

# MULTIMODAL MEDICAL CASE RETRIEVAL USING BAYESIAN NETWORKS AND THE DEZERT-SMARANDACHE THEORY

G. Quellec<sup>1,2</sup>, M. Lamard<sup>3,2</sup>, L. Bekri<sup>2,4</sup>, G. Cazuguel<sup>1,2</sup>, C. Roux<sup>1,2</sup>, B. Cochener<sup>3,2,4</sup>

1 : INSTITUT TELECOM; TELECOM Bretagne; UEB; Dpt ITI, Brest, F-29200 France;

2 : Inserm, U650, Brest, F-29200 France;

3 : Univ Bretagne Occidentale, Brest, F-29200 France;

4 : CHU Brest, Service d'Ophthalmologie, Brest, F-29200 France;

## ABSTRACT

In this paper, we present a Case Based Reasoning (CBR) system for the retrieval of medical cases made up of a series of images with semantic information (such as the patient age, sex and medical history). Indeed, medical experts generally need varied sources of information, which might be incomplete, uncertain and conflicting, to diagnose a pathology. Consequently, we derive a retrieval framework from Bayesian networks and the Dezert-Smarandache theory, which are well suited to handle those problems. The system is designed so that heterogeneous sources of information can be integrated in the system: in particular images, indexed by their digital content, and symbolic information. The method is evaluated on a classified diabetic retinopathy database. On this database, results are promising: the retrieval precision at five reaches 80.5%, which is almost twice as good as the retrieval of single images alone.

**Index Terms**— Case based reasoning, Image indexing, Bayesian networks, Dezert-Smarandache theory, Diabetic Retinopathy

## 1. INTRODUCTION

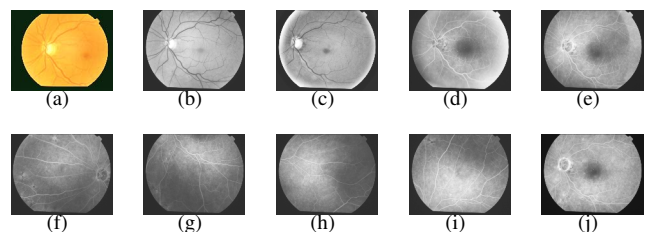
In medicine, the knowledge of experts is a mixture of text-book knowledge and experience through real life clinical cases. Consequently, there is a growing interest in case-based reasoning (CBR), introduced in the early 1980s, for the development of medical decision support systems [1]. The underlying idea of CBR is the assumption that analogous problems have similar solutions, an idea backed up by physicians' experience. In CBR, the basic process of interpreting a new situation revolves around the retrieval of relevant cases in a case database. The retrieved cases are then used to help interpreting the new one.

We propose in this article a CBR system for the retrieval of medical cases made up of a series of images with contextual information. The proposed system is applied to the diagnosis of Diabetic Retinopathy (DR). Indeed, to diagnose DR, physicians analyze series of multimodal photographs together with

contextual information like the patient age, sex and medical history.

When designing a CBR system to retrieve such cases, several problems arise. We have to aggregate heterogeneous sources of evidence (images, nominal and continuous variables) and to manage missing information. To solve these problems, we propose to express the different sources of information as probabilities and to model the relationships between each attributes with a Bayesian network. The Bayesian network may be used to fuse the sources of information. However, these sources may be uncertain and conflicting. As a consequence, we also applied the Dezert-Smarandache Theory (DSmT) of plausible and paradoxical reasoning, proposed in recent years [2], which is better suited than Bayesian approach to fuse uncertain, highly conflicting and imprecise sources of evidence.

## 2. DIABETIC RETINOPATHY DATABASE



**Fig. 1.** Photograph series of a patient eye

Images (a), (b) and (c) are photographs obtained by applying different color filters. Images (d) to (j) form a temporal angiographic series: a contrast product is injected and photographs are taken at different stages (early (d), intermediate (e)-(i) and late (j)).

Diabetes is a metabolic disorder characterized by sustained inappropriate high blood sugar levels. This progressively affects blood vessels in many organs, including the retina, which may lead to blindness. The database is made up of 63 patient files containing 1045 photographs altogether.

Patients have been recruited at Brest University Hospital since June 2003 and images were acquired by experts using a Topcon Retinal Digital Camera (TRC-50IA) connected to a computer. Images have a definition of 1280 pixels/line for 1008 lines/image. The contextual information available is the patients' age and sex and structured medical information (about the general clinical context, the diabetes context, eye symptoms and maculopathy). Thus, at most, patients records are made up of 10 images per eye (see figure 1) and of 13 contextual attributes; 12.1% of these images and 40.5% of these contextual attribute values are missing. The disease severity level, according to ICDRS classification [3], was determined by experts for each patient.

### 3. BAYESIAN NETWORKS AND THE DEZERT-SMARANDACHE THEORY

#### 3.1. Bayesian Networks

A Bayesian network [4] is a probabilistic graphical model that represents a set of variables and their probabilistic dependencies. It is a directed acyclic graph whose nodes represent variables, and whose arcs encode conditional independencies between the variables. Each arc in the graph is associated with a conditional probability matrix expressing the probability of a child variable given one of its parent variables. A directed acyclic graph is a Bayesian Network relative to a set of variables  $\{X_1, \dots, X_n\}$  if the joint distribution  $P(X_1, \dots, X_n)$  can be expressed as follows:  $P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | \text{parents}(X_i))$ . The network structure and conditional probability tables can be learnt automatically from data [5].

A Bayesian network is used to answer probabilistic queries about the variables; typically to find out updated knowledge of the state of a subset of variables when other variables (the evidence variables) are observed. This process of computing the posterior distribution of variables given evidence is called probabilistic inference. It can be used to fuse evidence from several sources of information.

#### 3.2. Dezert-Smarandache Theory

The Dezert-Smarandache Theory allows combining any types of independent sources of information represented in term of belief functions. It is more general than probabilistic (or Bayesian) fusion, discussed above, or Dempster-Shafer theory. It is particularly well suited to fuse uncertain, highly conflicting and imprecise sources of evidence [2].

Let  $\theta = \{\theta_1, \theta_2, \dots\}$  be a set of hypotheses under consideration for the fusion problem;  $\theta$  is called the frame of discernment. In Bayesian theory, a probability  $p(\theta_i)$  is assigned to each element  $\theta_i$  of the frame, such that  $\sum_{\theta_i \in \theta} p(\theta_i) = 1$ . More generally, in DSMT, a belief mass  $m(A)$  is assigned to each element  $A$  of the hyper-power set  $D(\theta)$ , i.e. the set of all composite propositions built from elements of  $\theta$  with  $\cap$  and

$\cup$  operators, such that  $m(\emptyset) = 0$  and  $\sum_{A \in D(\theta)} m(A) = 1$ . The belief mass functions specified by the user for each source of information, noted  $m_j$ ,  $j = 1..N$ , are fused into the global mass function  $m_f$ , according to a given rule of combination. Several rules have been proposed to combine mass functions, including the hybrid rule of combination or the PCR (Proportional Conflict Redistribution) rules [2]. It is possible to introduce constraints in the model [2]: we can specify pairs of incompatible hypotheses  $(\theta_a, \theta_b)$ , i.e. each subset  $A$  of  $\theta_a \cap \theta_b$  must have a null mass, noted  $A \in C(\theta)$ . Once the fused mass function  $m_f$  has been computed, a decision function is used to evaluate the probability of each hypothesis, one of these functions can be used: the credibility, the plausibility or the pignistic probability [2].

### 4. IMAGES IN THE BAYESIAN NETWORK

To include images in a Bayesian network, we associate a variable  $F_j$  with each imaging modality  $j$ . We have to define a finite number of states for these variables. In that purpose, we apply a principle similar to Content-Based Image Retrieval (CBIR) [6]. CBIR involves 1) building a signature for each image (i.e. extracting a feature vector summarizing their numerical content), and 2) defining a distance measure between two signatures. Thus, measuring the distance between two images comes down to measuring the distance between two signatures. Similarly, in a Bayesian network, we cluster similar image signatures (according to the defined distance measure) and associate a state of  $F_j$  for each image cluster.

In previous studies, we proposed to compute a signature for images from their wavelet transform (WT) [7]. These signatures model the distribution of the WT coefficients in each subband of the decomposition. The associated distance measure  $D$  [7] computes the divergence between these distributions. We used these signature and distance measure to cluster similar images.

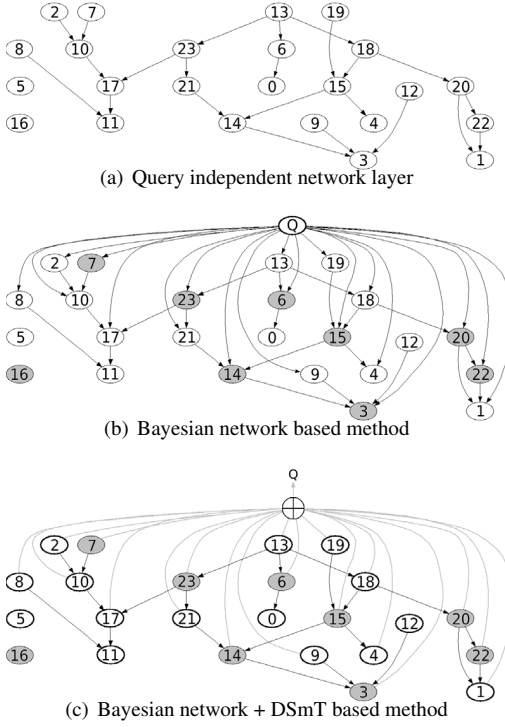
Any clustering algorithm can be used, provided that the distance measure between feature vectors can be specified. We used FCM (*Fuzzy C-Means*) [8], one of the most common algorithms, and replaced the Euclidian distance by  $D$ .

### 5. BAYESIAN NETWORK BASED RETRIEVAL

Let  $x_q$  be a case placed as a query. To assess the relevance of each case  $x$  in the database, we define a Bayesian network with the following variables: a variable  $F_j$ ,  $j = 1..N$ , representing each feature of  $x$  and a Boolean variable  $Q =$  "the query is satisfied" ( $\bar{Q} =$  "the query is not satisfied").

To build the network, we first learn the relationships between the feature variables  $F_j$ ,  $j = 1..N$ , from data [5]: we have thus built a sub-network, independent on both  $x_q$  and  $x$  (see figure 2 (a)).

$Q$  is then integrated in the network:  $x_q$  specifies which features should be found in the retrieved cases, so when the  $j^{th}$



**Fig. 2.** Evaluating a case  $x$  by the two proposed methods. Figure (a) describes the query independent network layer, learnt from data. Figure (b) (resp. figure (c)) describes the method presented in section 5 (resp. section 6). In this example, features 6, 7, 14, 15, 16, 20, 22 and 23 are available for  $x_q$ . Evidence nodes are grey. In figure (c),  $\oplus$  represents the fusion operator.

feature of  $x_q$  is available, we connect the two nodes  $Q$  and  $F_j$  (see figure 2 (b)). If a node  $F_j$  and  $Q$  are connected, we have to estimate the associated conditional probability matrix  $P(F_j = f_{jk}|Q)$ , where  $f_{jk}$  denotes the  $k^{th}$  possible state for  $F_j$ , according to  $x_q$ . To compute  $P(F_j = f_{jk}|Q)$ , we first estimate  $P(Q|F_j = f_{jk})$  by the procedure below and we apply Bayes theorem.

To estimate  $P(Q|F_j = f_{jk})$ , we use the membership degree of  $x_q$  to each state  $f_{jk}$  of  $F_j$ , noted  $\alpha_{jk}(x_q)$ . We assume that the state of the cases in the same class are predominantly in a subset of states for  $F_j$ . So, in order to estimate the conditional probabilities, we use a correlation measure  $S_{jk_1k_2}$  between two feature states  $f_{jk_1}$  and  $f_{jk_2}$ , regarding the class of the cases at these states. To compute  $S_{jk_1k_2}$ , we first compute the mean membership  $D_{jk_1c}$  (resp.  $D_{jk_2c}$ ) of cases in a given class  $c$  to the state  $f_{jk_1}$  (resp.  $f_{jk_2}$ ) (equation 1):

$$\left\{ \begin{array}{l} D_{jk_1c} = \beta \frac{\sum_x \delta(x,c) \alpha_{jk_1}(x)}{\sum_x \delta(x,c)} \\ \sum_{c=1}^C (D_{jk_1c})^2 = 1, \forall (j, k_1) \end{array} \right. \quad (1)$$

where  $\delta(x, c) = 1$  if  $x$  is in class  $c$ ,  $\delta(x, c) = 0$  otherwise, and  $\beta$  is a normalizing factor.  $S_{jk_1k_2}$  is given by equation 2:

$$S_{jk_1k_2} = \sum_{c=1}^C D_{jk_1c} D_{jk_2c} \quad (2)$$

In the proposed model, we choose  $P(Q|F_j = f_{jk})$  proportional to  $\sum_{l=1}^N \alpha_{jl}(x_q) S_{jkl}$ . The different cases in the database are then processed sequentially. To evaluate a case  $x$ , every available feature for  $x$  is processed as evidence to infer the posterior probability  $P(Q)$  (see figure 2 (b)). The cases are then ranked in decreasing order of  $P(Q)$ .

## 6. BAYESIAN NETWORK AND DSMT BASED RETRIEVAL

To extend the previous Bayesian network based method to the DSMT framework, we assign a belief mass not only to  $Q$  and  $\bar{Q}$ , but also to  $Q \cup \bar{Q}$  (not to  $Q \cap \bar{Q}$  because  $Q$  and  $\bar{Q}$  are incompatible hypotheses).

To compute the belief masses  $m_j$  for a given feature  $F_j$ , we define a test  $T_j$  on the degree of match  $dm_j(x, x_q)$  between  $x$  and  $x_q$ .  $dm_j(x, x_q)$  is defined as  $dm_j(x, x_q) = \sum_k P(Q|F_j = f_{jk}) \alpha_{jk}(x)$  and  $T_j$  is defined as “ $dm_j(x, x_q) \geq \tau_j$ ”,  $0 \leq \tau_j \leq 1$ . The sensitivity (resp. the specificity) of test  $T_j$  represents the degree of confidence in a positive (resp. negative) answer to the test. Whether the answer is positive or negative,  $Q \cup \bar{Q}$  is assigned the degree of uncertainty. The mass functions are then assigned according to  $T_j$ . If  $T_j$  is true:

$$\begin{array}{l} m_j(Q) = P(T_j|x \text{ relevant for } x_q) \rightarrow \text{sensitivity} \\ m_j(Q \cup \bar{Q}) = 1 - m_j(Q) \\ m_j(\bar{Q}) = 0 \end{array} \quad (3)$$

Otherwise:

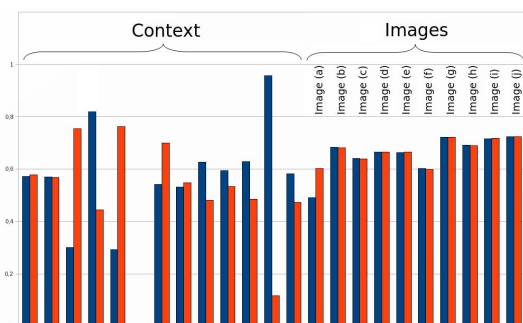
$$\begin{array}{l} m_j(\bar{Q}) = P(\bar{T}_j|x \text{ not relevant for } x_q) \rightarrow \text{specificity} \\ m_j(Q \cup \bar{Q}) = 1 - m_j(\bar{Q}) \\ m_j(Q) = 0 \end{array} \quad (4)$$

We want to define  $T_j$  so that is both sensitive and specific. As  $\tau_j$  increases, sensitivity increases and specificity decreases. So, we set  $\tau_j$  as the intersection of the two curves “sensitivity according to  $\tau_j$ ” and “specificity according to  $\tau_j$ ”.  $\tau_j$  is searched by the bisection method: for each value of  $\tau_j$  evaluated, sensitivity and specificity are estimated from the cases of the database.

Finally, to evaluate a case  $x$  with this model (see figure 2 (c)), every available feature for  $x$  is processed as evidence to estimate  $\alpha_{jk}(x) \forall k, j = 1..N$ . If the  $j^{th}$  feature of  $x_q$  is available, the degree of match  $dm_j(x, x_q)$  is computed and the belief masses are computed according to test  $T_j$ . The sources available for  $x_q$  are then fused with the PCR5 rule [2] and the pignistic probability of  $Q$ , noted  $betP(Q)$ , is computed. The cases are then ranked in decreasing order of  $betP(Q)$ .

## 7. RESULTS

The mean precision at five, i.e. the mean number of relevant cases among the top five results, reaches **69.5%** using the Bayesian network based system, and **80.5%** using the Bayesian network and DSMT based system. As a comparison, the mean precision at five obtained by CBIR (when cases are made up of a single image) with the same image signatures is 46.1% [7]. To evaluate the contribution of the proposed system for the retrieval of heterogeneous and incomplete cases, the proposed method is compared to a linear combination of heterogeneous distance functions, managing missing values [9], which is the natural generalization of classic CBR to the studied cases. Its extension to vectors containing images is based on the distance between image signatures (see section 4). A mean precision at five of 52.3% was achieved by this method. To evaluate the contribution of each attribute, we give in figure 3 the sensitivity and specificity of each test  $T_j$ . The method is robust regarding missing information: indeed for instance, the mean retrieval precision at five is 88.2% for examples with 22 available attributes out of 23, and 71.5% for examples with 12 available attributes.



**Fig. 3.** Influence of each descriptor. The sensitivity followed by the specificity of each test  $T_j$  is given for each descriptor (the same letter than in figure 1 are used to denote image modalities).

## 8. DISCUSSION AND CONCLUSION

In this article, we introduce a method to include image series and their numerical signatures, with contextual information, in CBR systems. In particular, a way to include image signatures in a Bayesian network was proposed. Two retrieval systems, based on the same principle, were proposed: a Bayesian network is used to model the relationships between case descriptors and thus handle missing information, and relevance information, coming from each descriptor are fused by either Bayesian fusion or DSMT. In this system, DSMT shows its superiority over Bayesian fusion. Bayesian Networks are however efficient for managing missing information. On this

database, the method outperforms our first CBIR algorithm by a factor of 175% in precision (80.5% as opposed to 46.1%). This stands to reason since an image alone is generally not sufficient for experts to correctly diagnose the disease severity level of a patient. However, figure 3 shows that each single image are relevant attributes. Besides, this non-linear retrieval method is 154% (80.5% as opposed to 52.3%) more precise than a simple linear combination of heterogeneous distances on the DR database. The proposed framework is also interesting for being generic: any multimedia database may be processed so long as a procedure to cluster cases is provided for each new modality (sound, video, etc).

## 9. REFERENCES

- [1] I. Bichindaritz and C. Marling, "Case-based reasoning in the health sciences: What's next?," *Artificial Intelligence in Medicine*, vol. 36, no. 2, pp. 127–135, january 2006.
- [2] F. Smarandache and J. Dezert, *Advances and Applications of DSMT for Information Fusion II*, American Research Press Rehoboth, 2006.
- [3] C. Wilkinson, F. Ferris, and R. Klein et al., "Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales," *Ophthalmology*, vol. 110, no. 9, pp. 1677–1682, 2003.
- [4] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, 1988.
- [5] J. Cheng, D. Bell, and W. Liu, "Learning bayesian networks from data: An efficient approach based on information theory," 1997.
- [6] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, December 2000.
- [7] M. Lamard, W. Daccache, G. Cazuguel, C. Roux, and B. Cochener, "Use of jpeg-2000 wavelet compression scheme for content-based ophtalmologic retinal retrieval," in *Proceedings of the 27th annual international conference of IEEE engineering in medicine and biology society*, september 2005.
- [8] J.C. Bezdek, *Fuzzy Mathematics in Pattern Classification*, Ph.D. thesis, Applied Math. Center, Cornell University, Ithaca, 1973.
- [9] D.R. Wilson and T.R. Martinez, "Improved heterogeneous distance functions," *Journal of Artificial Intelligence Research*, vol. 6, pp. 1–34, 1997.