

# Multi-agent-based modeling for extracting relevant association rules using a multi-criteria analysis approach

Addi Ait-Mlouk<sup>1</sup> · Fatima Gharnati<sup>1</sup> · Tarik Agouti<sup>1</sup>

Received: 12 May 2016 / Accepted: 28 May 2016 / Published online: 10 June 2016  
© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** Recently, association rule mining plays a vital role in knowledge discovery in database. In fact, in most cases, the real datasets lead to a very large number of rules, which do not allow users to make their own selection of the most relevant. The difficult task is mining useful and non-redundant rules. Several approaches have been proposed, such as rule clustering, informative cover method and quality measurements. Another way to selecting relevant association rules, we believe that it is necessary to integrate a decisional approach within the knowledge discovery process. Therefore, in this paper, we propose an approach to discover a category of relevant association rules based on multi-criteria analysis. In other side, the general process of association rules extraction becomes more and more complex, to solve such problem, we also proposed a multi-agent system for modeling the different process of our proposed approach. Therefore, we conclude our work by an empirical study applied to a set of banking data to illustrate the performance of our approach.

**Keywords** Data mining · Association rules · Multi-criteria analysis · Multi-agent system

## 1 Introduction

Data mining techniques have been designed to discover useful knowledge from database [1,2], in fact, data mining techniques play a vital role in many business analytics and predictive applications that come to complete data analysis

and predictive techniques. Seen the production increase of fast information, a large number of data are generated and stored in computer database in need to discover knowledge and useful information [3], by applying data mining techniques such as association rules to present the result as a valid element for further use.

Association rule is a technique that allows the user to discover the correlation between different object in databases. The result presented in the form of antecedent and consequence, for example, an association rule extracted from a transactional database:  $\text{Mouse} \wedge \text{Keyboard} \rightarrow \text{Computer}$

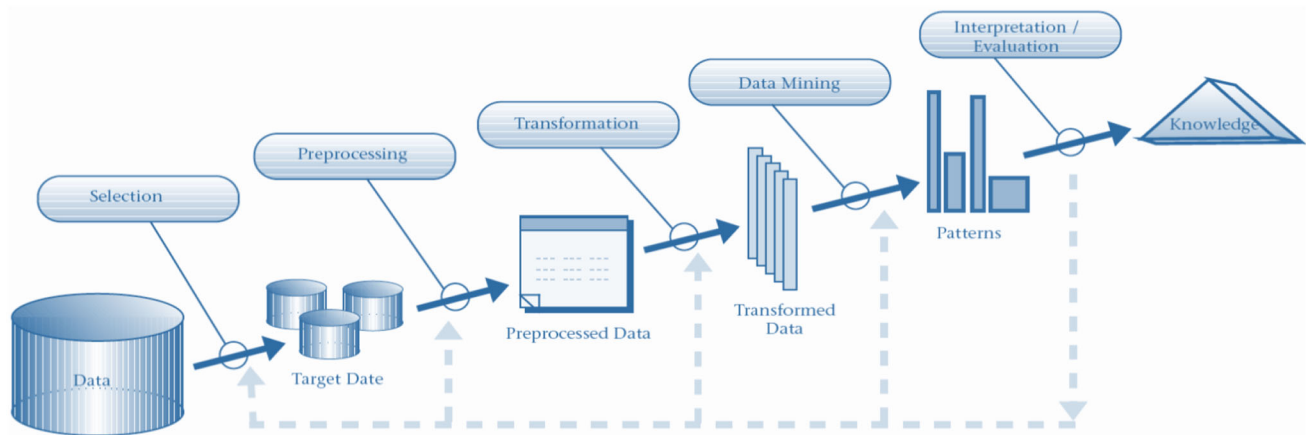
This rule indicates that the customer who buys mouse and keyboard together tends also to buy computer. The support defines the proportion of transactions that contains mouse and keyboard; therefore, the confidence is the proportion of transactions that contains mouse and keyboard, which also contains computer.

Although, association rule technique has many drawbacks such as a large number of discovered rules, redundancy, the production of non-interesting rules [5,6,35] and the complexity of rule extraction process. So, it is important to propose an approach that can help the user to make their own choice of interesting rules according to their specific needs. Its the case of our proposed approach of MCA within KDD process in which we are interested in evaluation of extracted rules by selecting the most relevant from the large number extracted, and provide a multi-agent approach to automate the process of extraction.

The remainder of this paper is organized as follows: in Sect. 2, we presented an overview of related work and concepts for extracting association rules. In Sect. 3, the contribution of the multi-criteria analysis approach is proposed. In Sect. 4, we describe the proposed architecture of the multi-agent system for modeling our proposed approach. In Sect. 5, we focus our attention on an empirical study to illustrate the

✉ Addi Ait-Mlouk  
aitmlouk@gmail.com

<sup>1</sup> Faculty of Science Semlalia, Cadi Ayyad University, Marrakech, Morocco



**Fig. 1** The knowledge discovery process

performance of our approach, finally this paper ended by a result discussion and concluding section.

## 2 Related work

In the literature, several approaches have been designed to manage the complexity of rules extraction, but these approaches are still limited and costly in terms of interestingness, redundancy, and the huge number of extracted rules. In this section, an overview of related work is presented such as KDD, Association rules, extraction algorithms, multi-agent system and quality measurements.

### 2.1 Knowledge discovery in databases

The process of KDD consists of five major steps, namely the extraction goal, data selection, data transformation, data mining techniques, and finally, the interpretation of results, these steps are given in Fig. 1.

**Objective of extraction** The first step of the KDD process is to understand the purpose of extraction to choose the appropriate techniques for solving the problem.

**Data selection** The selection of significant samples data, to minimize the mass of available data, and to facilitate the study of the main objective.

**Data processing** Make the necessary transformations of data using an extraction, transformation, and loading (ETL) techniques between different data sources.

**Application of data mining techniques** There are two types of models, classification models of organizing classes in data (identifiable sets) and regression models of determining variable dependency between them.

**Interpretation of results** Finally, the resulting information must be analyzed according to the specified objectives.

#### 2.1.1 Association rules concepts

Association rule was initiated by Agrawal [3], for the first time, to analyze transactional databases. It is a statement of the form  $A \rightarrow B$ , where  $A, B \subset I$  and  $A, B$  are non-empty set. The set  $A$  is called antecedent of the rule, the set  $B$  is called the consequent of the rule, and  $I$  is an itemset.

#### 2.1.2 Quality measurements

To evaluate the rules issued from extraction algorithms, the notion of interesting and relevance are introduced. Let  $A$  be an itemset,  $A \rightarrow B$ , a rule and  $T$  a set of transactions, we define the support, confidence, lift, and conviction as follows:

**Rule support** The support value is defined as the proportion of transactions in the database, which contains the itemset  $A$ :

$$\text{Supp}(A \rightarrow B) = \frac{|t(A \cup B)|}{t(A)} \quad (1)$$

**Confidence** The confidence determines how frequently items in  $B$  appear in transaction that contains  $A$ , the formal definition is:

$$\text{Confidence}(A \rightarrow B) = \text{Supp}(A \cup B) / \text{Supp}(A) \quad (2)$$

**Lift** The lift defined as the ratio of the observed support to that expected if  $A$  and  $B$  were independent.

$$\text{Confidence}(A \rightarrow B) = \text{Supp}(A \cup B) / \text{Supp}(A) \quad (3)$$

**Conviction** is another measure proposed to handle some of the weaknesses of confidence and lift, it is sensitive to rule direction:

$$\text{Confidence}(A \rightarrow B) = \text{Supp}(A \cup B) / \text{Supp}(A) \quad (4)$$

**Table 1** The set of quality measurement

Symbol	Pessimistic	Optimistic
PS	Piatetsky–Shapiro	[13]
LOE	Loevinger	[14]
ZHANG	Zhang	[15]
INDIMP	Implication index	[16]
LIFT	Lift	[17]
SURLP	Surprise	[18]
SEB	Sebag et Schoenauer	[17]
CONV	Conviction	[19]
IQC	Cohen quality index	[20]
IPD	Discriminative probabilistic index	[21]
LAP	Laplace	[22]

To establish the interestingness measure, Bayardo and Agrawal [28] considered that the interesting rule must reside along the initial parameters of support and confidence. In the other way, Piatetsky–Shapiro [7], proposed a new measure called Rule-Interest. Hilderman and Hamilton [8], proposed sixteen diversity measures. In addition, Carvalho et al. [9] evaluated eleven objective interestingness measures by interest. Huynh et al. [10] proposed a clustering approach to identify clusters. In addition, Gavrilov et al. [11] studied the similarity between the measures. In addition, computation techniques of suitable objective measures are proposed by Xuan-Hiep and Fabrice [12]. Moreover, the other measurements are given in Table 1.

### 2.1.3 Extraction algorithms of association rules

In the literature, several algorithms have been proposed to extract association rules, among them we find Apriori, the key algorithm proposed by Agrawal to extract the frequent itemset. These algorithms can be classified into three large categories namely frequent algorithms [23], maximum algorithms [3], and closed algorithms [24].

*Mining frequent itemsets* Mining frequent itemset is the basic technique of extraction rules it proposed in the first time by Agrawal to analyze the problem of shopping basket. For this category, the Apriori algorithm constituted the key algorithm for extracting frequent itemset. Moreover, it constitutes the basis of the majority algorithms that are coming to extract association rules, among them, we find AprioriTID, FP-Growth, Partition, DIC, Eclat, etc.

*Mining closed itemsets* An itemset is closed in a data set if there exists no superset that has the same support count as this original itemset. The extraction of frequent closed itemset is based on the closing of the Galois connection [25]. For this category, several algorithms have been proposed, among them, we can mention close algorithm, Pascal, etc.

*Mining maximal itemsets* An itemset is maximal frequent if none of its immediate supersets is frequent. Several algorithms have been designed to mine maximal itemset, among them, we find MaxMiner [28], Pincer Search [29], MaxEclat [25], etc.

## 3 Research methodologies

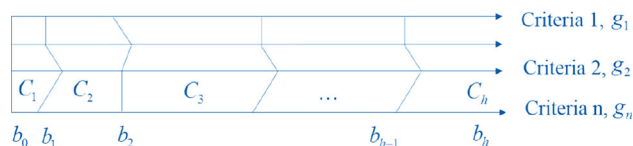
In this section, we discuss the various methodologies constructing our proposed approach, and then we start with the choice of Multi-Criteria Analysis method.

### 3.1 The choice of the MCA method

The MCA presented as an alternative to classical methods of optimization based on the definition of unique function reflects the consideration of several criteria. The interest of methods is to consider a different nature of criteria without necessarily turning them into economic criteria, either in a single function. This is not search of an optimum, but a compromise solution that can take various forms: choice, assignment or classification. In the literature [26,27], we encounter three problems like sorting, selection, and arrangement. In our context of a large number of extracted rules, we have to assign the extracted rules to the category of the most relevant, so we located at the assignment problem; therefore, the suitable method is ELECTRE TRI.

### 3.2 ELECTRE TRI method

Electre Tri is a suitable method adapted to simplify and solve complex decision problems of ranking type. The principal of this method is to assign a set of alternatives noted  $A = a_0, a_2, a_3, a_m$  on which the decision is based. We note  $F = 1, 2n$  the set of criteria. Each alternative of the set  $A$  will be evaluated by a real function expressing the evaluation of alternative for given criteria, we note  $G = g_0, g_2, g_3, , g_m$  the evaluation of the alternative of the criteria considered [31]. The alternatives are not compared with each other, but with thresholds reflecting the boundary between  $h$  classes predefined, noted  $C = C_1, C_2, C_3...C_h$ . Each alternative will be compared to the borders of each category, forming a profile  $B = b_1, b_2, b_3, b_h$ , an illustration is given in Fig. 2.



**Fig. 2** The illustration of the sort problem

The affectation of alternatives in categories is based on the concept of classification. An action  $a$  of set  $A$  outrank  $b_h$  noted  $b_h$ , if  $a$  as good as  $b_h$  on all criteria, and  $a$  not bad as  $b_h$  on the majority of the criteria. Electre Tri proceeds in two consecutive steps [32]:

**Step 1:** formulation of outranking relation  $S$  for the comparison of  $a$  to  $b_h$ .

Computation of partial concordance indices  $C_j(a, b_k)$ : it expresses to which the extent  $a$  outrank  $b_h$  or  $a$  is at least as good as  $b_h$ .

$$C_j(a, b_k) = \begin{cases} 0 & \text{if } g_j(b_k) - g_j(a) \geq p_j(b_k) \\ 1 & \text{if } g_j(b_k) - g_j(a) \leq q_j(b_k) \\ \text{otherwise} & \frac{g_j(b_k) + g_j(a) - g_j(b_k)}{p_j(b_k) - q_j(b_k)} \end{cases} \quad (5)$$

The computation of global concordance index  $C(a_h, b_k)$ :

$$C(a_h, b_k) = \frac{\sum_{j \in F} K_j C_j(a, b_h)}{\sum_{j \in F} K_j} \quad (6)$$

With:  $K_j$  Weight of criteria  $j$ ,  $C_j(a, b_h)$  The partial concordance indices of criteria  $j$

The computation of the discordance indices  $d_j(a, b_k)$ :

Computation of partial discordance index  $d_j(a, b_k)$ : it expresses to which extent criterion is opposed to the statement  $a$  outrank  $b_h$ .

$$d_j(a, b_k) = \begin{cases} 0 & \text{if } g_j(a_h) \leq g_j(b_h) + p_j(b_h) \\ 1 & \text{if } g_j(a_h) > g_j(b_h) + v_j(b_h) \\ \text{else} & \in [0, 1] \end{cases} \quad (7)$$

The computation of credibility index is based on the global concordance and the partial discordance index, see Fig. 3.

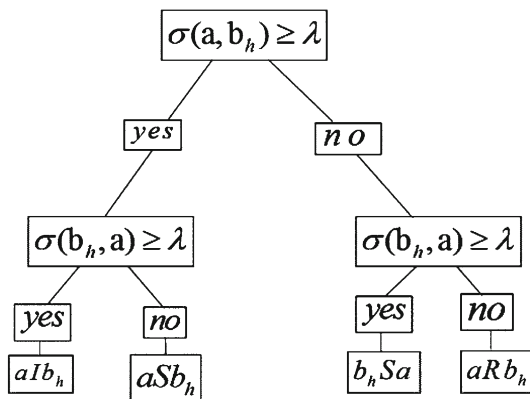


Fig. 3 The outranking relations

$$C(a, b_h) = \prod_{j \in \bar{F}} \frac{1 - d_j(a, b_h)}{1 - C(a, b_h)} \quad (8)$$

With:

$\bar{F} = j \in F : d_j(a, b_h) > C(a, b_h)$ ,  $C(a, b_h)$ : Global concordance index  $d_j(a, b_h)$ : Discordance indices

**Step 2:** assignment procedures.

Two assignment procedures, pessimistic and optimistic, are available to assign a set of actions to different categories.

The pessimistic assignment compares the alternative  $a$  to  $b_i$  for  $i = h, h - 1, \dots, 0$ , then assign alternative  $a$  to category  $C_{h+1}(a \rightarrow C_{h+1})$ .

The optimistic assignment: compare the alternative  $a$  to  $b_i$  for  $i = 1h$  and being the first profile such that  $b_h P a$  or  $b_h Q a$  is satisfied, ( $P, Q$  represent strongly or weakly preferred, then assign alternative  $a$  to category  $C_h(a \rightarrow C_h)$ .

### 3.3 Multi-agent system

A multi-agent system (MAS) is a collection of autonomous agents, which interact with each other or with their environments to achieve one or more objectives.

Data mining and MAS have been used for building a complex system [34], they are combined to produce automatic data mining system. In this work, we used MAS to modeling several autonomous intelligent agents: Mining Rules Agent (MR-Agent), Quality Measurement Agent (QM-Agent), Decision Support Agent (DS-Agent), Principal Agent (MCA-Principal Agent), Control agent (Control Agent) and the user interface agent. Each one has a specific task to achieve, the combination of the whole tasks will produce the main objective of the proposed approach.

## 4 The Proposed approach

The proposed approach is divided into four modules. The first is the association rules extraction module, for extracting rules from dataset using one of the efficacy algorithms of extraction. The second is the decision support module for making the appropriate choice of method according to the main objective and user specifications. The third is the quality measurement module of association rules [6,33], and the last one is the main module of association rules, quality measurement and decision support for evaluating the set of extracted rules. The details of the multi-agent approach are given in Fig. 4.

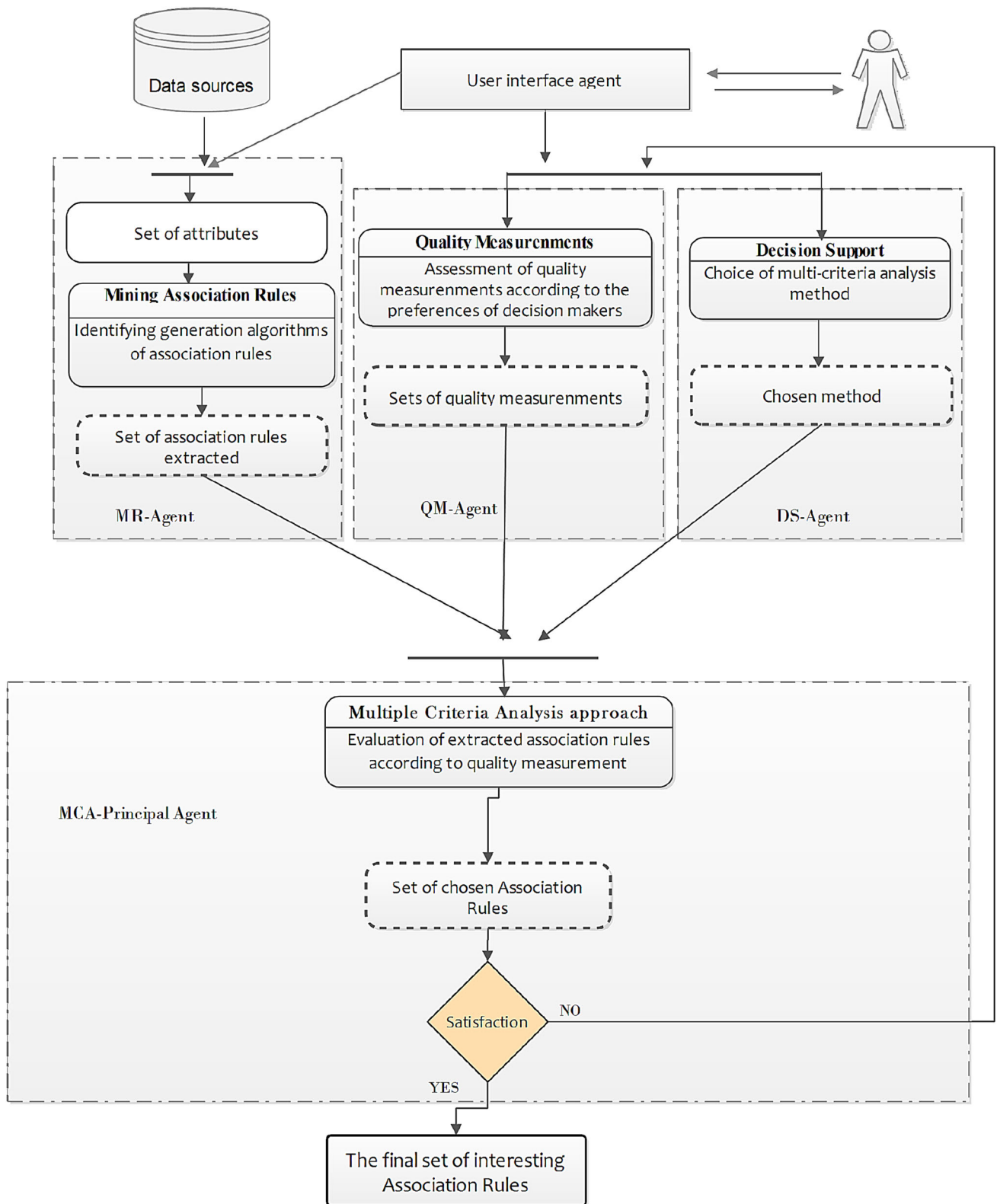


Fig. 4 The proposed multi-agent modeling

#### 4.1 Module 1: mining association rules

The extraction of association rules is processed in two different steps, the first is mining of all frequent itemsets, and the second is the extraction of association rules from frequent itemset. For testing the performance and the relevance of our proposed approach, we used Apriori algorithm on a set of attributes to generate an important number of rules.

#### 4.2 Module 2: decision support

The decision support is a process that uses a set of information available at a given time to formulate the problem and reach the decision on a specific object. In this module, we study the context and the main objective of the extraction problem to choose an appropriate method to be used in the next step.

#### 4.3 Module 3: quality measurement

The Apriori algorithms and its derivatives provide an elegant solution to the rules extraction problem, but produce a large number of rules, selecting certain rules without interest and ignorant of interesting rules. There must be other measures to complement the support and confidence measures. The measures of interest can play a key role to filter the rules, automatically extracted according to criteria adapted to user needs. Seen the active search on rules interestingness, we encounter many measurements in the literature [6] (See, Table 1 in related work for a list of some quality measures).

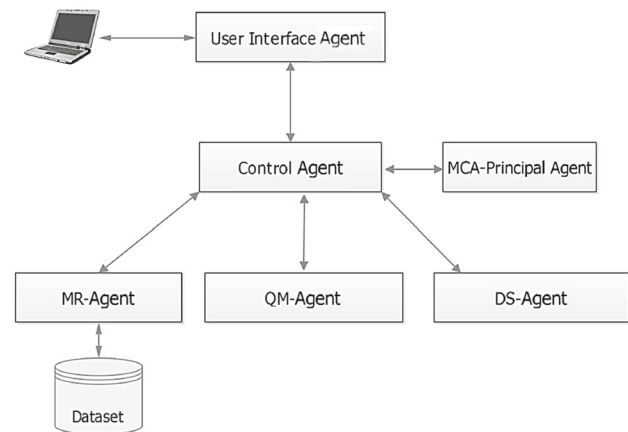
#### 4.4 Module 4: MCA module

The main module consists in evaluation of choice of extracted rules using the selected multi-criteria analysis method in which we use extracted rules as alternatives and quality measurement as criteria. This process will give us a set of chosen rules according to decision makers' preferences, in case of satisfaction we obtain the final set of relevant rules, if not, we change a set of thresholds and parameters to get the appropriate results.

### 5 Multi-agent-based modeling

Multi-agent systems today represent a new technology for a design and control of complex systems. It is composed of independent software and hardware entities called agents; this system usually has several important features such as parallelism, robustness and scalability [34].

Seen the active search in data mining, the KDD become more and more complex. To solve such problem, our proposed approach identified by several agents work in cooperation to achieve the task of relevant rules extraction, it is



**Fig. 5** The architecture of the proposed multi-agent system

composed by six agents: MR-Agent, QM-Agent, DS-Agent, MCA-Principal Agent, the user agent and the control agent. The details are given in Fig. 5.

The extraction of association rules task is performed by negotiation [34] between the control agent, MR-agents and user interface agents; these agents work in collaboration to achieve the required goals. The control agent receives a set of quality measurement, the selected method and a set of rules then send it to the principal agent to perform the evaluation task of association rules. All these agents should recurrently interact to share information and to perform tasks to achieve the process of extraction rules.

#### 5.1 User interface agent

This agent is responsible for interaction with users or domain experts to determine a set of specifications such as number of attributes, minimum support threshold value, and other detail information of the processed dataset. It is responsible for receiving a set of specifications from environment and send it to the control agent for further use by other agents.

#### 5.2 Control agent

The main function of this agent is to manage the communication between all components of the system, and control the data transmission among them. It is responsible for transferring the user needs from the user interface agent to MR-agent, QM-agent and DS-agent and sending back the result to the control agent to save it in the knowledge base. Once the result is saved in the knowledge base, the control agent, transfer it to the principal agent to begin the evaluation task of extracted rules.

### 5.3 MR-agent: mining association rules

The MR-agent receives a dataset and chooses the appropriate algorithm to extract association rules after receiving the minimum support from the control agent and sends the result to the knowledge base to be used by the principal agent.

### 5.4 DS-agent: decision support agent

In the literature, we encounter different methods of multi-criteria analysis. The DS-Agent is the responsible for choosing the most suitable method according to the studied problem, user needs and send it to the knowledge base to be used by the principal agent in the evaluation step.

### 5.5 QM-agent: the quality measurement agent

The objective of this agent is to filter the appropriate measures according to the user needs from those presented in the literature (See Table 1 in related work) and save them in the knowledge base for further use.

### 5.6 MCA-principal agent: multi-criteria analysis agent

The principal agent serves as a main agent, which facilitates the choice of relevant rules which have been extracted by MR-Agent and QM-Agent. It keeps track of the names and capacities of all registered agents in the system. Once the principal agent received the a of extracted rules from the control agent, and received a set of quality measurement from QM-Agent, it performs the extraction of relevant rules; the sequence of operation between different agents constituted the system given in Fig. 6.

This diagram shows the sequence of operations during the execution of the proposed multi-agent system. Moreover, it shows how the distribution of the mining tasks facilitate the process of relevant rules extraction.

## 6 Experimental results

Our research is based on a real dataset; after the preprocessing and the transformation of data, we select a set of records to facilitate the extraction process, and we focused on this study on extracting the association rules by applying Apriori algorithm. Moreover, the interesting and useful association rules are extracted using a multi-criteria analysis approach.

In this section, we applied the proposed approach on a real dataset commonly used in the field of KDD [36], which identifies the characteristics of a set of customers who filed a credit application file, as a case study to illustrate the performance of our proposed approach. In addition, according to decision makers preferences we used a threshold minimum support = 0.33, confidence = 0.75, max rule length = 3 and

lift = 1, for extracting frequent itemsets, then we obtain 27 extracted rules given in Table 2.

This table presents twenty-seven extracted rules in which some of them are redundant like the rules Rule 20, Rule 22 and Rule 24; in other ways, some other rules are not interesting for the user.

The Apriori algorithms and its derivatives provide an important solution to the rules extraction problem, but produce a large number of rules, selecting certain rules without interest and ignorant of interesting rules. There must be other techniques to complement the support and confidence measures. The measures of interest can play a key role to filter the rules automatically extracted according to criteria adapted to the decision makers preferences.

The usefulness and relevance of association rules extracted by the extraction algorithms are a critical problem. In fact, in most cases, the real datasets lead to a very large number of association rules, which does not allow users to make their own selection of the most relevant. In this context, we believe that the integration of MCA approach within the process of rules extraction combined with multi-agent system would be particularly useful for the decision makers who are suffering from the large number and the complexity of rules extraction process. Therefore, the next step is the application of our proposed approach to automate the process of relevant rules extraction. We used a set of rules previously extracted by MR-Agent (Table 2) as actions to be evaluated according to the chosen criteria (support, confidence, lift) given by QM-Agent, to conduct a comparative study between rules using Electre Tri methods.

The decision matrix is used to describe a multi-criteria decision analysis (MCDA) problem. An MCDA problem, where there are  $M$  alternatives (rules) and each one needs to be evaluated on  $N$  criteria, can be described by the decision matrix that has  $N$  rows and  $M$  columns; in our case study, the decision matrix is given in Table 3.

The next step is to define a set of profiles according to decision makers preferences that can be compared with the extracted association rules, the thresholds are given in Table 4.

As the assignment is done in three distinct categories, two profiles  $b_1$  and  $b_2$  are defined, where  $b_1$  is the border between the state class interested and status Medium, and  $b_2$  is the border between the state Medium and the state Not interested.

The importance of each criterion in the decision-making resulted in predefined thresholds, as given in Table 5, the definition of thresholds is given by decision makers according to their specific needs.

After the definition of thresholds, the next step is the computation of concordance index  $c(a, b_h)$  Eq. (5), and discordance index  $d_j(a, b_k)$  Eq. (7); thus, the formulation of outranking relation  $S$  by the computation of the degree of credibility.

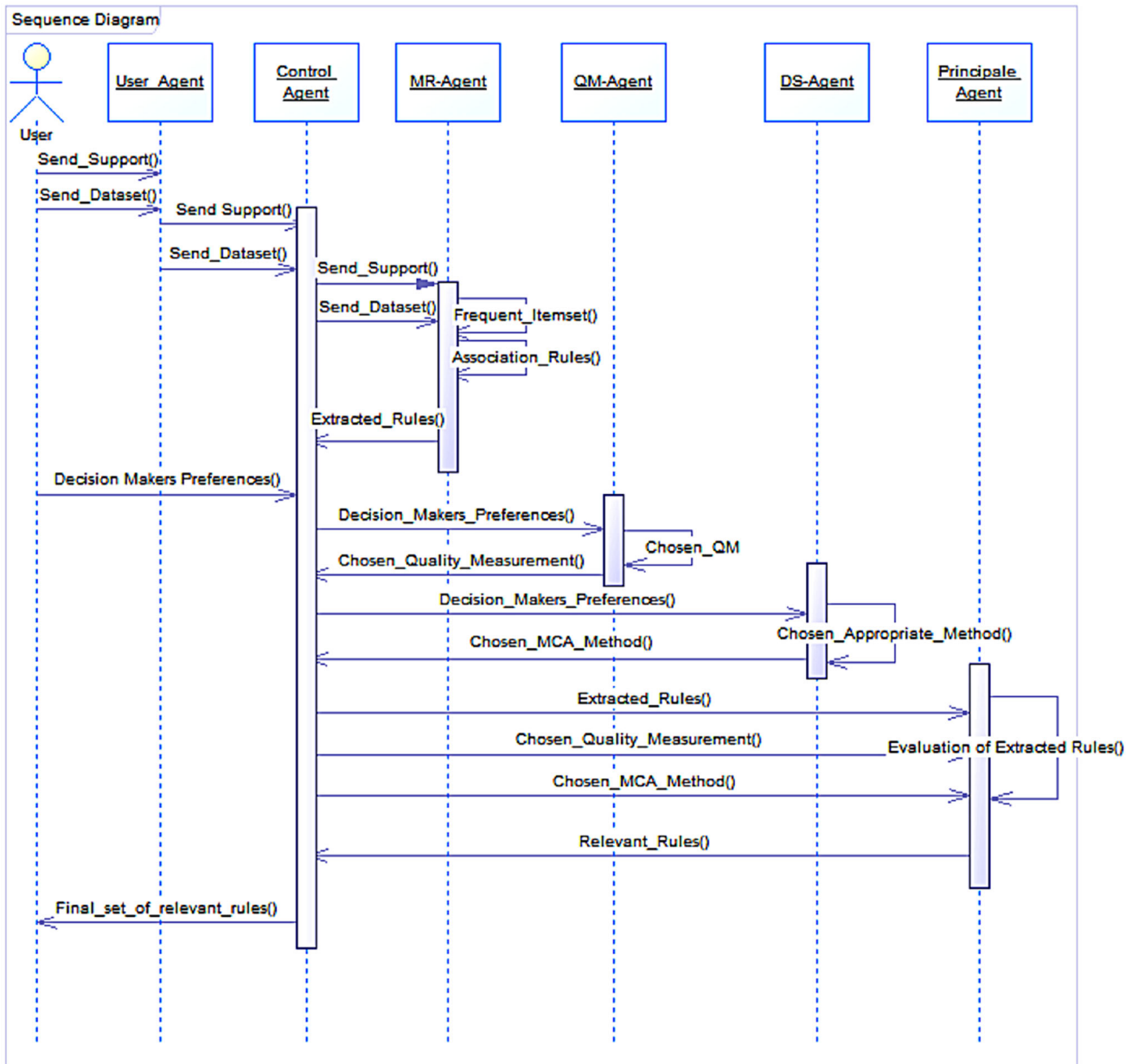


Fig. 6 The sequence diagram of the proposed system

The computation of credibility index is based on the global concordance and the partial discordance index. It presents the degree of credibility of the outranking relation, we define the default value  $\lambda$ -cut index as the parameter that determines the situation preferably between alternatives  $a$  and profiles  $b_h$ . The implementation of the case provides a preference relationship between rules and profiles. In addition, an assignment of association rules through two procedures: pessimistic and optimistic are given in Table 6 by considering that RL is the abbreviation of Rule.

ELECTRE Tri method provides a solution in the form of assignment categories by importance. According to these

two assignment procedures, we find that the relevant rules are RL9, RL1, RL14, RL15, RL16 and RL18, and the remainder belongs to the second category. For our case study of 27 rules, we conclude that the most relevant rules are given in Table 7.

According to the extracted relevant rules, the customers who benefited from credit are those who are tenant with middle grade.

### 6.1 Discussion

Based on this study, when applying multi-criteria analysis approach, especially Electre Tri method, on the set of previ-



**Table 2** Association rules

Rule	Antecedent (LHS)	Consequent (RHS)
1	Income=tranche_2	Habitat=tenant^ Grade=middle_grade
2	Income=tranche_2	Grade=middle_grade
3	Habitat=tenant^ Income=tranche_2	Grade=middle_grade
4	Grade=middle_grade^ Demand=consumption	Habitat=tenant
5	Demand=consumption	Habitat=tenant
6	Family_situation=married	Agreement=Yes
7	Demand=consumption	Habitat=tenant^ Grade=middle_grade
8	Habitat=tenant^ Children=zero	Grade=middle_grade
9	Habitat=tenant^ Agreement=Yes	Grade=middle_grade
10	Income=tranche_2	Habitat=tenant
11	Grade=middle_grade^ Agreement=Yes	Habitat=tenant
12	Agreement=Yes	Habitat=tenant^ Grade=middle_grade
13	Grade=middle_grade^ Income=tranche_2	Habitat=tenant
14	Habitat=tenant	Grade=middle_grade
15	Grade=middle_grade	Habitat=tenant
16	Agreement=Yes	Grade=middle_grade
17	Grade=middle_grade^ Children=zero	Habitat=tenant
18	Agreement=Yes	Habitat=tenant
19	Habitat=tenant^ Demand=consumption	Grade=middle_grade
20	Children=zero	Grade=middle_grade
21	Demand=consumption	Grade=middle_grade
22	Children=zero	Habitat=tenant^and Grade=middle_grade"
23	Grade=middle_grade^ Familsituation=married	Habitat=tenant
24	Children=zero	Habitat=tenant
25	Habitat=tenant^ Familsituation=married	Grade=middle_grade
26	Family_situation=married	Habitat=tenant
27	Family_situation=married	Grade=middle_grade

ously extracted rules by Apriori algorithm or its derivatives we obtained six relevant and significant rules after eliminating the redundant and non-interesting rules. The rest of the rules belongs to the other categories less interest within the first category.

As shown in the experiment results, when applying Apriori algorithm or its derivatives on a dataset, we obtained 27 extracted rules, and when applying multi-criteria analysis approach on a set of extracted rules, we obtained only six sig-

**Table 3** Decision matrix

Rules	Cr1	Cr2	Cr3
Rule1	1.07	0.41	0.86
Rule2	1.07	0.44	0.93
Rule3	1.07	0.41	0.92
Rule4	1.05	0.41	0.96
Rule5	1.05	0.47	0.96
Rule6	1.05	0.42	0.76
Rule7	1.04	0.41	0.84
Rule8	1.03	0.43	0.89
Rule9	1.05	0.41	0.96
Rule10	1.05	0.47	0.96
Rule11	1.05	0.42	0.76
Rule12	1.04	0.41	0.84
Rule13	1.05	0.41	0.96
Rule14	1.05	0.47	0.96
Rule15	1.05	0.42	0.76
Rule16	1.04	0.41	0.84
Rule17	1.05	0.41	0.96
Rule18	1.05	0.47	0.96
Rule19	1.05	0.42	0.76
Rule20	1.04	0.41	0.84
Rule21	1.07	0.44	0.93
Rule22	1.07	0.41	0.92
Rule23	1.05	0.41	0.96
Rule24	1.05	0.47	0.96
Rule25	0.97	0.43	0.84
Rule26	0.97	0.49	0.88
Rule27	0.94	0.45	0.81

**Table 4** Initial profiles and weight defining the category limits

Profiles	Cr1	Cr2	Cr3
Weight $k_j$	1.0	2.0	3.0
$b_1$	0.1	0.5	1.0
$b_2$	1.0	1.2	1.5

**Table 5** Initial values of indifference, preference and veto thresholds

Thresholds	Cr1	Cr2	Cr3
$q_j(b_1)$	0.1	0.2	0.1
$p_j(b_1)$	0.3	0.3	0.2
$v_j(b_1)$	0.4	0.5	0.4
$q_j(b_2)$	0.5	0.6	0.5
$p_j(b_2)$	0.6	0.7	0.8
$v_j(b_2)$	0.8	0.9	0.9

nificant and interesting rules. Moreover, the obtained results are always sensitive to the values of thresholds  $p_j, q_j, v_j$ , and the decision makers preferences.

**Table 6** Assignment procedures

Class	Pessimistic	Optimistic
Class 1	RL9,RL11,RL14, RL15,RL16,RL18	RL9,RL11,RL14, RL15,RL16,RL18
Class 2	RL21,RL23,RL24, RL26,RL1,RL2, RL3,RL4,RL5, RL8,RL10,RL13, RL17,RL19,RL20	RL21,RL22,RL23, RL24,RL25,RL26, RL27,RL1,RL2,RL3, RL4,RL5,RL6,RL7, RL8,RL10,RL12, RL13,RL17,RL19, RL20
Class 3	RL22,RL25,RL27, RL6,RL7,RL12	—

**Table 7** The most relevant rules

N	Rule
RL	Habitat=tenant $\wedge$ Agreement=Yes $\rightarrow$ Grade=middle_grade
RL11	Grade=middle_grade $\wedge$ Agreement=Yes $\rightarrow$ Habitat=tenant
RL14	Habitat=tenant $\rightarrow$ Grade=middle_grade
RL15	Grade=middle_grade $\rightarrow$ Habitat=tenant
RL16	Agreement=Yes $\rightarrow$ Grade=middle_grade
RL18	Agreement=Yes $\rightarrow$ Habitat=tenant

Compared to the previous studies [4–6], in which authors applied rules extraction algorithms and introduce some quality measurement, the results was performed well but always the large number of extracted rules is an obstacle for users to select the relevant rules according to their specific needs. On the other side, our proposed approach of MCA allows the decision makers to select the interesting rules according to their prespecific needs; in addition, the integration of the MCA approach allows users to solve the complex situation by selecting only the significant and useful rules. In addition, by applying multi-criteria analysis approach combined with multi-agent system, we reduce a set of extracted rules, then we eliminate the redundant rules, and the most important is the automation of rules extraction process using multi-agent system.

Finally, our proposed approach has several advantages:

Extract the relevant and useful association rules.

Automation of the rules extraction process.

Solve the complex situation of ranking problems.

Help users to choose their own specified rules.

Reduce the large number and redundancy of extracted rules.

## 7 Conclusion

In this work, we have discussed the usefulness and relevance problem issued from a KDD process, in terms of a large number of extracted rules, most of them are noisy, redun-

dant and not interesting. The deploying methods have been proposed. However, these methods produce an important quantity of extracted rules. To solve such problem, we proposed an approach that use a set of extracted rules and quality measurement within the multi-criteria analysis process to make correspondent recommendations of relevant rules. In other way, we integrate a multi-agent system to manage and model our proposed approach according to six agents working in cooperation to manage the complexity of KDD process.

Moreover, we studied twenty-seven rules according to chosen criteria for selecting the relevant class of association rules, then we find six pertinent rules as a final result. In the other side, the use of Electre Tri improved that the decision makers' preferences have a direct influence on selecting rules.

For further work, a new methodology combining this approach with the other optimization methods, applied to the big data, especially in the road prevention needs to be developed.

**Acknowledgments** The authors wish to acknowledge the contributions of other members of the department for their helpful discussions to make this work in its best condition. In addition, we would also like to thank reviewers for their remarks and suggestions.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Frawley, W., Piatetsky-Shapiro, G., Matheus, C.: Knowledge discovery in databases an overview. *AI Mag* **13**, 57 (1992)
2. Chun-Hsien, C., Li Pheng, K., Yih Tng, C., Xiao Feng, Y.: Knowledge discovery using genetic algorithm for maritime situational awareness. *Expert. Syst. Appl.* **41**(6), 2742–2753 (2014). doi:10.1016/j.eswa.2013.09.042
3. Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., Verkamo, A.: Fast discovery of association rule, advances in knowledge discovery and data mining, pp. 307–328. MIT express, Cambridge (1996)
4. Lallich, S., Vaillant, B., Lenca, P.: Parameterized measures for the evaluation of association rule interestingness. In: The 11th International Symposium on Applied Stochastic Models and Data Analysis, Brest, France, pp. 220–229
5. Lenca, P., Meyer, P., Vaillant, B., Picouet, P. et Lallich, S.: Evaluation et analyse multicritere des mesures de qualite des rgles d'association, *Revue des Nouvelles Technologies de l'Information, Mesures de Qualite pour la Fouille de Donnes, RNTI-E-1*, pp. 219–246 (2004) (ISBN 2-85428-646-4)
6. Lenca, P., Meyer, P., Picouet, P., Vaillant, B. et Lallich, S.: Critres d'valuation des mesures de qualite des rgles d'association, *Revue des Nouvelles Technologies de l'Information, Entreposage et Fouille de donnes (rdacteurs invits : Boussaid O. et Lallich S.)*, RNTI-1, pp. 123–134 (2003) (ISBN 2-85428-621-9)

7. Piatetsky-Shapiro, G., Frawley, W.: Discovery, analysis, and presentation of strong rules. *Knowl Discov Databases*, pp. 229–248 (1991)
8. Hilderman, J., Hamilton, H.J.: *Knowledge Discovery and Measures of Interestingness*. Kluwer Academic Publishers, Berlin (2001)
9. Carvalho, D.R., Freitas, A.A., Ebecken, N.F.F.: Evaluating the correlation between objective rule interestingness measures and real human interest. *PKDD'05, the 9th European conference on principles and practice of knowledge discovery in databases LNAI 3731*, pp. 453–461 (2005)
10. Huynh, H.X., Guillet, F., Briand, H.: Clustering interestingness measures with positive correlation. In: *ICEIS'05, Proceedings of the 7th International Conference on Enterprise Information Systems 2*, pp. 248–253 (2005)
11. Gavrilov, M., Anguelov, D., Indyk, P., Motwani, R.: Mining the stock market: which measure is best? In: *KDD'00, Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 487–496 (2000)
12. Xuan-Hiep, H., Fabrice, G., Henri, B.: A graph-based approach for comparing interestingness measures. In: *IEEE ICEIS'06, Proceedings of the 1st IEEE International Conference on Engineering of Intelligent Systems, Islamabad, Pakistan*, pp. 375–380 (2006)
13. Piatetsky-Shapiro, G., Frawley, W.J.: Discovery, analysis and presentation of strong rules. In: *Knowledge discovery in databases*. pp. 229–248. AAAI/MIT Press (1991)
14. Loevinger, J.: A systemic approach to the construction and evaluation of tests of ability. *Psychol Monogr* **61**(4) (1947)
15. Terano, T., Liu, H., et Chen, A.L.P.: *Association Rules, Lecture notes in computer science*. vol 1805. Springer, Berlin (2000)
16. Lerman, I.C., Gras, R., et Rostam, H., Elaboration d'un indice d'implication pour les donnes binaires, i et ii. *Mathmatiques et Sciences Humaines*, (74, 75) :5–35, 5–47 (1981)
17. Aze, J., Kodratoff, Y.: Evaluation de la rsistance au bruit de quelques mesures d'extraction de rgles d'association. *Extr Connaiss Apprentiss* **1**(4), 143–154 (2002)
18. Sergey, B., Rajeev M., et Craig, S.: Beyond, market baskets : generalizing association rules to correlations. In: *ACM SIGMOD/PODS '97 Joint Conference*, pp. 265–276 (1997)
19. Brin, S., Motwani, R., Ullman, J. D., Tsur, S.: Dynamic itemset counting and implication rules for market basket data. In: *Peckham J (ed) SIGMOD 1997, Proceedings ACM SIGMOD Int. Conference on Management of Data, Tucson, Arizona, USA*, p. 255–264. ACM Press, 05 (1997)
20. Cohen, J.: A coefficient of agreement for nominal scale. *Educ Psychol Meas* **20**, 37–46 (1960)
21. Lerman, I.C., Aze, J.: Une mesure probabiliste contextuelle discriminante de qualitt des rgles d'association. *RSTI-RIA* **1**(17), 247–262 (2003)
22. Good, I.J.: *The estimation of probabilities: an essay on modern Bayesian methods*. The MIT Press, Cambridge, MA (1965)
23. Gouda, K., et Zaki, M.: Efficiently mining maximal frequent itemsets. In: *Proceedings of 1st IEEE International Conference on Data Mining, San Jose* (2001)
24. Han, J., Kamber, M.: *Data mining: concepts and techniques*. The Morgan Kaufmann Series in data management systems (2000)
25. Zaki, M., Ogihara, M.: Theoretical foundations of association rules. In: *DMKD' 98 Workshp on Researche Issues in Data Mining and Knowledge Discovery*, ACM Press, pp. 1–8 (1998)
26. Ait-Mlouk, A., Agouti T., Gharnati F.: Comparative survey of association rule mining algorithms based on multiple-criteria decision analysis approach. In: *Control, Engineering and Information Technology (CEIT), 2015 3rd International Conference*. pp. 1–6, 25–27 (2015)
27. Ait-Mlouk, A., Agouti T., Gharnati F.: A choice of relevant association rules based on multi-criteria analysis approach. In: *2015 5th international conference on information and communication technology and accessibility (ICTA), Marrakech*, pp. 1–6 (2015)
28. Bayardo, R.: Efficiently mining long patterns from databases. In: *Proc. ACM-SIGMOD Int. Conf. Management of Data, Seattle*, pp. 85–93 (1998)
29. Lin D.-I., Kedem Z.M.: Pincer-Search: an efficient algorithm for discovering the maximum frequent set. *IEEE Trans Knowl Data Eng* **14**:553–566 (2002)
30. Zaki, M.J., Parthasarathy, S., Ogihara, M., Li, W.: New algorithms for fast discovery of association rules. In: *Proc. 3rd Int L Conf. on Knowledge Discovery in Databases and Data Mining*, pp. 283–286 (1997)
31. Roy, B.: Classement et choix en prsence de points de vue multiples (la mthode ELECTRE). *La Revue d'Informatique et de Recherche Oprationelle (RIRO)* **8**:57–75 (2002)
32. Mousseau, V., Slowinski, R., Zielniewicz, P.: *ELECTRE TRI 2.0 a methodological guide and user's manual* (1999)
33. Benites, F., Sapozhnikova, E.: Evaluation of hierarchical interestingness measures for mining pairwise generalized association rules, knowledge and data engineering, *IEEE Trans Knowl Data Eng* **26**(12):3012, 3025 (2014)
34. Kadhim, M.A., Afshar Alam, M., Kaur, H.: A Multi-Intelligent Agent for Knowledge Discovery in Database (MIAKDD): Cooperative Approach with Domain Expert for Rules Extraction, volume 8589 of the series *Lecture Notes in Computer Science*, pp. 602–614 (2014)
35. Chun-Wei, L., Tzung-Pei, H., Yi-Fan, C.: An integrated MFFP-tree algorithm for mining global fuzzy rules from distributed databases. *J Univ Comput Sci* **19**(4), 521–538 (2013)
36. Exemple d'application pour l'extraction des rgles d'association. <http://eric.univ-lyon2.fr/~ricco/tanagra/fr/tanagra.html#2005>. Accessed 2015