Correspondence

Fuzzy Set Theory in Medical Diagnosis

KLAUS-PETER ADLASSNIG

Abstract—Fuzzy set theory has a number of properties that make it suitable for formalizing the uncertain information upon which medical diagnosis and treatment is usually based. Firstly, it defines inexact medical entities as fuzzy sets. Secondly, it provides a linguistic approach with an excellent approximation to texts. Finally, fuzzy logic offers reasoning methods capable of drawing approximate inferences. These facts suggest that fuzzy set theory might be a suitable basis for the development of a computerized diagnosis system. This is verified by trials performed with the medical expert system CADIAG-2, which uses fuzzy set theory to formalize medical relationships and fuzzy logic to model the diagnostic process.

I. Introduction

It is widely accepted that the information available to the physician about his patient and about medical relationships in general is inherently uncertain. Nevertheless, the physician is still quite capable of drawing (approximate) conclusions from this information. This paper describes an attempt to provide a formal model of this process using fuzzy set theory, and to implement the model in the form of a computerized diagnostic system.

In medicine, the principle of "measuring everything measurable and trying to make measurable that which has not been measurable so far" (Galileo) is still practiced, although some fundamental limitations have been recognized during the course of this century. We now know that real-world knowledge is characterized by incompleteness (implying that the human process of cognition is infinite); inaccuracy (as stated in Heisenberg's Uncertainty Principle); and inconsistency (as may be anticipated by Gödel's Theorem).

Fuzzy set theory, which was developed by Zadeh [1], makes it possible to define inexact medical entities as fuzzy sets. It offers a linguistic approach that represents an excellent approximation to medical texts [2], [3]. In addition, fuzzy logic provides reasoning methods capable of making approximate inferences [4], [5]. These facts suggest that fuzzy set theory might be a suitable basis for the development of a computerized diagnosis system [6]. Current developments and applications of some medical expert systems on the basis of fuzzy set theory and fuzzy logic show that this is indeed the case [7]–[22]. CADIAG-2 (computer-assisted diagnosis), an expert system especially designed for internal medicine, which is presently being clinically tested, will be described in more detail in order to provide an example and report some results.

II. Real-World Knowledge

Precision exists only through abstraction. Abstraction may be defined as the ability of human beings to recognize and select the relevant properties of real-world phenomena and objects. This leads to the construction of conceptual models defining abstract classes of phenomena and objects. However, in actuality every real-world phenomenon and object is of course unique.

Abstract models of real-world phenomena and objects such as mathematical structures (circle, point, etc.), equalities ($a = b + c$), and propositions (yes, no) are artificial constructs. They represent ideal structures, ideal equalities, and ideal propositions.

Nevertheless, despite these caveats, abstraction forms the basis of human thought, and human knowledge is its result.

A. Incompleteness

Abstraction, however, is not a static concept. The process of abstraction is continuous and is constantly producing new results. The set of properties of real-world phenomena and objects under consideration is continually being enlarged and changed. Knowledge is therefore always and necessarily incomplete.

B. Inaccuracy

Unlimited precision is impossible in the real world. Anything said to be "precise" can only be considered as "precise to a certain extent."

The pursuit of maximum precision is still an important aim in science. Galileo, who is often credited with being the father of the quantitative scientific experiment, was certainly responsible for many scientific advances through his philosophy of "measuring everything measurable and trying to make measurable that which has not been measurable so far," although the limitations of this approach should be recognized.

Heisenberg's Uncertainty Principle [23] states the limits to accurate measurement very clearly. Of course, the principle applies only to the world of microphenomena and microobjects, but its philosophical implications go further. It shows that nature possibly is fundamentally indeterministic. And it seems meaningless to ask whether nature inherently lacks determinism or whether uncertainty stems only from experimentation.

C. Inconsistency

Abstraction does not always lead to the same results, which in turn are not always interpreted in the same way. Knowledge may differ according to nation, culture, religion, social status, education, etc., and information from different sources may therefore be inconsistent. To eliminate inconsistency from the information system is only possible in limited systems, and Gödel's Theorem [24] clearly demonstrates that contradictions within a system cannot be eliminated by the system itself.

III. Medical Information

In medicine, it is not necessary to deal with microphenomena and microobjects to run into the problems of incompleteness, uncertainty, and inconsistency. The lack of information, and its imprecise and sometimes contradictory nature, is much more a fact of life in medicine than in, say, the physical sciences. These problems have to be taken into account in every medical decision, where they may have important, even vital consequences for the object of medical attention: the patient.

A. Information About the Patient

Data about the patient can be divided into a number of different categories, which are all characterized by an inherent lack of certainty.

Manuscript received February 20, 1985; revised September 19, 1985. This work was supported in part by a Max Kade Fellowship for Scientific Research, in part by the National Science Foundation under grant NSF-IST 8320416, and in part by NASA under grant NCC2-275.

The author is with the Computer Science Division, Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720, USA, and the Department of Medical Computer Sciences, University of Vienna, Garnisongasse 13, A-1090 Vienna, Austria.

IEEE Log Number 8406786.

0018-9472/86/0300-0260$01.00

1) Medical History of the Patient: The medical history of the patient is given by the patient himself. It is highly subjective and may include simulated, exaggerated, or understated symptoms. Ignorance of previous diseases in himself or his family, failure to mention previous operations, and general poor recollection often raise doubts about a patient's medical history in the mind of the doctor. On the other hand, however, the information that finally leads to the correct diagnosis is very often found here.

2) Physical Examination: The physician subjects the patient to a physical examination from which he obtains more or less objective data. But of course, physicians can make mistakes, overlook important indications, or fail to carry out a complete examination. Furthermore, they may misinterpret other indications because the boundary between normal and pathological status is not always clearly defined.

3) Results of Laboratory Tests: The results of laboratory tests are considered to be objective data. However, measurement errors, organizational problems (mislabeling samples, sending them to the wrong laboratory, etc.), or improper behavior on the part of the patients prior to examinations can lead to imprecise and sometimes even totally incorrect data. Again, the boundaries between normal and pathological results are generally not strict: there are always borderline values that cannot be said to be either normal or pathological.

4) Results of Histological, X-ray, Ultrasonic, and Other Clinical Investigations: These results again depend on correct interpretation by medical or other staff. Such findings are often crucial, because they frequently indicate invasive therapy. In many cases, consideration of uncertainty is part of the evaluation procedure: for example, in cell counts, cell determination, picture analysis, etc.

B. Information on Medical Relationships

Medical knowledge consists of medical descriptions and assertions that are incomplete and uncertain. It has been built up step by step, and is based partly on theoretical studies (in areas such as anatomy and physiology) and partly on almost purely empirical observations (made in the course of surgery, for example). Medical knowledge may be said to comprise knowledge about causal relationships based in theory, statistical information, pure definitions, and personal judgment.

To add to the problem, the elements considered to form medical relationships differ according to place and time, vary between medical schools, and in some cases have not been studied to any significant extent.

C. Medical Inference

This is the process by which the physician uses his medical knowledge to infer a diagnosis from the symptoms displayed by the patient, his lab test results, and his medical history. It is a complex and partly uninvestigated process in which the physician is obviously able to work with uncertain and imprecise sets of data. To some extent it is a subconscious activity, which is why it is probably often called an art.

IV. Medical Expert System CADIAG-2

CADIAG-2 is intended to be an active assistant to the physician in diagnostic situations. In this way the experience, creativeness, and intuition of the physician is supplemented by the knowledge-based computational power of the computer. The general structure of CADIAG-2 is shown in Fig. 1.

A. Representation of Medical Knowledge

CADIAG-2 considers four classes of medical entities:

- symptoms, signs, test results, and findings (S_i)
- diseases, and diagnoses (D_j)
- intermediate combinations (IC_k)
- symptom combinations (SC_l).

Symptoms S_i take values μ_{S_i} in $[0,1] \cup \{\nu\}$. The value μ_{S_i} indicates the degree to which the patient exhibits symptom S_i (a value of ν implies that symptom S_i has not yet been examined). In the language of fuzzy set theory, μ_{S_i} expresses the grade of membership to which the patient's symptom manifestation S_i belongs to the patient. An example of this mode of representation is given in Table I.

A binary fuzzy relationship $R_{PS} \subset \Pi \times \Sigma$ is then established, defined by $\mu_{R_{PS}}(P_q, S_i) = \mu_{S_i}$ for patient P_q, where $P_q \in \Pi$ $(\Pi = \{P_1, \cdots, P_r\})$ and $S_i \in \Sigma$ $(\Sigma = \{S_1, \cdots, S_m\})$.

Diseases or diagnoses also take values in $[0,1] \cup \{\nu\}$. Fuzzy values $0.00 < \mu_{D_j} < 1.00$ represent possible diagnoses, while the values $\mu_{D_j} = 1.00$ and $\mu_{D_j} = 0.00$ correspond to confirmed and excluded diagnoses, respectively. Diagnoses that have not yet been considered take the value $\mu_{D_j} = \nu$. Formally, a relationship $R_{PD} \subset \Pi \times \Delta$ is established, defined by $\mu_{R_{PD}}(P_q, D_j) = \mu_{D_j}$ for patient P_q, where $D_j \in \Delta$ $(\Delta = \{D_1, \cdots, D_n\})$.

Intermediate combinations (fuzzy logical combinations of symptoms and diseases) were introduced to model the pathophysiological states of patients; symptom combinations are combinations of symptoms, diseases, and intermediate combinations.

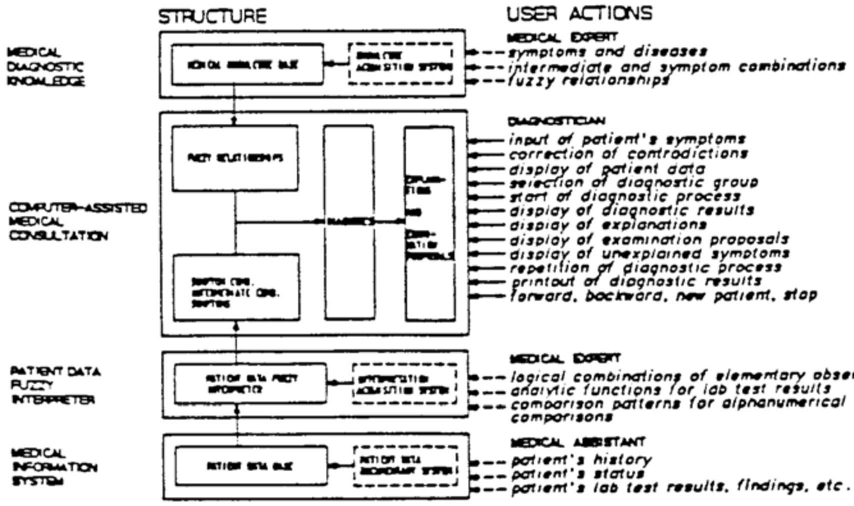

Fig. 1. Structure of CADIAG-2 with connection to a medical information system (dashed lines mark components effective before starting the individual consultation).

TABLE I
AN EXAMPLE OF THE REPRESENTATION OF MEDICAL INFORMATION ON THE PATIENT

Quantitative Value		Symptom	Fuzzy Value
Measured potassium level of 5.3 mmol/l	Fuzzy interpreter	Potassium, greatly reduced	$\mu_{S_1} = 0.00$
		Potassium, reduced	$\mu_{S_2} = 0.00$
		Potassium, normal	$\mu_{S_3} = 0.40$
		Potassium, elevated	$\mu_{S_4} = 0.60$
		Potassium, greatly elevated	$\mu_{S_5} = 0.00$

Both entities take their values μ_{IC_k} and μ_{SC_l} (respectively) in $[0,1] \cup \{\flat\}$, where $\flat$ implies that the actual value has not yet been determined. The relationship $R_{PSC} \subset \Pi \times K$ is defined by $\mu_{R_{PSC}}(P_q, SC_l) = \mu_{SC_l}$ for patient P_q, where $SC_l \in K$ ($K = \{SC_1, \cdots, SC_r\}$) formally describes the symptom combinations observed in the patient.

The fuzzy logical connectives are defined as follows.

Conjunction:

$$x_1 \wedge x_2 = \begin{cases} \min(x_1, x_2), & \text{if } x_1 \in [0,1] \text{ and } x_2 \in [0,1] \\ \flat, & \text{if } x_1 = \flat \text{ and/or } x_2 = \flat. \end{cases}$$

Disjunction:

$$x_1 \vee x_2 = \begin{cases} \max(x_1, x_2), & \text{if } x_1 \in [0,1] \text{ and } x_2 \in [0,1] \\ x_1, & \text{if } x_1 \in [0,1] \text{ and } x_2 = \flat \\ x_2, & \text{if } x_1 = \flat \text{ and } x_2 \in [0,1] \\ \flat, & \text{if } x_1 = \flat \text{ and } x_2 = \flat. \end{cases}$$

Negation:

$$\bar{x}_1 = \begin{cases} 1 - x_1, & \text{if } x_1 \in [0,1] \\ \flat, & \text{if } x_1 = \flat. \end{cases}$$

Note that the introduction of a value $\flat$ for missing operands, which has necessarily to be included into the definition of the connectives and the negation out of practical reasons, leads to violation of several laws (e.g., De Morgan's laws). These laws are valid for the classical fuzzy connectives $x_1 \wedge x_2 = \min(x_1, x_2)$, $x_1 \vee x_2 = \max(x_1, x_2)$ with respect to $\bar{x}_1 = 1 - x_1$ [1], but are now no longer valid. Further discussion on fuzzy connectives may be found in [25]–[27].

The following relationships between medical entities are considered in CADIAG-2:

- symptom–disease relationships ($S_i D_j$)
- symptom combination–disease relationships ($SC_i D_j$)
- symptom–symptom relationships ($S_i S_j$)
- disease-disease relationships ($D_i D_j$).

These relationships are characterized by two parameters:

- frequency of occurrence (o)
- strength of confirmation (c).

For a relationship between medical entities X and Y (where X and Y may be symptoms, diseases, or symptom combinations), the frequency of occurrence describes the frequency with which X occurs when Y is present. Similarly, the strength of confirmation reflects the degree to which the presence of X implies the presence of Y.

The relationships between medical entities are given in the form of relationship rules with associated relationship tupels. The general formulation of these rules is

IF (*antecedent*) THEN (*consequent*) WITH (o, c).

The relationship tuples (o, c) contain either numerical values μ_o and μ_c, or linguistic fuzzy values λ_o and λ_c, or both [3].

The definitions of the linguistic values λ_o and λ_c, the intervals that they cover, and their representative numerical values are given in Table II. Representative numerical values are useful in order to make fuzzy inferences easily possible (see Section IV-B). The way in which the linguistic fuzzy values, the numerical intervals and their representative numerical values were chosen is described in more detail in [8], [9]. Some examples of relationship rules are given below.

Example 1:
IF (ultrasonic of pancreas is pathological)
THEN (pancreatic cancer)
WITH (0.75 = often, 0.25 = weak).

Example 2:
IF (tophi)
THEN (gout)
WITH (0.25 = seldom, 1.00 = always).

Example 3:
IF (lower back pain $\wedge$ limitation of motion of the lumbar spine $\wedge$ diminished chest expansion $\wedge$ male patient $\wedge$ age between 20 and 40 years)
THEN (ankylosing spondylitis)
WITH ($\flat$, 0.90 = very strong).

TABLE II
LINGUISTIC FUZZY VALUES, NUMERICAL INTERVALS, AND REPRESENTATIVE NUMERICAL VALUES DESCRIBING FREQUENCY OF OCCURRENCE AND STRENGTH OF CONFIRMATION

Frequency of Occurrence			Strength of Confirmation		
Value λ_o	Interval	Representative Value μ_o	Value λ_c	Interval	Representative Value μ_c
Always	[1.00, 1.00]	1.00	Always	[1.00, 1.00]	1.00
Almost always	[0.98, 0.99]	0.99	Almost always	[0.98, 0.99]	0.99
Very often	[0.83, 0.97]	0.90	Very strong	[0.83, 0.97]	0.90
Often	[0.68, 0.82]	0.75	Strong	[0.68, 0.82]	0.75
Medium	[0.33, 0.67]	0.50	Medium	[0.33, 0.67]	0.50
Seldom	[0.18, 0.32]	0.25	Weak	[0.18, 0.32]	0.25
Very seldom	[0.03, 0.17]	0.10	Very weak	[0.03, 0.17]	0.10
Almost never	[0.01, 0.02]	0.01	Almost never	[0.01, 0.02]	0.01
Never	[0.00, 0.00]	0.00	Never	[0.00, 0.00]	0.00
Unknown		$\flat$	Unknown		$\flat$

The values μ_o and μ_c are interpreted as the values of the fuzzy relationships between antecedents and consequents. Thus

S_iD_j (occurrence relationship)	$R^o_{SD} \subset \Sigma \times \Delta$
S_iD_j (confirmation relationship)	$R^c_{SD} \subset \Sigma \times \Delta$
SC_iD_j (occurrence relationship)	$R^o_{SCD} \subset K \times \Delta$
SC_iD_j (confirmation relationship)	$R^c_{SCD} \subset K \times \Delta$
S_iS_j (occurrence relationship)	$R^o_{SS} \subset \Sigma \times \Sigma$
S_iS_j (confirmation relationship)	$R^c_{SS} \subset \Sigma \times \Sigma$
D_iD_j (occurrence relationship)	$R^o_{DD} \subset \Delta \times \Delta$
D_iD_j (confirmation relationship)	$R^c_{DD} \subset \Delta \times \Delta$.

B. Fuzzy Logical Inference

The compositional inference rule proposed by Zadeh [4] and introduced into medical diagnosis by Sanchez [28], [29] is adopted as an inference mechanism. It accepts fuzzy descriptions of the patient's symptoms and infers fuzzy descriptions of the patient's diseases by means of the fuzzy relationships described in the previous section.

Three such inference rules (compositions) are used to deduce the diseases D_j suffered by patient P_q from the observed symptoms S_i:

1) Composition for S_iD_j (hypotheses and confirmation):

$$R^1_{PD} = R_{PS} \circ R^c_{SD} \tag{1}$$

defined by

$$\mu_{R^1_{PD}}(P_q, D_j) = \max_{S_i} \min \left[\mu_{R_{PS}}(P_q, S_i); \mu_{R^c_{SD}}(S_i, D_j) \right].$$

2) Composition for S_iD_j (exclusion (by present symptoms)):

$$R^2_{PD} = R_{PS} \circ (1 - R^c_{SD}) \tag{2}$$

defined by

$$\mu_{R^2_{PD}}(P_q, D_j) = \max_{S_i} \min \left[\mu_{R_{PS}}(P_q, S_i); 1 - \mu_{R^c_{SD}}(S_i, D_j) \right].$$

3) Composition for S_iD_j (exclusion (by absent symptoms)):

$$R^3_{PD} = (1 - R_{PS}) \circ R^o_{SD} \tag{3}$$

defined by

$$\mu_{R^3_{PD}}(P_q, D_j) = \max_{S_i} \min \left[1 - \mu_{R_{PS}}(P_q, S_i); \mu_{R^o_{SD}}(S_i, D_j) \right].$$

The following diagnostic results are obtained. A diagnosis is confirmed if

$$\mu_{R^1_{PD}}(P_q, D_j) = 1.00. \tag{4}$$

A diagnosis is possible if

$$\epsilon < \mu_{R^1_{PD}}(P_q, D_j) < 0.99. \tag{5}$$

The boundary value ϵ is a heuristic value which precludes diagnoses with very low evidence (e.g., $\epsilon = 0.10$). A diagnosis is excluded if

$$\mu_{R^2_{PD}}(P_q, D_j) = 1.00 \tag{6}$$

or

$$\mu_{R^3_{PD}}(P_q, D_j) = 1.00. \tag{7}$$

Symptom combination–disease inferences (called compositions 4, 5, and 6) are carried out and interpreted in an analogous way. Symptom-symptom inferences (called compositions 7, 8, and 9) are computed in order to complete the patient's symptom patterns. Disease–disease inferences (called compositions 10, 11, and 12) are performed in order to confirm general disease categories from the presence of differential diagnoses, to exclude entire areas of differential diagnoses if a particular general disease category is definitely absent, and to exclude mutually exclusive diseases if one of these diseases is confirmed.

C. Acquisition of Medical Knowledge

The knowledge acquisition system is capable of acquiring information on medical entities and the relationships between them. In CADIAG-2, relationships are stored as numerical values in the range [0, 1]. Medical information can be acquired in two ways: 1) through numerical or linguistic evaluation by medical experts and 2) by statistical evaluation of a database containing medical data on patients with confirmed diagnoses.

Information on relationships can be gathered numerically or linguistically using predefined linguistic values to determine parameters such as frequency of occurrence o and strength of confirmation c (see Table II). Empirical, judgmental, and definitive knowledge can be acquired in this way.

CADIAG-2 relationships have the important property that they can be interpreted statistically. The values of the frequency of occurrence μ_o and the strength of confirmation μ_c are defined as follows:

$$\mu_o = \frac{F(S_i \cap D_j)}{F(D_j)} = F(S_i/D_j) \tag{8}$$

$$\mu_c = \frac{F(S_i \cap D_j)}{F(S_i)} = F(D_j/S_i) \tag{9}$$

where

$F(S_i \cap D_j)$	absolute frequency of occurrence of S_i and D_j
$F(D_j)$	absolute frequency of occurrence of D_j
$F(S_i)$	absolute frequency of occurrence of S_i
$F(S_i/D_j)$	conditional frequency of S_i given D_j
$F(D_j/S_i)$	conditional frequency of D_j given S_i.

With definitions (8) and (9), extended statistical evaluations of known medical relationships or as yet unidentified relationships can be carried out using data on patients with confirmed diagnoses.

D. The Diagnostic Process

1) Symptoms: The symptoms of the patient can be entered into CADIAG-2 in three ways (described in detail in [8]): a) by natural language input of symptoms S_i; b) by natural language input of keywords that trigger whole groups of symptoms S_i; and c) by accessing a database containing the patient's data and transferring information via a fuzzy interpreter.

Natural language input of symptoms S_i such as "high fever," "elevated GOT," or "blood stool positive" is achieved by a symptom search algorithm with an embedded word segmentation algorithm that allows the use of synonyms and abbreviations, orthographic variants, and different medical suffixes.

Input of keywords such as "present complaints," "previous complaints," "blood count," or "ultrasonic" causes whole sections of the symptom thesaurus to be displayed. Subsequently, fuzzy values can be linked with these symptoms by the physician.

The existence of a database that already contains the patient's symptoms suggests the automatic transfer of information from the database to CADIAG-2. During this transfer, the data is passed through a fuzzy interpreter, which contains instructions about the assignment of fuzzy values to observations, lab test results, and even simple alphanumeric texts.

After the patient's symptoms have been collected, symptom–symptom inferences are performed. The symptom list contains all necessary items of data, including fuzzy value, origin (measured; inferred), predefined symptom class (routine; specially requested; invasive or expensive), numerical value, units, and date of observation. The list of symptoms is then checked for contradictions.

2) Symptom Combinations: Intermediate combinations of symptoms are evaluated in the next step. Having passed the consistency check, fuzzy values for all symptom combinations are computed. The resulting lists are now as complete as possible and do not contain any contradictions.

3) Confirmed diagnoses: The fuzzy values $\mu_{D_j} = 1.00$, i.e., confirmed diagnoses D_j for patient P_q, are identified using the following equation:

$$\mu_{D_j} = 1.00, \quad \text{if} \begin{cases} \mu_{R^1_{PD}}(P_q, D_j) = 1.00 \\ \text{or} \\ \mu_{R^4_{PD}}(P_q, D_j) = 1.00. \end{cases} \tag{10}$$

4) Excluded Diagnoses: The fuzzy values $\mu_{D_j} = 0.00$, i.e., excluded diagnoses D_j for patient P_q, are identified using

$$\mu_{D_j} = 0.00, \quad \text{if} \begin{cases} \mu_{R^2_{PD}}(P_q, D_j) = 1.00 \\ \text{or} \\ \mu_{R^3_{PD}}(P_q, D_j) = 1.00 \\ \text{or} \\ \mu_{R^5_{PD}}(P_q, D_j) = 1.00 \\ \text{or} \\ \mu_{R^6_{PD}}(P_q, D_j) = 1.00. \end{cases} \tag{11}$$

Disease–disease relationships now allow the inference of further diagnoses (confirmed or excluded):

$$\mu_{D_j} = \begin{cases} 1.00, & \text{if } \mu_{R^{10}_{DD}}(P_q, D_j) = 1.00 \\ 0.00, & \text{if } \mu_{R^{11}_{DD}}(P_q, D_j) = 1.00 \\ 0.00, & \text{if } \mu_{R^{12}_{DD}}(P_q, D_j) = 1.00. \end{cases} \tag{12}$$

5) Possible Diagnoses: Method a): Fuzzy values μ_{D_j} such that $\epsilon \leq \mu_{D_j} \leq 0.99$ indicate possible diagnoses. These are determined as follows:

$$\mu_{D_j} = \max\left[\mu_{R^1_{PD}}(P_q, D_j); \mu_{R^4_{PD}}(P_q, D_j); \mu_{R^{10}_{DD}}(P_q, D_j)\right],$$

$$\text{if} \begin{cases} \epsilon \leq \mu_{R^1_{PD}}(P_q, D_j) \leq 0.99 \\ \text{and/or} \\ \epsilon \leq \mu_{R^4_{PD}}(P_q, D_j) \leq 0.99 \\ \text{and/or} \\ \epsilon \leq \mu_{R^{10}_{DD}}(P_q, D_j) \leq 0.99. \end{cases} \tag{13}$$

Method b): Because the values μ_{D_j} calculated by (13) are independent of the number of rules that can be used to support D_j, a heuristic function is introduced which considers the number of criteria present or partly present, which suggest but do not confirm disease D_j. The function then calculates the corresponding number of points PN_{D_j}. The values of PN_{D_j} are helpful in judging between the various possible diagnoses, although the ultimate aim should be to obtain a confirmed diagnosis. The number of points PN_{D_j} is calculated as follows:

$$PN_{D_j} = 100 \sum_{i=1}^{m^*} \left\{ \alpha \min\left[\mu_{R_{PS}}(P_q, S_i); \mu_{R^1_{SD}}(S_i, D_j)\right] + \beta \min\left[\mu_{R_{PS}}(P_q, S_i); \mu_{R^2_{SD}}(S_i, D_j)\right] \right\}, \tag{14}$$

where m^* is the number of symptoms exhibited by the patient P_q that occur in the definition of D_j, and $\alpha + \beta = 1.00$. At present, we generally take $\alpha = 0.09$ and $\beta = 0.91$, i.e., the strength of confirmation has ten times more influence than the frequency of occurrence on the value of PN_{D_j}. The multiplication of the sum with 100 is carried out in order to obtain easily readable and memorizable point numbers.

6) Explanation of Diagnostic Results: The physician's acceptance of CADIAG-2's diagnoses depends strongly on the ability of CADIAG-2 to explain its diagnostic output. On request, the information supporting confirmed diagnoses, excluded diagnoses, and possible diagnoses is presented; this takes the form of the names of the medical entities, their definitions, their measured and fuzzy values, and their relationships to the diagnostic output.

7) Proposals for Further Examination of the Patient: One of the main objectives of CADIAG-2 is to provide iterative consultations, starting with simple, easy-to-examine, and cheap data. A number of possible diagnoses can usually be inferred from these data, and further examinations are then necessary to confirm or exclude these hypotheses. CADIAG-2 uses the medical information stored in its knowledge base to propose what form these further examinations should take. The symptoms for further study are clearly those that would confirm or exclude a particular diagnosis. Additionally, those symptoms which enhance the position of the possible diagnosis in the ranked list of all possible diagnoses are also indicated.

8) Unexplained Symptoms: The confirmed diagnoses and any remaining possible diagnoses should together explain any pathological symptom, indication, or lab test result of the patient. Unexplained data (usually) indicates further diseases that should be investigated.

V. Results

A. Rheumatic Diseases

CADIAG-2/RHEUMA has undergone partial tests with data from patients at a rheumatological hospital. A study of 400 patients with rheumatoid arthritis, gout, Bechterew's disease, Sjögren's disease, systemic lupus erythematosus, Reiter's disease, and scleroderma showed that CADIAG-2 obtained the correct diagnosis in 94.5 percent of the cases considered. This figure was calculated by comparing the discharge diagnoses established by the consultant at the rheumatological hospital (assumed to be correct) with the confirmed and possible diagnoses made by CADIAG-2. Most of the cases in which clinical diagnoses could not be confirmed fell into two classes: 1) the patient was in hospital only temporarily to check the efficacy of drugs already administered and 2) the patient was in the early stages of one of the rheumatic diseases considered; in almost all of these cases a possible diagnosis was suggested.

B. Pancreatic Diseases

CADIAG-2/PANCREAS was tested with data from 47 patients. The discharge diagnoses of these patients were assumed to be correct.

Pancreatic cancer was confirmed three times. Confirmation was aided by the existence of a result "Specific abnormal pancreatic biopsy," which has a strength of confirmation $\mu_c = 1.00$ for pancreatic cancer.

Possible hypothese were generated for the other cases, and the heuristically determined number of points was taken as the basis for evaluation. The results are given in Table III.

TABLE III
COMPARISON OF CADIAG-2 POSSIBLE DIAGNOSES (44 CASES) WITH THE CLINICAL DIAGNOSES

Clinical Diagnosis	Percentage of cases
CADIAG-2's diagnosis with the highest number of points	84.1
CADIAG-2's diagnosis with the second highest number of points	6.8
No CADIAG-2 diagnosis	9.1

Acknowledgment

I gratefully acknowledge the contributions of G. Kolarz, M.D., and W. Scheithauer, M.D., in the medical documentation of rheumatic and pancreatic diseases. Without their cooperation, this work would not have been possible. I also would like to thank S. V. Ovchinnikov, Ph.D., for fruitful discussions. I am indebted to the referees for providing very helpful, detailed criticism and suggestions.

References

[1] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965.

[2] ——, "A fuzzy-algorithmic approach to the definition of complex or imprecise concepts," in *Systems Theory in the Social Sciences*, H. Bossel, S. Klaczko, and N. Müller, Eds. Stuttgart: Birkhäuser Verlag, 1976, pp. 202–282.

[3] ——, "Linguistic variables, approximate reasoning and dispositions," *Med. Inform.*, vol. 8, pp. 173–186, 1983.

[4] ——, "Outline of a new approach to the analysis of complex systems and decision processes," *IEEE Trans. Syst., Man, Cybern.*, vol. 3, pp. 28–44, 1973.

[5] R. E. Bellman and L. A. Zadeh, "Local and fuzzy logics," memo. ERL-M584, Electronics Research Laboratory, College of Engineering, University of California, Berkeley 94720, May 11, 1976.

[6] K. -P. Adlassnig, "A survey on medical diagnosis and fuzzy subsets," in *Approximate Reasoning in Decision Analysis*, M. M. Gupta and E. Sanchez Eds. New York: North-Holland, 1982, pp. 203–217.

[7] ——, "A fuzzy logical model of computer-assisted medical diagnosis," *Meth. Inform. Med.* vol. 19, pp. 141–148, 1980.

[8] ——, "Ein computerunterstütztes medizinisches Diagnosesystem unter Verwendung von fuzzy Teilmengen," Ph.D. dissertation, Technical University of Vienna, Vienna, 1983.

[9] K. -P. Adlassnig and G. Kolarz, "CADIAG-2: computer-assisted medical diagnosis using fuzzy subsets," in *Approximate Reasoning in Decision Analysis*, M. M. Gupta and E. Sanchez, Eds. New York: North-Holland, 1982, pp. 219–247.

[10] K. -P. Adlassnig, G. Kolarz, and W. Scheithauer, "Present state of the medical expert system CADIAG-2, *Meth. Inform. Med.*, vol. 24, pp. 13–20, 1985.

[11] G. Belforte *et al.*, "Selection and assessment of laboratory tests for the evaluation of liver functional impairment, *Meth. Inform. Med.*, vol. 24 pp. 39–45, 1985.

[12] G. Bortolan and R. Degani, "Comparitive evaluation of three models of fuzzy decision-making in electrocardiography," in *MEDINFO 83*, J. H. Van Bemmel, M. J. Ball, and O. Wigertz, Eds. New York: North-Holland, 1983, pp. 537–540.

[13] E. E. Kerre, "The use of fuzzy set theory in electrocardiological diagnostics, in *Approximate Reasoning in Decision Analysis*, M. M. Gupta and E. Sanchez, Eds. New York: North-Holland, 1982, pp. 277–282.

[14] M. Fieschi, *et al.*, "A program for expert diagnosis and therapeutic decision," *Med. Inform.*, vol. 8, pp. 127–135, 1983.

[15] L. Lesmo, M. Marzuoli, G. Molino, and P. Torasso, "An expert system for the evaluation of liver functional assessment," *J. Med. Syst.*, vol. 8, pp. 87–101, 1984.

[16] L. Lesmo, L. Saitta, and P. Torasso, "Learning of fuzzy production rules for medical diagnosis, in *Approximate Reasoning in Decision Analysis*, M. M. Gupta and E. Sanchez, Eds. New York: North-Holland, 1982, pp. 249–260.

[17] M. Milanese, *et al.*, "A sequential protocol for the optimization of diagnostic procedures in hepatology," *J. Med. Syst.*, vol. 8, pp. 73–85, 1984.

[18] L. Saitta and P. Torasso, "Fuzzy characterization of coronary disease," *Fuzzy Sets Syst.*, vol. 5, pp. 245–258, 1981.

[19] E. Sanchez, J. Gouvernet, R. Bartolin, and L. Vovan, "Linguistic approach in fuzzy logic of the W.H.O. classification of dyslipoproteinemias," in *Fuzzy Set and Possibility Theory—Recent Developments*, R. R. Yager, Ed. New York: Pergamon, 1982, pp. 582–588.

[20] P. Smets, "Fuzzy set theory for medical decision making, in *Objective Medical Decision-Making; Systems Approach in Acute Disease*, J. E. W. Beneken and S. M. Laveile, Eds. Berlin: Springer-Verlag, 1983, pp. 7–19.

[21] G. Soula and E. Sanchez, "Soft deduction rules in medical diagnostic processes, in *Approximate Reasoning in Decision Analysis*, M. M. Gupta and E. Sanchez, Eds. New York: North-Holland, 1982, pp. 77–88.

[22] G. Soula, B. Vialettes, and J. -L. SanMarco, "Protis: a fuzzy deduction-rule system: application to the treatment of diabetes," in *MEDINFO 83*, J. H. Van Bemmel, M. G. Ball, and O. Wigertz, Eds. New York: North-Holland, 1983, pp. 533–536.

[23] *Encyclopedia Britannica*, Chicago, 1975.

[24] E. Nagel and J. R. Newman, *Godel's Proof*. New York: New York University, 1973.

[25] D. Dubois and H. Prade, *Fuzzy Sets and Systems: Theory and Applications*. New York: Academic, 1980.

[26] S. Weber, "A general concept of fuzzy connectives, negations and implications based on *t*-Norms and *t*-Conorms," *Fuzzy Sets Syst.*, vol. 11, pp. 115–134, 1983.

[27] S. V. Ovchinnikov, "General negations in fuzzy set theory," *J. Mathematical Analysis Appl.*, vol. 92, pp. 234–239, 1983.

[28] E. Sanchez, "Compositions of fuzzy relations," in *Advances in Fuzzy Set Theory and Applications*, M. M. Gupta, R. K. Ragade, and R. R. Yager, Eds. New York: North-Holland, 1979, pp. 421–433.

[29] ——, "Medical diagnosis and composite fuzzy relations," in *Advances in Fuzzy Set Theory and Applications*, M. M. Gupta, R. K. Ragade, and R. R. Yager, Eds. New York: North-Holland, 1979, pp. 437–444.

GISMO: A Visual Problem-Structuring and Knowledge-Organization Tool

WILLIAM E. PRACHT

Abstract—**A visual problem-structuring and knowledge-organization technique applicable to complex unstructured decisionmaking situations is presented. The role of images in the human cognitive process is briefly examined to serve as a basis for a method that enables individuals to apply the power of the computer to problem structuring in a manner more closely resembling the natural though process. Five phases of interactive problem structuring / knowledge organization are described: 1) specifying the problem elements, 2) specifying the relationship between the elements, 3) specifying the strengths and time delays for these relationships, 4) constructing a visual representation of the problem called a digraph, and 5) refining and working directly with the digraph. The effects on decision makers of the visual problem structuring tool are discussed.**

Introduction

Modern computers have been successfully applied to the solution of problems for nearly half a century. From the calculation of new trajectory tables for different types of artillery in the 1940's to the landing of the space shuttle in the 1980's, computers have been used by professionals in their line of work. The introduction in 1975 of the microcomputer made the power of the computer readily available to small organizations and individuals. Accompanying the rapidly increasing number of microcomputers are basic changes in the philosophy and complexion of computer software. No longer is the typical user an expert in computers and so software is fast becoming "friendlier" and more attractive to the end user (i.e. more interactive and user-oriented). To put it another way, what this means is software that permits humans to use their visual thinking abilities to solve problems rather than the previously more common procedural, text-based approaches. Visicalc, the electronic visible calculator, is a well-known example of this type of software.

This visual computing environment does indeed enable users to apply the power of the computer in a manner which more closely resembles the natural thought processes involved in human problem solving. What is lacking, however, is an image-based tool to assist users' visual thinking process involved in problem *formulation*. A great deal of effort has been expended in developing techniques to assist individuals in solving problems. What is needed next are tools to help problem solvers conceptualize and formulate problems. It is the computer's capability for retaining and displaying images (i.e. knowledge related to problem situations) that gives it perhaps the greatest potential for assisting

Manuscript received July 21, 1985; revised November 24, 1985

The author is with the Department of Information Systems and Quantitative Sciences, College of Business Administration, Texas Tech University, Lubbock, TX 79409, USA.

IEEE Log Number 8407309.