Applying BN in CBR Adaptation-Guided Retrieval for Medical Diagnosis

Akila Djebbar[1] and Hayet Farida Merouani[2]

Computer Science Department, LRI laboratory, SRF equip,
Badji Mokhtar University, BP 12, Annaba 23000, Algeria
[1]aki_djebbar@yahoo.fr, [2]hayet_merouani@yahoo.fr

Abstract

The Case Based Reasoning (CBR) is an approach of solving problem which is based on the reuse, by analogy, of past experiences called case. It is based on the retrieval and adaptation of the old solutions to the new problems. This paper presents a Bayesian adaptation-Guided Retrieval phase for a CBR applied to the diagnosis of hepatic pathologies. The main idea consists in a modelling the case base by a Bayesian Network (BN). Its are excellent tools for modelling the uncertainty in terms of their clear graphic representation as well as the conditional probabilities laws defined on a graph. We are interested to retrieval and adaptation phases. The retrieval phase consists of selecting the most similar case of log linear model by the considering Bayesian Network as a log-linear model on the simplification of the probability. The adaptation phase means modifying solutions of retrieved cases to fit the current problem. The dependence between these two phases is defined by two measures: a similarity measure and an adaptation measure. The objective of this dependence is to guarantee the retrieved case which is the easiest to adapt and improve the performance of CBR. An example of the diagnosis of the hepatic pathologies will illustrate the presented approach.

Keywords: *Case Based Reasoning (CBR), adaptation, adaptation-guided retrieval, Bayesian Network (BN), medical diagnosis.*

1. Introduction

The Case-Based Reasoning is a methodology of resolution of problems based on the re-use of the past experiences, called "case" for the resolution of new problems. It solves problems by retrieving the most similar previous cases in a case base (*source cases*) and by re-using the knowledge and experiences from previous good quality solutions [1]. Our field of study is medical diagnosis where the uncertainly is a permanent companion in medical activity. The probabilistic models, such as Bayesian networks are largely used in domains based on decision [2]. These domains represent an uncertain knowledge of the treated domain and they are adaptable according to the variability of observations. A Bayesian Network [3] is a graphical representation used to model uncertainty. It captures causal and dependence relationships among the variables. The network structure and probabilistic values associated to variables are used with information obtained via CBR. A CBR system has to solve two main tasks [4]: The first one is the retrieval, which is the search for or the calculation of most similar cases. For this task much research has been undertaken. So, actually it has become correspondingly easy to find sophisticated CBR retrieval algorithms adequate for nearly every sort of application problem. The second task, the adaptation step must recognize differences between the new and retrieved problems, and refine the retrieved solution to reflect these

differences, as appropriate [5]. It means modifying solutions of former similar cases to fit a current problem. If there are no important differences between a current and a similar case, a solution transfer is sufficient. Since adaptation is even more difficult in medicine, we want to elaborate typical medical adaptation problems and we hope to show possibilities how to solve them [6].

In this work, we present the principle of the case base to search for similar cases in the target case. An original algorithm of retrieval and adaptation are proposed. The retrieval phase which consists of selecting for the most similar case via a log-linear model which models the structure of the case base in order to facilitate and improve the retrieval phase. After this phase, we describe our adaptation phase, it consist to select among the retrieved cases which one is the most adaptable. The rest of this paper is structured as follows. Section *2* presents related works. Section *3* describes the CBR cycle. Section *4* describes the presentation of Bayesian Network. In section *5* we describe the integration of Bayesian Network in CBR. In Section 6 we describe our proposed retrieval phase by defining the similarity metrics, therefore we present the approximate case retrieval algorithm. In Section *7* we illustrate our proposed adaptation phase, a discussion of our proposed approach in Section 8, Section 9 provides conclusion and the future research directions.

2. Related works

There are several ways to integrate BNs in CBR. However, a Bayesian retrieval in a CBR system and the use of log linear similarity measures in the retrieval step is the main contributions of this work. Below we have identified some works, which are presented with a brief description of how they combine BN and CBR. Such as Bayesian Case Reconstruction (BCR) [7] is a system for design of screening experiments for Macromolecular Crystallization. BCR is used to expand the coverage of the case base library. The BN is constructed from the domain experts and the content of the case library. The Retrieve phase selects the most probable cases, and they are disassembled in order to form new solutions. The Bayesian network contains the causal relations from the domain model that are well understood. In the Reuse phase, the BN is used to find the most probable new solutions. The result is a plausible solution, only. Another system that combines BN and CBR is used to choose the optimal parameters for an algorithm used in different domains [8]. The BN is learned and evaluated through experiments in the domains with the algorithms using different parameter settings. The retrieve phase selects the most similar cases. The reuse phase is used to calculate a reliability measure in addition to calculate the most probable arguments from the BN. The most reliable cases are those who have a high number of experiments and a large number of variations in the parameters used. [9] presents a technique that integrates Case-Based Reasoning and Bayesian Networks to build user profiles incrementally. Case-Based Reasoning provides a mechanism to acquire knowledge about user actions that are worth recording to determine his habits and preferences. Bayesian Networks provide a tool to model quantitative and qualitative relationships between items of interest. Information needed to build the BN is taken from cases stored in the case base. [10] examines probabilistic case-based reasoning by integrating Bayesian networks with CBR and proposes a probabilistic CBR framework for obesity prescription management (PCOPM) to assist health professionals to share their experiences of obesity exercise prescription online. The PCOPM ties together CBR and BN into a unified framework that includes both obesity experience and intelligent embodiment of decision making for obesity management. [11] focuses on the construction of diagnosis support tools for Acute Bacterial Meningitis ABM, reporting a comparative assessment of the quality of a Clinical Decision Support System (CDSS) resulting from the

application of CBR, to that of an existing CDSS system developed using a Bayesian expert system.

3. CBR Cycle

Case-based reasoning (CBR) [1] solves problems by retrieving the most similar previous cases in a case base (*source cases*) and by re-using the knowledge and experiences from previous good quality solutions. If necessary, the retrieved solutions are adapted by using domain knowledge so that they are applicable for the new problem. The case base is then updated by the new learned cases.
A general CBR cycle may be described by five processes [1] (Figure 1):

1- **Retrieves** case(s) most similar to the specified problem description.

2- **Reuse** the knowledge in the retrieved case(s) to solve the problem.

3- **Revise** the solution obtained from Step 2.

4- **Retain** the parts of this experience likely to be useful for future problem solving.

Generally "Reuse" and "Revise" are together referred to as the Case Adaptation process. Case adaptation is required for a CBR system because the previous occurrence problem definition does not usually match with the new problem definition.

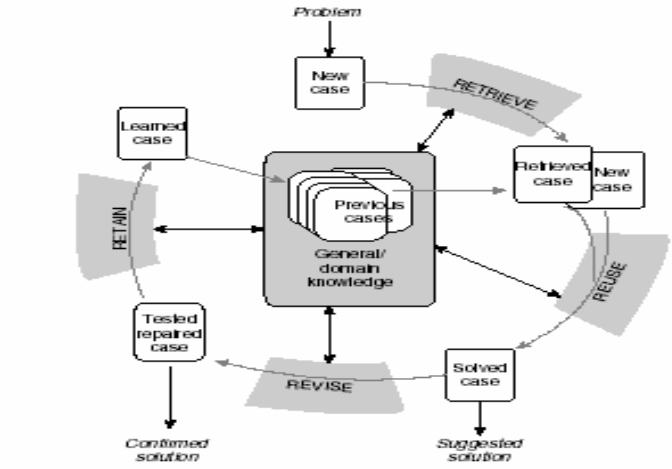

Figure 1. CBR Cycle Form

4. Presentation of Bayesian network

4.1. Definition of Bayesian Network

Bayesian network (BN) is one of the graph representations for modeling uncertainty. A Bayesian Network can depict a relationship between a disease and its symptoms, machine failure and its root causes or predict the most probable output passed through a noisy channel. In other words BNs are a structured, compact directed acyclic graphs (DAGs) which illustrate a problem that is too large or complex to be represented with tables or equations [3, 12].

A BN is a graph in which nodes represent random variables, and the links of the influences between variables. The graph is acyclic: it does not contain a buckle. Arrows represent relations between variables which are either determinists, or probability [3].

A Bayesian Network is defined by:

- A directed graph without circuit (DAG) G = (V, E), where V is all the nodes of G, and E is all the arcs of G;
- A space finished probabilities (Ω, Z, p);
- A set of random variables associated with the nodes of the graph and defined on (Ω, Z, p), such as:

$$p(V1, V2, \dots, Vn) = \prod_{i=1}^{n} p(Vi|C(Vi)) \quad (1)$$

Where C (V_i) is all the causes (parents) of V_i in the graph G. It is very exactly what we built on both examples above.

A BN is a graphic representation of the joint probability distribution P (x_1, x_2. x_n) .The Figure 2 presents a Bayesian network which models in a reasonable way the process of diagnosis of two diseases of the liver. The way of reasoning is represented here by a diagram of causality (BN). The probability is determined by the expert of the domain [13].

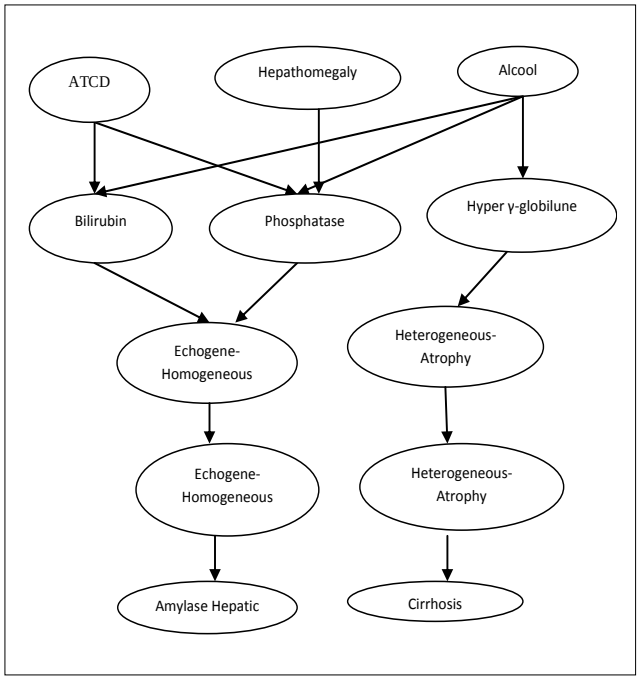

Figure 2. Example of a Bayesian Network with Two Diseases of the Liver

The construction of this network guarantees series of conditional independences verifying locally its semantics. It arises from the Algorithm 1.

Algorithm 1. Algorithm of construction of the Bayesian network.

Choose an order for these variables $X_1 \ldots X_n$
For i=1 to n **Do**
- Insert X_i to network
- Select this parents in $X_1, \ldots, X_{i-1}$ such as $P(X_i \mid Parents(X_i)) = P(X_i \mid X_1, \ldots, X_{i-1})$

End For

4.2. Inference

A Bayesian belief network is generally used for probabilistic inference, that is, for making probabilistic statements concerning the variables represented in the network. Since a belief network uniquely defines a joint probability distribution, it provides for computing any probability of interest. Several different algorithms have been designed for this purpose. In the present work, we apply the Pearl' Algorithm [3].

4.2.1 Pearl' algorithm

Pearl's basic algorithm for probabilistic inference takes the digraph of a belief network for its computational architecture. The vertices in the graph are viewed as autonomous objects and the arcs as bi-directional communication channels. Each vertex has a local processor that is capable of performing simple, pre-defined computations and a local memory in which its associated probabilities are stored. Through the communication channels the vertices send each other messages providing information about the represented probability distribution. Each vertex is able to compute the probabilities of its values from the probabilities stored in its memory and the information it receives from its neighbours [3]. When a vertex' true value becomes known, the messages this vertex sends to its neighbours are updated to reflect the evidence, forcing its neighbours to compute updated messages in turn. The impact of evidence thus spreads throughout the graph by message-passing between neighbours.

5. Integrating Bayesian Network in CBR

5.1 Network Variables Definition

In order to use the formalism of Bayesian networks, we deliberately shown the useful information for our problem representation. After several discussions with experts, we deduced that the most relevant information to be treated is divided into four levels [13]:

1. Clinical level: It consists of a set of possible factors for detecting the different hepatic illnesses. For example; alcohol, tobacco, dietetics, arterial high blood pressure (HTA), vascular abnormality, cancerological antecedent ...
2. Biological level: defines the pathology phase. It shows all possible pathology cases which may be detected from the previous level. For instance; cirrhosis, hepatic cell carcinoma, hepatic metastasis, lymphome (II) of liver, ...
3. Medical imaging level: contains the ultrasounded scan and scanning images of liver.
4. Diagnosis and therapy level: it shows the final diagnosis; hepatic cyst, hepatic metastasis abscess ...

These information are presented as discrete values variables, each variable is associated with a specific probability and a matrix that shows the assessment of the conditional probabilities.

5.2 Definition of Weight

In medicine, to diagnose a disease it is necessary to balance the symptoms to go out master symptom. For it attributing us to every level i describe previously a weight λ_i.

5.3 The Case Base Architecture

The work aims for constructing a Bayesian network to modeling the case base which consists of a set of information for the diagnosis. The network nodes represented as the variables of four levels, describe the dependence between the nodes, shown in the previous paragraph. Figure 3 shows the first step of modeling qualitatively the problem and determining the influence between the variables. Our Bayesian Network allows a representation of a qualitative and causal knowledge in addition to a quantitative knowledge that expresses the uncertainty. It contains four levels: clinical, biological, medical imaging and diagnosis-therapy. Each level has a table containing the different conditional probabilities in order to get a good final diagnosis. Each level is composed of a set of attributes; each attribute corresponds to a node of a network. Arcs describe the relationships between attributes as conditional probabilities of attributes in the case.

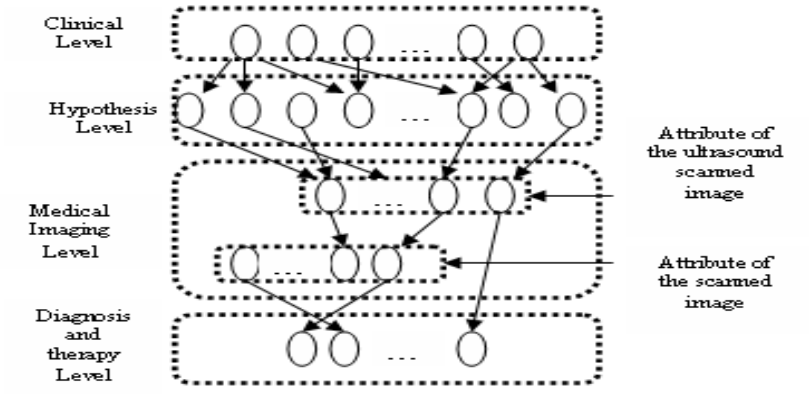

Figure 3. Modeling of the Base of Case by a Bayesian Network

5.4 Bayesian Case Description

The case is described in terms of the attribute dimension. These attributes are clinical data, hypotheses, characteristics extracted from the image of liver and the concluded diagnosis. Figure 4 shows a sample of a Bayesian case.

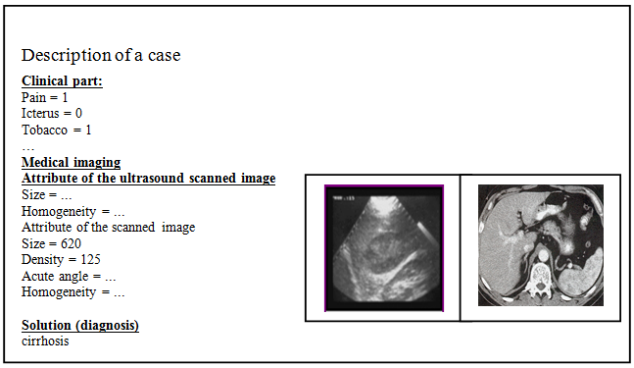

Figure 4. Bayesian Case Description

5.5 Log linear Models

The log-linear models can be also used to simplify the number of parameters of a law of conditional probability, or more generally a law of joined probability P (X_1, X_2, X_n) of a variable and its parents. The principle, very general, of these models is to decompose the logarithm of a law of probability into a sum of terms describing the interactions between variables. This decomposition is said saturated when all the terms are present in the decomposition, and not saturated when additional hypotheses are added, as for example the fact that certain variables are independent, to delete terms in the decomposition [14, 15, 16, 17].

In the case which interests us, we also know that the parents are mutually independent. Furthermore, [18] suggest keeping only the terms of interaction of order lower or equal to 2 (u, u_i, u_{0i}), arriving at the log linear model not saturated following:

$$\log P(Y, X_1, X_2, \dots, X_n) = u + \sum_i u_i(X_i) + \sum_i u_i'(X_i, Y) \quad (2)$$

Where u_i is the terms of interaction of lower order with n is number of terms. The determination of these terms of interaction passes by the resolution of a linear system, by using certain constraints as the fact as the sum of P(Y,X_1, ...,X_n) must be equal to 1. Supposing that the expert is questioned about all the marginal probability P (x_i), P (y), and on all the conditional probability P(y | x_i) and P(y|x_i). [3] show that it still stays 2^n-2n forced to satisfy to determine completely the parameters of the log-linear model. We consider the saturated log linear model. We note X_1, X_2,.., X_m, the variables of the model. These variables are discrete. The joined probability, taking into account all the possible interactions spells as follow: Our Bayesian Network establishes previously has at most n=2 parents. For this model containing a node, noted B with n similar, noted A1, it is necessary to define 2^n conditional probability and n a prior probability. If n> 5, the task of the expert becomes complex. The idea is to come down to a simple log linear model. This approach thus allows to obtain a modeling more general than both first ones, but requires more estimations on behalf of the expert when the number of parents of a variable is important.

6. The Proposed Retrieval Phase

The case base of our CBR system is a collection of vectors of case, noted C = {(c_1, p_1), (c_2, p_2). (c_l, p_l)}. Every case is described by two vectors C_k and P_k, such as C_k = (C_{k1}, C_{k2}, … C_{kn}), C_{ki} R a combination of the numerical values associated with every attribute of the case k and Pk = (P_{k1}, P_{k2}, … P_{kn}), P_{ki} [0,1] is a vector of probability which contains rough probability in attributes of the case. And as the construction of the probability in a Bayesian network is an often difficult stage, requires most part of the time of the opinions of the experts and to reduce the number of probability to ask the experts, we consider the BN as a log linear model [19].

This phase of retrieval passes by three processes (Figure 5):

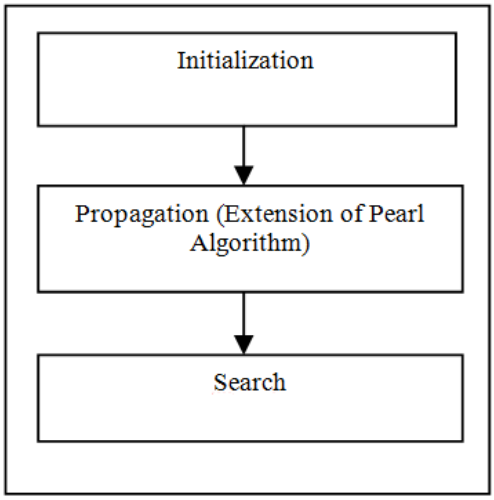

Figure 5. Steps of Retrieval Phase

6.1 Process of Initialization

This process allows to initialize the variables of the network and to attribute to every level i a weigh λ_i. These variables are the marginal probability in priori p of the first level (the clinical level) and the conditional probability of the other levels. P (S|D) is the posterior probability of the symptoms k {1..k} of the clinical, biological, medical imaging level that are given by the observation x.

6.2 Process of Propagation (Extension of Pearl Algorithm)

After variables are initialized by a set of probability, the process of distribution of message is made at each level of network until the level inferior (son) by a passage of message π which allows to entrainer for every variable an update of it probability, the message p which comes from the father towards the sons is noted for every variable as a current probability [3]. The variables of the first level the message π is calculated by the following formula:

$$\pi (x) = \lambda_i p(x) \quad (3)$$

The variables of the second level receive this message and update their probabilities for use the message π in the calculate of it current probability. A new message is obtained by the following formula:

$$\pi (X=x) = \lambda_i \sum z \, P (X=x\backslash Z=z) \, \pi x (Z=z) \quad (4)$$

Where:

π (x) is the coming message of x.

λ_i is the weight of the level.

p(x) is the marginal probability.

z is the parent.

P (X=x\Z=z) is the probability of X knowing his parents.

π x (Z=z) is the message coming of Z to his son X.

The same principle is made on the third level, for the variables of the last level, there is no message π to be sent, when all variable of the network update its probabilities in the light of the message π received, the process of distribution ends.

The result obtained at the last level allows to throw the process of retrieval which is going to supply the final result of the diagnosis.

6.3 Process of Retrieval

The retrieval is made according to the evaluation of the resemblance between the new case and the cases already known in the base. Let C the new case describes by a set of observations X = {X_1, X_2,.., X_n} who are clinical, biological signs and attributes of the medical imaging of the liver (Echography+TDM[1]). This phase consists in determining the diagnosis associated with this new case by basing their probabilities (obtained by the log-linear model) as measures of similarity. The retrieval for the similar case estimates by the quantity p (D|S) the probability that the descriptors of case S = S_1, S_2, S_j is the diagnosis of case D = D_1, D_2, D_J. We introduce the whole hierarchy A, then in practice the most likely diagnosis is considered.

6.3.1 The Use of Log Linear Model

In our case, we balanced each level of the network by a weight λ_i : the clinical level has a weight $\boldsymbol{\lambda}_1$=1, the biological level has a weight $\boldsymbol{\lambda}_2$=2, the medical imaging echography level has a weight $\boldsymbol{\lambda}_3$=3 and the medical imaging TDM level has a weight $\boldsymbol{\lambda}_4$=4. The weights granted to every level, corresponding to their level-headedness in the log-linear combination, establish the parameters of the model. After distribution of the algorithm of inference Pearl by weights $\boldsymbol{\lambda}_I$. The principle of the retrieval phase is a log-linear combination of several probability P (D|S) balanced by weights. The factor of level-headedness $\boldsymbol{\lambda}_i$ symptom of every level is useful to introduce master. It tries to maximize the following expression (Figure 6):

$$argmax \sum_{i=1}^{n} \lambda_i \log P(D = d|S)) \quad (5)$$

Where D: the diagnosis associated with the most similar case.

λ_i: The weight of the level i.

S: The symptoms.

Indeed, we select for the disease having the maximum probability knowing the probability and the weights λ_i.

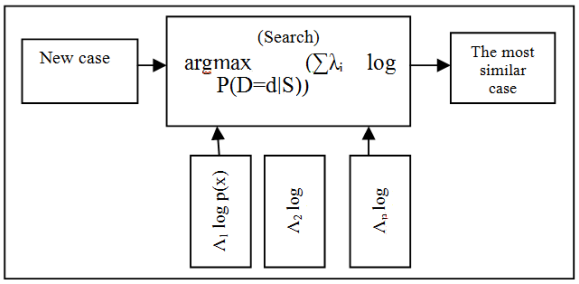

Figure 6. Modeling of the Retrieval Phase by a Log-linear Model

[1] TDM : *scanner* ou *tomodensitométrie*

6.3.2 The Retrieval Algorithm

The following paragraph summarizes our phase of retrieval in the Algorithm 2. This algorithm start by the selection of clinical, biological signs, and the attributes of medical imaging characterizing the target case. After the input of these information, every variable of the BN sends a message π of containing its current probability to all its sons except the variables of the last level which have not sons. In the reception of such a message, variables have to update their new probabilities in the light of this new information (the message π) then they send in their turn messages π to their sons. And at each sending we calculate the sum of the logarithm of every probability multiplied by its corresponding weight. At the end of this process and after all the sheets (variables of the last level) update its probability to obtain its in posteriori probability, the algorithm select for the maximal sum of the last level which will be the result of the diagnosis of the retrieved case.

Algorithm 2. Algorithm of the proposed retrieval phase.

```
Begin
Input :
BN : Bayesian Network (case base)
λi : the weigh of the level i ;
i : the numerous of the level
j : the numerous of the symptom
n : the number of  the level in BN
m : the number of  active node by level i
New case (set of  symptoms  S)
Output : source solution : max (Som i)
Begin
For i :=1 to N Do
   For j :=1 to M Do
                  Begin
                  Propagation of symptom nodes S ;
                  Som i := Som i + λi log P(D=d|S) ;
                  EndFor;
   EndFor;
D:=max (Somi); // D is an appropriate diagnosis of the similar case.
End.
```

7. The Proposed Adaptation Phase

The adaptation step means modifying solutions of former similar cases to fit current problem. If there are no important differences between a current and a similar case, a solution transfer is sufficient. Sometimes only few substitutions are required, but usually adaptation is a complicated process [4]. Since adaptation is even more difficult in medicine, we want to elaborate typical medical adaptation problems and we hope to show possibilities on how to solve them. The adaptation process depicted in Figure 7 starts by the retrieval phase (based on the different symptoms of the target case description and retrieved case description); this query is used to retrieve the most similar adaptation case. The similarity between the target and the retrieved case is again assessed. If the adaptation measure is higher than 3 the solution offered by the retrieved case is directly reused. This means that the solution of the current diagnosis case corresponds to the solution of the retrieved adaptation case. If the adaptation measure is lower than 3 then there is no adaptation.

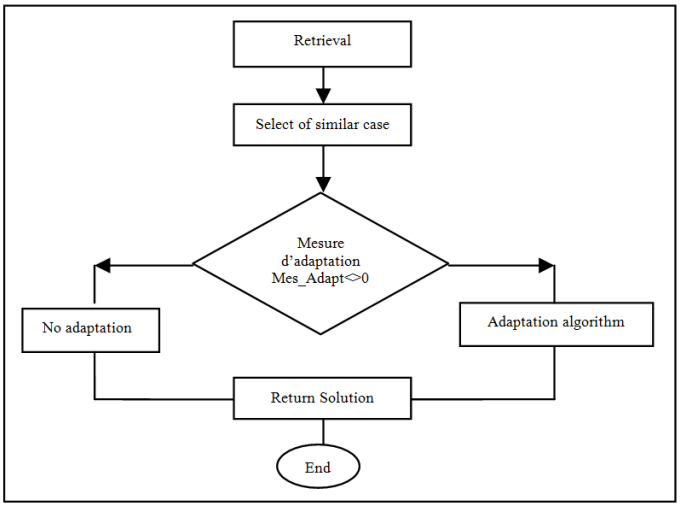

Figure 7. Adaptation Process

7.1 The Adaptation Measure Definition

The number of master symptom for each level of retrieved case is determined. This adaptation measure may be multiplied by a weighting factor. Then the sum of the master symptom of all attributes is calculated to provide a measure of the adaptation of the retrieved case to that of the target case. Defined by the following formula:

$$Adapt\_Meas(T, R) = \Sigma (\lambda_i * NB\_MA) \qquad (6)$$

T : Target case

S : Retrieved case

i : The number of the level from 1 to n

NB_MA : The number of master symptom for each level i

λ_i: weight of attribute i.

7.2 Algorithm of the Proposed Adaptation Phase

Let e an examination made for a given patient, where e_{ij} is the set of the clinical, biological and the medical imaging symptoms, such as i is the number of the level and j is the number of the symptom. After the application of the algorithm of the retrieval, we obtain the retrieved case. Then, we apply the adaptation measure *Adapt_Meas* :

If Adapt_Meas> 3 *then* the solution of the target case will be the solution of the retrieved case
Else the solution of the target case will be the solution of the corresponding case.
If Adapt_Meas=0 then " No. Adaptation ".

Algorithm 3. Algorithm of the proposed adaptation phase.

Input :

BN : Bayesian Network (case base) (Ps : Descriptors problem +Ss : descriptors solution)
i : the numerous of the level

```
j : the numerous of the symptom
n : the number of  the level in BN
m : the number of  active node by level i
dp : descriptors problem of retrieved case
Output: find the adaptation solution of target case
Begin
  Create a list containing master symptoms of every disease.
For i :=1 to N Do
      For j :=1 to M Do
Begin

    If (d1 is not clinical master symptom) and (d2 is not biological master
    symptom) and (d3 is not medical imaging master symptom) then no
    adaptation ;
    If (d1 is not clinical master symptom) and (d2 is not biological master
    symptom) and (d3 is a medical imaging master symptom) then copy the
    solution of the corresponding class;
    If (d1 is not clinical master symptom) and (d2 is a biological master
    symptom) and (d3 is not medical imaging master symptom) then copy the
    solution of the corresponding class;
    If (d1 is not clinical master symptom) and (d2 is a biological master
    symptom) and (d3 is a medical imaging master symptom) then copy the
    solution of the retrieved case;
    If (d1 is a clinical master symptom) and (d2 is not biological master
    symptom) and (d3 is not medical imaging master symptom) then « do
    biopsy » ;
    If (d1 is a clinical master symptom) and (d2 is not biological master
    symptom) and (d3 is a medical imaging master symptom) then copy the
    solution of the retrieved case;
    If (d1 is a clinical master symptom) and (d2 is a biological master
    symptom) and (d3 is not medical imaging master symptom) then copy the
    solution of the corresponding class;
    If (d1 is a clinical master symptom) and (d2 is a biological master
    symptom) and (d3 is a medical imaging master symptom) then copy the
    solution of the retrieved case;
End ;
End.
```

The Algorithm 3 treats the adaptation phase which follows the retrieval phase. It bases on the attribute "master symptom" as a measure of similarity of adaptation. Through this last one, we can adapt the solutions of the near cases in the case base. In our approach, when a retrieved case is found, it is necessary to search for all masters symptoms corresponding to this case and to make a comparison level by level (clinical, biological and imaging medical) between the retrieved case and the target case. If they are very near, copy the solution of the retrieved case in the target case. If the system does not manage to find a near problem in the case base, it is going to activate the biopsy of the liver, the obtained result will be a solution of the target problem.

7.3 Illustration of Adaptation Phase

Example 01: "Copy the solution of the retrieved case"
Let be the target case describing three levels of the Bayesian network and retrieval source case (Tab.1).

Table 1. The Adaptation Solution of Target Case A

	Descriptors	Target Case A	Retrieved Case
Description of the problem	Discovered Fortuitous		
	Icterus		0.1
	Hepathomegaly		
	High blood pressure portale	0.1	0.3
	Alcool	0.8	**0.9**
	Renal Insuffissance	0.8	**0.9**
	Viral Hepatitis	0.1	
	AgHg		
	Albuminemie, Phosphatase	-	0.1
	Bliribune		0.1
	Serology		
	Hyperglobulin	0.69	**0.7**
	Size of the liver		
	Homogeneity	0.5	**0.9**
	Atrophy		**0.8**
	Size of hurt		
	Density		
Solution description		---	Cirrhosis

We see that for every level, there is at least one master symptom, what implies to copy the solution of the source case.

- At the clinical level: the signs "Alcohol" and "renal Insuff" are masters symptoms because they carry a strong probability.
- At the biological level: the descriptor "Hyperglobulin" is master symptom.
- At the level of Medical imaging: the ultrasound image of the liver is considerable normal but it is heterogeneous and with scannographique images it is homogeneous and hyperdense.

We apply the measure of the proposed adaptation:

$Adapt\_Meas(T, R) = \Sigma (\lambda_i * NB\_MA)$

$Adapt\_Meas(T, R) = 2*1+2*1+3*1 = 7 > 3$

The solution of the adaptation is to copy the solution of the retrieved case: the disease is Cirrhosis.

Example 02: "Copy the solution of the corresponding class".

Let be the target case describing three levels of the Bayesian network and retrieval source case (Tab.2).

Table 2. The Adaptation Solution of Target Case B

	Descriptors	Target Case B	Retrieved Case
Description of the problem	Discovered Fortuitous	0.9	0.1
	Icterus		
	Hepathomegaly	0.3	
	Pain	0.1	

	Alcool		0.8
	ATCDF	0.8	**0.7**
	AgHg		
	Phosphatase		
	Bliribune		
	Serology		
	Renal Insuffissance	0.1	
	Size of the liver	1.0	0.9
	Homogeneity	0.5	0.9
	Cholestose		0.4
	Size of hurt		
	Density		
Solution description Solution		Liver polykystique Stop	Steatose Biopsy or Stop

We see at the clinical level and the level Medical imaging, there is no a master symptom but there exists only one master biological symptom, what implies that the solution of adaptation consists in copying the solution of the corresponding class.

We apply the measure of proposed adaptation:

$Adapt\_Meas(T, R) = \Sigma (\lambda_i * NB\_MA)$

$Adapt\_Meas(T, R) = 0*1+1*2+0*3 = 2 < 3$

The solution of the adaptation is to copy the solution of the corresponding case: the disease is Liver polykystique.

Example 03: "no adaptation"

If in three levels, there is no a descriptor which represents master symptom, i.e. there is no closer case in the target case; we will consider the opinion of the expert "doctor" not to propose irrelevant solutions of the target problem.

$Adapt\_Meas(T, R) = \Sigma (\lambda_i * NB\_MA)$

$Adapt\_Meas(T, R) = 0*1+0*2+0*3 = 0$

8. Discussion

In this work, we have presented the principle of the case base to search for similar cases in the target case. Firstly, an original algorithm of retrieval and a measure of similarity are proposed. We modeled the similarity measure by a log linear model which consists of expressing logarithms of the probability. This modeling allowed to decrease the number of necessary probability for the inference and it has the advantage to be hierarchical. It is thus possible to master the complexity of the model. The complexity of this retrieval algorithm is $O(\lambda_i \log P(D=d|S))$. The required space is O(N). This algorithm can be performed. What makes the retrieval more effective. Secondly, we have shown through the three examples that the most similar cases to the target case are not necessarily those chosen during the adaptation phase. And the adaptation measure *Adapt_Meas* has the advantage of facilitating the retrieved case of the adaptation by focusing on two essential parameters which are: NB_*MA* : number

of master symptom for each level i and λ_i: weight of attribute i. More the number of master symptom increases with a high weight adaptation will be more easy. That is to say the master symptoms of medical imaging level have a very important role to make the adaptation phase easier and more flexible.

9. Conclusion and Perspectives

In this paper we have presented a statistical modeling of the retrieval and the adaptation phase of CBR to help diagnosis of hepatic pathologies, by emphasizing the formalism of Bayesians networks. The approach which is described demonstrates the role of this modeling in the medical domain. As the knowledge of the domain is uncertain, this urged us to study a probability model of retrieval and more particularly the log linear model by basing it on a similarity metrics which allow an effective retrieval that supplies a more precise diagnosis. The adaptation step, will exploit the dependency relations between the target case and the retrieved case. In this frame, we have proposed an adaptation algorithm with an adaptation measure which support the parameters of the Bayesian network representing the retrieved case. We have come to the following results:

- A representation of the knowledge by a Bayesian network via a log linear model where we suppose conditional independences to reduce the number of probability facing the experts, to improve the retrieval phase and to offer the possibility of mastering the complexity of the model.
- A proposal of a new algorithm of retrieval of source cases similar to the new target case. The advantage of this algorithm is that the comparison is based on the route of the cases in a progressive way. It stops as they are not similar to any one which allows limiting the complexity of the retrieval phase in times of calculation.
- A proposal of a new adaptation algorithm, it unified the adaptation and retrieval step. It based on the dependency between the description problem of target case and the description problem of the retrieved case a through the retrieval measure and adaptation measure. This helps to select among the retrieved cases those which are most easily adaptable.

The possible perspectives are the following ones:

- To validate our proposed model, it is necessary to make a complete medical experiment and to confront our modeling in a real situation.
- To establish a comparative study with regard to the other models.

References

[1] A. Aamodt and E. Plaza, "Case Based Reasoning: Foundational issues, methodological variations, and system approaches", AI communications, IOS press, Vol.7: 1, pp. 39-59, **(1994)**.

[2] O. Agnieszka, J. Malek, J. Druzdzel and H. Wasyluk, "A Probabilistic Causal Model for Diagnostic of Liver Disorders", Intelligent Information Systems VII, Proceedings of the Workshop held in Malbork, **(1998)**, Poland.

[3] J. Pearl, "Probabilistic reasoning intelligent systems: networks of plausible inference", Morgan Kaufman, **(1988)**.

[4] J. Koldner, "Case-based Reasoning", Morgan Kaufmann, **(1993)**.

[5] B. Smyth and M. T. Keane, "Retrieving Adaptable Cases. The role of adaptation knowledge in case retrieval", In proceedings of EWCBR'93, LNAI 837, éd. Springer, **(1993)**.

[6] M. K. Haouchine, B. Chebel-Morello and N. Zerhouni, "Adaptation-Guided Retrieval for a Diagnostic and Repair Help System Dedicated to a Pallets Transfer", In 3rd European Workshop on Case-Based Reasoning and Context-Awareness, 9th European Conference on Case-Based Reasoning, ECCBR, **(2008)**, Trier, Germany.

[7] D. Hennessy, B. Buchanan and J. Rosenberg, "Bayesian Case Reconstruction", Lecture Notes in Computer Science, pp.148–158, **(2002)**.

[8] R. Pavon, F. Diaz, R. Laza and V. Luzon, "Automatic parameter tuning with a Bayesian case-based reasoning system", A case of study. Expert Systems with Applications, 36 (2P2), pp. 3407–3420, **(2009)**.

[9] N. Silvia, Schiaffino and A. Amandi, "User profiling with Case-Based Reasoning and Bayesian Networks", In Open Discussion Track Proceedings, International Joint Conference, IBERAMIA-SBIA, **(2000)**, Atibaia, Brazil.

[10] D. Dong, S. Zhaohao and G. Feng, "PCOPM: A Probabilistic CBR Framework for Obesity Prescription Management", In Proceedings of ICIC (2), **(2010)**, pp. 91-99.

[11] E. Ocampo, M. Maceiras, S. Herrera, C. Maurente, D. Rodríguez and M. Sicilia, "Comparing Bayesian inference and case-based reasoning as support techniques in the diagnosis of Acute Bacterial Meningitis", Elsevier journal, **(2011)**.

[12] P. Naïm, P.H. Wuillemin, Ph. Leray, O. Pourret and A. Becker, Réseaux Bayésiens. Eyrolles, **(2004)**.

[13] A. Djebbar and H. F. Merouani, Une modalisation de la base de cas par un réseau bayésien: application à l'aide au diagnostic médical. In proceedings of MOSIM, **(2006)** 3-5 Avril, Rabat- Maroc.

[14] R. Christensen, "Log-Linear Models and Logistic Regression", Springer, **(1997)**.

[15] P. Blunsom, K. Kocik and J. R. Currran, "Question classification with Log-Linear Models", In proceeding of SIGIR, **(2006)**, Seattle, Washington, USA.

[16] D. Keysers and R. Paredes, "Comparaison of Log-linear Models and Weighted Dissimilarity Measure", **(2002)**.

[17] D. Déchelotte, Traduction automatique de la parole par méthodes statistiques. Thèse de doctorat, **(2007)**, pp. 16-23.

[18] F. Corset, Optimisation de la maintenance à partir de réseaux bayésiens et fiabilité dans un contexte doublement censuré. PhD thesis, Université Joseph Fourier, **(2003)**.

[19] A. Djebbar and H. F. Merouani, Une modélisation bayésienne de la remémoration pour l'aide au diagnostic. ICITeS, **(2011)**, April 10-12, Sousse, Tunisia. ww.dline.info/content.htm.

Authors

Akila DJEBBAR received her engineer in Computer science in 2002 from Annaba University. She received a Magister in "Distributed Artificial Intelligence" in 2006. She prepares a doctorate in "Integrating CBR with BN in medical diagnosis", and she is currently searcher at the department of computer science in Annaba University (Algeria). Her research intere sts are: Case Based Reasoning, Bayesian Network, Retrieval, Adaptation, and Similarity Measure.

Hayet Farida MEROUANI received her Eng.Degree from Annaba University, Algeria in 1984, the PhD Degree from Robert Gordon University, Aberdeen, UK. Actualy, she is full Associate Professor at Badji Mokhtar University, Annaba. She also leads Research group of Pattern recognition, as a National program research of breast Cancer. Her current works focus on the computer vision, medical imaging and Biometry.