I.J. Information Technology and Computer Science, **2014, 12, 24-34**
Published Online November 2014 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijitcs.2014.12.04

MAROR: Multi-Level Abstraction of Association Rule Using Ontology and Rule Schema

Salim Khiat
University of Sciences and Technology-Mohamed Boudiaf (USTOMB)/ Computer Sciences and Mathematics Faculty/ Computer Sciences Department Oran, 31000, Algeria
Email: salim.khiat@univ-usto.dz

Hafida Belbachir, Sid Ahmed Rahal
University of Sciences and Technology-Mohamed Boudiaf (USTOMB)/ Computer Sciences and Mathematics Faculty/ Computer Sciences Department Oran, 31000, Algeria
Email: {h_belbach, rahalsa2001}@yahoo.fr

Abstract—Many large organizations have multiple databases distributed over different branches. Number of such organizations is increasing over time. Thus, it is necessary to study data mining on multiple databases. Most multi-databases mining (MDBM) algorithms for association rules typically represent input patterns at a single level of abstraction. However, in many applications of association rules – e.g., Industrial discovery, users often need to explore a data set at multiple levels of abstraction, and from different points of view. Each point of view corresponds to set of beliefs (and representational) commitments regarding the domain of interest. Using domain ontologies, we strengthen the integration of user knowledge in the mining and post-processing task. Furthermore, an interactive and iterative framework is designed to assist the user along the analyzing task at different levels. This paper formalizes the problem of association rules using ontologies in multi-database mining, describes an ontology-driven association rules algorithm to discoverer rules at multiple levels of abstraction and presents preliminary results in petroleum field to demonstrate the feasibility and applicability of this proposed approach.

Index Terms— Multi-Level Rule Synthesis, Local Pattern Analysis, Global Rules, Exceptional Rules, Rule Schema.

I. Introduction

Because of the rapid growth in information and communication technologies, a company's data may be spread over several continents. For an effective decision-making process, knowledge workers need data, which may be geographically spread in different locations. In such circumstances, the multi-database mining using local patterns analysis plays a major role in the process of extracting knowledge from different data sources. Given this model of mining multiple databases, each branch of a company requires to mine the local database by utilizing some traditional data mining technique. Afterwards, each branch forwards the discovered pattern base to the central office where they will be synthesized in the global, exceptional and majority patterns and eventually makes decisions in central office. A pattern can be a frequent itemset or an association rule. Interesting research papers on multi-database mining have been presented in [1] [2] [3] [4] [5] [6]. However in the real world, the structure of an interstate company is usually more complex where each branch can also have sub-branches and so on. In order to discover the interesting patterns in such organization we propose a new process called *multi-databases mining multi-level*. It can be defined as the process of synthesizing frequent patterns from different data sources at multiple levels of abstraction to form global, majority, exceptional and local rules. In industrial discovery applications, because users often need to examine data in different contexts from different perspectives and at different levels of abstraction, there is no single universal belief representation that can serve all users, or for that matter, even a single user, in every context. Hence, methods for association rules from ontologies and data are needed to support knowledge acquisition from heterogeneous distributed data. Making ontological and rules schema commitments (that are typically implicit in a data set) explicit enable users to explore data from multiple perspectives, and at different levels of abstraction. In this paper, we propose an active system framework of multi-database mining for association mining that utilizes user preference in attempt to provide automatic triggering of a mining process with the involvement of a user's query formulation by the rule schema.

The remaining of this paper is organized as follows: Section II introduces the related work in ontology driven-association rule. Section III gives detail of the proposed approach for multi-Level rule synthesis. A case study is described in the section IV and in the last section we terminate with conclusion and shows directions for future research.

II. Related Work

Traditional data mining is based on the objective measures which can be the frequency of occurrence of instances and co-occurrence of items in transaction. The meaning of each item or instance is not taken into consideration. The semantic content extracted from the

Copyright © 2014 MECS

ontologies allows inserting more intelligence and knowledge in data mining, improving their quality.

Ontologies introduced in data mining for first time in early 2000, can be used in several ways [7]: Domain and Background Knowledge Ontologies, Ontologies for Data Mining Process, or Metadata Ontologies. Background Knowledge Ontologies organize domain knowledge and play important roles at several levels of knowledge discovery process. Ontologies for Data Mining Process codify mining process description and choose the most appropriate task according to the given problem; meanwhile, Metadata Ontologies describe the construction process of items [8].

In this study, we are interested in Domain and Background Knowledge Ontologies and we will present past studies related to them. The first idea of using Domain Ontology was introduced by Srikant and Agrawal with the concept of Generalized Association Rules [9]. The authors present Cumulative and EstMerge algorithms to find associations between items at any level by adding all ancestors of each item to the transaction. In this research, items of different levels are added to candidates during the mining.

Češpivová et al. suggested in [10] the introduction of medical ontology and other background knowledge into the process of association mining. The inventory used consisted of the LISp-Miner tool, the UMLS ontology, the STULONG dataset on cardiovascular risk, and a set of simple qualitative rules. The experiment suggested that ontology may bring benefits to all phases of the Knowledge Discovery in Databases (KDD) cycle as described in CRISP-DM.

Euler.T, and Scholz.M presented in [11] a metamodel of KDD preprocessing chains that contains ontology for describing conceptual domain knowledge. This metamodel is operational, yet abstract enough to allow the reuse of successful KDD applications in similar domains. For finding the right representation of data in the KDD process they proposed the MiningMart system which offers support for modeling conceptual knowledge about the domain. The description of available data in a higher level representation language is the first step towards an understandable and operational case study in this framework. Relevant concepts of a domain are structured by domain ontology, allowing to structure concepts by means of inheritance and to represent different kinds of relationships between concepts. This kind of domain model allowed structuring the relevant concepts better than the facilities available in relational databases, and allowed for a convenient handling of views in different contexts. With the MiningMart meta model M4 it is possible to set up operational sequences of preprocessing steps, making use of the conceptual descriptions of the data, only. This result is an increase of the interpretability and reusability of KDD processes.

Charest.M and Delisle.S proposed in [12] the realization of a hybrid intelligent data mining assistant, based on the synergistic combination of both declarative (Description Logic) and procedural (SWRL Rules) ontology knowledge in order to empower the non-specialist data miner throughout the key phases of the CRISP-DM data mining process. They developed an ontology-guided method for data mining using case-based reasoning. The method is based on having an expert system assistant to help non-expert data miners.

Zhou.X and Geller.J proposed in [13] a Raising method that used ontology to perform a preprocessing step on the input datasets before data mining. They proposed a novel data mining processes applied to the input datasets of any data mining algorithm. This Raising method takes advantage of the hierarchy structure of ontology and collects instances at the lower levels of the hierarchy to enrich the derivation of the association rules which involves the ancestors of these instances. The difference with Generalized Association Rules is that this solution proposes to use a specific level for raising and mining. In their experiments, the support values of rule sets were greatly increased, up to 40 times. The effects of Raising on the confidence values were also analyzed based on each type of the possible derived rules. Thus Raising resulted in better rules with higher support and confidence values.

Fuzzy sets theory [14] has been applied a lot on data mining, especially on mining association rules. Although generalized association rule mining approaches based on fuzzy ontology express semantically richer information, they may result in a great amount of redundant rules. Thus, redundancy treatment has been an interesting research topic. There are many algorithms for mining fuzzy ontology association rules. The SSDM (Semantically Similar Data Miner) algorithm [15] considered not only exact matches between items, but also the semantic similarity between them. SSDM uses fuzzy logic concepts to represent the similarity degree between items, and proposes a new way of obtaining support and confidence for the association rules containing these items.

After date Escovar.E et al. extended in [16] the SSDM algorithm in order to obtain from a fuzzy ontology the semantic relations between items. As a consequence, the generated rules can be more understandable, improving the utility of the knowledge supplied by them. Therefore, Extended SSDM can reuse consensual and shared knowledge, easing the process of acquiring semantic information. The NARFO algorithm [17] proposed a new algorithm for mining non-redundant and generalized association rules based on fuzzy ontologies. Fuzzy ontology is used as background knowledge, to support the discovery process and the generation of rules. One contribution of this work is the generalization of non-frequent itemsets that helps to extract important and meaningful knowledge. NARFO algorithm also contributes at post-processing stage with its generalization and redundancy treatment. Their experiments showed that the number of rules had been reduced considerably, without redundancy, obtaining 63.63% average reduction in comparison with XSSDM [16] algorithm.

Mansingh.G et al. proposed and illustrated in [18] a new hybrid method for processing association rules in

Copyright © 2014 MECS

order to reduce the cognitive burden on the users. The method involved the use of both domain knowledge (representing by ontologies) and objective measures to extract and partition interesting patterns and knowledge from databases. They partitioned knowledge into four partitions: Ω_{Known},Ω_{Novel}, $\Omega_{Contradictory}$, $\Omega_{Missing}$ a set of known, novel, contradictory and missing association rules. They applied this method to a medical domain dataset, and through this means they demonstrated that the hybrid method provided a mechanism for reusing and automatically updating a knowledge base.

The MOAL: Multi-Ontology data mining at All Levels algorithm [19] uses the structure and relationships of a Genetic Ontology to mine multi-ontology multi-level association rules. They introduce two interestingness measures: Multi-ontology support and Multi-ontology confidence customized to evaluate multi-ontology multi-level association rules. They also describe a variety of post-processing strategies for pruning uninteresting rules.

Neves and Ana presented in [20] a study on the effects, in terms of precision and recall, of using a data preparation technique, called SemPrune, which is built on domain ontology. SemPrune is intended for pre- and post-processing phases of data mining. Identifying generalization/specialization relations, as well as composition/decomposition relations, is the key to successfully applying SemPrune.

Liu.B et al. proposed in [21] a new framework to allow the user to explore the discovered rules to identify those interesting ones. This framework has two components, an interestingness analysis component, and a visualization component. The interestingness analysis component analyzes and organizes the discovered rules according to various interestingness criteria with respect to the user's existing knowledge. The visualization component enables the user to visually explore those potentially interesting rules. After this interestingness analysis component was developed by [8] where she proposed a new approach to prune and filter discovered rules. She addressed two main issues: The integration of user knowledge in the discovery process and the interactivity with the user. The first issue requires defining an adapted formalism to express user knowledge with accuracy and flexibility such as ontologies in the Semantic Web. Second, the interactivity with the user allows a more iterative mining process where the user can successively test different hypotheses or preferences and focus on interesting rules. For that she proposed a new rule-like formalism, called Rule Schema, which allows the user to define his expectations regarding the rules through ontology concepts. She applied the proposed framework successfully over the client database provided by Nantes Habitat.

In this paper, we report our recent work in addressing the association rule mining at multiple-levels of abstraction. We develop the rule schema proposed by [8] in order to represent user belief at different level in the organization. We propose also a new-like formalism, called Rule Schema Multi-Level which allows the user of different level in the organization to define their expectations regarding the rules through ontology concepts. In addition, we propose a new set of operators over each Rule Schema for interactive processing that these users can choose. We also propose a synthesizing process with the exact rule support which will solve the problem of the estimated rules support in multi-database mining process. Finally, we propound a novel strategy in maintenance of equipment failure in industrial fields called proactive maintenance. This one can be used successfully in the maintenance process for detecting equipments failure which will reduce the cost of the maintenance as demonstrated in the case study section.

III. Proposal Multi-Level Rules Synthesis Framework

MAROR(Multi-level Abstraction of Association Rule using Ontology and Rule schema) is the proposed association rule filtering approach. It proposes to select only the association rules that are interesting for the user at different levels of abstraction. MAROR works in local and global step of the MDBM process. In this context, MDBM process is executed in two steps: first, intra-site processing where association rules are generated by integrating the user knowledge in the rules mining process, next, inter-site processing selects only the interesting ones for each organization level. Fig.1 presents the proposed MAROR framework. In a first instance, we focus on user knowledge. This part consists in a knowledge base allowing formalizing user prior knowledge.

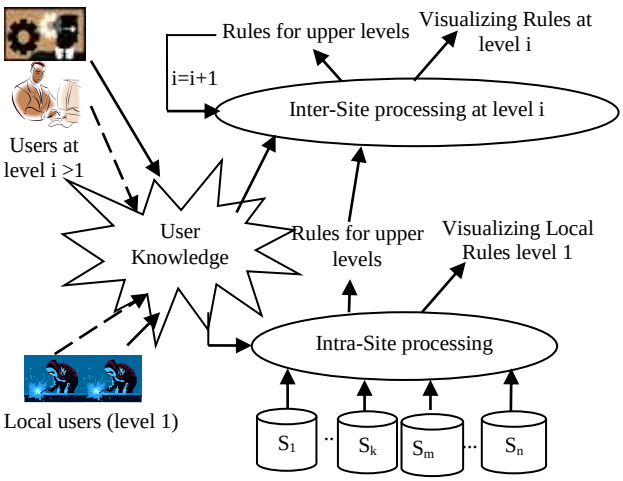


Fig. 1. MAROR Framework

We propose a model to represent user knowledge. This model must take part of the multi-level organization of the company. First, ontology allows the user to express his domain knowledge by means of a high semantic model. Second, we propose a new rule-like formalism, called Rule Schema Multi-Level, which allows the user to define his expectations regarding the rules through ontology concepts. Last, the user can choose among a set of operators for interactive processing the one to be applied over each Rule Schema (i.e. conforming, unexpectedness).

Copyright © 2014 MECS

Ontologies

Ontology is a conceptualization of a specification. It specifies the terms or concepts and relationships among terms and their intended correspondence to objects and entities that exist in the world [22].

Definition 1:

Formally, an ontology is specified by a collection of names for concept $C=\{C_1,C_2,\ldots,Cp\}$ and a relation types $R=\{R_1,R_2,\ldots,R_r\}$ organized in a partial ordering by the type-subtype relation [23] and a directed acyclic graph (DAG) over concepts defined by the subsumption relation (is-a relation, $\leq$) between concepts. We say that C_2 is-a C_1, $C_2 \leq C_1$, if the concept C_1 subsumes the concept C_2.

In this approach, we propose a domain knowledge model based on ontologies connecting ontology concepts over database items. In this scenario, it is fundamental to connect the ontology to the database, each concept and each instance being instantiated in one/several items.

Considering that the set of concepts C is defined as the union of three concepts subsets $C=C_0 \cup C_1 \cup C_2$:

- C_0 is defined as the set of leaf-concepts of the ontology connected in the easiest way to database,

$$C_0=\{c_0 \in C \mid \exists c' \in C, c' \leq c_0\} \quad (1)$$

In this manner, each concept from C_0 is associated to an item in the database.

$$f_0: C_0 \rightarrow I$$
$$\forall c_0 \in C_0,\ i \in I,\ i=f_0(c_0) \quad (2)$$

- C_1 is described as the set of generalized concepts in the ontology. A generalized concept is connected to database through its subsumed concepts. That means that, recursively, only the leaf-concepts subsumed by a generalized concept contribute to its database connection.

$$f: C_1 \rightarrow 2^i$$
$$\forall c \in C_1,\ f(c)=\{i= f_0(c_0) \mid c_0 \in C_0,\ c_0 \leq c\} \quad (3)$$

- More generally, we propose the definition of ontology concepts by logical expressions defined over items, organized in the C_2 subset. In a first attempt, we base the description of the logical expression on the OR logical operator. Thus the defined concept associated could be connected to a disjunction of items.

$$f: C_2 \rightarrow 2^i,\ \forall c \in C_2,$$
$$c \rightarrow E(c)$$
$$f(c)=\{ f(c') \mid c' \in E(c)\} \quad (4)$$

Rule Schemas Multi-Level

To improve association rule selection, we propose a rule filtering model, called Rule Schema Multi-Level. In other words, a rule schema describes, in a rule-like formalism, the user expectations in terms of interesting rules. As a result, rule schemas act as a rule grouping, defining rule families.

The base of rule Schema formalism is the user representation model introduced by Liu & All in [21] composed of: General Impression, Reasonably Precise Concepts and Precise knowledge. We propose to develop two of them: General Impression and Reasonably precise concepts. Thus, rule schemas bring the complexity of ontologies in rule mining combining not only item constraints, but also ontology concept constraints. We develop the formalism in [8] in order to take part of the multi-level organization. The proposed rule schema is presented as follow:

Definition 2: A rule schema is defined as:

$$<(X_1, X_2, \ldots X_m) \rightarrow (Y_1, Y_2, \ldots Y_k)\ (T)(N)>$$

Where:

- X_i and Y_j are ontology concepts and the implication '$\rightarrow$' is optional.
- T ={L,M,E,G} is the type of knowledge which can be local(L), majority(M), exceptional(E) and global rules(G).
- N is the level of the rule schema which indicates the level of users that formulate this one. The lower level (n=1) exposes the decision maker's belief in the lower organization level. The upper level n ($n \geq 0$) expresses the decision maker's belief in the head quarter of the organization.

If the implication '$\rightarrow$' is mentioned in the rule schema we say that the rule schema is an implication rule schema, it defines the reasonably precise concepts. Meanwhile, If we do not keep the implication '$\rightarrow$' we define non implicative rules schemas generalizing general impressions.

For example, a rule schema $< (C_1, C_3 \rightarrow C_2)\ (M)\ (2) >$

Correspond to "all Majority Association Rules whose condition verifies C_1 and C_3 and whom conclusion verifies C_2 at level 2".

Operators

The post-processing task that we design is based on operations applied over rule schemas allowing to user to perform several actions over the discovered rules. We propose two kinds of operators: intra-site and inter-site operators. The intra-site operators are applied to the antecedent and the consequence of the rule schema. So that, we propose three operators: conforming, unexpectedness consequence and unexpectedness antecedent. However, the inter-site operators are applied to the antecedent, consequence and the type of the rule schema. In addition of the three operators we propose a new operator called unexpectedness type. These two kinds of operators will be presented in the inter-site and intra-site processing section.

The filtering technique of the association rules is based on the idea of comparing association rules with the rule schemas. Therefore, we use as comparison technique a modified version of the syntactical method which is defined as follows.

Definition 3

Let us consider an ontology concept C associated in the database to $F(C) = \{y_1,\ldots,y_n\}$

Where $\{y_1,\ldots,y_n\} \in I$ and an itemset $X = \{x_1,\ldots,x_m\}$.

We say that the itemset X is conforming to the concept C if conf (X ,C) = TRUE, where:

$$conf(X,C)=\begin{cases} TRUE & if\ \exists y \\ FALSE & otherwise \end{cases} \quad (5)$$

In other words, an itemset is conforming to an ontology concept if the latter is associated to at least one item of the itemset.

A. Inter-Site Step

The process of *MAROR* inter-site framework (presented in Fig.2) aims to guide the user through the post-processing phase. Several steps are suggested as follows:

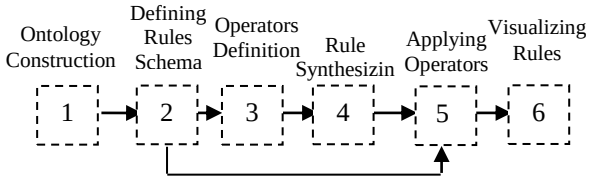


Fig. 2. Inter-site Process

1. Ontology construction: This first phase consists in developing the first part of the user knowledge base. Starting from the database, and eventually from existing ontology, the user develops ontology of database items. It is important to encourage the user to propose a wide range of knowledge, and, if possible, to complete with different information comparing to those found in the database;

2. Rule Schemas definition. The second phase consists in creating the second part of the user knowledge base. The user uses ontology concepts in order to express his local goals and expectations concerning the discovered association rules;

3. Operators definition. The user choose the operators that will be applied over the Rule Schemas which represents an important point in the liberty of the users in this framework, because choosing the operators is choosing the actions to be performed. Once the operators are selected, selecting filters are generated and applied over the set of rules;

In the following we describe the four operators that we integrate in MAROR inter-site which are: confirming, unexpectedness antecedent, unexpectedness consequence and unexpectedness type.

Conforming: Applied over a rule schema, the conforming operator, C(RS), confirms an implication or finds the implication between several concepts. For an association rule to be selected by the Conforming operator over a rule schema, the following set of conditions should be completed in function of the different types of the rule schema:

- If the rule schema is not an implication, an association rule is conformed to it if the itemset created by the union of the antecedent and the conclusion itemsets of the association rule is conforming to each concept composing the rule schema.

Let us consider the following global association rule

$$A \rightarrow B$$

where A and B are itemsets, and a rule schema

$$RS(< M > <T><N>)$$

where

$$M = \{C_1, \ldots, C_k\} \text{ and } T = \{ \text{ and } N > 1$$

We say that the association rule is selected by the Conforming operator, in other words, the association rule is conforming to the rule schema if :

$$\forall C_i \in M, conf(A \cup B, C_i) = TRUE \quad (6)$$

- If the rule schema is defined as an implication, an association rule is conformed to if the antecedent and the consequent itemsets of the association rule are conforming to the each antecedent concept and, respectively, to each consequent concept of the rule schema.

In order to formalize this definition, let us consider the following global association rule

$$A \rightarrow B$$

And the rule schema:

$$RS(< M_A \rightarrow M_B > <T><N>)$$

Where

$$M_A = \{C_1, \cdots, C_K\} \text{ and } M_B = \{C_1', \cdots, C_K'\}$$
$$And\ T = \{L, G, M, E\} et\ N > 1$$

We say that the global association rule is selected by the conforming operator, in other words, the association rule is conforming to the rule schema if:

$$\forall C_i \in M_A, conf(A, (\text{AND}$$
$$\forall C'_i \in M_B, conf(B, (\quad T=G \quad (7)$$

Unexpectedness: With a higher interest for the user, the unexpectedness operator, U(RS), proposes to filter a set of rules with a surprise effect for the user. This type of rules interests the user more than the conforming one since, generally, a decision maker searches to discover new knowledge with regard to his/her prior knowledge.

Moreover, several types of unexpected operators are proposed:

- Antecedent unexpectedness operator, $U_A(RS)$ – a rule is selected by this operator if it is not conformed to the rule schema by its antecedent;
- Consequent unexpectedness operator, $U_C(RS)$ – a rule is selected by this operator if it is not conformed to the rule schema by its consequent;
- And type unexpectedness operator, $U_T(RS)$ – a rule is selected by this operator if it is not conformed to the rule schema by its type.

Next, due to space limit and to repetitiveness of the definitions, in the following, we will detail only the antecedent and type unexpectedness operator applied over implicative rule schemas.

Antecedent Unexpectedness (U_A): Given a rule schema, an association rule is unexpected regarding the antecedent if the antecedent itemset of the association

rule is not conforming to each antecedent concept of the rule schema, and if the consequent itemset of the association rule is conforming to each concept in the consequent of the rule schema. In order to formalize this definition, let us consider the following majority association rule

$$A \rightarrow B$$

and a rule schema:

$$RS(< M_A \rightarrow M_B > <T><N>)$$

Where

$$M_A = \{C_1, \cdots, C_K\} \text{ and } M_B = \{C_1', \cdots, C_K'\}$$
$$And\ T = \{L, G, M, E\}\ AND\ N > 1$$

We say that the association rule is selected by the antecedent unexpectedness operator, in other words, that the association rule is conforming to the rule schema if:

$$\forall C_i \in M_A, conf(A, C_i) = FALSE\ AND$$
$$\forall C'_{i'} \in M_B, conf(B, C \quad (8)$$

Type Unexpectedness (U_T): Given a rule schema, an association rule is unexpected regarding the Type if the type of the association rule is not conforming to the type of the rule schema, and if the antecedent and the consequent itemset of the association rule are conforming to each concept in the Antecedent and the consequent of the rule schema. In order to formalize this definition, let us consider the following exceptional association rule

$$A \rightarrow B$$

and a rule schema:

$$RS(< M_A \rightarrow M_B > <T><N>)$$

Where

$$M_A = \{C_1, \cdots, C_K\} \text{ and } M_B = \{C_1', \cdots, C_K'\}$$
$$And\ T = \{L, G, M, E\}\ AND\ N > 1$$

We say that the exceptional association rule is selected by the type unexpectedness operator, in other words, that the association rule is conforming to the rule schema if :

$$\forall C_i \in M_A, conf(A, C_i) = TRUE\ AND$$
$$\forall C'_{i'} \in M_B, conf(B, C \quad (9)$$

4. Rule Synthesizing: The rule synthesizing process should generate meaningful rules which make sense with respect to the user's knowledge. It is proposed to get a Global(G), Majority(M) and Exceptional(E) set of synthesized rules, which are potentially useful for a multi-level organization in the decision-making process from the local rules.

The synthesizing process is based on the user knowledge and expectations into the synthesizing process. Only interesting rules are synthesized into three groups: Global, Majority and Exceptional rules. The construction of these groups is based on the rule schema and the operators as described in the fig.3.

Majority rules [24] can grasp the distribution of rules in local ones and reflect the "commonness" of branches in their voting. High-vote rules are useful for global applications of interstate companies.

Exceptional rules [5] can grasp the individuality of branches. It often present as more glamorous than high-vote rules in such areas as marketing, science discovery and information safety.

Global rules can grasp the globality of rules and reflect the distribution of the rules supports. It detects the global rules instead the mono-database mining (put all databases in a huge database and apply a classic mining algorithm). In other words, it reflects the global rules which are tailed with the mono-database mining. Our framework allows extracting the set of global rules exactly lossless rules.

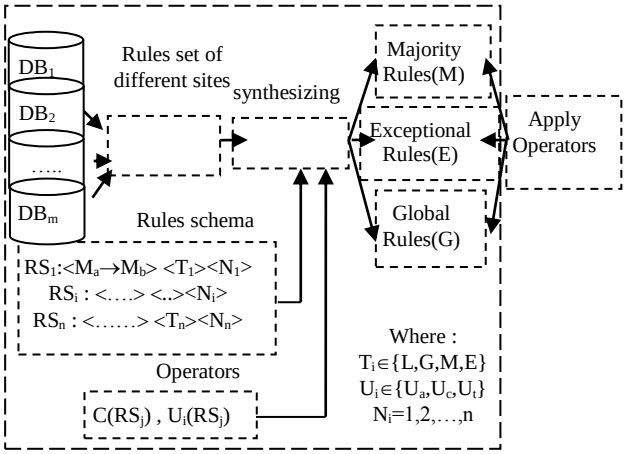


Fig. 3. Synthesizing process

Definition 4

Let the set of rules schema RS_1:<$M_a \rightarrow M_b$> <T_1> <N_1>, , … , RS_n:<$M_{a'} \rightarrow M_{b'}$> <$T_n$><$N_1$> and a set of conforming and unexpectedness operators $C(RS_1)$,…, $C(RSn)$; $U_i(RS_k)$,…, $U_i(RS_m)$ where $U_i=\{U_a,U_c\}$, clusters generated by the synthesized process are T_1 ,…,T_n.

Definition 5

Let rules schema RS_1:<$M_a \rightarrow M_b$> <T_1><N_1> ,… , RS_n:<$M_{a'} \rightarrow M_{b'}$> <$T_n$><$N_1$> and a set of confirmation and unexpectedness operators $C(RS_1)$,…, $C(RS_n)$; $U_i(RS_k)$,…, $U_i(RS_m)$ where $U_i=\{U_a,U_c,U_t\}$. All the three clusters are generated by the synthesized process.

5. Applying Operators: This phase consists of applying the four operators described below for each rule cluster.

6. Visualization: The visualization phase is very important, proposing to the user the result of his actions.

B. Intra-Site Step

The process of MAROR intra-site framework (presented in fig.4) aims to guide the user through the mining rules process phase. Several steps are suggested as follows:

1. Local Rules schema: This phase describes the local user's belief.

2. Set Rules Schema of upper levels: This step describes the user's belief for the upper level hierarchy.

Copyright © 2014 MECS

3. Applying Operators: This phase has the same role as in the inter-site processing expected the type unexpectedness operator.

4. Candidates Rules Generation: Candidates rules are all possible rules that are conforming to the specified schemas and operations. After generation, a pass through the database is performed in which the support and the confidence of candidate rules are computed. In order to be present in the output, rules must comply with the support and confidence requirements specified and the others rules which do not satisfy the support and confidence are transferred into the uppers level.

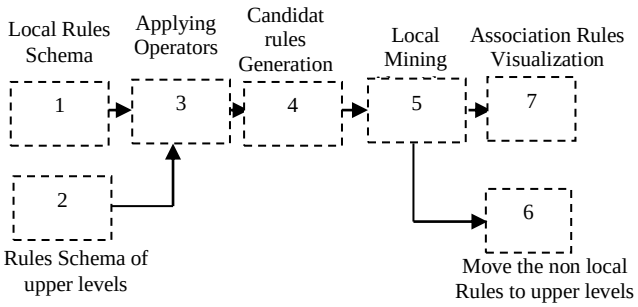


Fig. 4. Intra-Site Process

Next, due to space limit, in the following, we will detail only the generation of the candidate rules based on the confirmation operator. The pseudo code of the confirmation operator algorithm is presented as follow:

Input: Set of rules schema RS per level described by Ant antecedent, Cons consequent, N level and I set of items.

Output:

Itemsetlist: the list of itemset with their supports per level that will be used in the upper levels.

rulelist: the list of locale rules confirming the local rule schema.

1. rulelist = ϕ, for i =2.. N do itemsetlist[i]=ϕ
2. Let a set of items not in {antecedent, consequent} RQ = {I-Ant-Cons}
3. For each Rule schema RS[N]
4. For each side Pd in SR[N] {Ant, Cons}
5. For each subset SE of RQ do
6. if N≠1 then /*upper itemsets*/
7. Add to itemsetliste[N] two new
8. candidats itemsets (ISC)
ISC[N](Ant U SE)
ISC[N](Ant U Cons U (RQ))
9. Else /*Local association rule*/
Add to rulelist a new candidat rule
RC (Ant U SE → Cons U (RQ-SE))
10. For each candidats itemsets ISC[N]∈ itemsetliste[N] and a candidat rule RC∈ rulelist
11. calculate the support s of itemsets in ISC[N]
12. Verify the support s and confidance c for RC
13. Remove from a liste rulelist all rules with s<minsup
14. Return rulelist and itemsetlist

5. Local Mining Algorithm: Association rule mining is widely used data mining approach for discovering patterns and relationships between variables from data. Apriori algorithm [25] is one of the most commonly used methods for Association Rule. By an incremental approach, Apriori finds all frequent itemsets–all itemsets that have a support above a certain threshold. On the basis of the frequent itemsets, the algorithm builds all rules that have a confidence value above a given threshold.

MAROR intra-site approach extract only interesting rules for that it integrates user knowledge and expectations into the rule mining process.

In this approach, the search for interesting rules is done locally, in the neighborhood of rules and associations that the user believes to be true, specified by means of the Rule Schemas. Instead of generating all rules (by means of frequent itemsets), and filtering those that are conform to user knowledge, the new approach consists of first generating locally all candidate rules, based on the rule schemas of all the upper level and operators, and then checking their support and confidence against the transaction database.

6. Move the non local Rules at upper levels: These rules are important for detecting the global rules that are not locally frequent. So, we don't need to estimate the non locally rules and the globally rules will be tailed with the mono-base mining results.

7. Association Rules Visualization: The visualization phase is very important, proposing to the user the result of his research.

IV. Case Study

In this section, the MAROR model is illustrated along three case studies, taken from Petroleum Company more specially the maintenance field that has modeled about the work request concerning the equipment failure. Fig.5 describes the multi-level organization of the petroleum company. The upper level (Central Unit) consists of the head quarter of the company and the lower level (U_i) consists of the operational unites and the middle level constitute the branches of the company. Each operational unit is connected to the maintenance database. This one consists in a daily transaction work request for the failure of equipment since 1998 performed by operator maintenance.

Each Unite have the same structure of the databases represented in fig.6.

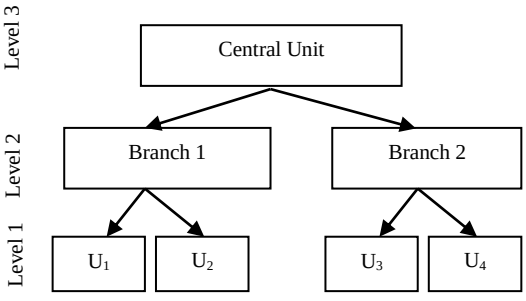
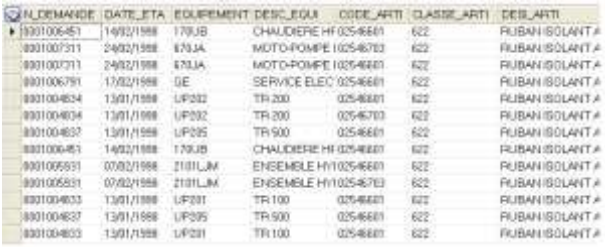


Fig. 5. Petroleum Organization

The number of transaction about all the units is among 100 000 and 130 000 and 100 items for each unites.



N_DEMANDE	DATE_ETA	EQUIPEMENT	DESC_EQUI	CODE_ARTI	CLASSE_ARTI	DES_ARTI
0001006451	14/01/1998	170UB	CHAUDIERE HP	02546601	622	RUBAN ISOLANT A
0001007311	24/02/1998	670JA	MOTO-POMPE I	02546703	622	RUBAN ISOLANT A
0001007311	24/02/1998	670JA	MOTO-POMPE I	02546601	622	RUBAN ISOLANT A
0001006791	17/02/1998	GE	SERVICE ELEC	02546601	622	RUBAN ISOLANT A
0001004834	13/01/1998	UP202	TR 200	02546601	622	RUBAN ISOLANT A
0001004834	13/01/1998	UP202	TR 200	02546703	622	RUBAN ISOLANT A
0001004837	13/01/1998	UP205	TR 500	02546601	622	RUBAN ISOLANT A
0001006451	14/01/1998	170UB	CHAUDIERE HI	02546601	622	RUBAN ISOLANT A
0001005931	07/02/1998	2101LJM	ENSEMBLE HY	02546601	622	RUBAN ISOLANT A
0001005931	07/02/1998	2101LJM	ENSEMBLE HY	02546703	622	RUBAN ISOLANT A
0001004833	13/01/1998	UP201	TR 100	02546601	622	RUBAN ISOLANT A
0001004837	13/01/1998	UP205	TR 500	02546601	622	RUBAN ISOLANT A
0001004833	13/01/1998	UP201	TR 100	02546601	622	RUBAN ISOLANT A

Fig. 6. Database example

For example the first transaction describes that in *14/01/1998* there is an request number *0001006451* for the equipment code *170UB* untitled «*CHAUDIERE HP* » and specially the component code *02546601* untitled «*RUBAN ISOLANT* » which belong to the class code *622*.

A. Ontology Structure

Ontology is defined basically by two elements: a set of concepts (C) hierarchized by the subsumption relation and a set of relation (R) over concepts.

We propose ontology composed of two mains parts, as shown in fig.7.

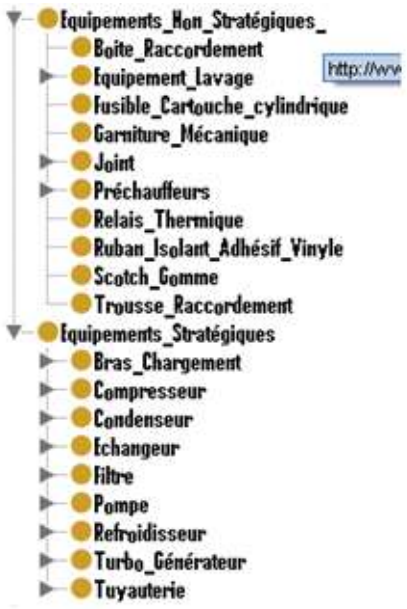


Fig. 7. Ontology structure in OWL

The first one is a database items organization with the root defined by the Attributes concept. The items are organized among the thematically structure of equipments in the maintenance databases. For instance, considering the «*equipements_stratégiques* » concept: it regroups «*Bras_Chargement* », «*compresseur* »,…. Concepts. Where the concepts «*Bras_Chargement* » regroups «*Bras_Chargement_Propane* » and «*Bras_Chargement_GPL_liquide_jeté* ».

To describe the ontology we use the Web Semantic representation language, OWL-DL [1]. Based on description logics, OWL-DL language permits, along with the ontological structure, to create concepts using necessary and sufficient conditions over other concepts. Also, we use the software to edit the ontology.

B. Ontology-Databases Mapping

Part of rule schema definition, ontology concepts are mapped to a several items in the database. Thus, several ontology-database connection types can be conceived.

Firstly, the simplest ontology-database mapping is the direct one. It connects one leaf-concept of the attribute hierarchy to a set of items.

Considering the concept C_1= «*Joint* » of the ontology, it is associated to the attribute I_1= «*Joint* 5 », I_2= «*Join spirale avec anneau de centrage* », I_3= «*joint plat à face surelevee* », I_4= «*joint torique* ». Furthermore, the concept C_1 is instantiated in the ontology by 4 instances describing the concept C_1 with 4 possible articles.

C. Rule Schemas

A rule schema allows user expectation representation and permit to the user to supervise association rule mining, meanwhile operators guide the intra-site and inter-site processing by filtering discovered rules. For example, let us consider the set of rule schemas with the operator presented in table 1.

Table 1. Operators and Rule Schemas

	Rule Schema	Operator
RS_1	<Equipement_Lavage><L><1>	Conforme
RS_2	<Prechauffeurs> <G><3>	Conforme
RS_3	<Pompe, Joint> <M><2>	Conforme
RS_4	<Joint → Echangeur><G><2>	Conforme
RS_5	<Joint → Pompe><E><2>	Unexpectedness Type

The first non implicative rule schema defined at lower level (level 1) by operator expresses possible local relationship between the «*Equipement_Lavage* » product and others products for local database.

The second non implicative rule schema defined at upper level (central unit) by decision maker expresses possible global relationship between the «*Prechauffeurs* » product and others products.

The third non implicative rule schema defined at middle level (level 2) by the decision maker of branch expresses possible majority relationship, without knowing the direction of the association, between the «*pompe* » and «*joint* » of equipment.

The fourth implicative rule schema defined at middle level (level 2) by the decision maker of branch expresses possible global relationship, with knowing the direction of the association, between the «*Joint* » and «*Echangeur* » of equipment.

The last implicative rule schema defined at middle level (level 2) by the decision maker of branches expresses possible non majority relationship, with knowing the direction of the association, between the «*Joint* » and «*Pompe* » of equipment.

[1] http://www.w3.org/TR/owl-features

Copyright © 2014 MECS

D.Case Study 1

In the first case study, the analyst (decision maker at branch 1 or 2) is interested on the failure of «*Pompe* » and «*Joint* » equipments in the majority of sites. The analyst already doesn't know the direction of the implication and creates a Rule Schema RS_3.

The application of a conformed operator over this Rule Schema has the following majorities rules output:

62 results:

Joint,GarnitureSD-3REF→ Pompe [S=24%C=60,8%]

Joint,Prechauffeurs→ Pompe [S=21%C=59,8%]

…

The decision maker must give directives to the operator of the maintenance to change the «*joint* » equipment before the preventive maintenance. Because the failure of this «*joint*» equipment can destroy the failure of expensive equipment which is the «*pompe* » equipment.

E. Case Study 2

In this second example, the expert (decision maker at branch 1 or 2) is interested on the failure of «*Echangeur* » equipment in the union of the sites which is responsible. The analyst already knows the direction of the implication and creates a Rule Schema RS_4.

The application of a conformed operator over this Rule Schema has the following global rules output:

10 results:

Joint,GarnitureSD-3REF→Echangeur [S=25%C=75,8%]

Joint,Pompe→Echangeur [S=20%C=70,8%]

….

The first rule describe that the failure of the «*joint* » equipment can destroy the «*Echangeur* » equipment. The analyst can deduce that the failure of some money can induce of the failure of the equipment of the million of $.

Than the operator of the maintenance of all the sites should give more attention to the «*joint* » equipment and change it before the preventive maintenance.

F. Case Study 3

In this last study, the expert is interested on the failure of «*Joint* » and «*Pompe* » equipments. The analyst already knows the direction of the implication and creates a Rule Schema RS_5.

The application of an unexpectedness type operator over this Rule Schema has the following majority rules output:

20 results:

Joint→Pompe [S=29%C=78%]

Joint,GarnitureSD-3REF→Echangeur [S=25%C=75,8%]

Joint,Pompe→*Echangeur* [S=20%C=70,8%]

….

The result shows, that the failure of the «*joint* » equipment can destroy the «*pompe* » equipment. The analyst can deduce that the failure of some money can induce of the failure of the equipment of the million of $. In addition this rule is in the majority of sites. So, the operator of the maintenance of all sites should give more attention to the «*joint* » equipment and change it before the preventive maintenance.

G.Proactive Maintenance (PA.M)

In industrial maintenance, the maintenance of equipment is based on four strategies: Curative, preventative, predictive and scheduled stopped maintenance.

The primitive curative maintenance (CU.M) consists with the intervention after the appearance of the break downs or anomalies.

The preventative maintenance (PR.M) has been based primarily on lifetime estimates for particular parts and then replacing those parts at scheduled intervals before they exceed their lifetime estimate.

The predictive maintenance (PD.M) consists of application some measure techniques on the servicing equipment. These equipments make it possible to diagnose the state of the equipments in order to judge the advisability of launching the preventative maintenance action or of deferring it on rational bases.

The scheduled stopped (S.ST) is one of the forms of maintenance the most used in the petroleum industry. It concerns in most of the time equipment that's their maintenance is impossible during their operation.

With our framework we introduce a novel strategy in maintenance called Proactive Maintenance (PA.M) Fig.8. It uses our proposed method to predict parts that are likely to fail. PA.M uses a data mining tool that finds affinities between repairs or affinities between reports and subsequent repairs. We have shown in the case study that according to the history of the breakdowns of the equipment, one can extract from the relations or correlations between the breakdowns which go sets. This makes it possible to reduce the cost of maintenance enormously.

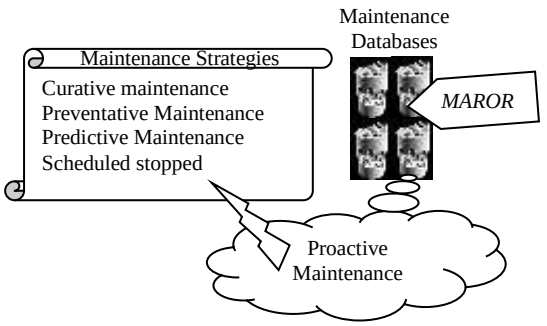


Fig. 8. Proactive Maintenance

V. Conclusion

Our work in this paper addresses the main issues: the integration of user knowledge in the multi-database mining process, without losing knowledge. For this purpose we take advantage from the research carried out in the Semantic Web field, and more precisely from representation language developed in order to be used as

user knowledge representation in the Multi-Database Mining process. We integrate the knowledge user in the two steps of multi-database mining for keeping only the interesting association rules. For this purpose, we have proposed a new formalism for representing the user beliefs in such environment based on the rule schema and proposed new set of operators applied over this rule schema. In intra-site phase we need only one scan over each database for extracting local association rules and global rules candidate. In the inter-site processing these local rules are synthesized into Global, Majority and Exceptional rule. We have shown that this synthesized process was driven by the user knowledge. Further the domain experts who participate in this study also reported that the method facilitated their examination of the generated association rules. Indeed the case study describes some interesting results for the decision makers in the maintenance field. We can say that the proposed strategy called proactive maintenance based on our framework can help the users of maintenance for reducing the cost of the maintenance.

We intend to improve this approach in three directions:

- Developing others appropriate operators applied on the rules schemas.
- Applying MAROR framework in production fields.
- Developing a visualization framework for visualizing the association rules for different users in the multi-level organization.

REFERENCES

[1] Ramachandrarao Pralhad, Animesh Adhikari, Witold Pedrycz, « Developing Multi-Database Mining Applications », Advanced Information and Knowledge Processing; Springer-Verlag London Limited.2010

[2] Ramkumar Thirunavukkarasu, Srinivasan Rengaramanujam; «Modified Algorithms for Synthesizing High-Frequency Rules from Different Data Sources », Knowl Inf Syst 17:313–334. (2008) Springer

[3] Ramkumar Thirunavukkarasu, Srinivasan Rengaramanujam; «Multi-Level Synthesis of Frequent Rules from Different Data-Sources », International Journal of Computer Theory and Engineering, Vol. 2, No. 2 April, 2010

[4] Zhang Chengqi, Meiling Liu, Wenlong Nie, and Shichao Zhang, «Identifying Global Exceptional Patterns in Multi-database Mining », IEEE Computational Intelligence Bulletin Feburuary 2004 Vol.3 No.1.

[5] Zhang Shichao, Chengqi Zhang, Jeffrey Xu Yu, «An Efficient Strategy for Mining Exceptions in Multi-Databases ». Article in press; An international journal information Science. Elsevier.2003

[6] Xindong Wu, Shichao Zhang, Chengqi Zhang; « Multi-Database Mining » IEEE Computational Intelligence Bulletin Vol.2 No.1 2003.

[7] Nigro H.O., S.E Gonzalez Cisaro, and Xodo: « Data Mining With Ontologies: Implementations, Findings and Frameworks ». Idea Group Reference.

[8] Marinica Claudia « Association Rule Interactive Post-processing using Rule Schemas and Ontologies – ARIPSO » These de doctorat en philosophy de "Ecole poltechnique de l'Universit_e de Nantes" Department d'informatique. 26 October 2010

[9] Srikant.R and Agrawal.R. «Mining Generalized Association Rules ». Proceedings of the 21st International Conference on Very Large Databases, pages 407–419. Zurich, Swizerland 1995.

[10] Češpivová.H, J. Rauch, V. Svátek, M. Kejkula, and M. Tomečková. «Roles of Medical Ontology in Association Mining CRISP-DM Cycle ». Workshop Knowledge Discovery and Ontologies in ECML/PKDD.(2004)

[11] Euler.T, and Scholz.M. (2004) Using Ontologies in a KDD Workbench ». In Workshop on Knowledge Discovery and Ontologies at ECML/PKDD.

[12] Charest M, Delisle S. «Ontology-Guided Intelligent Data Mining Assistance:Combining Declarative and Procedural knowledge ». In Artificial Intelligence and. Soft Comput 2006:9–14

[13] Zhou.X and Geller.J. (2007). «Raising, to Enhance Rule Mining in Web Marketing with the Use of an Ontology ». Date Mining with Ontologies: Implementations, Findings and Frameworks, pages 18-36.

[14] Zadeh L.A, «Fuzzy Sets », In: Fuzzy Sets and Applications: Select Papers by L.A. Zadeh, Edited by R. R. Yager; S. Ovchinnikov, et al, Wiley-Interscience, 29-44.1987

[15] Escovar Eduardo L.G, M.Biajiz, and M.T.P. Vieira. «SSDM: A Semantically Similar Data Mining Algorithm ». In XX Simposio Brasileiro de Banco de Dados (SBBD), pages 265–279, Uberl ândia, MG, Brasil(2005).

[16] Escovar Eduardo L.G., Yaguinuma, C.A., Biajiz, M. «Using Fuzzy Ontologies to Extend Semantically Similar Data Mining ». In: 21st Brazilian Symposium of Databases, Florian ópolis, Brazil, October 16-20 (2006)

[17] Miani Rafael Garcia, Cristiane A. Yaguinuma, Marilde T.P. Santos, and Mauro Biajiz «NARFO Algorithm: Mining Non-redundant and Generalized Association Rules Based on Fuzzy Ontologies ». Springer-Verlag Berlin Heidelberg pp. 415–426,2009.

[18] Mansingh Gunjan, Kweku-Muata Osei-Bryson and Han Reichgelt « Using Ontologies to Facilitate Post-Processing of Association Rules by Domain Experts » Information Sciences 181 (2011) 419-434 ELSEVIER.

[19] Manda.P, McCarthy F, Bridges S. « Interestingness measures and Strategies for Mining Multi-Ontology Multi-Level Association Rules from Gene Ontology Annotations for the Discovery of New GO Relationships ». Journal of Biomedical Informatics: Available online 11 July 2013 - http://dx.doi.org/10.1016/j.jbi.2013.06.012 DOI:10.1016/J.JBI.2013. 06. 012

[20] Neves Inha úma Ferraz1 and Ana Cristina Bicharra Garcia: «Ontology in Association Rules ». SpringerPlus 2013.

[21] Liu Bing, Wynne Hsu, Ke Wang and Shu Chen, 1999: « Visually Aided Exploration of Interesting Association Rules ». Proceeding of the Third Pacific-Asia Conference on Methodologies for Knowledge Discovery and Data Mining. Lecture Notes In computer science, Vol, 1574, Springer-Verlag, pages 26-28.

[22] Gruber. T.R: « A translation Approach to Portable Ontology Specification » Knowledge acquisition, 5(2):199-220 (1993).

[23] Sowa.J: «Knowledge Representation ». Logical, philosophical, and Computational Foundations. Books Cole Publishing Co..Pacific grove, CA (2000)

[24] Zhang Shichao, Chengqi Zhang, Jeffrey Xu Yu, «Identifying Interesting Patterns in Multi-Databases ». Studies in Computational Intelligence (SCI) 4.91-112. Springer-Verlag Berlin Heidelberg 2005.

[25] Agrawal Rakesh, Ramakrishnan Srikan « Fast Algorithms for Mining Association Rules » In VLDB'94 pp 487-499.

Copyright © 2014 MECS

Authors' Profiles

Salim Khiat holds a post graduation degree in computer science from University of sciences and technology– Mohamed Boudiaf Oran USTOMB Algeria in 2007. He teaches courses in undergraduate and graduate composition, at National School Polytechnic Oran Algeria. He is memberships in Signal, System and Data Laboratory (LSSD).

His current research interests include the databases, multi-database mining for software engineering, Ontology, grid and cloud computing.

Hafida Belbachir Received PH.D degree in Computer Science from University of Oran, Algeria in 1990. Currently, she is a professor at the Science and Technology University USTO in Oran, where she heads the Database System Group in the LSSD Laboratory. Her research interests include Advanced DataBases, DataMining and Data Grid.

Sid Ahmed Rahal is Doctor in computer science since 1989 in Pau University France. He is memberships in professional activities are:
-Member of LSSD (Laboratory Signal, System and Data)
-Interest in DataBases, Data Mining, Agent and expert systems.

How to cite this paper: Salim Khiat, Hafida Belbachir, Sid Ahmed Rahal,"MAROR: Multi-Level Abstraction of Association Rule Using Ontology and Rule Schema", International Journal of Information Technology and Computer Science(IJITCS), vol.6, no.12, pp.24-34, 2014. DOI: 10.5815/ijitcs.2014.12.04

Copyright © 2014 MECS