MANAGEMENT OF UNCERTAINTY IN AN EXPERT SYSTEM

Lucimar F. de Carvalho[1], Silvia M. Nassar [2], Fernando M. Azevedo[2], Cristiane Koehler[2], Hugo T.de Carvalho[1], Fernanda Ramos [3]

[1]Curso de Ciência da Computação, Dpto. de Informática e Estatística, Universidade de Passo Fundo (UPF), Brasil, 9999191-111 Fone (0XX54)3111177/33, Fax (0XX54)3113347

[2]Curso de Pós-Graduação em Ciência da Computação e Engenharia Biomédica, Dpto. de Engenharia Elétrica, Universidade Federal de Santa Catarina (UFSC), Brasil, 88040-900
Fone (0XX48)3317747

[3]Alunos Bolsistas – CNPq - PIBIC/UPF

fossatti@upf.tche.br, nassar@inf.ufsc.br, azevedo@gpeb.ufsc.br, koehler@inf.ufsc.br, hcarvalho@saude.upf.tche.br, 22443@lci.upf.tche.br,

ABSTRACT

The objective of this paper is the application of two numeric approaches used in the treatment of uncertainty and its respective validation. This research accomplished a study on the probabilistic reasoning in an Expert System: Bayesian Networks and Certainty Factors, presenting as application domain the Clinical Diagnosis of the Epileptic Seizures. The system was developed with KADS methodology and is based on the classification of the type of seizures of the International League Against Epilepsy /ILAE81 [1]. The concurrent validation showed that the Sensibility and Negative Predicted Value was more precise through the Bayesian Networks while Specificity and Affirmative Predicted Value showed similarity in both approaches.

Key words: Artificial Intelligence, Expert Systems, Uncertainty, Epileptic Seizures.

1. INTRODUCTION

The objective of this paper is to represent through computational techniques the present uncertain reasoning in a Knowledge Based System (KBSs). The uncertainty and the imprecision in the information can be understood as the lack of apropriate
information for decision taking. Some theories are being used in the treatment of uncertainty: bayesian networks, certainty factors, fuzzy sets of Zadeh and Dempster Shafer's theory [2], [3], [4], [5] e [6].

The implemented system is of support to the Clinical Diagnosis of Epileptic Seizures: Simple Partial Seizures, Secondarily Widespread Partial Seizures and the Widespread Seizures. The system was implemented with the help of the KADS methodology (Knowledge Acquisition Design Structuring) [7].

To do the validation of the approaches two computerized systems were implemented: the first, was the implementation of KBSs of Support to Clinical Diagnosis of the Epileptic Seizures that uses Certainty Factors (KBSs_CF) and the second, consists of the implementation of the Probabilístic Expert System that uses Bayesian Networks (PES_BN) to represent uncertainty.

2. THE KADS METHODOLOGY

The KADS methodology was developed at the University of Amsterdam in 1983 to give support to the development of Expert System through the - ESPRIT (European Strategic Programme in Research in Information Technology). The KADS methodology provides a flexible life-cycle model by structuring and controlling the development process, rules and guidelines , it guides the knowledge engineer throgh the life-cycle model, methods, tools and techniques. The life-cycle model of KADS presents the following phases: analysis of the domain area, design, implementation, installation, use, maintenance and knowledge refinement. In the specification phase, a deep analysis of the domain was implemented. To give support to the analysis phase, the KADS methodology works with the four layers theory known as expertise model composed of basic elements present in the expert knowledge: domain, task, inference and strategic layer. This model of four layers supplies a base for the knowledge acquisition. The bottom layer describes all domain-specific knowledge in terms of concepts, relations and structures. The other three layers describe the generic problem-solving method in a domain-independent way.

In the Domain Layer there is the static knowledge representing concepts, relationships between concepts and structures of these relationships. In this layer there was described the knowledge area in which the system was implemented. In the Inference Layer there is the knowledge of the different inference types that can be worked. The Task Layer defines as the knowledge that is found in the two previous layers will be used in direction the final end of the system. The Strategic Layer contains the strategic knowledge to combine the tasks that will be used in the problem solution. In this layer the tasks structure can be altered to supply flexibility to the model.

A KBSs developed through the KADS methodology tries to embrace organization aspects, user x system interaction, besides expertise model.

3. IMPLEMENTATION OF CERTAINTY FACTORS (CF)

The KBSs_CF integrates a Knowledge Base (KB) composed of the rules base and facts base. The Knowledge Base (see Figure 1) was implemented with the shell Kappa-PC guided to objects[11]. The approach of CF had as pioneer the MYCIN system [12], bacteriological infections diagnosis support system. The original formula of MYCIN works with the difference between faith and disbelief: CF[h,e] = MC[h,e] - MD[h,e].

Starting from this original formula, other formulas were proposed. The choice between a formula and another will depend on the reality that the system proposes.

In the implementation of the Clinical Diagnosis System of Epileptic Seizures the following formula was used [12]: (1)

$$CF[h,e] = CF[h,e_1] + CF[h,e_2] * (1 - CF[h,e_1]) + CF[h,e_3] * (1- CF[h,e_1] + CF[h,e_2] * (1 -CF[h,e_1]))$$ [12]

The CF can be calculated with one evidence, two or three evidences. This formula was chosen because it adapted better to the Epileptic Seizures domain where, to determine a hypothesis, it was necessary to prioritize the order of evidences occurrence, not existing values for disbelief. In the case of Epileptic Seizures, the first symptoms generally can define a diagnosis.

4. IMPLEMENTATION OF THE PROBABILISTIC EXPERT SYSTEM

A Probabilistic Expert System (PES), besides the present components in an expert system (knowledge base, inference machine, knowledge acquisition subsystem, explanation subsystem and interface), its knowledge base should be represented by a probabilities distribution, see Fig. 1. The base of this distribution should involve a group of discreet variables with attributes. The Bayes's Theorem has as main function to modernize and to revise the conditional probabilities based on the basic concepts about Conditional Probability, that is, P(H|e) - the probability of a hypothesis "*h*" to happen given an evidence "*e*"[5].

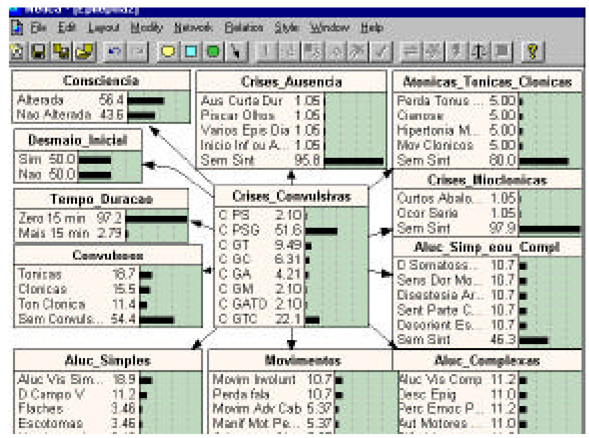

Fig. 1 - Bayesian Network

5. RESULTS

Initially, there will be described the stage of validation of knowledge Base of Widespread Secondarily Partial Seizures for PES_BN and KBSs_CF. For comparation effects the following partial results were found until the moment: 95 patients were appraised, of these 95 patients, the expert validated 43 patients as carriers of Widespread Secondarily Partial Seizures; the PES_BN found 42 patients and KBSs_CF found 39 patients with Widespread Secondarily Partial Seizures presenting the following percentiles, where: DSaffirmative /negative denotes the event System's Diagnosis is affirmative/negative and DEaffirmative/negative denotes the Expert's Diagnosis is respectively affirmative/negative [13].

Sensibility = Prob(DSaffirmative/ DEaffirmative).

Sensibility (PES_BN) = 42 / 43 = 0,97. that is, 97 percent of the patients with WSPS are diagnosed by the system as patients that present WSPS.

Sensibility (KBSs_CF)=39 / 43 = 0,90.

That is, 90 percent of the patients with WSPS are diagnosed by the system as patients that present WSPS.

Specificity = Prob(DSnegative / DEnegative)

Specificity (PES_BN) = 52 / 52 = 1. That is to say, 100 percent of the patients that don't have WSPS are diagnosed by the system as not being carriers of WSPS.
Specificity (KBSs_CF) =52 / 52 = 1. That is to say, 100 percent of the patients that don't have WSPS are diagnosed by the system as not being carriers of WSPS.

APV = Prob(DEaffirmaive / DSaffirmative)
APV (PES_BN) = 42 / 42 = 1. If the system diagnoses a patient as a carrier of WSPS the probability of the patient really being a carrier of a WSPS is of 100 percent .
APV (KBSs_CF) = 39 / 39 = 1. If the system diagnoses a patient as a carrier of WSPS the probability of the patient really being a carrier of a WSPS is of 100 percent .

NPV = Prob((DEnegative / DSnegative)
NPV (PES_BN) = 52 / 53 = 0,98. That is, if the system diagnoses a patient as non carrier of WSPS the probability of the patient really not having WSPS is of 98 percent.
NPV (KBSs_CF) = 52 / 56 = 0,92. That is, if the system diagnoses a patient as non carrier of WSPS the probability of the patient really not having WSPS is of 92 percent.
The concurrent validation showed that the Sensibility and Negative Predicted Value to be more precise through the Bayesians Networks while Specificity and Affirmative Predicted Value showed similarity in both approaches, see Fig. 2.

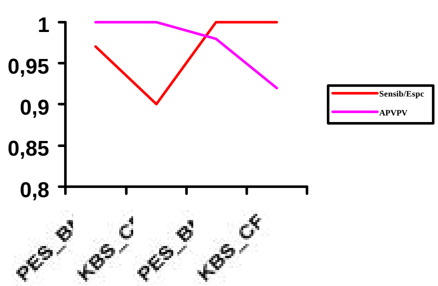

Fig. 2 - Graphic Representation

6. CONCLUSIONS

In this work, it was sought, initially, to validate two approaches that work with uncertainty: Bayesian Networks and Certainty Factors. The validation work is difficult and complex. Even so this validation is evidencing the need and importance of accomplishing an internal validation in all the developed systems. It is from this moment that both implemented systems can be improved to obtain a more precise and true result. Next some advantages/disadvantages found in both approaches were listed.

One of the main advantages of use of Bayesian Network in this system was the form in which the theory transmits graphically the expert's knowledge, although the obtaining of the probabilities of the hypothesis *Hi* and the conditional *P(Hi|E)* has been a difficult task. The use of Bayesians Networks showed to be accurate and coherent when the expert's knowledge was formalized.
To develop the Bayesian Network it was necessary to obtain a group of probabilities a *priori* that integrate the pattern space, resulting in a group of probabilities a *posteriori*. In the implementation of CF it was not necessary to obtain a data base, which turned the implementation simpler. Only to validate the system, the data base was present.
The acquisition of data in the Bayesian Network was accomplished in an objective way, due to the fact that the pattern space is known. Unlike the data acquisition with CF: accomplished in a subjective way, there didn't exist an only formalism in the obtaining of the probabilities. This way, the use of Bayesian Networks is useful far solving more complex problems, provided that one knows the pattern space.
The modernizations of the probabilities in the Bayesians Networks happened through an only step: it was enough to alter the value of the evidence and automatically the value of the hypothesis was altered. With this, *P(h/e)* can be obtained, which facilitated the time spent on the implementation. With CF these modernizations happened through two stages: first it was necessary to alter the value of the evidence and next the value of the hypothesis was calculated again.
One of the difficulties found was the inexistence of some evidences relative to the hypotheses that were not explicit in the two bases. And, also, it was necessary to do small alterations in the occurrence probabilities. With this, a larger time was taken to improve the performance of the two systems.
It was verified that each one of the approaches possesses better performance to different types of situation, that is, the existence of a pré-defined data base and the complexity of the system to be implemented (choice of the domain). In both approaches there is the independence of the evidences, that is, an evidence doesn't depend on the other to be defined as a probable hypothesis.
The result found in the validation through the Bayesians Networks showed to be more precise, for the fact that the net manages, through the occurrence probabilities, to approach more the objective, supplying a more refined result. A fact which doesn't happen to CF, although the occurrence probabilities have been calculated through one, two or three symptoms (in agreement with the occurrence).

REFERENCES

[1] Commission on Classification and Terminology of the International League Against Epilepsy – Proposal for revised clinical and electroencephalographic classification on epilepsies seizures. Epilepsia (1981) (22):489-501.
[2] Collazos, K.S.: Sistema Especialista Nebuloso para Diagnóstico Médico – Tese de Mestrado. Universidade Federal de Santa Catarina. Florianópolis (1997).
[3] Gordon, J., Shortliffe, E.: The Dempster-Shafer tehory of evidence: Rule-Based Expert Systems. (1985) 272-292.

[4] Hartmut K. et al.: Um Sistema Inteligente de Apoio à Decisão Estratégica Baseado em Inferência Probabilística. In: Anais do VII Congresso Nacional de Investigação Operacional, Aveiro Portugal (1996) 1-3.
[5] Pearl J.:Probabilist Reasoning in Intelligent Systems: Networds of Plausible Inference, California: Morgan Kaufmann (1988).
[6] Koehler, C.: Uma Abordagem Probabilística para Sistemas Especialistas. Tese de Mestrado. Universidade Federal de Santa Catarina. Florianópolis (1998).
[7] Hickman F.R.: Analysis for Knowledge - Based Systems - a practical guide to the KADS methodology, England: Ellis Horwood Limited (1989).
[8] Nassar, S.M.: Sistema Estatístico Inteligente para Apoio a Pesquisas Médicas - Tese de Doutorado, Universidade Federal de Santa Catarina/SC, (1995).
[9] Chadwick D.: Diagnosis of Epilepsy. Lancet, (1990) (336):291-295.
[10] Commission on Classification and Terminology of the International League Against Epilepsy (1985) – Proposal for classification of epilepsy and epileptic syndromes. Epilepsia, 26, 268-278.
[11] Booch G.: Object Oriented Design with Applications. California: RedWood City, (1991).
[12] Shorliffe, E.H., Buchanan, B.G.: Rule-Based Expert Systems: The Mycin Experiments of the Stanford Heuristic Programming Project. Addison Wesley Publishing Company (1984).
[13] Armitage, P. & Berry G.: Statistical Methods in Medical Research. London: Blackwell Scientific Publications (1991).

MANAGEMENT OF UNCERTAINTY IN AN EXPERT SYSTEM

ABSTRACT

The objective of this paper is the application of two numeric approaches used in the treatment of uncertainty and its respective validation. This research accomplished a study on the probabilistic reasoning in an Expert System: Bayesian Networks and Certainty Factors, presenting as application domain the Clinical Diagnosis of the Epileptic Seizures. The system was developed with KADS methodology and is based on the classification of the type of seizures of the International League Against Epilepsy /ILAE81. The concurrent validation showed that the Sensibility and Negative Predicted Value was more precise through the Bayesian Networks while Specificity and Affirmative Predicted Value showed similarity in both approaches.