# FUZZY BOOLEAN REASONING FOR DIAGNOSIS OF DIABETES

Mohamed Benamina, Baghdad Atmani and Sofia Benbelkacem

Laboratoire d'Informatique d'Oran (LIO)
University of Oran 1 Ahmed Ben Bella
BP 1524, El M'Naouer Es Senia, 31 000 Oran, Algeria

***ABSTRACT***

*The classification by inductive learning finds its originality in the fact that humans often use it to resolve and to handle very complex situations in their daily lives. However, the induction in humans is often approximate rather than exact. Indeed, the human brain is able to handle imprecise, vague, uncertain and incomplete information. Also, the human brain is able to learn and to operate in a context where uncertainty management is indispensable. In this paper, we propose a Boolean model of fuzzy reasoning for indexing the monitoring sub-plans, based on characteristics of the classification by inductive learning. Several competing motivations have led us to define a Boolean model for CBR knowledge base systems. Indeed, we have not only desired experiment with a new approach to indexing of cases by fuzzy decision tree, but we also wanted to improve modelling of the vague and uncertain of the natural language concepts, optimize response time and the storage complexity.*

***KEYWORDS***

*Boolean Modelling, Cellular Machine, Case-Based Reasoning, Diabetes Diagnosis, Fuzzy Reasoning, Planning.*

## 1. INTRODUCTION

The problem of planning and scheduling of tasks is one of the most complex problems in the field of Artificial Intelligence. The best-known situations include crisis management, production management, project management, robotics, medical, etc. The goal of planning is to provide a system (robotics, computer, human, ...) the capacity to reason to interact with its environment in an autonomous manner, in order to achieve the objectives that have been assigned.

Scheduling is organized in time a set of tasks. Historically, scheduling problems were discussed initially in the field of operational research (graph dynamic programming, linear programming, methods of combinatorial optimization theory), but quickly showed their limits in terms of expressiveness. Artificial intelligence and knowledge-based systems are then addressed the problem, renewing techniques through a richer representation of the domain knowledge (problems of satisfaction of constraints, constraints propagation algorithms, constraint programming languages). Among knowledge-based systems we looked on the reasoning from case (CBR). The CBR based on artificial intelligence techniques is an approach to problem solving that uses past experiences to solve new problems by finding similar cases in its knowledge base and adapting them to the particular case. All the experiences form a case basis. Each case is represented by a knowledge experience. This experience is a lesson for the CBR system to solve problems of various kinds. The CBR consists of five phases: (1) Elaboration of

Dhinaharan Nagamalai et al. (Eds) : CSEIT, NCS, SPM, NeTCoM - 2018
pp. 27–44, 2018. © CS & IT-CSCP 2018 DOI : 10.5121/csit.2018.81803

the case; (2) Retrieval; (3) Adaptation; (4) Review; (5) Memory. For our project we are interested in the second phase: retrieval.

Therefore our contribution in this area is double, on the one hand it offers a reactive planning module based on a CBR for the optimization of the scheduling, and on the other hand it offers a classification induction graph [1] for the acceleration of the indexing of cases: remembering. The classification issue is to assign the various observations to categories or predefined classes [2] [3]. In general classification methods consist in several stages. The most important step is to develop the rules of classification from a priori knowledge; it is the learning phase [4]. The classification by inductive learning finds its originality in the fact that humans often use it to resolve and to handle very complex situations in their daily lives [5]. However, the induction in humans is often approximate rather than exact. Indeed, the human brain is able to handle imprecise, vague, uncertain and incomplete information [6]. Also, the human brain is able to learn and to operate in a context where uncertainty management is indispensable. In this paper, we propose a Boolean model of fuzzy reasoning for indexing the sub-plans [13], based on characteristics of the classification by inductive learning in humans [7].

This article is structured as follows. Section 2 presents a state of the art about the use of fuzzy decision tree in the retrieval step of CBR, and also work about cellular automaton and Boolean modelling. Section 3 is devoted to the proposed approach Fuzzy-BML-CBR. Section 4 presents results of experimentation. Finally, we present the guidance of our contribution and experimentation and we conclude in section 5.

## 2. LITERATURE REVIEW

We present the state of the art in two ways. First, we quote work which combine fuzzy reasoning with decision tree in the retrieval step of CBR. Then, we give works about cellular automaton and Boolean modeling.

### 2.1. FUZZY DECISION TREE FOR RETRIEVAL

Fuzzy decision tree have been applied in various areas and specifically in medicine. Boyen and Wehenkel [8] describe a new algorithm able to infer fuzzy decision trees in domains where most of the input variables are numerical and output information is best characterized as a fuzzy set. It comprises three complementary steps: growing for selecting relevant attributes and fuzzy thresholds; pruning for determining the appropriate tree complexity; refitting for tuning the tree parameters in a global fashion. Begum et al. [9] presented a case-based decision support system to assist clinicians in stress diagnosis. Case-based reasoning is applied as the main methodology to facilitate experience reuse and decision explanation by retrieving previous similar temperature profiles. Further fuzzy techniques are also employed and incorporated into the case-based reasoning system to handle vagueness, uncertainty inherently existing in clinicians reasoning as well as imprecision of feature values. The work of Barrientos and Sainz [10] provides support for decision making about resource planning of an emergency call center in order to reach its mandatory quality of service. This is carried out by the extraction of interpretable knowledge from the activity data collected by an emergency call center. A linguistic prediction, categorization and description of the days based on the call center activity and information permits the workload for each category of day to be known. This has been generated by a fuzzy version of an unsupervised decision tree, merging decision trees and clustering. Levashenko and Zaitseva [11] proposed a decision making support system based on fuzzy logic for oncology disease diagnosis. The decision making procedure corresponds to the classification of the new case by analyzing a set of instances for which classes are known. Solved cases are defined as fuzzy classification rules that are formed by different fuzzy decision trees. Three types of fuzzy decision trees are considered in the paper: non-ordered, ordered and stable. Induction of these fuzzy decision trees is based on cumulative information estimates. Adidela [12] proposed a Hybrid Classification System to predict the occurrence of diabetes. The system adopts three phases. In the first phase, clustering of the data using EM-algorithm is performed. The second phase carries out the classification of the obtained individual clusters using fuzzy ID3. As of the second phase of the process, adaptation rules are obtained. These rules are essential in the prediction of diabetes. In the third phase the test tuple is supplied to the rules to predict the class label. Benamina et al [13] combined fuzzy logic and decision tree to improve the response time and the accuracy of the retrieval of similar cases. The proposed Fuzzy case-based reasoning is composed of two complementary parts, a classification by fuzzy decision tree and a CBR part. The aim of this approach was to reduce the complexity of calculating similarity degree between diabetic patients.

### 2.2. CELLULAR AUTOMATON AND BOOLEAN MODELLING

The objective of our approach is double: first, it provides a reactive planning module based on CBR for scheduling optimization; secondly it generates a classification decision tree to accelerate the indexing of sub-plans. The second step uses the Boolean modeling. So, the cellular machine allows to reduce the size of decision tree and to optimize automatically the generation of symbolic rules [3].

Amrani et al. [14] proposed an approach based on cellular automata for regulation and reconfiguration of urban transportation systems. Barigou et al. [15] proposed a Boolean modeling approach which uses a boolean inference engine based on a cellular automaton to do extraction. Atmani et al. [16] proposed a boolean modeling of the fuzzy reasoning and used the characteristics of induction graph classification. The retrieval phase of CBR was modeled in the form of a database with membership functions of fuzzy rules. Brahami et al. [17] exploited different data sources for improving the process of acquisition of explicit knowledge on an organization by producing inductive Boolean rules. Benfriha et al. [18] proposed a new text categorization framework based on a cellular automaton for Symbolic Induction. Aissani et al. [19] exploited a Job Shop scheduling log and simulations to extract knowledge enabling to create rules for the selection of priority rules. These rules are implemented in a CASI cellular automaton. First, symbolic modeling of the scheduling process is exploited to generate a decision tree. Then, decision rules are extracted to select priority rules. Finally, the rules are integrated in CASI which implements the decisional module of agents in a distributed manufacturing control system.

## 3. PROPOSED APPROACH FUZZY-BML-CBR

The architecture of the proposed Fuzzy-BML-CBR for diabetes application is given in figure 1. The main aim of the proposed framework is on improving the accuracy of Diabetes classification. The followings are the main contributions of this paper:

- The proposed framework is a novel combination of different techniques that perform classification to Diabetes patients using Fuzzy data mining, Boolean modeling and CBR;
- Fuzzy decision tree classifier is used to generate a crisp set of rules;
- Fuzzy modeling is used to deal with the uncertainty related to the medical reasoning;
- Boolean modeling is used for fuzzy rules optimization and inference;
- Case based reasoning.

It is composed of two complementary parts, the fuzzy boolean modeling part using Fispro and the case based reasoning part using JColibri platform. FisPro has been used for various modeling projects [20], and we hope that the approach presented in this paper will help in new modeling tasks. JColibri [21] is an object-oriented tool dedicated to the development of case-based reasoning applications. It is an open source tool that allows the user to customize the classes and methods of the platform according to specific needs.

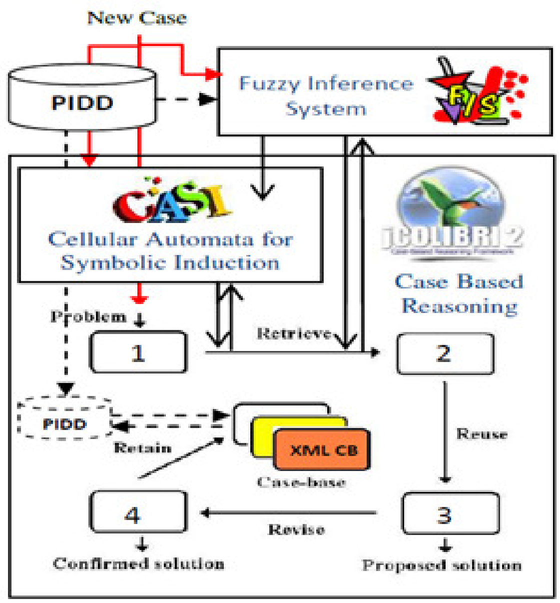


Figure 1. Fuzzy-BML-CBR for diabetes application

The main steps of the proposed framework are :

- Construction by Fispro of the fuzzy decision tree and extraction of the fuzzy rule base;
- From the fuzzy decision tree CASI begin the boolean modelling for the construction of the boolean fuzzy decision tree.
- Finally, JColibri combine the fuzzy inference system and CASI to improve the response time and the accuracy of the retrieval of similar cases.

## 3.1. CELLULAR AUTOMATON AND BOOLEAN MODELING

### 3.1.1. PIMA INDIANS DIABETES DATABASE

The Pima Indian Diabetes Dataset (PIDD) has been taken from the UCI Machine Learning repository. The input variable are Plasma glucose concentration in 2-hours OGTT(Glucose), 2-hour serum insulin(INS), Body mass index(BMI), Diabetes pedigree function(DPF), Age(Age) and the output variable are Diabetes Mellitus(DM) (Table 1). The data with the age group from 25-30 are taken to test the Fuzzy Inference Mechanism Framework [22].

Table 1. Attributes of PIDD.

| Abbreviation | Full name | UoM |
|---|---|---|
| Pregnant | Number of times pregnant | - |
| Glucose | Plasma glucose concentration in 2-hours OGTT | mg/dl |
| DBP | Diastolic blood pressure | mmHg |
| TSFT | Triceps skin fold thickness | mm |
| INS | 2-hour serum insulin | mu U/ml |
| BMI | Body mass index | Kg/m2 |
| DPF | Diabetes pedigree function | - |
| Age | Age | - |
| DM | Diabetes Mellitus where 1 is interpreted as tested positive for diabetes | |

### 3.1.2. FUZZY MODELING USING FISPRO

Figure 2 illustrates the fuzzy inference system. Firstly, a crisp set of input data are gathered and converted to a fuzzy set using fuzzy linguistic variables, fuzzy linguistic terms and membership functions. This step is known as fuzzification. Afterwards, an inference is made based on a set of rules. Lastly, the resulting fuzzy output is mapped to a crisp output using the membership functions, in the defuzzification step.

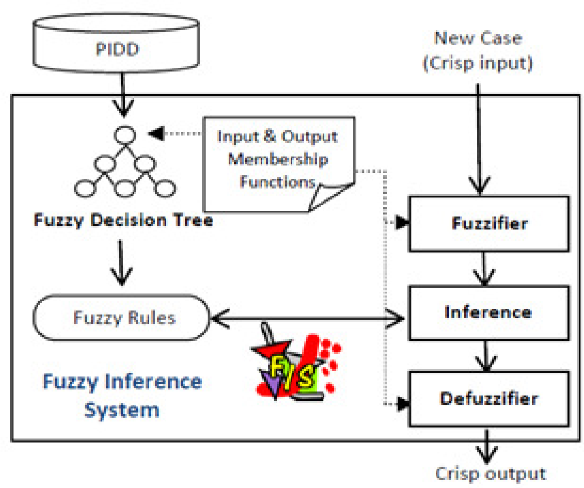


Figure 2. Fuzzy Inference System

The process of fuzzy inference mechanism is explained in Algorithm 1.
Algorithm 1 Fuzzy logic algorithm [13]:

1. Define the linguistic variables and terms (initialization).
2. Construct the membership functions (initialization).
3. Construct the fuzzy decision tree and the fuzzy rule base (initialization).
4. Convert crisp input data to fuzzy values using the membership functions (fuzzification).
5. Evaluate the rules in the fuzzy rule base (inference).
6. Combine the results of each rule (inference).
7. Convert the output data to non-fuzzy values (defuzzification).

**Fuzzification :** The conversion from crisp to fuzzy input is known as fuzzification [23]. Each crisp input is converted to its fuzzy equivalent using a family of membership function. Additionally, an interface is offered to tune and validate the parameters of the built fuzzy numbers. The parameter is fixed with Minimum value, Mean, Standard Deviation, Maximum value for each variable. Then the membership function µ (x) of the triangular fuzzy number is given by :

$$\mu(x) = \begin{cases} 0, x \le a \\ (x - a)/(b - a), a < x \le b) \\ (c - x)/(c - b), b < x \le c \\ 0, x > c \end{cases}$$

The parameters of fuzzy numbers are listed in Table 2.

Table 2. Parameters of triangular membership functions [22].

| **Fuzzy variables** | **Fuzzy Numbers** | **Fuzzy Triangular numbers** |
|---|---|---|
| Glucose (Plas) | Low | [0, 88.335, 121.408] |
| | Medium | [88.335, 121.408, 166.335] |
| | High | [121.408, 166.335 ,199] |
| INS (Insu) | Low | [0 ,17.276, 173.175] |
| | Medium | [17.276, 173.175, 497] |
| | High | [173.175,497, 846] |
| BMI (Mass) | Low | [0, 0, 27.792] |
| | Medium | [0, 27.792, 38.864] |
| | High | [27.792, 38.864, 67.1] |
| DPF (Pedi) | Low | [0.078, 0.272, 0.682] |
| | Medium | [0.272, 0.682, 1.386] |
| | High | [0.62, 1.386, 2.42] |
| Age (Age) | Young | [21, 25.475, 40.537] |
| | Medium | [25.475, 40.537, 57.798] |
| | Old | [40.537, 57.798, 81] |
| DM (Class) | Very low | [0, 0, 0.25] |
| | Low | [0, 0.25, 0.5] |
| | Medium | [0.25, 0.5, 0.75] |
| | High | [0.5, 0.75, 1] |
| | Very high | [0.75, 1, 1] |

**Fuzzy inference engine :** In our study, we make use of the Fuzzy Decision Trees (FDT), which are an extension of classical decision trees [1] [24] and constitute a popular elaborate application of region based methods. The FDT proposed in FisPro are based on the algorithm presented in [25]. The FisPro implementation relies on a predefined fuzzy partition of the input variables, which is left untouched by the tree growing algorithm.

FisPro is an open source tool for creating fuzzy inference systems (FIS) to be used for reasoning purposes, especially for simulating a physical or biological system [20]. It includes many algorithms (most of them implemented as C programs) for generating fuzzy partitions and rules directly from experimental data. In addition, it offers data and FIS visualization methods with a java-based user-friendly interface. We make use of the Fuzzy Decision Trees (FDT) [26] algorithm provided by FisPro.

**Defuzzification :** Defuzzification process is conducted to convert aggregation result into a crisp value for DM output. In this process a single number represents the outcome of the fuzzy set. The final combined fuzzy conclusion is converted into a crisp value by using the centroid method [13].

## 3.2. Cellular Automata for Symbolic Induction

### 3.2.1. Classification by Inductive Learning

In a context of diabetic patients monitoring [27], setting up tools for accident detection is not possible without considering the necessary role that the physician must have. The aim is to design a system for assisted monitoring and diagnosis that will provide specialists with the necessary information for identifying the diabetes type of patients.

Let $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$ be the population of diabetic patients taken into account for the training. An attribute is associated with this population, called endogenous variable (also called explicative variable or class attribute), denoted $C$.

A class $C(\omega)$ can be associated with every individual $\omega$. The endogenous variable $C$ takes its values in the set $IC$ of class identifiers.

$$C:\ \Omega \rightarrow IC = \{c_1, c_2,...,c_m\}$$

$$\omega_i \rightarrow C(\omega_i) = c_j$$

The data are taken from PIDD, the input variable are Plasma glucose concentration in 2-hours OGTT(Glucose), 2-hour serum insulin(INS), Body mass index(BMI), Diabetes pedigree function (DPF), Age (Age) and the output variable are Diabetes Mellitus (DM). This will be designated by an endogenous variable.

$$C: \Omega \rightarrow IC = \{c_1, c_2\}.$$

The objective is to define a function φ for predicting the class C, thus the diagnosis of diabetes. The determination of the prediction model φ, which is the goal of the training, is bound to the hypothesis that the values taken by the endogenous variable $C$ are not at random, but depend upon certain individual situations, called exogenous variables that are determined by the expert.
The exogenous variables concerning an individual constitute a tuple of attributes:

$$X = \{X_1, X_2,...,X_p\}$$

The exogenous variables take their values in a set $IM$ of mode identifiers:

$$X:\ \Omega \rightarrow IM = \{c_1, c_2,...,c_m\}$$
$$X(\omega) = \{X_1(\omega), X_2(\omega),...,X_p(\omega)\}$$

The value taken by $X_j(\omega)$ is called the modality of the attribute $X_j$ for $\omega$. In our case the exogenous variables are summarized in Table 3.

Table 3. Exogenous variables, semantics and possible modality.

| Exogenous Var. | Semantics | Modality |
|---|---|---|
| X1 | Glucose | low, medium, high |
| X2 | DBP | low, medium, high |
| X3 | TSFT | low, medium, high |
| X4 | INS | low, medium, high |
| X5 | BMI | low, medium, high |
| X6 | DPF | low, medium, high |
| X7 | Age | young, medium, old |
| Y | DM | very low, low, medium, high, very high |

Updating φ requires two samples denoted Ωa and Ωt, which are subsets of Ω. The first one, Ωa, used for training, will serve for the construction of φ. The second one, Ωt, used for test, will serve for testing the validity of φ. For all patients ω∈(Ωa ∪ Ωt) we assume that both the values $X(\omega)$ and the class $C(\omega_i)$ are known. We also define Ωe, the set of individuals in Ωt (patients) not correctly classified during the test of the symbolic training. The data with the age group from 25-30 are taken to test the Fuzzy Decision Tree [22].

### 3.2.2. General Process of Training

The general process of training followed by our cellular system CASI (Cellular Automaton for Symbolic Inference) is organized in three stages :

1. Boolean generation and optimization of the decision tree by the cellular automaton (BOG-CASI);

2. Fuzzy conjunctive rules inference by the cellular automaton (BIE-CASI);

3. Validation by the cellular automaton (BV-CASI).

Figure 3 summarizes the general diagram of our system CASI.

**Boolean Generation and Optimization (BOG) of the Decision Tree :** In this section, we present the principles of construction, by boolean modelling [15] [23] [28] [29] [30] of induction decision tree in the problems of discrimination and classification [3] [28]: we want to explain the class taken by one variable to predict categorical Y, attribute class or endogenous variable; from a series of p variables $X = \{X_1, X_2,...,X_p\}$ denoted variable predictive (descriptors) or exogenous, discrete or continuous. According to the terminology of machine learning, we are therefore in the context of supervised learning.

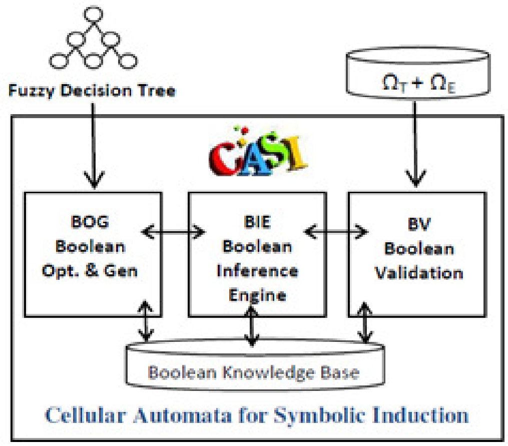


Figure 3. Cellular Automaton for Symbolic Inference (CASI)

From the sample $\Omega_a$ we begin the symbolic treatment for the construction of the decision tree (method ID3).

1. Initialize the parameters and the initial partition $S_0$;

2. Use the ID3 method to pass of partition $S_i$ to $S_{i+1}$ and generate the decision tree.

3. Finally, generation of prediction rules.

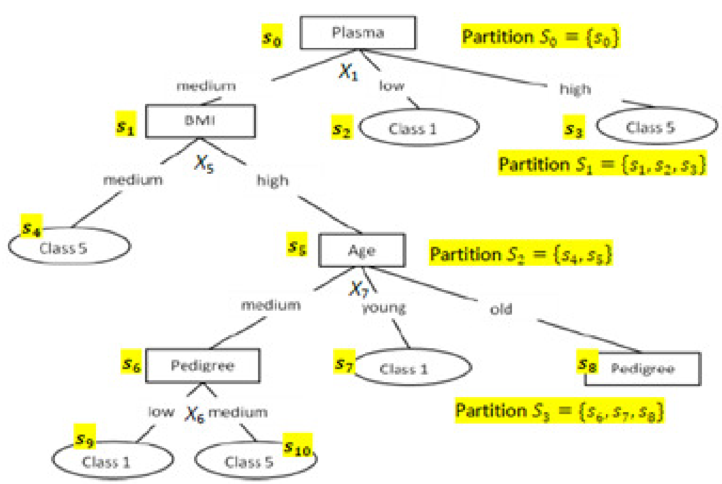


Figure 4. Example of a fuzzy decision tree

The initial partition $S_0$ has one noted $s_0$ element, which includes the entire sample learning. The next partition $S_1$ is generated by the variable $X_1$ after fuzzification and individuals in each node $s_j$ are defined as follows: $s_1 = \{\omega \in \Omega a \ / \ X_1(\omega) = \text{medium}\}$ ; $s_2 = \{\omega \in \Omega a \ / \ X_1(\omega) = \text{low}\}$ and $s_3 = \{\omega \in \Omega a \ / \ X_1(\omega) = \text{high}\}$.

As well as in the $s_0$ node, there are in $s_1$, $s_2$ and $s_3$, individuals of the classes$\{c_1, c_2\}$. The figure 5 summarizes the steps of construction of $s_0$, $s_1$, $s_2$ and $s_3$. The $S_1$ partition, the process is repeated looking for a $S_2$ score which would be better. We use the three arcs $A_1$, $A_2$ and $A_3$ to reach the vertices $s_1$, $s_2$ and $s_3$. Similarly $A_4$ and $A_5$ to reach the vertices $s_4$ and $s_5$.

To illustrate the architecture and the operating principle of the BOG module, we consider figure 4 with the $S_0$={$s_0$} partitions and $S_1$={$s_1$, $s_2$, $s_3$}. Figure 5 shows how the knowledge extracted from this graph database is represented by the *CELFACT* and *CELRULE* layers.

| *CELFACT* | *EF* | *IF* | *SF* |
|---|---|---|---|
| $s_0$ | 1 | 1 | 0 |
| $X_1 = medium$ | 0 | 1 | 0 |
| $s_1$ | 0 | 1 | 0 |
| $X_1 = low$ | 0 | 1 | 0 |
| $s_2$ | 0 | 1 | 0 |
| $X_1 = high$ | 0 | 1 | 0 |
| $s_3$ | 0 | 1 | 0 |
| $X_5 = medium$ | 0 | 1 | 0 |
| $s_4$ | 0 | 1 | 0 |
| $X_5 = high$ | 0 | 1 | 0 |
| $s_5$ | 0 | 1 | 0 |

| *CELRULE* | *ER* | *IR* | *SR* |
|---|---|---|---|
| $A_1$ | 0 | 1 | 1 |
| $A_2$ | 0 | 1 | 1 |
| $A_3$ | 0 | 1 | 1 |
| $A_4$ | 0 | 1 | 1 |
| $A_5$ | 0 | 1 | 1 |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

Figure 5. Boolean partitions modeling $S_0$ and $S_1$

Initially, all entries in cells in the *CELFACT* layer are passive (*EF* = 0), except for those who represent the initial basis of facts (EF= 1). In the case of an induction decision tree, *IF* = 0 corresponds to a Fact of the type node (*si*), *IF*=1 corresponds to a Fact of the type attribute = *value* ($X_1$=*medium*, for example).

In figure 6 are, respectively, represented the impact of input matrices $R_E$ and exit $R_S$ the Boolean model.

| $R_E$ | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $A_5$ |
|---|---|---|---|---|---|
| $s_0$ | 1 | 1 | 1 | | |
| $X_1 = medium$ | | | | | |
| $s_1$ | | | | 1 | 1 |
| $X_1 = low$ | | | | | |
| $s_2$ | | | | | |
| $X_1 = high$ | | | | | |
| $s_3$ | | | | | |
| $X_5 = medium$ | | | | | |
| $s_4$ | | | | | |
| $X_5 = high$ | | | | | |
| $s_5$ | | | | | |

| $R_S$ | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $A_5$ |
|---|---|---|---|---|---|
| $s_0$ | | | | | |
| $X_1 = medium$ | 1 | | | | |
| $s_1$ | 1 | | | | |
| $X_1 = low$ | | 1 | | | |
| $s_2$ | | 1 | | | |
| $X_1 = high$ | | | 1 | | |
| $s_3$ | | | 1 | | |
| $X_5 = medium$ | | | | 1 | |
| $s_4$ | | | | 1 | |
| $X_5 = high$ | | | | | 1 |
| $s_5$ | | | | | 1 |

Figure 6. Input/Output incidences matrices

The relationship entry, denoted $iR_E j$, is formulated as follows [28][30]:

$$\forall i=\{1,...,l\}, \forall j=\{1,...,r\}, \text{ if (the fact } i \in \text{ to the premise of the rule } j\text{) then } R_E(i,j) \leftarrow 1.$$

The relationship of output, denoted $i\, R_S\, j$ , is formulated as follows:

$$\forall i=\{1,...,l\}, \forall j=\{1,...,r\}, \text{ if (the fact } i \in \text{ to the conclusion of the rule } j\text{) then } R_S(i,j) \leftarrow 1.$$

Incidence matrices $R_E$ and $R_S$ represent the relationship input/output of the facts and are used in forward-chaining [23] [28]. You can also use $R_S$ as relationship of input and $R_E$ as relationship of output to run a rear chaining inference. Note that no cell in the vicinity of a cell that belongs to CELFACT (at CELRULE) does not belong to the layer CELFACT (at CELRULE).

The dynamics of the cellular automaton CASI [28][30], to simulate the operation of an Inference engine uses two functions of transitions $\delta_{fact}$ and $\delta_{rule}$, where $\delta_{fact}$ corresponds to the phase of assessment, selection and filtering, and $\delta_{rule}$ corresponds to the execution phase [15] [28]. To set the two functions of transition we will adopt the following notation: *EF*, *IF* and *SF* to designate CELFACT: *E*, *I* and *S*; Respectively *ER*, *IR* and *SR* to designate CELRULE: *E*, *I* and *S*.
The transition function $\delta_{fact}$:

$$(EF, IF, SF, ER, IR, SR) \xrightarrow{\delta_{fact}} \left(EF, IF, \boldsymbol{EF}, \boldsymbol{ER} + \left(\boldsymbol{R_E^T} \cdot \boldsymbol{EF}\right), IR, SR\right)$$

The transition function $\delta_{rule}$ :

$$(EF, IF, SF, ER, IR, SR) \xrightarrow{\delta_{rule}} (\boldsymbol{EF} + (\boldsymbol{R_S} \cdot \boldsymbol{ER}), IF, SF, ER, IR, \overline{\boldsymbol{ER}})$$

In order to illustrate decision tree optimization and rules generation by the cellular method using BOG, Figure 4 shows some possible useless splitting cases. The majority class is associated with each terminal node in the decision tree. We obtain as many rules as there are terminal nodes and, in each rule, as many conjunctions as there are branches back to the root.

In knowledge discovery from database, the rules are generated from a training sample and have a double objective of characterizing the classes of concepts, and assigning a class to an example whose class is unknown. In the production rules which we wish to generate, the condition is a conjunction of elementary propositions made of an attribute, an operator (=, ≥, ≠, ...) and an attribute value. The conclusion consists of a particular proposition where the attribute relates to the class (for example diabetic or not). It is possible to associate with each rule a coefficient which defines the certainty, or probability, with which a class is predicted. After data exploration, the cellular automaton assists the Fuzzy ID3 method to generate a decision tree. This graph is represented using only $R_E$ because, for such a type of graph, the output matrix $R_S$ is elementary and does not even require an internal representation.

**Boolean Inference Engine (BIE) :** To automatically generate conjunctive rules we use same $\delta_{fact}$ and $\delta_{rule}$ transition functions with the permutation of input matrices $R_E$ and exit $R_S$. We suppose that all the facts of the form $X_1$=*value* are established (EF=1). Going from the terminal nodes back to the root, $s_0$, and launching the cellular inference engine (BIE) in back chaining with a depth asynchronous mode imposed by the form of $R_S$. At the end of the symbolic training by the Fuzzy ID3 method (Fispro), we can generate the fuzzy rules coming from the fuzzy decision tree. Let us consider the figure 7 as if it was a final rules base. At that point, we can deduce five prediction rules $R_1$, $R_2$, $R_3$, $R_4$ and $R_5$ that have the form if condition then conclusion, where condition is a logical expression in conjunctive form and conclusion is the majority class in the node reached by the condition. For example, in figure 4, the majority class of $s_2$ is *very low* (class 1), but the majority class of $s_3$ is *very high* (class 5).

**Generation of Conjunctive Rules :** We proceed in the same way with the decision tree generated by Fispro and we obtain the following conjunctive rules (Figure 7):

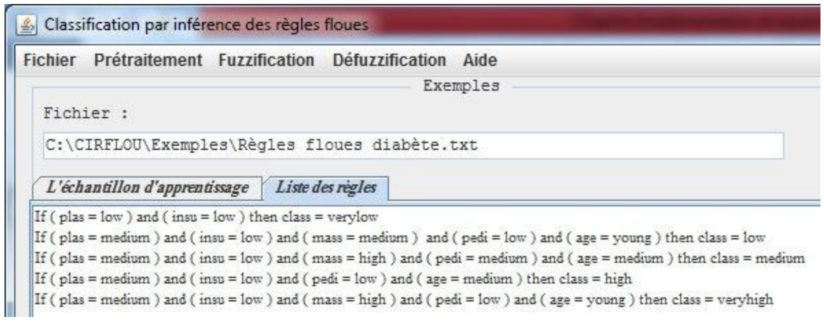


Figure 7. Generation of conjunctive rules

The representation of this knowledge base by the cellular machine is illustrated in figure 8.

| CELFACT | EF | IF | SF |
|---|---|---|---|
| $X_1$ = low | 0 | 1 | 0 |
| $X_1$ = medium | 0 | 1 | 0 |
| $X_1$ = high | 0 | 1 | 0 |
| $X_4$ = low | 0 | 1 | 0 |
| $X_4$ = medium | 0 | 1 | 0 |
| $X_4$ = high | 0 | 1 | 0 |
| $X_5$ = low | 0 | 1 | 0 |
| $X_5$ = medium | 0 | 1 | 0 |
| $X_5$ = high | 0 | 1 | 0 |
| $X_6$ = low | 0 | 1 | 0 |
| $X_6$ = medium | 0 | 1 | 0 |
| $X_6$ = high | 0 | 1 | 0 |
| $X_7$ = young | 0 | 1 | 0 |
| $X_7$ = medium | 0 | 1 | 0 |
| $X_7$ = old | 0 | 1 | 0 |
| $Y$ = very low | 0 | 1 | 0 |
| $Y$ = low | 0 | 1 | 0 |
| $Y$ = medium | 0 | 1 | 0 |
| $Y$ = high | 0 | 1 | 0 |
| $Y$ = very high | 0 | 1 | 0 |

| CELRULE | ER | IR | SR |
|---|---|---|---|
| $R_1$ | 0 | 1 | 1 |
| $R_2$ | 0 | 1 | 1 |
| $R_3$ | 0 | 1 | 1 |
| $R_4$ | 0 | 1 | 1 |
| $R_5$ | 0 | 1 | 1 |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

| $R_E$ / $R_S$ | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ |
|---|---|---|---|---|---|
| $X_1$ = low | 1 | | | | |
| $X_1$ = medium | | 1 | 1 | 1 | 1 |
| $X_1$ = high | | | | | |
| $X_4$ = low | 1 | 1 | 1 | 1 | 1 |
| $X_4$ = medium | | | | | |
| $X_4$ = high | | | | | |
| $X_5$ = low | | | | | |
| $X_5$ = medium | | 1 | | | |
| $X_5$ = high | | | 1 | | 1 |
| $X_6$ = low | | 1 | | 1 | 1 |
| $X_6$ = medium | | | 1 | | |
| $X_6$ = high | | | | | |
| $X_7$ = young | | 1 | | | 1 |
| $X_7$ = medium | | | 1 | 1 | |
| $X_7$ = old | | | | | |
| $Y$ = very low | 1 | | | | |
| $Y$ = low | | 1 | | | |
| $Y$ = medium | | | 1 | | |
| $Y$ = high | | | | 1 | |
| $Y$ = very high | | | | | 1 |

Figure 8. Boolean knowledge base of the figure 7

**Boolean Validation (BV) :** Upon completion of this process, the cellular machine is ready to launch the validation phase. By using the same guiding principle of an inference engine and the same $\delta_{fact}$ and $\delta_{rule}$ transition functions (figure 9), the cellular automaton advances from a configuration to the next, for finally generating the set Ωe [3].

| *CELFACT* | *EF* | *IF* | *SF* |
|---|---|---|---|
| $X_1 = low$ | 0 | 1 | 0 |
| $X_1 = medium$ | 0 | 1 | 0 |
| $X_1 = high$ | 0 | 1 | 0 |
| $X_4 = low$ | 0 | 1 | 0 |
| $X_4 = medium$ | 0 | 1 | 0 |
| $X_4 = high$ | 0 | 1 | 0 |
| $X_5 = low$ | 0 | 1 | 0 |
| $X_5 = medium$ | 0 | 1 | 0 |
| $X_5 = high$ | 0 | 1 | 0 |
| $X_6 = low$ | 0 | 1 | 0 |
| $X_6 = medium$ | 0 | 1 | 0 |
| $X_6 = high$ | 0 | 1 | 0 |
| $X_7 = young$ | 0 | 1 | 0 |
| $X_7 = medium$ | 0 | 1 | 0 |
| $X_7 = old$ | 0 | 1 | 0 |
| $Y = very\ low$ | 0 | 1 | 0 |
| $Y = low$ | 0 | 1 | 0 |
| $Y = medium$ | 0 | 1 | 0 |
| $Y = high$ | 0 | 1 | 0 |
| $Y = very\ high$ | 0 | 1 | 0 |

| *CELRULE* | *ER* | *IR* | *SR* |
|---|---|---|---|
| $R_1$ | 0 | 1 | 1 |
| $R_2$ | 0 | 1 | 1 |
| $R_3$ | 0 | 1 | 1 |
| $R_4$ | 0 | 1 | 1 |
| $R_5$ | 0 | 1 | 1 |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

| $w_1$ | $R_E$ / $R_S$ | $R_1$ | $R_2$ | $R_3$ | $R_4$ | $R_5$ |
|---|---|---|---|---|---|---|
| | $X_1 = low$ | 1 | | | | |
| 1 | $X_1 = medium$ | | 1 | 1 | 1 | 1 |
| | $X_1 = high$ | | | | | |
| 1 | $X_4 = low$ | 1 | 1 | 1 | 1 | 1 |
| | $X_4 = medium$ | | | | | |
| | $X_4 = high$ | | | | | |
| | $X_5 = low$ | | | | | |
| | $X_5 = medium$ | | 1 | | | |
| 1 | $X_5 = high$ | | | 1 | | 1 |
| | $X_6 = low$ | | 1 | | 1 | 1 |
| 1 | $X_6 = medium$ | | | 1 | | |
| | $X_6 = high$ | | | | | |
| | $X_7 = young$ | | 1 | | | 1 |
| 1 | $X_7 = medium$ | | | 1 | 1 | |
| | $X_7 = old$ | | | | | |
| | $Y = very\ low$ | 1 | | | | |
| | $Y = low$ | | 1 | | | |
| | $Y = medium$ | | | 1 | | |
| | $Y = high$ | | | | 1 | |
| | $Y = very\ high$ | | | | | 1 |

Figure 9. Boolean validation

## 3.3. FUZZY BOOLEAN MODELING

According to [5], founder of fuzzy logic, the limits of the classical theories applied in artificial intelligence come because they require and manipulate only accurate information. Fuzzy logic provides approximate reasoning modes rather than accurate. It is mainly the mode of reasoning used in most cases in humans.

### 3.3.1. BOOLEAN FUZZIFICATION OF EXOGENOUS VARIABLES

Fuzzy-BML modelling deals with the fuzzy input variables and provides results on output variables themselves blurred. Fuzzification, illustrated by the following example, is the step that consists of fuzzy quantification of actual values of a language variable. Fuzzifier to: the universe of discourse, i.e. a range of possible variations of the corresponding entry. A partition interval fuzzy from this universe, for the identification of the cost we partitioned space of variables to 7 with a Boolean modeling on 3 bits of 000 to 110 (Figure 10), finally, the duties of membership classes.

Valeurs floues | Paramètres flous

| | 000 | 000 | 001 | 001 | 010 | 010 | 011 | 011 |
|---|---|---|---|---|---|---|---|---|
| PLAS = LOW | 0 | 17.344 | 17.344 | 34.688 | 34.688 | 52.032 | 52.032 | 69.376 |
| PLAS = MEDIUM | 88.335 | 99.478 | 99.478 | 110.621 | 110.621 | 121.764 | 121.764 | 132.906 |
| PLAS = HIGH | 121.408 | 132.493 | 132.493 | 143.577 | 143.577 | 154.662 | 154.662 | 165.746 |
| INSU = LOW | 0 | 24.739 | 24.739 | 49.479 | 49.479 | 74.218 | 74.218 | 98.957 |
| INSU = MEDIUM | 17.276 | 39.879 | 39.879 | 62.482 | 62.482 | 85.085 | 85.085 | 107.688 |
| INSU = HIGH | 173.175 | 269.293 | 269.293 | 365.411 | 365.411 | 461.529 | 461.529 | 557.646 |
| MASS = LOW | 0 | 3.970 | 3.970 | 7.941 | 7.941 | 11.911 | 11.911 | 15.881 |
| MASS = MEDIUM | 0 | 5.552 | 5.552 | 11.104 | 11.104 | 16.656 | 16.656 | 22.208 |
| MASS = HIGH | 27.792 | 33.407 | 33.407 | 39.023 | 39.023 | 44.638 | 44.638 | 50.254 |

Figure 10. Boolean modeling of the fuzzy triangular numbers.

### 3.3.2. BOOLEAN DEFUZZIFICATION

Output the Fuzzy-BML modeling cannot communicate to the user of the fuzzy values. The role of the defuzzification is therefore to provide accurate values.

During this step, the system will perform tests to define the range of proven goal. This test will depend on the number of rules candidates and the de facto number of each rule that participated in the inference according to the following principle illustrated by figure 11 :

- Cases for a single rule and a single fact:
  if (*fact*) then (*conclusion*).
  $CELFACT_{IF}(conclusion) = Minimum(CELFACT_{IF}(fact), CELRULE_{IR}(rule))$

- Cases for a single rule with several facts:
  if ($fact_1$) and ($fact_2$) and ($fact_3$) ... then (*conclusion*).
  $CELFACT_{IF}(conclusion) = Minimum(CELFACT_{IF}(fact_1), CELFACT_{IF}(fact_2),...)$
  The *'Minimum'* operator in Boolean logic represents the logical AND.

- Several rules :
  $CELFACT_{IF}(goal) = Maximum(CELRULE_{IR}(rule_1), CELRULE_{IR}(rule_2),...)$
  The *'Maximum'* operator in Boolean logic represents the logical OR.

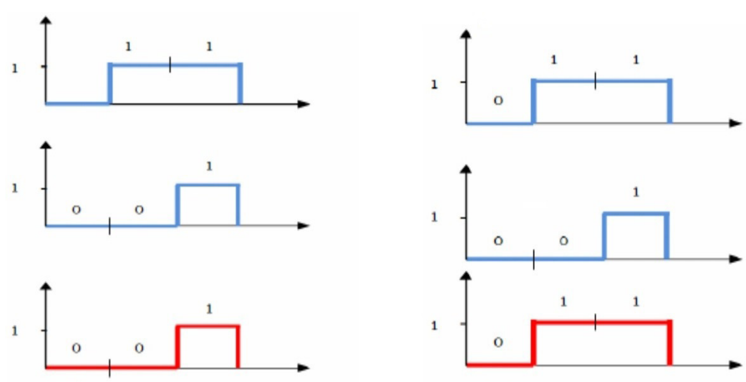

Figure 11. Boolean operator for the defuzzification

## 4. EXPERIMENTATION

We have evaluated our approach on a case basis of Diabetes. Diabetes is a disease in which the body does not properly process glucose or sugar. We exploited the Pima Indians Diabetes dataset of the UCI database library. This dataset consists of 768 instances characterized by 8 descriptors: During this step, the system will perform tests to define the range of proven goal. This test will depend on the number of rules candidates and the de facto number of each rule that participated in the inference according to the following principle:

$X_1$*: Number of times pregnant;*
$X_2$*: Plasma glucose concentration a 2 hours in an oral glucose tolerance test;*
$X_3$*: Diastolic blood pressure;*
$X_4$*: Triceps skin fold thickness;*
$X_5$*: 2-Hour serum insulin;*
$X_6$*: Body mass index;*
$X_7$*: Diabetes pedigree function;*
$X_8$*: Age.*

We compare the proposed approach Fuzzy-BML with k-NN [31] and decision tree on the same case base. We show in Table 4 the rate of correctly classified instances with each method using the supervised mode of discretization.

Table 4. Results of experimentation.

| k-NN | Decision tree | Fispro Fuzzy-DT [13] | Fuzzy-BML |
|---|---|---|---|
| 66% | 73% | 81% | 81% |

The rate of correctly classified instances is 66% with k-NN, 73% with decision tree, 81% with Fuzzy-DT and Fuzzy-BML-CBR. From the obtained results, we note that the Fuzzy-BML-CBR method has provided better results with a rate of 81% of well classified instances. In this paper Fuzzy BML approach has been applied to optimize response time and the storage complexity.

## 5. CONCLUSIONS AND PERSPECTIVES

Several competing motivations have led us to define a Boolean model for CBR knowledge base systems. Indeed, we have not only desired experiment with a new approach to indexing of cases by decision tree, but we also wanted to improve modeling of the vague and uncertain of the natural language concepts, optimize response time and the storage complexity.

For the calculation of the similarity in the retrieval (cases indexing) phase, typically used k-nearest neighbours. So we compared our Fuzzy Boolean Model with k-nearest neighbours (k-NN) and decision tree. We noticed that the indexing of cases for the choice of a plan is significantly better with Fuzzy-BML. Finally, we can say that the structure of the cases that we have used is quite simple. We have described the part problem of cases by age, weight and an antecedent. By adding other constraints could subsequently used a slightly more complex representation.

As a future perspective of this work, we propose to improve the other steps of the CBR process for the proposed approach.

## REFERENCES


[1] Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. (1984) 'Classification and regression and trees', Technical report, Wadsworth International, Monterey, CA.

[2] Kodratoff, Y. (1997) 'The extraction of knowledge from data, a new topic for the scientific research', Magazine electronic READ.

[3] Atmani, B. & Beldjilali, B. (2007) 'Neuro-IG: A Hybrid System for Selection and Elimination of Predictor Variables and non Relevant Individuals', Informatica, Vol. 18, No. 2, pp163-186.

[4] Fayyad, U., Shapiro, G.P. & Smyth, P. (1996) 'The KDD process for extraction useful knowledge from volumes data', Communication of the ACM.

[5] Zadeh, L.A. (1997) 'Some reflections on the relationship between AI and fuzzy logic: A heretical view', Workshop on Fuzzy Logic in Artificial Intelligence, pp1-8.

[6] Zadeh, L.A. (1968) 'Probability measures of fuzzy events', Journal of Mathematical Analysis and Applications, Vol. 23, pp421-427.

[7] Zighed, D.A. & Rakotomalala, R. (2000) Graphs of induction, Training and Data Mining, Hermes Science Publication, pp21-23.

[8] Boyen, X. & Wehenkel, L. (1999) 'Automatic induction of fuzzy decision trees and its application to power system security assessment', Fuzzy Sets and Systems, Vol. 102, pp3-19.

[9] Begum, S., Ahmed, M.U., Funk, P., Xiong, N. &Von Schéele, B. (2009) 'A case-based decision support system for individual stress diagnosis using fuzzy similarity matching', Computational Intelligence, Vol. 25, pp180-195.

[10] Barrientos, F. & Sainz, G. (2012) 'Interpretable knowledge extraction from emergency call data based on fuzzy unsupervised decision tree', Knowledge-Based Systems, Vol. 25, pp77-87.

[11] Levashenko, V. & Zaitseva, E. (2012) 'Fuzzy decision trees in medical decision making support system', Proceedings of the Federated Conference on Computer Science and Information Systems, pp213-219.

[12] Adidela, D.R., Devi, L., Suma, J. & Allam, A.R. (2012) 'Application of fuzzy Id3 to predict diabetes', International Journal of Advanced Computer and Mathematical Sciences, Vol. 3, No. 4, pp541-545.

[13] Benamina, M., Atmani, B., & Benbelkacem, S. (2018) 'Diabetes Diagnosis by Case-Based Reasoning and Fuzzy Logic'. International Journal of Interactive Multimedia and Artificial Intelligence, (In Press).

[14] Amrani, F., Bouamrane, K., Atmani, B., & Hamdadou, D. (2011). 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.

[15] Barigou, F., Atmani, B. & Beldjilali, B. (2012) 'Using a cellular automaton to extract medical information from clinical reports', Journal of Information Processing Systems, Vol. 8, No. 1, pp67-84.

[16] Atmani, B., Benbelkacem, S. & Benamina, M. (2013) 'Planning by case-based reasoning based on fuzzy logic', First International Conference on Computational Science and Engineering, Dubai, UAE, pp53-64.

[17] Brahami, M., Atmani, B. & Matta, N. (2013) 'Using rules to enhance evolution of knowledge mapping: Application on Healthcare', International Journal of Computer Science Issues, Vol. 10, No. 3, pp261-270.

[18] Benfriha, H., Barigou, F. & Atmani, B. (2013) 'Lattice-cell: Hybrid approach for text categorization', Second International Conference on Software Engineering and Applications, Dubai, UAE.

[19] Aissani, N., Atmani, B., Trentesaux, D. & Beldjilali, B. (2014) 'Extraction of priority rules for boolean induction in distributed manufacturing control', In Borangiu, T, Trentesaux, D. and Thomas, A. (Ed.), 'Service orientation in holonic and multi-agent manufacturing and robotics', Studies in computational intelligence, Vol. 544, Springer, pp127-144.

[20] Guillaume, S. & Charnomoridc B. (2011) 'Learning interpretable fuzzy inference systems with FisPro', Information Sciences, Elsevier, Vol. 181, No. 20, pp4409-4427.

[21] Bello-Tomas, J., Gonzalez-Calero, P. & Diaz-Agudo, B. (2004) 'Jcolibri: An object oriented framework for building CBR systems', Advances in Case-Based Reasoning, 7th European Conference on Case-Based Reasoning, pp32–46.

[22] Kalpana, M. & Kumar A. (2011) 'Fuzzy expert system for diabetes using fuzzy verdict mechanism', International Journal of Advanced Networking and Applications, Vol. 3, No. 2, pp1128-1134.

[23] Beldjilali, A. & Atmani, B. (2009) 'Identification du type de diabète par une approche cellulo-floue', Quatrième atelier sur les systèmes décisionnels, Jijel, Algeria, pp203-218.

[24] Quinlan, J.R. (1986) 'Induction of decision trees', Machine Learning, Vol. 1, pp81-106.

[25] Weber, R. (1992) 'Fuzzy-Id3: A class of methods for automatic knowledge acquisition', 2nd International conference on fuzzy logic and neural networks, pp265-268.

[26] Alonso, J.M., Cordon, O., Quirin, A. & Magdalena, L. (2011) 'Analyzing interpretability of fuzzy rule-based systems by means of fuzzy inference-grams', World Conference on Soft Computing, San Francisco, USA.

[27] Zulj, S., Seketa, G., Dzaja, D., Sklebar, F., Drobnjak, S., Celic, L. & Magjarevic R. (2017) 'Supporting diabetic patients with a remote patient monitoring systems', In: Torres I., Bustamante J. and Sierra D. (Ed.) VII Latin American Congress on Biomedical Engineering CLAIB 2016, Bucaramanga, Santander, Colombia, October 26th -28th, 2016, IFMBE Proceedings, Vol. 60, Springer, Singapore.

[28] Atmani, B. & Beldjilali, B. (2007) 'Knowledge discovery in database: Induction graph and cellular automaton', Computing and Informatics Journal, Vol. 26, No. 2, pp171-197.

[29] Benamina, M. & Atmani, B. (2008) 'WCSS: un système cellulaire d'extraction et de gestion des connaissances', Troisième atelier sur les systèmes décisionnels, Mohammadia, Morocco, pp223-234.

[30] Brahami, M., Atmani, B. & Matta, N. (2013) 'Dynamic knowledge mapping guided by data mining: Application on healthcare', Journal of Information Processing Systems, Vol. 9, No. 1, pp01-30.

[31] Guo, G., Wang, H., Bell, D., Bi, Y. & Greer, K. (2003) 'Knn model-based approach in classification', International Conference on Ontologies, Databases and Applications of Semantics, pp986– 996.

## AUTHORS

**Mohamed Benamina** is currently a PhD candidat at the University of Oran 1 Ahmed Ben Bella and affiliated researcher in Laboratoire d'Informatique d'Oran, Algeria. He received his Master degree in Computer Science in 2010 from University of Oran, Algeria. His research interests include data mining algorithms, expert systems and decision support systems.

**Baghdad Atmani** received his PhD in computer science from the University of Oran (Algeria) in 2007. He is currently a full professor in computer sciences. His interest field is artificial intelligence and machine learning. His research is based on knowledge representation, knowledge-based systems, CBR, data mining, expert systems, decision support systems and fuzzy logic. His research are guided and evaluated through various applications in the field of control systems, scheduling, production, maintenance, information retrieval, simulation, data integration and spatial data mining.

**Sofia Benbelkacem** is currently a PhD candidate at the University of Oran 1 Ahmed Ben Bella and affiliated researcher in Laboratoire d'Informatique d'Oran, Algeria. Her research interests include data mining, planning, case-based reasoning, medical decision support systems, machine learning.