb	Scordom
8 9 21 ==> 3	41	1	2,463	0,901	1	0,241	1	1	8912	0	1,000000000048
4 9 12 16 22 ==> 1 2 8	16.0	1.0	5.05	1.619	1.0	0.127	1.0	0.894	7686	80	0,9930337862319
8 9 14 21 ==> 0 3 6	18	0,562	2,840	1,044	0,222	0,115	0,391	0,711	4809	118	0,9897248351702
6 8 ==> 7 9	27	0,574	1,234	0,210	0,259	0,050	0,402	0,574	767	1173	0,8978578897113
8 ==> 3 7 9 12 14 21	26	0,313	1,216	0,196	-1,191	0,045	0,181	0,559	389	1695	0,8524033440129
2 5 7 12 ==> 8 11 15 24	13	0,928	7,214	1,976	0,923	0,110	0,866	0,963	633	2667	0,7677638459999
6 9 12 14 17 ==> 0 3 7 8 21	12	0,923	2,453	0,897	0,916	0,073	0,857	0,539	359	3161	0,7247474740630
0 3 7 8 21 ==> 6 9 14	17	0,447	1,964	0,675	-0,235	0,082	0,288	0,575	571	3802	0,6689306474330
8 12 ==> 0 3 6 9 21	16	0,216	1,091	0,087	-2,625	0,013	0,121	0,415	167	5773	0,4973005919428
9 ==> 2 5 8	11	0,137	0,578	-0,547	-5,272	-0,079	0,073	0,251	0	11483	8,70780359197e-05

TABLE 9. RESULT OF A SAMPLE OF 10 RULES GENERATED FROM MUSHROOM DATASET

Association rules	SUP	CONF	LIFT	GI	ECR	PS	JRD	COS	NDom	NDomb	Scordom
2 59 85 ==> 90	3272	0,898	0,975	-0,025	0,887	-0,010	0,816	0,626	2090	0	1,000000000010
36 ==> 24 85 90	3488	0,512	0,943	-0,057	0,047	-0,025	0,344	0,636	1316	0	1,000000000007
34 85 ==> 23 36 86	3272	0,413	1,026	0,0261	-0,418	0,010	0,260	0,643	421	4	0,9984928410000
34 76 85 86 ==> 59	3280	0,748	1,174	0,160	0,663	0,059	0,597	0,688	412	8	0,9969856820000
34 85 ==> 28 86	3328	0,420	1,026	0,026	-0,378	0,010	0,266	0,648	360	64	0,9758854560000
34 85 ==> 39 59 86	3400	0,429	1,026	0,026	-0,327	0,010	0,273	0,655	335	144	0,9457422760000
34 85 ==> 36 63 86	3992	0,504	1,026	0,026	0,017	0,012	0,337	0,710	230	316	0,8809344455861
34 85 ==> 39 86	5402	0,682	1,023	0,022	0,534	0,015	0,518	0,824	218	488	0,8161266083222
34 ==> 85 86	7906	0,998	1,024	0,023	0,998	0,023	0,997	0,998	3	504	0,8100979650008
36 ==> 85	6812	1	1	0	1	0	1	0,915	0	2205	0,1691785980003

TABLE 10. RESULT OF A SAMPLE OF 10 RULES GENERATED FROM MONKS1 DATASET

Association rules	SUP	CONF	LIFT	GI	ECR	PS	JRD	COS	NDom	NDomb	Scordom
13 ==> 1	108	1	2	0,693	1	0,125	1	0,707	370	0	1,000000002303
1 9 ==> 13	54	0,5	2	0,693	0	0,062	0,333	0,5	36168	5	0,9998686350035
1 4 ==> 7 18	24	0,333	2	0,693	-1	0,027	0,2	0,333	30243	105	0,997241343000013
1 9 13 ==> 7	24	0,333	2	0,693	-1	0,027	0,2	0,333	22542	1422	0,9626399030542
5 11 ==> 0 18	12	0,25	1	0	-2	0	0,142	0,166	18057	4675	0,8771740840260
0 15 ==> 9 12	12	0,166	1	0	-4	0	0,090	0,166	13431	6208	0,8368976930003
0 8 ==> 4 16	12	0,111	1,333	0,287	-7	0,006	0,058	0,192	10173	8911	0,7658819820000
6 11 ==> 14 17	6	0,125	1	0	-6	0	0,066	0,117	7908	21631	0,4316904000280
12 18 ==> 1 3 9	6	0,083	1	0	-10	0	0,043	0,117	4382	26350	0,3077084760002
18 ==> 0 5 9 10	6	0,027	1	0	-34	0	0,014	0,117	36	36220	0,04839472400000

These experiments show that our proposed method can rank all AR and make it easy for the expert to choose the best rules from a huge number of AR. Let's now, show the ability of our approach. We compare the results of our experiments to another one based on Data Envelopment Analysis DEA. The first method [13] based on Data Envelopment Analysis (DEA), aims to estimate and rank the efficiency of association rules with multiple criteria: Support, confidence and two other subjective measures: itemset value and Cross-selling. Toloo et al. [14] proposes a new DEA-based methodology for ranking units, which identifies the best efficient unit by considering only output data of DMUs. An example of market basket data is adopted from Chen et al. [13]. Association rules are first discovered by the Apriori algorithm, in which minimum support and minimum confidence are set to 1.0 and 10.0, respectively. Forty-six rules then are identified and presented in Table 11.

By applying our algorithm to data presented in table 11 we obtain ndom,ndomb, scordom correspondant for each rule. Table 12 presents results of ranking efficient rules in comparison to Chen′s method [13] and Toloo et al.′s method [14]. Table 12 shows that the results of proposed method are different from results of previous methods. Obviously, proposed method provides decision makers with more accurate and simple results as its main advantage to previous methods. In comparison to previous works, our method is computationally efficient and also ranks all association rules.

6. Conclusion

The knowledge post-processing phase becomes very challenging in association rules mining process, In this paper, we proposed a new approach for selecting, sorting and ranking association rules with multiple objective and subjective criteria, using a method based on dominance relation. The proposed approach provides more insights into the rules discovered and can assist rule evaluation and selection. The efficiency and the applicability of the proposed method is illustrated by comparing its results with those of the previous methods. In future studies we will try to ameliorate our approach to be able to rank the rules by combining dominance relation and other methods and we will try to apply the proposed algorithm to test it in the context of big data.

TABLE 11. DATA OF ASSOCIATION RULES AND SUMMARY OF RESULTS OF OUR APPROACH

Association Rule	Support	Confidence	Itemset value	Cross-selling	Ndom	Ndomb	Scordom
1	3.87	40.09	337.00	25.66	3	0	1,00000000000003
2	1.42	18.17	501.00	11.63	0	7	0,847826087000343
3	2.83	17.64	345.00	11.29	0	0	1,00000000000050
4	2.34	30.83	163.00	19.73	0	3	0,934782609000436
5	2.63	23.90	325.00	15.3	0	4	0,913043478000418
6	1.19	55.65	436.00	35.61	2	0	1,00000000000330
7	1.19	47.42	598.00	30.35	1	0	1,00000000521227
8	1.19	15.70	436.00	52.91	2	8	0,826086962480702
9	1.19	10.82	598.00	36.45	0	10	0,782608696000267
10	1.19	12.32	436.00	20.08	0	21	0,543478261000047
11	1.19	12.32	598.00	40.04	2	7	0,847826090437451
12	3.87	38.08	337.00	103.97	4	0	1,00000000000011
13	1.18	15.09	710.00	41.19	0	1	0,978260870000469
14	2.44	15.22	554.00	41.56	3	1	0,978260870000014
15	2.14	28.21	372.00	77.02	4	0	1,00000000000011
16	2.51	22.81	534.00	62.26	8	0	1,00000000571202
17	1.19	50.92	436.00	139.02	5	0	1
18	1.19	45.25	598.00	123.52	9	0	1,00000000000006
19	1.19	11.70	436.00	43.54	0	16	0,652173913000098
20	1.19	11.70	598.00	62.50	2	4	0,913043478690980
21	1.42	13.99	501.00	61.16	2	4	0,913043478690980
22	1.18	12.23	710.00	53.45	0	0	1,00000000000050
23	1.50	13.64	698.00	59.59	4	0	1,00000000000011
24	2.83	27.82	345.00	78.17	4	0	1,00000000000011
25	2.44	25.27	554.00	71.0	14	0	1,00000000004252
26	1.25	15.97	718.00	44.87	5	0	1
27	1.22	34.89	339.00	98.04	0	2	0,956521739000436
28	1.30	35.12	435.00	98.68	2	1	0,978260870000824
29	1.42	33.81	534.00	95.01	7	1	0,978260870064673
30	1.91	25.26	380.00	70.97	2	2	0,956521739027664
31	1.43	37.14	618.00	104.35	14	0	1,00000000004252
32	2.38	21.63	542.00	60.78	7	0	1,00000000041041
33	1.18	30.24	366.00	84.98	0	5	0,891304348000375
34	1.23	29.36	626.00	82.51	6	0	1,00000000083827
35	1.58	22.65	354.00	63.64	0	5	0,891304348000375
36	2.34	22.99	163.00	22.76	0	6	0,869565217000371
37	2.14	22.14	372.00	21.92	1	4	0,913043477999992
38	1.91	11.94	380.00	11.82	0	7	0,847826087000343
39	2.03	18.42	360.00	18.23	0	7	0,847826087000343
40	1.19	30.73	436.00	30.43	1	6	0,869565217000016
41	2.63	25.87	325.00	67.52	2	2	0,956521739027664
42	2.51	25.98	534.00	67.81	11	0	1,00000000247323
43	1.50	19.16	698.00	50.02	5	0	1
44	2.38	14.85	542.00	38.75	2	3	0,934782609068463
45	2.03	26.73	360.00	69.78	2	1	0,978260870000824
46	1.19	30.73	598.00	80.22	7	2	0,956521739049919

TABLE 12. RANKING OF PROPOSED METHOD IN COMPARISON TO CHEN'S METHOD AND TOLOO ET AL.'S METHOD

Ranking	Association Rules			
Ranking	Chen's Method	Toloo et al.'s Method	Proposed method(NDom)	Proposed method(scordom)
1	26	18	31	31
2	22	23	18	18
3	18	26	26	43
4	17	12	17	17
5	7	31	43	26
6	23	43	23	23
7	6	22	12	12
18	43	6	1	1
9	31	17	6	6
10	12	1	7	7
11	1	7	22	22

REFERENCES

[1] R. AGRAWAL, R. SRIKANT: *Fast algorithms for mining association rules*, In Proceeding VLDB '94 Proc. 20th Int. Conf. Very Large Data Bases, **1215** (1994), 487–499.

[2] N. PASQUIER, Y. BASTIDE, R. TAOUIL, L. LAKHAL: *Discovering frequent closed item-sets*, In ICDT, (1999), 398–416.

[3] M. M. J. KABIR, S. XU, B. H. KANG, Z. ZHAO: *A New Evolutionary Algorithm for Extracting a Reduced Set of Interesting Association Rules*, In: Arik S., Huang T., Lai W., Liu Q. (eds) Neural Information Processing, Lecture Notes in Computer Science, Springer. Cham., **9490** (2015), 133-142.

[4] A. SILBERSCHATZ, A. TUZHILIN: *What Makes Pattens Interesting in Knowledge Discovery Systems*, IEEE Trans. Knowl. Data Eng., **8**(6) (1996), 970-974.

[5] H. X. HUYNH, F. GUILLET, T. Q. LE, H. BRIAND: *Ranking objective interestingness measures with sensitivity values*, VNU J. Sci. Nat. Sci., **24** (2008), 122-132.

[6] M. OHSAKI, Y. SATO, S. KITAGUCHI, H. YOKOI, T. YAMAGUCHI: *Comparison between Objective Interestingness Measures and Real Human Interest in Medical Data Mining*, In: Orchard B., Yang C., Ali M. (eds) Innovations in Applied Artificial Intelligence. IEA/AIE 2004. Lecture Notes in Computer Science, Springer, **3029** (2004), 1072-1081.

[7] G. GARIMA, J. A. VASHISHTHA: *Study of Interestingness Measures for Knowledge Discovery in Databases—A Genetic Approach*, In: Jain L., Behera H., Mandal J., Mohapatra D. (eds) Computational Intelligence in Data Mining - Volume 2. Smart Innovation, Systems and Technologies, Springer, **32** (2015), 69-79.

[8] J. BLANCHARD, F. GUILLET, P. KUNTZ: *Semantics-based classification of rule interestingness*, In: Zhao Y, Zhang C,Cao L, eds. Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction. Hershey, PA: IGI, (2009), 56-79.

[9] M. GAVRILOV, D. ANGUELOV, P. INDYK, R. MOTWANI: *Mining The Stock Market: Which Measure Is Best?*, Acm. Sigkdd, (2000), 487-496.

[10] P. N. TAN, V. KUMAR, J. SRIVASTAVA: *Selecting the Right Interestingness Measure for Association Patterns*, In Proc. eighth ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. - KDD '02, **2** (2002), 32-41.

[11] P. LENCA, P. MEYER, B. VAILLANT, S. LALLICH: *On selecting interestingness measures for association rules: User oriented description and multiple criteria decision aid*, Eur. J. Oper. Res., **184**(2) (2008), 610-626.

[12] A. DAHBI, M. MOUHIR, Y. BALOUKI, T. GADI: *Classification of association rules based on K-means algorithm*, in: 4th IEEE International Colloquium on Information Science and Technology (CiSt), IEEE, Tangier, Morocco, (2016), 300-305.

[13] M. C. CHEN: *Ranking discovered rules from data mining with multiple criteria by data envelopment analysis*, Expert Syst. Appl., **33**(4) (2007), 1110-1116.

[14] M. TOLOO, B. SOHRABI, S. NALCHIGAR: *A new method for ranking discovered rules from data mining by DEA*, Expert Syst. Appl., **36**(4) (2009), 8503-8508.

[15] A. DAHBI, Y. BALOUKI, T. GADI: *A new method for selecting interesting association rules using genetic algo-rithm with multiple criteria*, ARPN Journal of Engineering and Applied Sciences, **12**(19) (2017), 5405-5412.

[16] A. DAHBI, S. JABRI, Y. BALOUKI, T. GADI: *A New Method to Select the Interesting Association Rules with Multiple Criteria*, International Journal of Intelligent Engineering and Systems, **10**(5) (2017), 191-200.

[17] B. LIU, W. HSU, S. CHEN, Y. MA: *Analyzing the subjective interestigness of association rules*, IEEE Intell. Syst., **15**(5) (2000), 47-55.

[18] L. CAO, D. LUO, C. ZHANG: *Knowledge actionability: satisfying technical and business interestingness*, Int. J. Bus. Intell. Data Min., **2**(4) (2007), 496-514.

[19] S.O. REZENDE, E.A. MELANDA, M.L. FUJIMOTO, R.A. SINOARA, V.O. DE CARVALHO: *Combining data-driven and user-driven evaluation measures to identify interesting rules*, Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction, (2009), 38-55.

[20] R. PAUL, T. GROZA, J. HUNTER, A. ZANKL: *Semantic interestingness measures for discovering association rules in the skeletal dysplasia domain*, J. Biomed. Semantics, **5** (2014), 1-13.

[21] S. ANTOINETTE AROUL JEYANTHI, S. PANNIRSELVAM: *Post Analysis Based AKD Using Domain Ontology for Disease Prediction*, IOSR Journal of Computer Engineering (IOSR-JCE), **19**(2) (2017), 19-26.

[22] L. GENG, H. J. HAMILTON: *Interestingness measures for data mining*, ACM Comput. Surv., **38**(3) (2006), 1-32.

[23] H. X. HUYNH, F. GUILLET, T. Q. LE, H. BRIAND: *Ranking objective interestingness measures with sensitivity values*, VNU Journal of Science, Natural Sciences and Technology, **24** (2008), 122-132.

[24] A.K. SEN: *Collective Choice and Social Welfare*, Elsevier, US, 1979.
[25] J. FIGUEIRA, S. GRECO, M. EHROGOTT, M. EHRGOTT: *Multiple Criteria Decision Analysis: State of the Art Surveys*, Springer-Verlag, New York, US, 2005.
[26] B. ROY: *Multicriteria Methodology for Decision Aiding*, Springer US, 1996.

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
HASSAN 1ST UNIVERSITY, FST SETTAT
SETTAT, MOROCCO
Email address: azdine.im@gmail.com

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
HASSAN 1ST UNIVERSITY, FST SETTAT
SETTAT, MOROCCO
Email address: jabrisiham14@gmail.com

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
HASSAN 1ST UNIVERSITY, FST SETTAT
SETTAT, MOROCCO
Email address: balouki.youssef@gmail.com

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
HASSAN 1ST UNIVERSITY, FST SETTAT
SETTAT, MOROCCO
Email address: gtaoufiq@yahoo.fr