

Contribution to the Association Rules Visualization for Decision Support: A Combined Use Between Boolean Modeling and the Colored 2D Matrix

Fatima Zohra Benhacine*, Baghdad Atmani, Fawzia Zohra Abdelouhab

Laboratoire d'Informatique d'Oran (LIO), University of Oran 1 Ahmed Benbella (Algeria)

Received 22 February 2018 | Accepted 14 September 2018 | Published 24 September 2018



ABSTRACT

In the present paper we aim to study the visual decision support based on Cellular machine CASI (Cellular Automata for Symbolic Induction). The purpose is to improve the visualization of large sets of association rules, in order to perform Clinical decision support system and decrease doctors' cognitive charge. One of the major problems in processing association rules is the exponential growth of generated rules volume which impacts doctor's adaptation. In order to clarify it, many approaches meant to represent this set of association rules under visual context have been suggested. In this article we suggest to use jointly the CASI cellular machine and the colored 2D matrices to improve the visualization of association rules. Our approach has been divided into four important phases: (1) Data preparation, (2) Extracting association rules, (3) Boolean modeling of the rules base (4) 2D visualization colored by Boolean inferences.

KEYWORDS

Association Rules, Boolean Modeling, Cellular Automaton, Clinical Decision Support System (CDSS), Colored 2D Matrix, Data Mining.

DOI: 10.9781/ijimai.2018.09.002

I. INTRODUCTION

THE decision support is a wide area. Since the 90's the decision making became a major activity that requires the establishment of effective dedicated systems. The Decision support is intended to assist the decision maker and to support him on his understanding of the decision situation by providing him the reason of the selected choices, allowing him to assess the risks when he adopts any strategy [1]. Decision support can be encountered in many application domains as economy, industry, agriculture and medicine. In medicine, the decision is regarded as the center of the medical act. The medical decision process is to make a diagnosis, and also propose a treatment. So, a large number of decision support systems (DSS) have been developed in this area. These DSS are destined to support health workers in their decision making [2]. The clinical decision support system (CDSS) is defined as an information technology tool whose main objective is to assist clinicians to organize, store, extract, and exploit medical knowledge [3].

We contribute to an interactive approach to clinical decision support. The objective of such CDSS is to provide interactive help to users which face similar medical decision problems.

Research in the field of CDSS has resulted in the appearance of new technologies on storage, processing and analysis of data and information required for decision process. The development of such systems requires a real knowledge of the application area, which gave rise to Knowledge Discovery from Databases approach [3]. The decision value is in data history. The purpose of clinical decision support system (CDSS) based on Knowledge Discovery from Databases (KDD) is to capture medical data that can provide decision

support at the time of the treatment. Using the techniques of KDD, the large databases become potentially rich and reliable information sources for knowledge generation and validation. The Data Mining represents the center phase of KDD and to apply the machine-learning algorithms on data to extract models [3].

In our work, we will focus on association rules, constituting one of the powerful models in Data Mining. Association rules are used to treat and discovering interesting rules from large collection of data as $X \rightarrow Y$ where X and Y are set of items [4]. Association rules have been successfully used in many areas as economic planning support [5], diagnosis assistance and medical research [6]. They are able to detect the trends and hidden relationships and exploring correlations from data. However the significant number of rules generated by this method makes it difficult for human eye. Given these large volumes, the text representation mode is not suitable and the interpretation is impossible. To overcome this problem various visualization techniques [7], [8], [9], [10], [11], [12], [13] have been proposed to improve interpretation of extracted knowledge, clearly and precisely.

The objective of our work is to improve visualization of association rules to increase performance of CDSS and reduce the physician's cognitive loading using 2D colored matrices [14] and Cellular Automaton for Symbolic Induction CASI [15]. We propose a solution that provides a reasoning and essentially guarantees, storage space optimization and execution time.

Our contribution deals mainly with two aspects:

- Firstly, we propose an Intelligent Clinical Decision Support System for immunization using Knowledge Discovery from Databases to provide interactive help to physicians.
- Secondary, we use jointly the CASI cellular machine and the colored 2D matrix to improve visualization of association rules to increase performance of ICDSS and reduce the physician's

* Corresponding author.

E-mail address: benhacine.fatima@gmail.com

cognitive loading.

This article is structured as follows: In Section II, an overview of CDSS, association rules and techniques for their visualization are presented. In Section III, the proposed approach is described in detail. The results of the approach are explained in Section IV. Finally, we present our conclusions and our future research perspectives in section V.

II. RELATED WORK

In a patient environment, the medical decision process is to choose an investigative mode. Thence, medical decision support is defined as information management techniques helping partially or completely, the physician in decision processes [16]. The clinical decision support system (CDSS), are playing increasingly important roles in medical practice by helping physicians or other medical professionals making clinical decisions. CDSS are having a greater influence about the care process. They are expected to improve the medical care quality; their impact should intensify due to increasing capacity for more efficient data processing [17]. So, great numbers of decision support systems have been developed in this domain [18], [19]. These applications are made to support health services in their deciding making [2].

There are three main categories of CDSS, the first category is indirect decision support systems or documentary assistance systems whose objective is to facilitate access to relevant information rapidly; but these systems have no reasoning method [20]. The second category is about the automatic reminder systems that prevent doctors from making mistakes or remind important elements to be taken into account for the decision. Assistance provided is not a reasoning help but rather a reminder providing useful information in an easy defined situation [21]. These systems, like the previous ones do not reason [20]. Finally, the third category is consulting systems that provide the user with reasoned conclusions according the reasoning methods used. The conception is more satisfactory than precedents. Developers are primarily interested in this category [20]. This category includes medical expert systems [21] and systems based on Data Mining [22]. Our contribution is part of the CDSS based on Data Mining.

Data Mining uses a variety of methods to process large amounts of data and information to discover useful knowledge for decision. It is a decision support tool in various sectors like health sector facing increasing pressures to improve the quality of health care while reducing costs. It is therefore not surprising that health organizations have been interested in data mining to improve physician practices, disease management and the resource utilization. Hence the progressive use of data mining in the medical domain.

Throughout our study, we identify different data mining techniques [23], [24], [25] such as association rules [23], [24], neural networks and naive bayes [24], Decision Tree and Case-Based Reasoning. The competing objective leading us to propose association rules is their simplicity and unlike other data analysis methods, they provide simple and easily interpretable results.

Since its introduction in [4] the task of association rules has received a great deal of attention. Today the association rules are still one of the most popular pattern discoveries in KDD. An association rule is an expression $R:X \rightarrow Y$ where X and Y are sets of items. X is called antecedent or left-hand-side (LHS) and Y , consequent or right-hand-side (RHS). The association rules search expresses, from data contained in a relational database, implicit trends between attributes of antecedent and consequent. In order to select interesting rules from the set of all possible rules; the best-known indicators are support and confidence [4].

One of the recurring problems for association rules is the exponential growth of generated rules volume which impacts doctor's

adaptation. The expert must be able to move as quickly as possible to the essentials during an evaluation. That is why we are interested in graphical representation techniques of association rules. In fact, some approaches are pretty clear for very small data sets; but less so when these quantities increase. In addition, if an expert searches particular information, his eye has a high chance to mask some important information for his analysis [12].

There are many works that deal with the association rules visualization in Data Mining [7], [8], [9], [10], [11], [12], [13].

Different applications have been developed to visualize association rules as tables [9]. Each row corresponds to a rule, and each column represents a rule characteristic whose first column element represents the antecedent, the second column represents the consequent and the following columns represent quality measures of rules. These applications have a rule filtering interface, by specifying the items or setting the support and/or confidence. These applications allow you to classify the rules in ascending or descending order using support or confidence measure [26]. The main limitation of its application is the textual presentation, which does not suit studying large quantities of rules.

On the other hand, the graph visualization [27] is used to represent relationships between items. They are therefore adapted to visualize association rules connecting their antecedent and consequent with an arc. The graph may be directed or undirected. In the undirected graph, in [28], authors used two colors to identify the antecedent and consequent. If an itemset is at the same time the antecedent of one rule and the consequent of another rule, its node has two colors. Three other colors are used to express rule support; light color for low supports, a dark color for high supports and a black color for supports with value 1. The rule confidence is represented by arc length. High confidence is represented by a long arc [26]. An association graph can quickly turn into a tangled display with as few rules. Mosaic plot was introduced by [7] to visualize contingency tables and have been arranged for association rules in [29]. Each rule is represented by a rectangle, whose width is the support and height is the confidence. The main limit to this method is that it can only represent a section of the association rules that have the same consequent and have the same attributes in [26].

In [8], author introduced parallel coordinates as a multidimensional data representation method and they are adapted by [28] to visualize association rules. Variables are represented by parallel axes on which items are distributed and each rule is represented by a broken line which cuts the parallel axes at the level of the items it contains [26].

Matrices are certainly the most used means to represent rules. We identify two types of matrices: Two-Dimensional Matrix (2D) [14], [10], [11], [12]) and three-dimensional (3D) matrices [10], [11].

The work of [14] proposes a Two-Dimensional Matrix (2D). A rule is represented by a cell, the antecedent is displayed on row and consequent is displayed in column or inversely. Visualization of association rule with 3D matrices was introduced with MineSet tools [30]. These matrices known as «item to item» are composed of three axes whose two are used to display items of antecedent or consequent. The third axis is used to represent indicators as support or confidence. A rule is represented by a bar whose gradual color expresses its support and the height corresponds to his confidence [26]. This visualization technique has been improved in matrices «rule-to-item» where each row represents a consequent item and each column has an antecedent item. A rule between two items is represented in the intersection box by a 2D/3D object that indicates the presence or not of the item in the rule. The color of these objects indicates the presence of the item in antecedent or consequent. The matrix is completed by two lines that indicate support and confidence by the height of the bars in three-dimensional. The main disadvantage of this representation is that it

reaches considerable sizes for important sets of rules and become illegible.

The majority of visualization methods are not adapted to represent large sets of patterns. They become unusable; when the pattern number is too important and few methods give an overview of the pattern sets. Finally, no method is adapted to explain how to deduce some rules or the presence of other rules. We are interested in our work on the Two-Dimensional Matrix (2D) for the simplicity to represent the rules on two colors, blue for antecedent and red for consequent.

Nonetheless, in the presence of a significant number of association rules this representation also becomes illegible, and the rules overlap. Occultation problems of such representation become inevitable. To resolve this problem, we have opted for rule optimization using boolean modeling offered by the Cellular Automaton (CASI) [15], and its inference engine to explain the reasoning of some deductions.

In our approach we based on some aspects of the CASI Cellular Automaton: its simplicity to express knowledge in rules and facts, its efficiency in optimizing storage space and execution time. The latter is a particular model of dynamic and discrete systems able to acquire, represent and process extracted knowledge in Boolean form. On the other hand, the machine exploitation as a Cellular Automaton, in the CDSS area, is a novel idea within the team and also at the scientific community. The Cellular Automaton showed its evidence in several research studies in Data Mining: [31] proposed an approach using cellular automaton for the regulation and the reconfiguration of urban transportation systems, [32] proposed a new approach based on cellular automata to reduce the classification time of knearest neighbors algorithm, and [33] proposed a new mapping approach based on the Boolean modeling of critical domain knowledge and also, on the use of different data sources via the data mining technique for the purpose of improving the process of acquiring knowledge explicitly. The problem addressed by [34] is data mining of biological Mycobacterium Tuberculosis responsible for tuberculosis. The author proposed a process of data-enough to generate new knowledge that will be profitable for extraction of particular patterns in the rules of association and they are modeled by the Boolean principle adopted by the cellular machine CASI (Cellular Automaton for Symbolic Induction). [35] proposed a new text categorization framework using concepts lattice and cellular automata. Finally, [36] developed a new approach to boolean fusion ontologies using the CASI cellular machine to optimize the complexity of classical fusion algorithms. The originality of our system is essentially in a proposal of a Intelligent Clinical Decision Support System for immunization using Knowledge Discovery from Databases and the combination

between the 2D matrix and the CASI Cellular Automaton to prove its effectiveness in a new field which is visualization.

III. PROPOSED APPROACH

In order to handle the large volume of generated rules by an extraction algorithms, our approach operates as a real Knowledge Discovery Process with an important focus on the post-treatment phase. Indeed, our contribution is attempting to provide an effective means for the association rules analysis by graphical visualization.

Our objective is the involvement of visualization techniques in the medical decision-making process guided by data mining and Boolean modeling of CASI. Our contribution is a CDSS based on Data Mining named VisuelAR (Visuel Association Rules). The latter has an important role in medical practice, by helping physicians or other medical professionals making clinical decisions, using visualization techniques. Our process is divided into four important phases 1) Data preparation, (2) Extracting association rules, (3) Boolean modeling of the rule base (4) 2D visualization colored by Boolean inferences. (Fig. 1).

A. Preprocessing Module

Preprocessing is an crucial step in the Knowledge Discovery from Databases (KDD). Results obtained at the end of this phase depend mostly on the quality of the data used. Pre-treatment steps relate to data access to build tables named individual-variable tables, grouping observations (explicit data). Depending on data type (numeric, symbolic), pre-treatment methods structure data, clean them, process missing data, and select the attributes when they are numerous: selection of the most informative attributes then sampling. This phase is very important because it will condition the model quality established during Data Mining. Finally, these choices are intended to emerge information contained in the data set [37].

B. Data Mining Module

Data mining module extracts association rules using the APRIORI algorithm [4]. The user can modify configuration settings such as support and confidence as needed. At the output of this extraction algorithm, the association rule sets are simple text lists.

C. Boolean Modeling Module

CASI [15] is a cellular automaton that simulates basic functioning of an inference engine within an expert system, considering a cellular automaton made of two finite arbitrary long layers of finite-

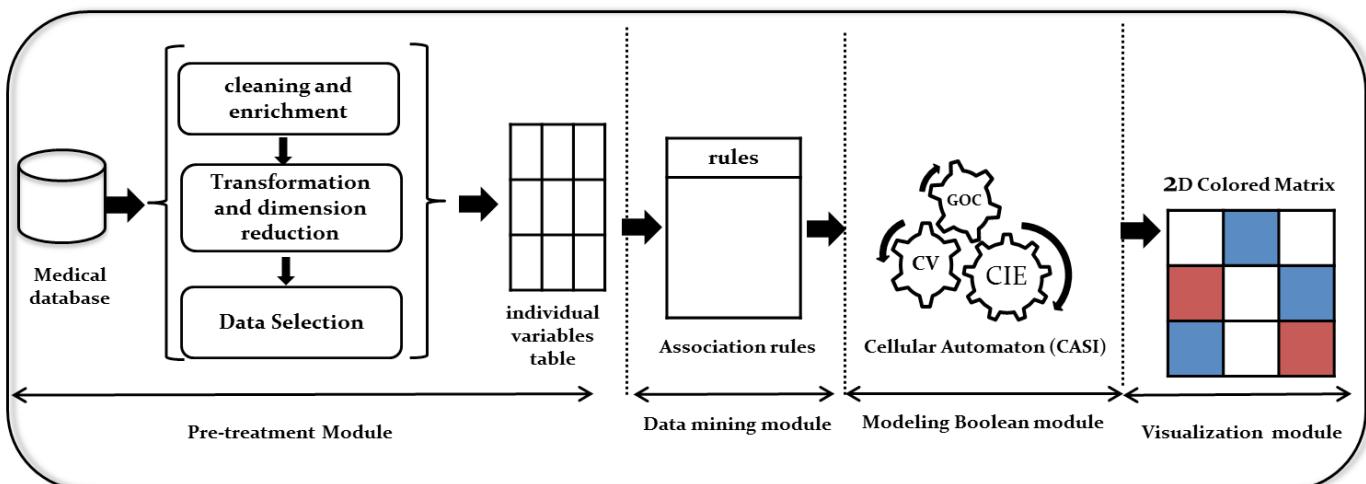


Fig. 1. VisualAR System Architecture.

state machines (cells), all identical. The operation of the system is synchronous, and the state of each cell at time $t+1$ depends only on the state of its vicinity cells, and on its own state at time t . The behavior of a knowledge base can be represented by such a cellular automaton with two layers. A first layer known as CELFACT is the fact base, and a second layer, known as CELRULE, is the rule base. In each layer, the cell contents determine whether and how it participates in each inference step: at every iteration, a cell can be active or passive, can participate in the inference or not. We suppose that there are l cells in the layer CELFACT, and r cells in the layer CELRULE. The cell states are composed of three parts: the input EF, internal state IF and output parts SF of the CELFACT cells and ER, IR and SR which are the input, internal state and output parts of the CELRULE cells, R_E and R_S are the input and the output incidence matrices, respectively:

- The input relation, noted RE, is formulated as follows: $\forall i \in [1, l], \forall j \in [1, r]$, if (fact i is antecedent of rule j) then $RE(i, j) \leftarrow 1$.
- The output relation, noted RS, is formulated as follows: $\forall i \in [1, l], \forall j \in [1, r]$, if (fact i is consequent of rule j) then $RS(i, j) \leftarrow 1$.

The incidence matrices R_E and R_S represent the input/output relation of the facts and are used in forward chaining. One can also use R_E as output relation and R_S as input relation for backward chaining. Finally, since there are l cells in the layer CELFACT, EF, IF and SF will be considered as l -dimensional vectors ($EF, IF, SF \in \{0, 1\}$). Similarly, since there are r cells in the layer CELRULE, ER, IR and SR will be considered as r -dimensional vectors ($ER, IR, SR \in \{0, 1\}$). The cellular automaton dynamics implements the CIE module as a cycle of an inference engine made up of two local transitions δ_{fact} and δ_{rules} where δ_{fact} corresponds to the evaluation, selection and filtering phases, and δ_{rules} corresponds to the execution phase.

δ_{fact} transition function

$$(EF, IF, SF, ER, IR, SR) \xrightarrow{\delta_{fact}} (E^t, F^t, E^t, R^t + (R_E^t \bullet E^t), R^t, R^t) \quad (1)$$

δ_{rules} transition function

$$(EF, IF, SF, ER, IR, SR) \xrightarrow{\delta_{rules}} E^t + (RS \cdot ER), IF, SF, ER, IR, ^ER \quad (2)$$

Where the matrix R_E^t is the transpose of R_E .

We consider G_0 as the initial cellular automaton configuration and the $\Delta = \delta_{fact} \circ \delta_{rules}$ as a global transition function: $(G_0) = G_1$ if $G_0 \xrightarrow{\delta_{fact}} G'_0$ and $G'_0 \xrightarrow{\delta_{rules}} G_1$

$\Delta(G_0) = G_1$ si $G_0 \xrightarrow{\delta_{fact}} G'_0$ et $G'_0 \xrightarrow{\delta_{rules}} G_1$ Let $G \{G_0, G_1, G_2, \dots, G_q\}$ be the configuration set of our cellular automaton. The automaton evolution in discrete time steps from one generation to the next is defined by the configuration sequence $G_0, G_1, \dots, G_q \xrightarrow{\Delta} G_{i+1} = \Delta(G_i)$

Integration of an inference engine in 2D colored matrix gives it the possibility to justify the presence of some rules and facilitates expert decision-making.

D. Visualization Module

This module uses a 2D matrix visualization technique «rule-to-item», where each row represents a consequent item and each column has an antecedent item. The cell color indicates the itemset presence in the R_E or R_S matrix of CASI. The presence in the R_E matrix is represented by the blue and its presence in the R_S matrix is represented by the red. This technique reduces the physician's cognitive loading and improving their visual perception results.

IV. RESULTS AND DISCUSSION

In our approach, we are interested in a preventive medicine area which concerns more than half of the world, namely vaccination.

In the world in general, and particularly in Africa, immunization became, through the different Expanded Program on Immunization, one of the most effective strategies for controlling viral and bacterial diseases. In the late 1970s, it was erected as a health program by the World Health Organization (WHO). This program is to protect children aged 0 to 11 months against the six (6) most deadly diseases. These include tuberculosis, diphtheria, tetanus, whooping cough, polio and measles. Indeed, these diseases are responsible for more than two (2) million deaths per year globally, according to WHO. Yet effective ways to protect children against these diseases have been popularized around the world. Unfortunately, a related problem is that all children presented at the first immunization one week after their birth do not receive all of the vaccines after 11 months for many reasons. They are the lost ones. A true problem for the country's health authorities who are always looking to reduce the causes of lost ones problems.

Through the various surveys funded by the WHO in African countries such as Cameroon, Niger, Burkina Faso, Benin [38] [39], [40], [41]: one of the causes of this low immunization coverage is the poor quality of the service offered. This quality of service can be appreciated from the knowledge viewpoint, staff attitudes and practices, the resource availability, the services organization, satisfaction and the beneficiaries' knowledge. Several studies have revealed that immunization service quality is often poor: it is characterized by the administration age of the different vaccines, high wait times, the many missed occasions, poor reception of providers and the appearance of abscesses after vaccination.

Several studies have shown the correlation between the levels of knowledge mothers have about vaccination and the lost ones [41]. In this context, we are interested in this work to implement our approach on the detection of lost ones and abundance causes in relation to the mother's socio-economic characteristics.

A. The questionnaire

To collect the data about the mother's socio-economic characteristics we have developed an online questionnaire: <https://goo.gl/forms/mqDTqjXjs659AlvG3> with 9 questions in total about the mothers' socio-economic profile using the Google Forms platform. We opted for a majority of single-answer multichotomous questions (7 of 9 questions) in order to have a rapid response and data exploitation. The questionnaire results obtained have been saved in an Excel file (.csv).

Study Population Characteristics

In Table I, we present the socio-economic characteristics of the mothers interviewed. A total of 368 mothers replied to the questionnaire. Mothers aged above 30 represented the largest age group with 64.2% of the total. The university level represents 61.6% of the education level; the main occupation was state function work with 42.5%. 43.7% of these mothers had only 1 child. The most common means of travel for the immunization session was personal transportation in 72% of cases, with travel time less than 30 minutes for 62.3% of mothers. 66.4% of women are dissatisfied with the start time of the vaccination session and 76.1% of women respect vaccination dates.

TABLE I. SAMPLE DESCRIPTION

	EFFECT	Pourcentage
The number of mothers interviewed	268	100%
Distribution by Age		
Between 18 and 30 years old	96	35.8%
More than 30 years old	172	64.2 %
Education level		
Primary	3	1.1 %
Medium	4	1.5 %
Secondary	22	8.2 %
Academic	165	61.6 %
Post-Graduation	74	27.6 %
Occupation		
Student	21	7.8 %
Housewife	91	34 %
State function	114	42.5 %
Liberal function	42	15.7 %
Number of children		
1 only	117	43.7 %
2	85	31.7 %
3	44	16.4 %
4	16	6 %
5 and more	6	2.2 %
Means of travel for vaccination		
Walking	58	21.6 %
Public transport	6	2.2 %
A taxi	11	4.1 %
Personal transportation	193	72 %
Travel time for vaccination		
Less than half an hour	167	62.3 %
Between 30 minutes and 1 hour	68	25.4 %
More than an hour	33	12.3 %
Satisfaction of the immunization session		
Yes	147	54.9 %
No	121	45.1 %
Satisfaction of the start time for the immunization session?		
Yes	178	66.4 %
No	90	33.6 %
Do you respect vaccination dates and times?		
Yes, at each session	204	76.1 %
Yes, occasionally	54	20.1 %
No, never	10	3.7 %

To ensure data quality data pretreatment has been developed:

- *Data selection*

We can select data now that already exist in our initial file; we have selected the following attributes: Breakdown by age, occupation, number of children, education level, Means of travel, Travel time, Respect for vaccination.

- *Data cleansing and enrichment*

Data cleaning starts immediately after data selection [42]. If enrichment is required a second cleaning step is essential. In cleaning, we find the processing of missing data.

- *Data transformation*

This step consists of transforming one attribute A to another A' [42], in our case, to simplify the nomenclature we have used indexation to designate the different answers obtained and prepare the individual/variable tables (Table II).

TABLE II. UNITS FOR MAGNETIC PROPERTIES

ATTRIBUT	INDEXATION
Distribution by Age	Age
Between 18 and 30 years old	18-30
More than 30 years old	>30
Education level	NvEtud
Primary	Pr
Medium	Moy
Secondary	Sec
Academic	Univ
Post-Graduation	PG
Occupation	Occup
Student	Etud
Housewife	FmFoy
State function	FctEt
Liberal function	FctLib
Number of children	NbEnf
1 only	1
2	2
3	3
4	4
5 and more	<=5
Means of travel for vaccination	MoyDep
Walking	Marche
Public transport	TransPb
A taxi	Taxi
Personal transportation	TransPer
Travel time for vaccination	TempDep
Less than half an hour	<30mn
Between 30 minutes and 1 hour	30-60mn
More than an hour	>60mn
Do you respect vaccination dates and times?	RespVacc
Yes	Oui
No	Non

B. Extraction of Association Rules

We implemented the Apriori algorithm in JAVA language on a Windows platform. We conducted a series of experiments on our mother's socio-economic database to choose the right model (Table III). We have fixed a suitable and sufficient number for our experience equal to 500 rules. An excerpt from the rules is given in Fig. 2.

TABLE III. ASSOCIATION RULES SERIES

Test	Min Support	Min Confidence	Rules number	Execution time
Series 1	10%	90%	15	0.0s
Series 2	10%	80%	87	0.2s
Series 3	9%	80%	110	0.2s
Series 4	7%	80%	359	0.5s
Series 5	5%	80%	500	1s

```

R1={NbEnf=4, Occup=FctLib}---->[RespVacc=Non]
R2={NbEnf=4, NvEtud=Univ, Occup=FctLib}---->[RespVacc=Non]
R3={NbEnf=4, Occup=FctLib}---->[NvEtud=Univ, RespVacc=Non]
R4={MoyDep=TransPer, NbEnf=4, Occup=FctLib}---->[RespVacc=Non]
R5={NbEnf=4, Occup=FctLib}---->[MoyDep=TransPer, RespVacc=Non]
R6={NbEnf=4, Occup=FctLib}---->[MoyDep=TransPer, NvEtud=Univ]
R7={NbEnf=4, Occup=FctEt}---->[RespVacc=Oui]
R8={MoyDep=Marche, NbEnf=3}---->[RespVacc=Oui]
R9={MoyDep=TransPer, NbEnf=3, TempDep=<30mn}---->[RespVacc=Oui]
R10={NbEnf=4, NvEtud=Univ, RespVacc=Oui}---->[Occup=FctLib]

```

Fig. 2. Extract from the Association Rules of the 5th Series.

C. CASI and the 2D Colored Matrix

The integration of an inference engine in the 2D color matrix gives it the opportunity to justify the presence of certain rules and facilitate expert decision-making. We will use the sample of the preceding association rules to explain how the cellular Automaton CASI works especially backward chaining. The expert can launch CASI on all iterations and visualize all rules that he/she can deduce from a certain fact. Based on formal mathematical reasoning, the resulting rules can be automatically validated. On the other hand, the expert can visualize the rules obtained, step by step, and stop the system any time. This shows the interactivity between the expert and the visualization matrix which is the expected contribution of our approach.

Inference by Purpose

If inference by Purpose, the user can interact with the inference engine so that he can stop it at any iteration level following a specific purpose (Fig. 3, Fig. 4, Fig. 5):

Fact	EF	IF	SF	Rule	ER	IR	SR
NbEnf=4	0	1	0	R1	0	1	1
Occup=FctLib	0	1	0	R2	0	1	1
MoyDep=TransPer	0	1	0	R3	0	1	1
NvEtud=Univ	0	1	0	R4	0	1	1
RespVacc=Non	0	1	0	R5	0	1	1
RespVacc=Oui	0	1	0	R6	0	1	1
MoyDep=Marche	0	1	0	R7	0	1	1
NbEnf=3	0	1	0	R8	0	1	1
Occup=FctEt	0	1	0	R9	0	1	1
TempDep=<30mn	0	1	0	R10	0	1	1
CELFAC				CELRULE			

Fig. 3. Initial cellular automaton configuration CELFACT/CELRULE.

R_E	R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	R ₇	R ₈	R ₉	R ₁₀
NbEnf=4	1	1	1	1	1	1				1
Occup=FctLib	1	1	1	1	1					
MoyDep=TransPer				1				1	1	
NvEtud=Univ		1								1
RespVacc=Non										
RespVacc=Oui										1
MoyDep=Marche							1			
NbEnf=3						1	1	1		
Occup=FctEt					1					
TempDep=<30mn							1	1		
R_S	R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	R ₇	R ₈	R ₉	R ₁₀
NbEnf=4										
Occup=FctLib										1
MoyDep=TransPer					1	1				
NvEtud=Univ			1			1				
RespVacc=Non	1	1	1	1	1					
RespVacc=Oui							1	1	1	
MoyDep=Marche								1		
NbEnf=3							1	1	0	
Occup=FctEt						1				
TempDep=<30mn							1	1		

Fig. 4. Input/output incidences matrices.

Fact	EF	IF	SF	Rule	ER	IR	SR
NbEnf=4	0	1	0	R1	1	1	1
Occup=FctLib	0	1	0	R2	1	1	1
MoyDep=TransPer	0	1	0	R3	0	1	1
NvEtud=Univ	0	1	0	R4	1	1	1
RespVacc=Non	1	1	1	R5	0	1	1
RespVacc=Oui	0	1	0	R6	0	1	1
MoyDep=Marche	0	1	0	R7	0	1	1
NbEnf=3	0	1	0	R8	0	1	1
Occup=FctEt	0	1	0	R9	0	1	1
TempDep=<30mn	0	1	0	R10	0	1	1
CELFAC				CELRULE			

Fig. 5. Configuration obtained with δ_{fact} .

Application of δ_{fact}

The 2D matrix displays the applicable rules step by step, we note that the rules applicable in this step are R1, R2 R4 (Fig. 6.). Fig. 7 and Fig. 8 represent the 2nd configuration G2. The visualization of this configuration is represented by Fig. 9.

	R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	R ₇	R ₈	R ₉	R ₁₀
NbEnf=4										
Occup=FctLib										
MoyDep=TransPer										
NvEtud=Univ										
RespVacc=Non										
RespVacc=Oui										
MoyDep=Marche										
NbEnf=3										
Occup=FctEt										
TempDep=<30mn										

Fig. 6. Visualization of the applicable rules.

Application de δ_{rules}

Fact	EF	IF	SF	Rule	ER	IR	SR
NbEnf=4	1	1	0	R1	1	1	0
Occup=FctLib	1	1	0	R2	1	1	0
MoyDep=TransPer	1	1	0	R3	0	1	1
NvEtud=Univ	1	1	0	R4	1	1	0
RespVacc=Non	1	1	1	R5	0	1	1
RespVacc=Oui	0	1	0	R6	0	1	1
MoyDep=Marche	0	1	0	R7	0	1	1
NbEnf=3	0	1	0	R8	0	1	1
Occup=FctEt	0	1	0	R9	0	1	1
TempDep=<30mn	0	1	0	R10	0	1	1
CELFAC				CELRULE			

Fig. 7. Configuration G1 = $\Delta(G_0)$ obtained with $\delta_{fact}(G_0)$ and $\delta_{rules}(G_0)$.

Application of δ_{fact}

Fact	EF	IF	SF	Rule	ER	IR	SR
NbEnf=4	1	1	1	R1	1	1	0
Occup=FctLib	1	1	1	R2	1	1	0
MoyDep=TransPer	1	1	1	R3	1	1	1
NvEtud=Univ	1	1	1	R4	1	1	0
RespVacc=Non	1	1	1	R5	1	1	1
RespVacc=Oui	1	1	1	R6	1	1	1
MoyDep=Marche	0	1	0	R7	0	1	1
NbEnf=3	0	1	0	R8	0	1	1
Occup=FctEt	0	1	0	R9	0	1	1
TempDep=<30mn	0	1	0	R10	1	1	1
CELFAC				CELRULE			

Fig. 8. The 2nd configuration G2.

	R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	R ₇	R ₈	R ₉	R ₁₀
NbEnf=4										
Occup=FctLib										
MoyDep=TransPer										
NvEtud=Univ										
RespVacc=Non										
RespVacc=Oui										
MoyDep=Marche										
NbEnf=3										
Occup=FctEt										
TempDep=<30mn										

Fig. 9. Visualization of the 2nd iteration.

Automatic inference

In automatic inference case, the user does not interfere in the inference engine process so that iterations are started automatically until no rules are applicable (Fig. 10 and Fig. 11).

Fact	EF	IF	SF	Rule	ER	IR	SR
NbEnf=4	1	1	1	R1	1	1	0
Occup=FctLib	1	1	1	R2	1	1	0
MoyDep=TransPer	1	1	1	R3	1	1	0
NvEtud=Univ	1	1	1	R4	1	1	0
RespVacc=Non	1	1	1	R5	1	1	0
RespVacc=Oui	1	1	1	R6	1	1	0
MoyDep=Marche	1	1	1	R7	1	1	0
NbEnf=3	1	1	1	R8	1	1	0
Occup=FctEt	1	1	1	R9	1	1	0
TempDep=<30mn	1	1	1	R10	1	1	0
CELFAC				CELRULE			

Fig. 10. Final iteration.

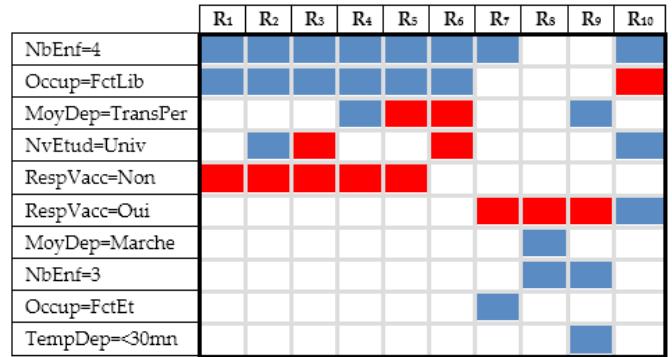
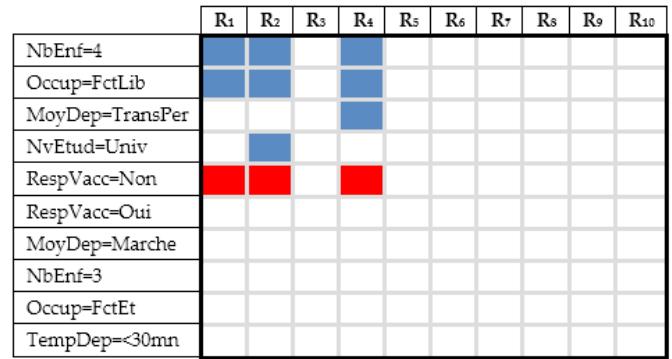


Fig. 11. Visualization of the last iteration.

V. DISCUSSION AND VALIDATION

For this experiment we used the 5th series (Table III) with 500 interesting rules (Fig. 12) which we cannot all interpret. In order to retrieve the rules that can meet our objective which is the detection of the causes of non-vaccination we will use the Boolean modeling and colored 2D matrix visualization.

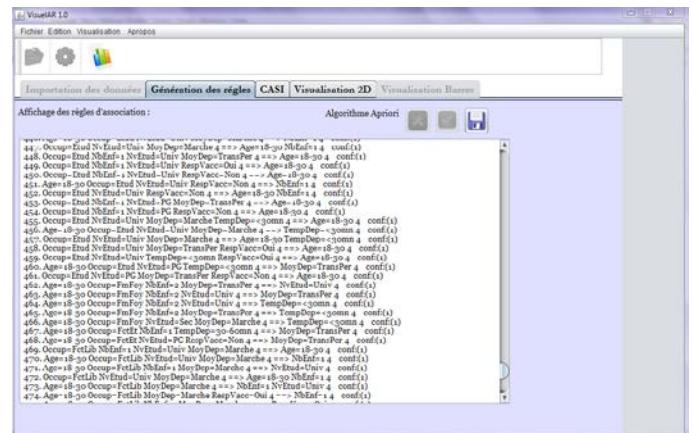


Fig. 12. Textual rules list.

In our case the fact (RespVacc=non) represents the initial fact to be determined (Fig. 13 and Fig. 14).

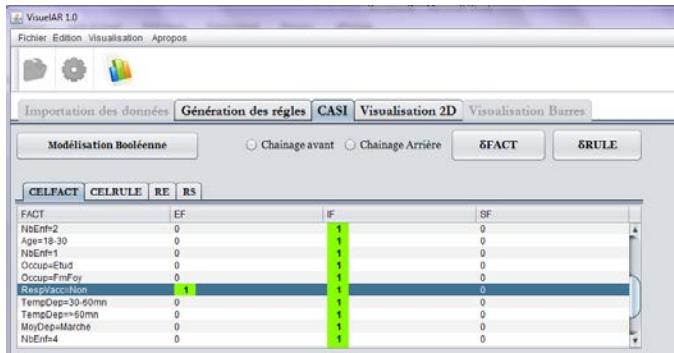


Fig. 13. Initial cellular automaton configuration CELFACT.

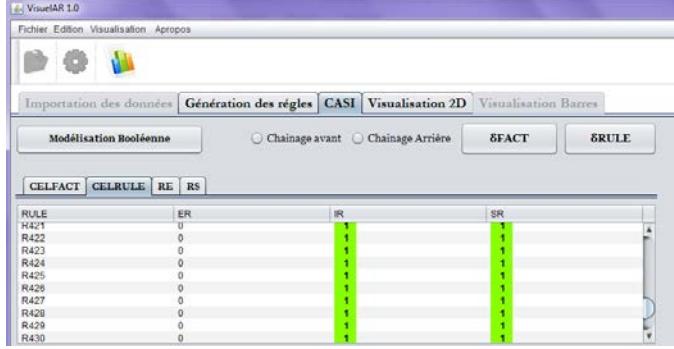


Fig. 14. Initial cellular automaton configuration CELRULE.

The incidence matrices RE (Fig. 15) and RS (Fig. 16) represent the input/output relation of the facts and are used in forward chaining from antecedent to consequent. One can also use in transition function δ_{fact} and δ_{rules} , RS as input relation and RE as output relation for backward chaining, from consequent to antecedent. In our case we will choose a backward chaining to detect the facts that are responsible for the non-vaccination.

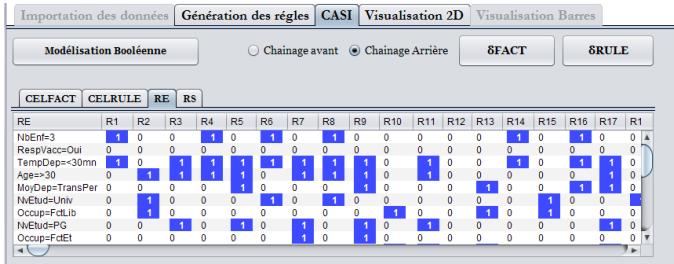


Fig. 15. Input incidence matrix R_E .

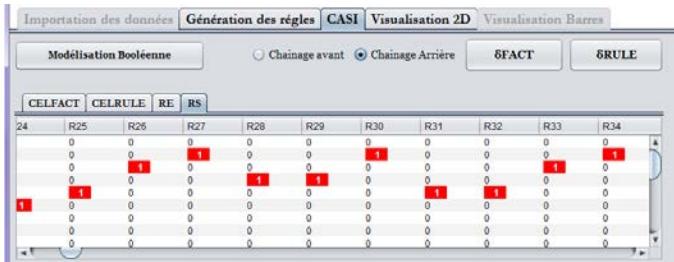


Fig. 16. Output incidence matrix R_s .

Fig. 17 and Fig. 18 illustrate configuration $G1 = \Delta(G_0)$ obtained with δ_{fact} and δ_{rules} (CELFAC).

CELFAC	CELRULE	RE	RS
FACT	EF		
NbEnf=2	0	1	0
Age=18-30	0	1	0
NbEnf=1	0	1	0
Occup=FctLib	0	1	0
Occup=FmFoy	0	1	0
RespVacc=Non	1	1	1
TempDep=30-60mn	0	1	0
TempDep=>60mn	0	1	0
MoyDep=Marche	0	1	0
NbEnf=4	0	1	0

Fig. 17. Configuration $G1 = \Delta(G_0)$ obtained with δ_{fact} and δ_{rules} (CELFAC).

CELFAC	CELRULE	RE	RS
RULE	ER		
R421	0	1	1
R422	0	1	1
R423	0	1	1
R424	0	1	1
R425	0	1	1
R426	0	1	1
R427	0	1	1
R428	0	1	1
R429	0	1	1
R430	0	1	1
R431	0	1	1
R432	0	1	1
R433	0	1	1
R434	0	1	1
R435	0	1	1

Fig. 18. Configuration $G1 = \Delta(G_0)$ obtained with δ_{fact} and δ_{rules} (CELRULE).

	R378	R429	R430	R433	R434
NbEnf=3
RespVacc=Oui
TempDep=<30mn
Age=>30
MoyDep=TransPer
NvEtud=Univ
Occup=FctLib
NvEtud=PG
Occup=FctEt
NbEnf=2
Age=18-30
NbEnf=1
Occup=Etud
Occup=FmFoy
RespVacc=Non
TempDep=30-60mn
TempDep=>60mn
MoyDep=Marche
NbEnf=4
NvEtud=Sec
NbEnf=>5
MoyDep=Taxi
NvEtud=Moy

Fig. 19. Visualization of the applicable rules.

The CASI cellular Automaton gave us the opportunity to retrieve 5 important rules from our series of 500 rules (Fig. 19). The rules, presented in Fig. 19, are R378, R429, R430, R433, R434.

These rules were presented to an expert who helped us interpret them.

The rule 378: {NbEnf=4, Occup=FctLib} \rightarrow {RespVacc=Non} indicates that women who have at least 4 children and exercise a liberal profession (presented by a blue color), do not respect the vaccination date (presented by a red color).

The rule 429: {NbEnf=4, NvEtud=Univ, Occup=FctLib} \rightarrow {RespVacc=Non} indicates that women who have at least 4 children and exercise a liberal profession; have a university education (presented by a blue color), do not respect the vaccination date (presented by a red color).

The rule 433: {MoyDep=TransPer, NbEnf=4, Occup= FctLib} \rightarrow {RespVacc=Non} indicates that women who have personal transportation, at least 4 children, exercise a liberal profession, have a university education (presented by a blue color), do not respect the vaccination date (presented by a red color).

The rest of the rules can be interpreted in the same way and it is clear that this result depends on the population and sample size of the study.

System validation

The complexity of every system is calculated by two factors: the temporal factor, it must go faster as possible, and a spatial factor, it must consume less of memory.

The incidence matrix, RE, facilitates the rules transformation into boolean expressions and makes it possible, to use elementary boolean algebra to test different simplifications. Boolean modeling proves that this can be achieved by reducing the amount of storage and execution time. Indeed, this is due to the use of the Boolean representation of the RE and RS matrices, and the Boolean multiplication used by the transition functions δ_{fact} and δ_{rules} . The two intensive computation processes in CASI are the storage and Boolean multiplication of the RE and RS incidence matrices.

- Storage in R_E and R_S memory; the latter, being Boolean matrices, can be expressed as follows: two vectors of various binary sequences (in Hexa). The amount of memory necessary to store Boolean matrices is in the order of $O(q)$ when using q sequences of r bits or $O(r)$ when r sequences of q bits are used. Such matrices can be processed in $q \times r$ steps. On the other hand, these matrices are in all the hollow iterations (contain a lot of zeros), it is enough to store only the other values equal to 1.
- The standard algorithm ($R_E^T \cdot EF$), for example, used by the transition function δ_{fact} can be expressed by a sequential algorithm of boolean vector-matrix multiplication executed in a time $O(rq)$ where q is the dimension of the vector EF and $r \times q$ the matrix R_E^T dimension. The multiplication of R_E^T and EF can be executed using the boolean matrix vectorization technique, in a time $\approx O(r \log q)$, where the internal product of a REt. line with the vector EF is reduced to the product of the parity bit (bit wise And). We can therefore conclude that boolean modeling by cellular Automaton, can be a powerful tool for exploration of research spaces efficiently and effectively. It represents an algorithmic alternative of less complexity that facilitates scaling.

VI. CONCLUSION

This paper presents a new solution to the complex problem of decision visualization. We are interested in improving the visualization of large sets of association rules to increase system performance while reducing the user's cognitive load. In this article we have essentially tried to highlight the combined use of visualization and data mining techniques in an intelligent clinical decision support system. We proposed an approach combining a colored 2D matrix with Boolean modeling, this brings a deduction strategy that makes it expressive. Indeed, the integration of an inference engine on the 2D colored matrix can justify the presence of certain rules and facilitate the expert's decision-making. The result obtained illustrates this intention. The expert can choose to run the Automaton on all its iterations and see all of the rules it can infer from a certain fact. Using formal mathematical reasoning, the resulting rules can be automatically validated. On the other hand, the expert can visualize the rules obtained step by step and stop the system at will. This reflects the interactivity between the expert and the visualization matrix, which is the contribution of our approach. However, some improvements can be envisaged to make our approach more complete: set grayscale to colors based on confidence and support rules; submit to experts, sets of rules from medical areas to validate our approach. The results presented in this paper provide the basis for future research in several areas. Firstly, we propose to evaluate our approach in other application areas and with other visualization

techniques. Secondly, one future direction of our work is to evaluate association rules with the different measures [43].

REFERENCES

- [1] A. Akharraz, J. Montmain, G. Mauris, "Elucidation and risk expressions of a movie recommendation based on a multi-criteria aggregation with a Choquet integral". In IIPMU 2004, 10th International Conference on Information processing and Management of uncertainty in Knowledge-Based Systems, 2004.
- [2] S. Benbelkacem, B. Atmani, A. Mansoul, "Planification guidée par Raisonnement à base de cas et Datamining : Remémoration des cas par Arbre de décision", Atelier aide à la Décision à tous les Etages (Aide@EGC2012), 2012.
- [3] H. Ltifi, C. Kolski, M. B. Ayed & A. M. Alimi, "A human-centred design approach for developing dynamic decision support system based on knowledge discovery in database". Journal of Decision Systems, 22(2), 69-96 , 2013.
- [4] R. Agrawal, T. Imielinski, A. N. Swami, "Miningassociation rules between sets of items in large databases". In PeterBuneman and SushilJajodia, editors, Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207-216, Washington, D.C , 1993.
- [5] D. Dahmani, "Data Mining Using Association Rules Conducted By Pre-selected Rules Patterns and Ontology", ICCESS 2010
- [6] N. Pasquier, "Extraction de bases pour les Règles d'Association à partir des Itemsets Fermés Fréquents", INFORSID'2000 Congress, (pp. 56-77), 2005.
- [7] J. Hartigan, B. Kleiner "Mosaics for contingency ". Proceedings of the 13th symposium on the interface. Springer US, 1981.
- [8] A. Inselberg, "N-dimensional graphics, Part I—lines and hyperplanes ", Scientific Center Report G320-2711, IBM, Los Angeles, CA, 1981.
- [9] S. Chakravarthy, H. Zhang. "Visualization of association rules over relational dbmss", SAC, 2003.
- [10] S. Ben Yahia, E. Nguiifo, "Emulating a cooperative behavior in a generic association rule visualization tool with Artificial Intelligence ". ICTAI 2004. 16th IEEE International Conference 2004.
- [11] K. Zhao, B. Liu, M. Thomas, Tirpak, X. Weimin, "Opportunity map: a visualization framework for fast identification of actionable knowledge. In CIKM'05 ": Proceedings of the 14th ACM international conference on Information and knowledge management. ACM Press, New York, NY, USA, 2005.
- [12] O.Couturier, J. Rouillard, V. Chevrin, "Une approche hybride pour une meilleure visualisation de grands ensembles de règles d'association ", ERGOIA'06, 2006.
- [13] M. Ounifi, S. Amdouni, H. Elhoussine, R. Slimane. "New 3D Visualization and Validation Tool for Displaying Association Rules and Their Associated Classifiers ". In Information Visualization (IV). 20th International Conference (pp. 152-158). IEEE, 2016.
- [14] J. Boulicaut, P. Marcel, C. Rigotti , "Query driven knowledge discovery in multidimensional data ", DOLAP '99 : Proceedings of the 2nd ACM international workshop on Data warehousing and OLAP, pp. 87-93, New York, NY, USA. ACM Press 1999.
- [15] B. Atmani, , B. Beldjilali, "Knowledge Discovery in Database : Induction Graph and Cellular Automaton ", Computing and Informatics Journal, 26(2), 171-197, 2007.
- [16] S. Darmoni, "Titres et travaux, Informatique de santé ", Sciences et Technologies de l'Information et de la Communication 2003.
- [17] G. Leroy, H. Chen. "Introduction to the special issue on decision support in medicine", Decision Support Systems, 43 (4), 1203-1206, 2007.
- [18] C. R. Chen, , Y. H. Huang, C. T. Bau, , S. M. Chen, "A recommendation system based on domain ontology and SWRL for anti-diabetic drugs selection". Expert Systems with Applications, 39(4), 3995-4006, 2012.
- [19] A. Khdega Yosef Galala, "An Expert System for Diagnosis of Ear Problems in Children, International Journal of Advanced Research in Computer and Communication Engineering 4(7), 2015.
- [20] M. Cleret, P. Le Beux, F. Le Duff, "Les systèmes d'aide à la décision médicale ", Les Cahiers du numérique, 2(2/2001), 125-154, 2001.
- [21] P. Degoulet M. Fieschi : "Traitement de l'information médicale : méthodes et applications hospitalières ", Masson, 1991.
- [22] González-Ferrer, A., Seara, G., Cháfer, J., & Mayol, J. "Generating Big Data Sets from Knowledge-based Decision Support Systems to Pursue

- Value-based Healthcare.” International Journal of Interactive Multimedia and Artificial Intelligence, Special Issue on Big Data and e-Health 4(7), 42-46, 2018.
- [23] V. Chevrin, , O. Couturier, , E. Mephu Nguifo, , and J. Rouillard, “Recherche anthropocentré de règles d’association pour l’aide à la décision. Revue d’Interaction Homme-Machine, 8(2), 2007.
- [24] J. Soni, , U. Ansari, , D. Sharma, , S. Soni, “Predictive data mining for medical diagnosis: An overview of heart disease prediction ”. International Journal of Computer Applications, 17(8), 43-48, 2011.
- [25] B. diri, & A. Napoli, Découverte de règles d’association pour l’aide à la prévision des accidents maritimes. In Revue des Nouvelles Technologies de l’Information (No. RNTI-E-23, pp. pages-243). Hermann Editio, 2012, January.
- [26] A. Achouri . “ Extraction de relations d’associations maximales dans les textes : représentation graphique”, PhD thesis, University of Québec Trois-Rivières, 2012.
- [27] J. Han, Y. Fu, W. Wang, J. Chiang, W. Gong, K. Koperski, D. Li, Y. Lu, A. Rajan, N. Stefanovic, B. Xia, O. Zaïane , “Dbminer : A system for mining knowledge in large relational databases ”. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 250–255,1996.
- [28] D. Bruzzese, P. Buono, “Combining Visual Techniques for Association Rules Exploration”, Proceedings of the International Conference Advances Visual Interfaces, Gallipoli, Italy, May 25-28, 2004.
- [29] H. Hofmann, A. Wilhelm, “Visual comparison of association rules”, Computational Statistics 16(3), 399-415, 2001.
- [30] C. Brunk, J .Kelly, and R. Kohavi, “Mineset : An integrated system for data mining ”. KDD, (135-138), 1997.
- [31] F. Amrani, K. Bouamrane, B. Atmani, and D. Hamdadou, “Une nouvelle approche pour la régulation et la reconfiguration spatiale d’un réseau de transport urbain collectif”. Journal of Decision Systems 20(2), 207-239, 2011.
- [32] F. Barigou, , B. Atmani, , Y. Bouziane, and N. Barigou, “Accélération de la méthode des K plus proches voisins pour la catégorisation de textes ”. EGC*2013 , 241-246, 2013.
- [33] M. Brahami, B. Atmani, N. Matta, “Dynamic knowledge mapping guided by data mining : Application on Healthcare ”. JIPS 9(1), 1-30, 2013.
- [34] A. Mansoul, B. Atmani, S. Benbelkacem, “A hybrid decision support system: application on healthcare”. CoRR abs/1311.4086, 2013
- [35] H. Benfriha, F. Barigou , B. Atmani, “A text categorisation framework based on concept lattice and cellular automata ”. Int. J. of Data Science, 1(3), 227 – 246, 2016.
- [36] F. Abdelouhab, and B. Atmani, “Fusion cellulaire des ontologies”. Journal of Decision Systems ISSN: 1246-0125 (Print) 2116-7052 (Online) Journal homepage: <http://www.tandfonline.com/loi/tjds20>, 2016.
- [37] D.A Zighed, R.RaKotomalala, “Graphs of induction, Trainining and datamining”, Hermes science publication (Edition Herme), 2000.
- [38] S. Savadogo. Etude des causes des abandon de vaccination des enfants de 0 – 11 mois, DIPLOME INTERUNIVERSITAIRE (DIU) Organisation et Management des systèmes Publics de Prévention Vaccinale dans les Pays en Développement, 2003.
- [39] N. Ganda, Etude des causes de non vaccination des enfants de 12-23 mois en zone semi nomade dans le district sanitaire de N’Guigmi au Niger en 2007, DIPLOME INTERUNIVERSITAIRE (DIU) 3ème Cycle, Organisation et Management des Systèmes Publics de prévention vaccinale dans les Pays en Développement. 2009
- [40] V. D Sodjinou. Etude des déterminants de non vaccination en milieu holli dans la zone sanitaire de pobe au bénin, Organisation et Management des Systèmes Publics de prévention vaccinale dans les Pays en Développement, 2008.
- [41] JR. Balo Evaluation de la qualité des prestations de vaccination dans le district de sante de Mbarkomo, Cameroun, Master Economie et Management de la Santé Publique dans les Pays en Développement, Université de Cocody Abidjan et Université de Paris Dauphine, 2010.
- [42] U G., Fayyad, Piatetsky-Shapiro et P. Smyth. Knowledge Discovery and Data Mining: Towards a Unifying Framework», Proc. 2nd International Conference on KDD & DM, Simoudis E. and Han J. (Eds.), AAAI Press, Menlo Park CA, P. 82-87, 1996.
- [43] S. Lalllich, & O. Teytaud. Évaluation et validation de l’intérêt des règles d’association. Revue des Nouvelles Technologies de l’Information, 1(2), 193-218, 2004.



Fatima Zohra Benhacine

She is currently a PhD candidate at the University of Oran 1 Ahmed Ben Bella and affiliated researcher in Laboratoire d’Informatique d’Oran, Algeria. She received her Master on ID-IHM Degree from university of Oran 1 Ahmed BenBella, Algeria in 2013. Her research topics include data mining, Vizualisation, medical decision support systems, machine learning.



Baghdad Atmani

He is a Professor of Computer Science at the University of Oran 1 Ahmed Benbella. His field of interests is Data Mining and Machine Learning Tools. His research is based on Knowledge Representation, Knowledge-based Systems and CBR, Data and Information Integration and Modelling, Data Mining Algorithms, Expert Systems and Decision Support Systems. His research is guided and evaluated through various applications in the field of control systems, scheduling, production, maintenance information retrieval, simulation, data integration and spatial data mining.



Fawzia Zohra Abdelouhab

She is a Master of Conference and teacher at the department of computer Science, University of Oran1, Ahmed Ben Bella, Algeria. She received her Magister in Computer Science from Oran University. She is currently a Ph.D. in the Computer Science Department at the same university. Her research interests focus on Knowledge Engineering, Ontologies Matching, and information retrieval areas.