International Journal of Mechatronics, Electrical and Computer Technology

Special Issue 2014, pp 1228-1246, **ISSN**: 2305-0543
Available online at: http://www.aeuso.org
© Austrian E-Journals of Universal Scientific Organization

Artificial Intelligence and Expert Systems in Medicine and Their Ability to Prediction as Therapy Planning Systems by CADIAG-2 Algorithm

Mohammad Madadpour Inallou[1*], Zeinab Ajurlou[2] and Bahman Mehri[2]

[1] Young Researchers and Elite Club, West Tehran Branch, Islamic Azad University, Tehran, Iran

[2] Department of Mathematics, Sharif University of Technology, Tehran, Iran

*Corresponding Author's Email: *Madadpour@yahoo.com*

Abstract

Expert Systems in Medicine is a *collection, storage, retrieval, communication* and *processing* of medical data for the purposes of interpretation, inference, decision-support, research and so other purposes in medicine. Expert System is an interactive computer-based decision tool that uses both facts and heuristics to solve difficult decision problems based on knowledge acquired from an expert. Expert systems provide expert advice and guidance in a wide variety of activities, from computer diagnosis to delicate medical surgery and so more other. CADIAG-2 is a well known rule-based medical expert system aimed at providing support in medical diagnose in the field of internal medicine. This paper employs CADIAG-2 fuzzy set theory and fuzzy logic to formalize medical entities and relationships and uses a large collection of *IF-THEN* rules to represent *uncertain* relationships between distinct medical entities. This paper is organized as follows: At first, the introduction and fundamental of CADIAG-2, in second step notation and preliminary definitions is regarded, and in third part the Algorithm will be consider, and the final step is discussed about satisfiablity and Unsatisfiable of CADIAG-2 sets.

Keywords: CADIAG-2, Expert Systems, Fuzzy Logic, Clinical Decision Support Systems (CDSS) and Therapy Planning Systems.

1- Introduction

Knowledge is the enemy of disease. The application of what we know will have a bigger impact than any drug or technology likely to be introduced in the next decade. Many science-based fields are facing a “knowledge crisis”, in that knowledge is expanding explosively while economic

Special Issue 2014, pp 1228-1246, ISSN: 2305-0543
Available online at: http://www.aeuso.org
© Austrian E-Journals of Universal Scientific Organization

resources and the human capacity to apply it remain finite. Medicine is perhaps the most prominent example, the field is the focus of a vast and productive research effort, but new tools and techniques are not always either quickly disseminated or effectively used [7]. In the last thirty years, pioneering clinicians and computer scientists have developed remarkable information technologies that could hold the key to managing that ever-growing complexity. Inspired by that commitment, research on computerized Clinical Decision Support Systems (CDSS) has blossomed into a global movement, with applications in most areas of medicine [19]. This paper is the introduction to Expert Systems-CADIAG-2 (Computer Assisted DIAGnosis).

Its fundamental and the ability to assist in decision making by making predictions, based on existing data and so can help in therapy planning, information seeking, the generation of alerts and etc. Clinical decision support systems are also more commonly known as Expert Systems [12]. By definition, an expert system is a computer program that simulates the thought process of a human expert to solve complex decision problems in a specific domain. In other hand, "An *expert system* is an interactive computer-based decision tool that uses both facts and heuristics to solve difficult decision problems based on knowledge acquired from an expert" [4]. Expert systems provide expert advice and guidance in a wide variety of activities, from computer diagnosis to delicate medical surgery and so more other [5]. The knowledge base CADIAG-2 (Computer Assisted DIAGnosis) consists of a set of *IF-THEN* rules also known in the literature as *production* rules– intended to represent relationships between distinct medical entities: symptoms, findings, signs and test results on the one hand and diseases and therapies on the other [1].

Such assistance may be to an inexperienced doctor regarding diagnosis, or helping even an experienced physician in a situation which is too complex for them to handle alone [12]. CADIAG-2 is a well known rule-based medical expert system aimed at providing support in medical diagnose in the field of internal medicine. Its knowledge base consists of a large collection of *IF-THEN* rules that represent *uncertain* relationships between distinct medical entities [19]. Given this uncertainty and the size of the system it has been challenging to validate its consistency. Recent attempts to partially formalize CADIAG-2's knowledge base into decidable Gödel logics have shown that, on that formalization, the system is inconsistent. In this paper will be adopted an alternative, more

Special Issue 2014, pp 1228-1246, ISSN: 2305-0543
Available online at: http://www.aeuso.org
© Austrian E-Journals of Universal Scientific Organization

expressive formalization of CADIAG-2's knowledge base as a set of probabilistic conditional statements and apply the state-of-the-art probabilistic logic solver to confirm its inconsistency and to compute all its conflicting sets of rules under a slightly relaxed interpretation of them [17].

2- Fundamental of CADIAG-2

2-1. Heuristics

Heuristics are strategies for solving problems that do so by trial and error with regards to some goal, whose achievement is the aim. Procedural methods of problem solving use some predefined sequence of operations and whilst they may work sometimes, there is no guarantee that they will work all the time. In contrast to this, an approach using heuristics will naturally try a number of different possibilities and so offer greater flexibility and wider scope for finding a solution [11]. CADIAG-2 consists of two fundamental pieces:

- Knowledge in the form of facts / *rules* stored in a knowledge base.
- Inference engine.
- Explanation facilities

Knowledge: is represented as rules in decision support systems, which represent what the system "knows" about the world we are modeling with it or what it can "deduce" from a given situation [8]. Here is an example of a rule:

Example1:

IF suspicion of liver metastases by liver palpation
THEN pancreatic cancer
With degree of confirmation 0.3

Example2:

IF the stain of the organism is gram negative,
AND the morphology of the organism is rod
AND the aerobicity of the organism is anaerobic

International Journal of Mechatronics, Electrical and Computer Technology

Special Issue 2014, pp 1228-1246, ISSN: 2305-0543
Available online at: http://www.aeuso.org
© Austrian E-Journals of Universal Scientific Organization

THEN there is strongly suggestive evidence 0.8 that the class of the organism is Enterobacter iaceae [15].

In CADIAG-2 such rules as the one above specify the outcome for a particular situation. The IF part of the rule consists of a conjunction of facts representing some situation that if true, causes this rule to be chosen when solving some problem. The *THEN* part is like the logical consequence of the situation we are currently in. So, in the above example, "there is strongly suggestive evidence (0.8) that the class of the organism is Enterobacter iaceae "is the consequence of the conjunction of facts being true. So from the current situation, the system draws that conclusion. So, the user would provide knowledge about the current situation and the system, using heuristics would use this knowledge and its set of rules to reach some conclusion (a suggestion of what action should be taken) or infer new clinical knowledge [16].

***Inference engine*:** The mechanism of the CADIAG-2 that takes the knowledge and then searches through the rules to draw some conclusion is called the inference engine. It is the intelligence of the system. The inference engine can work in one of two ways:

1. Forward chaining
2. Backward chaining

With forward chaining, the inference engine starts from the knowledge it currently possesses and works towards the goal by means of a *"using all you know"* approach. Backwards chaining on the other hand involves starting from the main goal and deducing what would need to be true to achieve it and making that the new goal and achieving that in a similar manner until there is nothing left to achieve. It is perhaps more efficient in that at each stage we deduce a new goal knowing that if we can satisfy it, then our original goal is achieved [9]. Forward chaining may introduce many different possibilities, each of which is not guaranteed to lead to the main goal. Some of the knowledge inferred may be irrelevant and a waste of time and as the new knowledge that has been deduced is used the next time around, it increases the weight of the knowledge base during inference, which may affect performance [17].

International Journal of Mechatronics, Electrical and Computer Technology

Special Issue 2014, pp 1228-1246, ISSN: 2305-0543
Available online at: http://www.aeuso.org
© Austrian E-Journals of Universal Scientific Organization

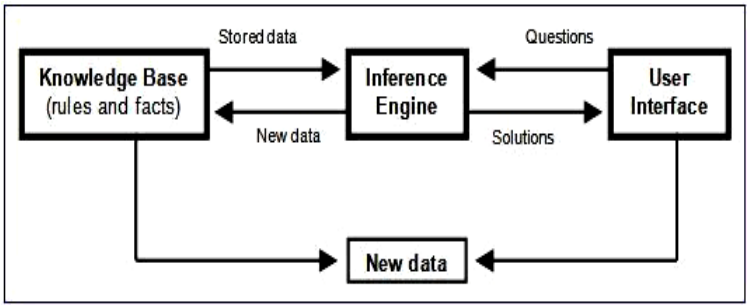

Figure1: General Framework for Expert Systems-CADIAG-2 Knowledge [21].

***Explanation facilities*:** Clinicians often have to give justifications for decisions they have made or will make in cases, for example, where the decision could be potentially controversial, at the request of the patient, or to make the patient aware of any risks and the alternative treatments available. This leads on to another feature of decision support systems, namely explanation. This refers to the ability to explain *how* it reached a particular conclusion (the logic behind it) [6]. The system may also ask the user questions in the event that it doesn't have enough information to proceed and so it can also explain *why* it is asking the question. For example, the system may suggest a treatment based on details of the patient's condition that it is aware of. The clinician may then wish to know why such a treatment will treat the patient [18].

2-2. Notation and Preliminary Definitions

Throughout we will be working with a finite propositional language $L = \{p_1, ..., p_l\}$. We will denote by *SL* its closure under Boolean connectives. Within the context of CADIAG-2 the language *L* will represent the set of medical entities $S \cup D$ in the system.

Definition 1: Let $\omega : SL \to [0,1]$. We say that ω is a probability function on *L* if the following two conditions hold, for all $\theta, \phi \in SL$:

- If ‘ θ then $\omega(\theta) = 1$.
- If ‘ $\neg(\theta \wedge \phi)$ then $\omega(\theta \vee \phi) = \omega(\theta) + \omega(\phi)$

International Journal of Mechatronics, Electrical and Computer Technology

Special Issue 2014, pp 1228-1246, ISSN: 2305-0543
Available online at: http://www.aeuso.org
© Austrian E-Journals of Universal Scientific Organization

Definition 2: We say that the probability function ω on L *satisfies* $(\Delta, v) \in \mathrm{PK}$ if, for all $\phi \mid \theta \in \Delta$, we have that $\omega(\theta) > 0$ and $\omega(\phi \mid \theta) \in v(\phi \mid \theta)$. (*satisfiable* or *consistent*) [3].

Definition 3: We say that (Δ, v) is a *minimal unsatisfiable set* (or minimal inconsistent set) if (Δ, v) is not satisfiable and, for all $\Delta' \subset \Delta$, $(\Delta', v_{|\Delta'})$ is satisfiable.

Definition 4: An interpretation I of L is a pair $(D^{\mathrm{I}}, v^{\mathrm{I}})$ where D^{I} is a finite non-empty domain and v^{I} is a map from $L \times D^{\mathrm{I}}$ to [1], [9]. An interpretation I is said to be *classical* if $v^{\mathrm{I}}(p, a) \in \{0,1\}$ for all $(p, a) \in L \times D^{\mathrm{I}}$. It is said to be *rational* if $v^{\mathrm{I}}(p, a) \in [0,1] \cap \mathbb{Q}$ for all $(p, a) \in L \times D^{\mathrm{I}}$. Given an interpretation I of L, we will refer to the elements in D^{I} by Latin characters a, b, c… [3], [10].

2-3. The Knowledge Base of CADIAG-2

We can classify CADIAG-2's binary rules (Φ_{CB}) into three different types:

- Rules in which both conditioning and uncertain event are medical entities in S (*symptom-symptom*, $\Phi_{S|S}$).
- Rules in which both conditioning and uncertain event are medical entities in D (*disease-disease*, $\Phi_{D|D}$).
- Those in which the conditioning event is a medical entity in S and the uncertain event an entity in D (*symptom-disease*, $\Phi_{D|S}$) [20].

Let $\langle q \mid p, \eta \rangle \in \Phi_{CB}$ be a rule in CADIAG-2's binary knowledge base, with $p, q \in L$ and $\eta \in [0,1] \cap \mathbb{Q}$. The value η is intended to quantify the *degree* to which p (the conditioning event) *confirms* q (the uncertain event), claimed in most of the literature on CADIAG-2 to have been calculated from a certain database or interpretation I as follows:

International Journal of Mechatronics, Electrical and Computer Technology

Special Issue 2014, pp 1228-1246, ISSN: 2305-0543
Available online at: http://www.aeuso.org
© Austrian E-Journals of Universal Scientific Organization

$$\frac{\sum_{a\in D^{\mathrm{I}}} min\{v^{\mathrm{I}}(p,a), v^{\mathrm{I}}(q,a)\}}{\sum_{a\in D^{\mathrm{I}}} v^{\mathrm{I}}(p,a)} = \eta . \tag{1}$$

Definition 5. We say that I is a model of Φ (denoted ‘$_{\mathrm{I}}$ Φ) if $(q \mid p)_{\mathrm{I}} \in \Omega$ for all $\langle q \mid p, \Omega\rangle \in \Phi$.

Proposition 1. $\Phi \in \mathrm{P}\,\mathrm{K}_L$ Has a classical model if and only if it has a rational model.

Proof. The right implication follows trivially from the fact that every classical interpretation is also rational. In order to prove the left implication let us assume that $\mathrm{I} = (D^{\mathrm{I}}, v^{\mathrm{I}})$ is a rational interpretation such that ‘$_{\mathrm{I}}$ Φ [2].

3- Algorithms

In this subsection PSAT briefly will be sketched and conflict-finding algorithms implemented in *Pronto* within the frame of classical propositional logic (since for Φ_{CB} formalization in terms of the probabilistic description logic language is not necessary) [13].

Probabilistic satisfiability algorithm: For the sake of clarity and brevity we will consider the case of PSAT for sets of probabilistic statements of the form $\langle \phi \mid \bullet, \eta\rangle$ on L (a finite propositional language), with $\phi \in SL$ and where η represents the probability assigned to it (i.e., all probabilistic statements considered are unconditional and assigned point-valued probabilities. It is straightforward, but technically awkward and space consuming, to generalize the procedure to handle conditional interval statements for a more detailed account [9]. A collection of probabilistic conditional statements of such form, say Φ, is satisfiable if and only if the objective value of the following linear program is equal to 1:

Maximize $\sum_{\alpha\in W} x_\alpha$ given the following constraints:

$$\sum_{\alpha\in W_\phi} x_\alpha = \eta \times \sum_{\alpha\in W} x_\alpha, \text{ for each } \langle \phi \mid \bullet, \eta\rangle \in \Phi, \tag{2}$$

Special Issue 2014, pp 1228-1246, ISSN: 2305-0543
Available online at: http://www.aeuso.org
© Austrian E-Journals of Universal Scientific Organization

$$\sum_{\alpha \in W} x_\alpha \leq 1 \text{ and all } x_\alpha \geq 0, \tag{3}$$

where x_α is the assignment to the possible world $\alpha \in W$. Let A denote the matrix of linear coefficients in (2). At every step of the simplex algorithm, A is represented as a combination (B, N) where B and N are the submatrices of the *basic* and *non-basic* variables, respectively. Values of non-basic variables are fixed to zero and the solver proceeds by replacing one basic variable (i.e., column in A) by a non-basic one until the optimal solution is found. As the size of A is exponential in the size of our language L, one should determine the entering column without representing A explicitly [14]. This is done using the column generation technique in which entering columns are computed by optimizing a subproblem, sometimes referred to as the *pricing-out problem* (or POP).

Observe that the above system of linear inequalities always admits a solution (e.g. $x_\alpha = 0$ for all $\alpha \in W$), even if Φ is unsatisfiable, which facilitates the column generation process. Note, however, that the actual linear programs solved in *Pronto* are considerably more involved (in particular, they include slack and stabilization variables to improve convergence) [8]. The critical step is to formulate linear constraints for the pricing-out problem such that every solution (a column) corresponds to a possible world in W. In the propositional case this can be done by employing a well known formulation of SAT as a mixed-integer linear program. In the case of expressive languages, such as those in description logics, there appears to be no easy way of determining a set of constraints H for the pricing-out problem such that its set of solutions is in one-to-one correspondence with W [15], [17].

Pronto implements a novel *hybrid* procedure to compute H iteratively via interaction with a description logic reasoner. The main idea of the algorithm is that every column produced as a solution to the pricing-out problem is converted to a description logic concept expression which is then checked for satisfiability by the description logic reasoned [6], [13], [21]. If the expression is satisfiable, it means that the column corresponds to a possible world (in W in our context) and can be added to (2). Otherwise, the justifications of unsatisfiability are converted into linear constraints

Special Issue 2014, pp 1228-1246, ISSN: 2305-0543
Available online at: http://www.aeuso.org
© Austrian E-Journals of Universal Scientific Organization

and added to the pricing-out problem, which is then re-optimized. Finally, either an entering column is found or the pricing-out problem becomes infeasible (which implies that the system (2) is optimal). A detailed description of the PSAT algorithm is beyond the scope of this paper and is left as the core of a future publication [1], [5], [9].

Conflict-finding Algorithm: A satisfiability algorithm is generally not sufficient for a comprehensive analysis of an inconsistent knowledge base (either classical or probabilistic). Typically users need to identify those fragments of the knowledge base which cause the inconsistency in order to repair them. Such fragments are typically required to be *minimal* so that the user can choose a repair strategy with minimal impact on the rest of the knowledge base. We are interested in determining the minimal unsatisfiable subsets of a certain collection of probabilistic conditional statements in L, say Φ [12], [17]. We apply the classical approach to finding minimal unsatisfiable sets based on hitting sets which dates back to Reiter. Reiter's hitting set tree (HST) algorithm requires, as a subroutine, a satisfiability procedure which can extract one minimal unsatisfiable set from the knowledge base. It then systematically removes each axiom from that minimal unsatisfiable set and applies the satisfiability procedure again to generate a new minimal unsatisfiable set [2], [20].

By being systematic in the ‘repairs’, the procedure finds all minimal unsatisfiable sets in the knowledge base. We reduce the problem of finding a single minimal unsatisfiable subset of Φ to the problem of finding a minimal infeasible subset of inequalities in the corresponding linear system of the form (2) above. Such subsets are known as *Irreducible Infeasible Systems* (IIS) in the linear programming literature. However, given that the system (2) is never represented in its full version the application of the Ryan and Parker's algorithm is far from straightforward. If the optimal value of the system (2) is less than 1 then some inequalities have non-zero dual values [11], [19]. Such inequalities correspond to conflicting constraints in Φ but are not guaranteed to be minimal (though is typically quite small and close to the minimal set). We then do a brute force trial and error search to remove all superfluous constraints [13].

International Journal of Mechatronics, Electrical and Computer Technology

Special Issue 2014, pp 1228-1246, ISSN: 2305-0543
Available online at: http://www.aeuso.org
© Austrian E-Journals of Universal Scientific Organization

4- Checking Satisfiability of CADIAG-2

One of the main goals of Expert Systems is to determine satisfiability or unsatisfiability of Φ_{CB} and, therefore, solve the PSAT problem (Probabilistic SATisfiability) for Φ_{CB}. Provided that Φ_{CB} is unsatisfiable (which turns out to be the case) we also want to compute all its minimal unsatisfiable subsets. Despite the fact that Φ_{CB} is formalized as a propositional knowledge base and the existence of propositional PSAT algorithms we use *Pronto*, this pointed PSAT solver designed for more expressive probabilistic description logics for this purposes [14], [18]:

- First, unlike propositional solvers, it treats classical and probabilistic knowledge separately and scales well with respect to the amount of the former. Φ_{CB} Contains many classical rules (for example, the number of rules in $\Phi_{D|D}$ is over 200) and so, given the scalability limits of PPL solvers (about 1000 statements), they are likely to be unable to handle a sufficient number of uncertain *symptom-disease* rules in addition to $\Phi_{D|D}$ and (a fragment of) $\Phi_{S|S}$ [20].
- Second, *Pronto* has *pinpointing* capabilities for finding all minimal unsatisfiable subsets in an unsatisfiable knowledge base. This feature is critical in the context of this work given the size of Φ_{CB} and, as we will see shortly, the number of potentially overlapping inconsistencies. It must be noted that finding *all* minimal unsatisfiable subsets is by no means a trivial extension of the PSAT algorithm (for example, its naïve implementation using a PSAT solver as a black-box reasoner is not practical due to the hardness of PSAT). Even with our implementation, the number of unsatisfiable subsets in a range of fragments of Φ_{CB} presents a significant challenge to the reasoned [9], [12].
- Last but not least, we are interested in evaluating our algorithms (see the next subsection) on such a large and naturally occurring knowledge base as Φ_{CB} [6], [14].

International Journal of Mechatronics, Electrical and Computer Technology

Special Issue 2014, pp 1228-1246, ISSN: 2305-0543
Available online at: http://www.aeuso.org
© Austrian E-Journals of Universal Scientific Organization

4-1. Decomposition of Φ_{CB}

To our knowledge, none of the existing probabilistic solvers can solve PSAT for Φ_{CB} taken as a whole within a reasonable amount of time (see the table below for a precise account of the size of Φ_{CB}). However, Φ_{CB} has a certain structure that allows splitting it into fragments that can be examined independently. A crucial property of our probabilistic formalization is that Φ_{CB} is satisfiable if and only if all of the fragments are individually satisfiable, as shown below [15], [16].

Table 1: individually satisfiable fragments [5]

Number of distinct symptoms	**1761**
Number of distinct diseases	**341**
Number of *symptom-symptom* rules	**720**
Number of *disease-disease* rules	**218**
Number of *symptom-disease* rules	**17573**

We can regard Φ_{CB} as a directed graph where the nodes are the medical entities in L and the edges are given by the rules in Φ_{CB} (i.e., a rule of the form $\langle q \mid p, \eta \rangle$ in Φ_{CB} would correspond to an edge directed from p to q). Let $p \in L$. We denote by $\Phi_p \subseteq \Phi_{CB}$ the set of rules that yield a directed edge in a path from p to any other medical entity in L or a directed edge in a path from any medical entity in L to p. For the next two results let us consider two medical entities $p_1, p_2 \in S$ and assume that there is no path from p_1 to p_2 in Φ_{CB} or vice versa and that there is no medical entity $p \in L$ from which there exists a path to both p_1 and p_2 [2], [3], [8].

Proposition 4. If Φ_{p_1} and Φ_{p_2} are satisfiable then $\Phi_{p_1} \cup \Phi_{p_2}$ is satisfiable.

Proof. Let I_1 and I_2 be models of Φ_{P_1} and Φ_{P_2} respectively. We can assume without loss of generality that $D^{\mathrm{I}_1} \cap D^{\mathrm{I}_2} = \varnothing$. We can construct a model I of $\Phi_{p_1} \cup \Phi_{p_2}$ from I_1 and I_2 in a

International Journal of Mechatronics, Electrical and Computer Technology

Special Issue 2014, pp 1228-1246, ISSN: 2305-0543
Available online at: http://www.aeuso.org
© Austrian E-Journals of Universal Scientific Organization

pretty trivial way by setting $D^{\mathrm{I}} = D^{\mathrm{I}_1} \cup D^{\mathrm{I}_2}$ and $\nu^{\mathrm{I}}(p,a) = 1$ if and only if $\nu^{\mathrm{I}_1}(p,a) = 1$ or $\nu^{\mathrm{I}_2}(p,a) = 1$ and 0 otherwise, for all $(p,a) \in L \times D^{\mathrm{I}}$. In can be easily seen that I thus defined satisfies $\Phi_{p_1} \cup \Phi_{p_2}$ [17], [19].

Corollary 2. If Φ is a minimal unsatisfiable set of rules in $\Phi_{p_1} \cup \Phi_{p_2}$ then Φ needs to be contained in Φ_{p_1} or Φ_{p_2}.

Proof. It follows trivially from Proposition 4. Notice that if Φ were a minimal unsatisfiable subset of $\Phi_{p_1} \cup \Phi_{p_2}$ and that it were neither contained in Φ_{p_1} nor in Φ_{p_2} then we could define satisfiable subsets from Φ_{p_1} and Φ_{p_2} of the same structure (possibly by removing rules from other minimal unsatisfiable subsets of Φ_{p_1} and Φ_{p_2}, but none from Φ). Although trivial, it is worth mentioning that the previous propositions also hold for any alternative interpretation of the rules in terms of probabilistic intervals (i.e., by taking certain real intervals in place of point-valued probabilities). Φ_{CB} has the following properties which will enable us to decompose it into a set of fragments:

- All conditional statements contain only atomic medical entities (i.e., entities in L).
- *Uncertain* rules (i.e., non-classical, $\Phi_{D|S}$) condition only on symptoms.
- The graph of $\Phi_{S|S}$ contains many disconnected subgraphs [21].

We split Φ_{CB} into a set of fragments of the form Φ_p, where $p \in S$ is a symptom such that there is no rule in Φ_{CB} of the form $\langle p \mid q, \eta \rangle$. For simplicity we include the entire $\Phi_{D|D}$ in each fragment since it is decomposable to a much less extent than $\Phi_{S|S}$. The largest fragments have around 200 conditional statements that normally relate two or three connected symptoms to diseases. Corollary 2 guarantees that all minimal unsatisfiable sets of rules in Φ_{CB} can be found by computing such sets for each fragment. Thus our methodology is simply a systematic analysis of the sets of the

International Journal of Mechatronics, Electrical and Computer Technology

Special Issue 2014, pp 1228-1246, ISSN: 2305-0543
Available online at: http://www.aeuso.org
© Austrian E-Journals of Universal Scientific Organization

form Φ_p, which involves a PSAT test and (if the fragment is unsatisfiable) the computation of all minimal unsatisfiable sets in it (see the algorithms in the two previous subsections) [3], [7].

4-2. Minimal Unsatisfiable Sets in Φ_{CB}

We present here results concerning the satisfiability check of Φ_{CB} when considering a slightly relaxed interpretation of Φ_{CB} by replacing each rule of type *symptom-disease* of the form $\langle q \mid p, \eta \rangle \in \Phi_{CB}$, for some $p, q \in L$ and $\eta \in (0,1) \cap \square$, by $\langle q \mid p, \Omega_\eta \rangle$, with $\Omega_\eta = [\eta - 0.01, \eta + 0.01] = [\underline{\eta}, \overline{\eta}]$. (The degrees of confirmation of the rules in Φ_{CB} are all of the form $k/100$, for some $k \in \{0,1,...,100\} \subset \square$, and thus Ω_η is well defined) [4], [14]. We have opted for checking satisfiability of this slightly relaxed interpretation of the rules in Φ_{CB} against a point-valued interpretation (i.e., the standard interpretation with point-valued probabilities) because of time constraints. The implementation of our algorithms for the relaxed interpretation of Φ_{CB} completes the task of finding all minimal unsatisfiable subsets in a reasonable amount of time (around one hour).

It is a well-known fact in model-diagnosis theory that computing *all* minimal unsatisfiable subsets of a certain knowledge base requires a number of satisfiability tests (in our case, PSAT tests) that is exponential in the number of unsatisfiable subsets [10]. Our relaxed interpretation of Φ_{CB} already contains a high number of minimal unsatisfiable sets (which also happen to be minimal under a point-valued interpretation) and a point-valued interpretation adds up more. Furthermore, some of the unsatisfiable sets that are present in the point-valued interpretation and not in our relaxed one are relatively large (some contain 7 rules) and do not overlap with other unsatisfiable sets. Such facts bring the algorithm's running time closer to its worst case. An example of a type of minimal unsatisfiable set detected under a precise interpretation of the rules but not under our relaxed version is the one that follows:

$$\langle q_1 \mid p_1, \eta_1 \rangle, \langle q_2 \mid p_1, \eta_2 \rangle, \langle q_1 \mid p_2, \eta_3 \rangle, \langle q_3 \mid p_2, \eta_4 \rangle, \langle q_3 \mid q_1, 1 \rangle, \langle q_3 \mid q_2, 1 \rangle, \langle p_2 \mid p_1, 1 \rangle, \quad (4)$$

International Journal of Mechatronics, Electrical and Computer Technology

Special Issue 2014, pp 1228-1246, ISSN: 2305-0543
Available online at: http://www.aeuso.org
© Austrian E-Journals of Universal Scientific Organization

for $p_1, p_2 \in S$, $q_1, q_2, q_3 \in D$, $\eta_1, \eta_2, \eta_3, \eta_4 \in [0,1]$, with $\eta_3 = \eta_4$ and $\eta_1 < \eta_2$. Notice that the rules $\langle q_1 \mid p_2, \eta_3 \rangle$ and $\langle q_3 \mid p_2, \eta_4 \rangle$ along with $\langle q_3 \mid q_1, 1 \rangle$ intuitively claim that the set of patients with symptom p_2 and disease q_2 coincides with the set of patients with symptom p_2 and disease q_3 when assuming $\eta_3 = \eta_4$ [1]. Under such an assumption the rules $\langle q_1 \mid p_1, \eta_1 \rangle$ and $\langle q_2 \mid p_1, \eta_2 \rangle$ along with the remaining classical rules generate an inconsistency whenever $\eta_1 < \eta_2$. Notice also that, for example, for $\eta_3 < \eta_4$ the set would not be unsatisfiable and thus our relaxed interval interpretation would yield this set consistent (assuming $\eta_3, \eta_4 < 1$). For the sake of simplicity we will adopt the same notation for the rules of type *symptom-disease* of the form $\langle q \mid p, \eta \rangle$, with $\eta \in \{0,1\}$. We will write $\langle q \mid p, \Omega_\eta \rangle$, with $\Omega_\eta = [\eta, \eta] = [\underline{\eta}, \overline{\eta}]$. We list the different types of minimal unsatisfiable sets encountered in Φ_{CB} under this relaxed interpretation of the rules [8]:

Type 1. Our first type of minimal unsatisfiable set in Φ_{CB} is given by a collection of rules of the form $\langle q_1 \mid p, \Omega_\eta \rangle$, $\langle q_2 \mid p, \Omega_\zeta \rangle$, $\langle q_2 \mid q_1, 1 \rangle$,

for $p \in S$, $q_1, q_2 \in D$, $\eta, \zeta \in [0,1]$ and $\overline{\zeta} < \underline{\eta}$.

By $\overline{\zeta} < \underline{\eta}$ we are intuitively assuming that the number of patients that have both symptom p and disease q_1 is greater than the number of patients with both symptom p and disease q_2, which contradicts $\langle q_2 \mid q_1, 1 \rangle$ (i.e., the assumption that all patients that have disease q_1 have also disease q_2) [18].

Type 2. Our second type of minimal unsatisfiable set in Φ_{CB} is given by a set of rules of the form $\langle q_1 \mid p, \Omega_\eta \rangle$, $\langle q_2 \mid p, \Omega_\zeta \rangle$, $\langle q_2 \mid q_1, 0 \rangle$, for $p \in S$, $q_1, q_2 \in D$, $\eta, \zeta \in [0,1]$ and $\underline{\eta} + \underline{\zeta} > 1$.

Notice that the rule $\langle q_2 \mid q_1, 0 \rangle$ assumes *disjointness* between q_1 and q_2 (intuitively, there cannot be a patient with both disease q_1 and q_2), which rules out the possibility of consistency whenever $\underline{\eta} + \underline{\zeta} > 1$ [16].

International Journal of Mechatronics, Electrical and Computer Technology

Special Issue 2014, pp 1228-1246, ISSN: 2305-0543
Available online at: http://www.aeuso.org
© Austrian E-Journals of Universal Scientific Organization

Type 3. The third type of minimal conflict set in Φ_{CB} is given by a set of the form $\langle q \mid p_1, \Omega_\eta \rangle$, $\langle q \mid p_2, \Omega_1 \rangle$, $\langle p_2 \mid p_1, 1 \rangle$, for $p_1, p_2 \in S$, $q \in D$, $\eta \in [0,1]$ and $\bar{\eta} < 1$. Intuitively, the rule $\langle p_2 \mid p_1, 1 \rangle$ says that all patients with symptom p_1 also have symptom p_2. The rule $\langle q \mid p_2, \Omega_1 \rangle$ intuitively says that all patients with symptom p_2 have disease D. These two facts together imply that patients with symptom p_1 should all have disease D (i.e., $\bar{\eta} = 1$).

Type 4. The fourth and last type of minimal unsatisfiable set is given by a collection of rules of the form $\langle q_1 \mid p, \Omega_\eta \rangle$, $\langle q_2 \mid p, \Omega_\zeta \rangle$, $\langle q_3 \mid p, \Omega_\lambda \rangle$, $\langle q_3 \mid q_1, 1 \rangle$, $\langle q_3 \mid q_2, 1 \rangle$, $\langle q_2 \mid q_1, 0 \rangle$, With $p \in S$, $q_1, q_2, q_3 \in D$, $\eta, \zeta, \lambda \in [0,1]$, $\bar{\lambda} < \underline{\eta} + \underline{\zeta} \leq 1$ and $\underline{\zeta}, \underline{\eta} \leq \bar{\lambda}$ (to guarantee minimality). Intuitively, assuming $\langle q_1 \mid p, \Omega_\eta \rangle$, $\langle q_2 \mid p, \Omega_\zeta \rangle$ and $\langle q_2 \mid q_1, 0 \rangle$, the proportion of patients that, having symptom p, have either disease q_1 or q_2 is at least $\underline{\eta} + \underline{\zeta}$. On the other hand, assuming $\langle q_3 \mid q_1, 1 \rangle$ and $\langle q_3 \mid q_2, 1 \rangle$, we have that all patients with disease either q_1 or q_2 have also disease q_3. Thus, under such assumptions, satisfiability requires that $\bar{\lambda} \geq \underline{\eta} + \underline{\zeta}$. The following table shows the amount of minimal unsatisfiable sets found for each type and the number of rules involved in them [4], [6].

Table 2: amount of minimal unsatisfiable sets found for each type and the number of rules involved

Type	Amount	Number of rules involved
1	420	**3**
2	5	**3**
3	1	**3**
4	**269**	**6**

A thorough analysis of these types of inconsistencies in connection with the whole knowledge base and with possible repair strategies and in relation to other sets of inconsistencies obtained under alternative interpretations of Φ_{CB} (as briefly pointed above, under the standard point-valued

Special Issue 2014, pp 1228-1246, ISSN: 2305-0543
Available online at: http://www.aeuso.org
© Austrian E-Journals of Universal Scientific Organization

interpretation of the rules or when regarding η in $\langle q \mid p, \eta \rangle \in \Phi_{CB}$ as a lower-bound threshold) is an ongoing research challenge [11].

5- EVALUATING INCONSISTENCY IN Φ_{CB}

While we have determined that Φ_{CB} is inconsistent and have found various sets of statements and patterns to account for that inconsistency, it is not entirely clear how to interpret this information. In particular, we may wish to determine *how* inconsistent Φ_{CB} is. We have just seen that Φ_{CB} contains a large number of minimal inconsistent subsets. That, on a straightforward reading, makes Φ_{CB} *highly* inconsistent as each inconsistent subset is *an inconsistency* and thus Φ_{CB} contains a lot of inconsistencies [8]. Such a reading is the basis of several approaches in the literature to *measuring* the amount of inconsistency of a database (or even of each particular statement in the database based on the number and composition of its minimal inconsistent subsets. Clearly, under some of these approaches Φ_{CB} would certainly be regarded as *highly* inconsistent. However, while appealing, this is not the only way to measure inconsistency. We consider here an alternative approach to measuring the degree of inconsistency by considering *how far* the knowledge base is from consistency, with the ultimate goal of giving an alternative evaluation of the amount of inconsistency of Φ_{CB} and providing a tool for determining possible, *minimal* repair strategies for Φ_{CB} and similar knowledge bases [5], [7].

5-1. Measuring $DC(\Phi_{CB})$

The computation of *DC* is certainly not a trivial matter for the general case. The main difficulty stems from our definition of satisfiability (i.e., from the fact that conditioning events are required to have a strictly positive probability). Some heuristics can be of much help in this matter. In fact, in order to compute $DC(\Phi_{CB})$, we can take advantage of its *simple* structure, as we will shortly show. For the next proposition let $(\Delta, v) \in \mathrm{PK}^c$ be as follows:

- $\phi, \theta \in L$ for all $\phi \mid \theta \in \Delta$.

International Journal of Mechatronics, Electrical and Computer Technology

Special Issue 2014, pp 1228-1246, ISSN: 2305-0543
Available online at: http://www.aeuso.org
© Austrian E-Journals of Universal Scientific Organization

- (Δ, v) is, as a directed graph, acyclic.
- $v(\phi \mid \theta) \in (0,1) \cap \mathbb{Q}$, for all $\phi \mid \theta \in \Delta$ (5)

Proposition 5. (Δ, v) is satisfiable.

Proof. By Proposition 3 the pair (Δ, v) will be satisfiable if and only if it has a classical model. We aim at constructing a classical model of (Δ, v). Let $C = \{v(\phi \mid \theta) \mid (\phi \mid \theta) \in \Delta\}$. By assumption on (Δ, v), C will be a set of rational values. Let Q be the least common multiple of the denominators of all the values in C. For the construction of I we consider first the set of *nodes* in (Δ, v) that have no *parents*, when dealing with (Δ, v) as a graph. Let p be a node in (Δ, v) with no parents. We set a collection of Q elements for p, labeled $D_p = \{a_1^p, ..., a_Q^p\}$. At the nth step in the construction of I, we select a node q in (Δ, v), with parents $q_1, ..., q_k$ (for some $k \in \mathbb{N}$) and assume, without loss of generality, that the sets $D_{q_1}, ..., D_{q_k}$ have been defined at a previous step. We set a collection of $Q(\sum_{i=1}^{k} |D_{q_i}|)$ elements for q, labeled

$$D_q = \{a_1^q, ..., a^q_{Q(\sum_{i=1}^{k} |D_{q_i}|)}\} \tag{6}$$

We impose some restrictions on the sets of the form D_p, for $p \in L$. We construct I in a way that, for $p, q \in L$, $D_p \cap D_q \neq \varnothing$ if and only if one of $p|q$, $q|p$ is in Δ or there exists $r \in L$ such that $p \mid r, q \mid r \in \Delta$. (In general, we could impose the condition that distinct sets of the form D_p, for $p \in L$, should have as little domain elements in common as possible). That this can be achieved is clear given the restrictions on the cardinality of such sets and the fact that $v(\phi \mid \theta)$ is at most $(Q-1)/Q$ for all $\phi \mid \theta \in \Delta$. For $p \in L$, with parents $q_1, ..., q_k$ (for some $k \in \mathbb{N}$) and sets $D_{q_1}, ..., D_{q_k}$ already defined, we will set D_p to have exactly $|D_{q_i}| v(p \mid q_i)$ elements of D_{q_i}, for each $i \in \{1, ..., k\}$. The elements of D_{q_i} chosen for D_p could also be elements of D_{q_j},

International Journal of Mechatronics, Electrical and Computer Technology

Special Issue 2014, pp 1228-1246, ISSN: 2305-0543
Available online at: http://www.aeuso.org
© Austrian E-Journals of Universal Scientific Organization

for $j \in \{1,...,i-1,i+1,...,k\}$, (depending on whether q_i and q_j are related in Δ by a conditional statement or not) but of no other sets thus far defined. That a suitable definition of D_p can be achieved (considering that the sets $D_{q_1},...,D_{q_k}$ might not be pair wise disjoint) is clear given the restrictions on the cardinality of such sets and the fact that $1/Q \le v(\phi \mid \theta) \le (Q-1)/Q$ for all $\phi \mid \theta \in \Delta$. We define $D^{\mathrm{I}} = \bigcup_{p \in L} D_p$ and v^{I} on $L \times D^{\mathrm{I}}$ as expected. For $p \in L$ and $a \in D^{\mathrm{I}}$,

$$v^{\mathrm{I}}(p,a) = \begin{cases} 1 & \text{if } a \in D_p \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

This completes the proof.

Proposition 6. $DC(\Phi_{CB}) = \varepsilon$.

Proof. We know from results in previous sections that Φ_{CB} is inconsistent and thus that $DC(\Phi_{CB}) > 0$ (i.e., at least ε). On the other hand Φ_{CB} is, when regarded as a graph, acyclic and with nodes in L. Replacing 0 and 1 by any values in $(0,1) \cap \square$ as close as desired to 0 and 1 in rules of the form $\langle q \mid p, 0\rangle, \langle q \mid p, 1\rangle \in \Phi_{CB}$ respectively guarantees satisfiability, by Proposition 5. Therefore we can conclude that $DC(\Phi_{CB}) = \varepsilon$ [8], [15].

CONCLUSION

Knowledge base must be clear because any Inconsistency in a knowledge base, can effect on information and their integrity. For reaching this aim, an *Expert System* is noticeable. CADIAG-2 is different in kind. It has a knowledge base and inference engine. While knowledge base is formalized as a probabilistic logic theory, will be highly unsatisfiable. Also inference engine can't be extended when it face an Inconsistency. So we need an inference mechanism (Expert System) to analysis the inconsistency to embody the knowledge base. Thus CADIAG-2 is referred (PSAT). By using the existent roles in CADIAG-2, the inference engine can detect and resolve the conflicts, by internal automated reasoning services in CADIAG-2, to have a clear and transparent knowledge base.

International Journal of Mechatronics, Electrical and Computer Technology

Special Issue 2014, pp 1228-1246, ISSN: 2305-0543
Available online at: http://www.aeuso.org
© Austrian E-Journals of Universal Scientific Organization

References

[1] P. Klinov, B. Parsia, D. Muiño, "The Consistency of the Medical Expert System CADIAG-2: A Probabilistic Approach", Journal of Information Technology Research, 4(1), 1-20, January-March, 2011.

[2] D.J. Power, "What is a DSS?, The Journal for Data-Intensive Decision Support", October 21,2007:Vol. 1, No. 3.

[3] E. Coiera, P.Y. Arnold, "Guide to health informatics, On-Line Executive", January 2007 Vol.6. No.3.pp. 25-39.

[4] P. Szolovits, R.S. Patil, "The roll of AI in medicine, Annals of Internal Medicine", Vol.108; No.1, pp. 80-87. January 2009, Cambridge and Boston, Massachusetts.

[5] K. Darlington, "South Bank University, Basic Expert Systems", London, 2007, Vol.6 No.5, P. 177-217.

[6] J. Andrew, D. Savon, University of Science & Engineering, "Expert Systems Characteristics", pp. 13-27. June 2008, Sami's Publisher, Austria.

[7] B. William, M.D. Schwartz, "Artificial Intelligence in Medical Diagnosis", the Department of Electrical Engineering and Computer Science, Copyright 2008, American College of Physicians.

[8] C.S. Krishnamoorthy; S. Rajeev, "Artificial Intelligence and Expert Systems for Engineers", CRC Press, CRC Press LLC, ISBN: 0849391253, 2d Edition May 2009, pp. 128-153.

[9] S. Rao, C., Krishnamoorthy, and S. Rajeev, "Generic tool for engineering design synthesis (GENSYNT)", Research report R-94-2,Department of Civil Engineering, Indian Institute of Technology, Madras,2007.

[10] S. Fenves, and Baker, N. Spatial, "functional representation language for structural design, Expert Systems in Computer-Aided Design", Ed. Gero, J., Amsterdam: Elsevier, 511, 1987.

[11] P. Klinov, B. Parsia and D.P. Muiño, The University of Manchester, Oxford Road, Manchester, M139PL,UK, and Institut für Diskrete Mathematik und Geometrie,Wiedner Hauptstrasse 8. A-1040 Vienna, Austria.

[12] A. Bilvak, and K. D, "The Rules and Signs of Medical Expert Systems", University of Michigan, 2004.

[13] M. Gray, D. Noulla, University of Medical Informatics-Munich, Motivation and Opportunity, 2008.

[14] T. Kiseliova, "A theoretical comparison of disco and CADIAG-II-like systems for medical diagnoses", Kybernetika, Vol. 42 (2006), No. 6, 723-748.

[15] M. Baaz, A. Ciabattoni, C. G. FermÄuller, "Cut-Elimination in aSequents-of-Relations Calculus for GÄodel Logic", in: Proceedings of the 31st International Symposium on Multiple-Valued Logic (Warsaw 2001), IEEE Computer Society; 181{186, 2001.

[16] http://www.lcs.mgh.harvard.edu/dxplain.htm Lab of Computer Science (MGH).

[17] http://www.bmis.org UK Health Informatics Society (UKHiS).

[18] http://www.hisa.org.au Health Informatics Society of Australia (HISA).

[19] http://www.nursing-informatics.com/kwantlen/nrsg3110.html Medical Expert Systems.

[20] http://www.mieur.nl/mihandbook Handbook of Medical Informatics.

[21] http://www.emergisoft.com Emergency Department Information Systems.