UNIFORM REPRESENTATION OF VAGUENESS AND IMPRECISION IN PATIENT'S MEDICAL FINDINGS USING FUZZY SETS

Klaus-Peter Adlassnig
Department of Medical Computer Sciences
(Director: Prof. Dr. G. Grabner)
University of Vienna
Garnisongasse 13, A - 1090 Vienna, Austria

ABSTRACT. Based on the mathematical theory of fuzzy sets, a uniform representation of patient's symptoms, signs, test results, and clinical findings considering their inherent vagueness and imprecision is proposed. The uniformly represented patient data form the point of departure for the medical expert system's subsequent diagnostic inferences.

1. INTRODUCTION

Patient records usually contain a variety of different medical data: history items, results from physical and psychological examinations, laboratory test results, data from clinical investigations such as US, ECG, X-ray, endoscopy, NMR, and many others. Exploiting these data for inferring diagnostic conclusions automatically by means of an expert system makes a uniform representation of these data necessary. Inherent vagueness of many of the medical terms contained in the expert system's knowledge base such as *elevated glucose level in serum* or *reduced leukocytes in blood cell count* demands mathematical modeling which takes their vagueness and imprecision into account.

This paper proposes a uniform representation which is applied in CADIAG-2 [1-4], a data-driven, fuzzy medical expert system integrated into the medical information system WAMIS [5] of the Vienna General Hospital.

2. REPRESENTATION OF PATIENT'S MEDICAL DATA

2.1. Data abstraction and reasoning

In the medical information system WAMIS, the central patient records contain usually medical data collected and stored in that form in which they occur; that is, either as quantitative results of laboratory tests or as qualitative pieces of information gained from patient's history, from physician's evaluation of patient's state, and from clinical investigations.

To give some examples, an *erythrocyte sedimentation rate (after 1 hour) of 40 mm* is a quantitative laboratory test result; qualitative medical data are *swelling of the second distal interphalangeal joint of the left hand* obtained from a detailed physical examination done with patients at the rheumatological unit of the hospital and *ankylosis of the small vertebral joints in X-ray* as result of an X-ray interpretation by a radiologist.

R. Trappl (ed.), Cybernetics and Systems '88, 685–692.

The medical expert system CADIAG-2, however, reasons from patient information being on a higher semantic level than that in WAMIS. CADIAG-2 was designed to model the human reasoning process trying to infer diagnoses from abstracted data such as *blood sedimentation rate is increased* or *finger joints are affected.* The reasoning process itself, than, employs known associations between those abstract concepts and diseases.

An example of this kind of knowledge—as is commonly found in textbooks—is the following: *Very high amylase activities (about 5 times the standard) are almost confirming acute pancreatitis.* By analyzing this sentence, we find the abstract concept of a finding, '*very high amylase activities*', vaguely defined by '*about 5 times the standard*', showing a medical relationship '*almost confirming*' to the disease concept 'acute *pancreatitis*'.

In order to be used in the reasoning process of CADIAG-2, patient data collected and stored in the patient data base of WAMIS have to be abstracted. In this context, we will distinguish two kinds of abstraction (cf. [6]):

- qualitative abstraction of quantitative data, done with respect to some normal or expected value (*130 mg/dl glucose level in serum* is an *elevated serum glucose level*);
- definitional abstraction of binary data, based on essential, necessary features of a concept (*redness, swelling, and tenderness in one or more of the finger joints* means *affected finger joints*).

Data abstraction is usually made with certainty. Belief thresholds and qualifying conditions are chosen so the abstraction is categorical. In medicine, however, vagueness is commonly inherent part of the established abstract concepts; that is, the boundaries of these concepts are imprecise and transitions from one concept to adjacent ones are gradual rather than sharp. This is due to biological variety that inhibits the application of precise models to formally describe concepts and relationships between them.

For CADIAG-2, the formal modeling of abstract concepts is done by applying the theory of fuzzy sets [7-12]. Vague concepts are modelled as fuzzy sets and degrees of compatibility indicate to which extend given data are compatible with the concept under consideration.

Data abstraction in CADIAG-2 is realized by an interface program, called *patient data fuzzy interpreter*, connecting WAMIS and CADIAG-2. Given the identification of a patient, the patient data fuzzy interpreter accesses the patient data base and abstracts patient's medical data according to instructions included in the interpreter. Afterwards, the abstracted findings are passed on to the reasoning part of CADIAG-2.

2.2. Qualitative abstraction of quantitative data

In WAMIS, quantitative test results are either transferred automatically from laboratory automata or entered manually by medical technicians. Only certain values are admissible values for the respective laboratory tests. These values will be referred to as universe of discourse U of the laboratory test. The universe of discourse is usually a segment of the real line denoted by $U = [b_l, b_u]$, where b_l and b_u are the lower and upper physiological boundary, respectively. Values outside the physiological boundaries cannot be observed in a living human being and are prevented from being entered by the medical documentation system of WAMIS. For the *glucose level in serum*, for example, we have a universe of discourse $U = [10\ mg/dl, 1,500\ mg/dl]$.

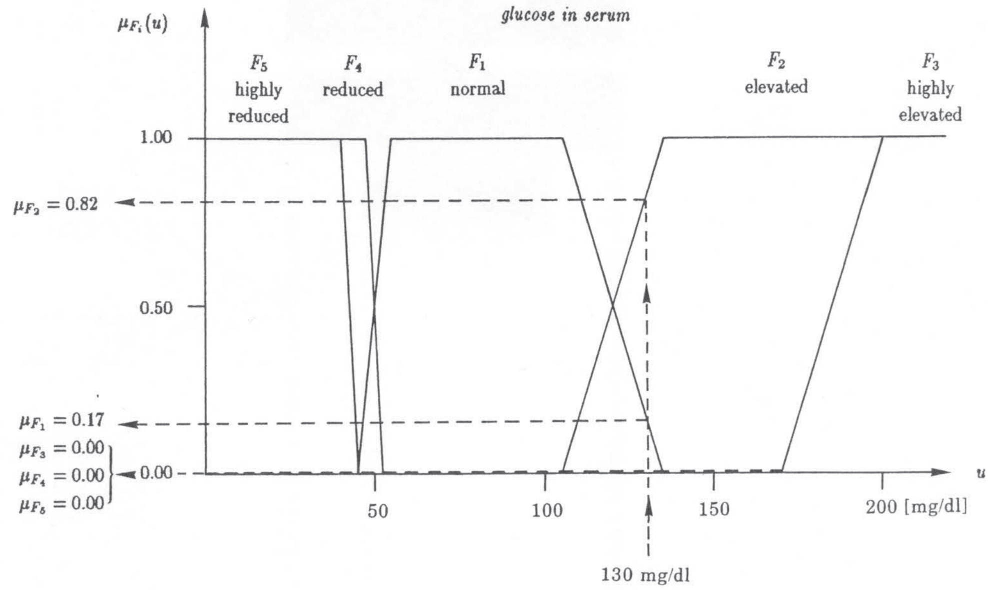

Figure 1: Qualitative abstraction of quantitative laboratory test results.

The point of departure for the abstraction process is the respective stored numerical value, for example, 130 mg/dl for patient's *glucose level in serum*. The result of the abstraction are abstracted findings, for example, *glucose level in serum* is *elevated*, which can be interpreted as the abbreviated form of *glucose level in serum* is *elevated* is 'yes'. The abstracted findings are determined according to certain ranges for the possible numerical values of the laboratory test defining all the related concepts such as *normal*, *elevated*, and *reduced*. Yet, there is common sense understanding in medicine that the ranges for normal and pathological are not crisp in nature, but rather exhibit a gradual transition from one to another.

In CADIAG-2, medical concepts such as *normal*, *elevated*, and *reduced* are considered to be fuzzy sets. Fuzzy sets are defined by fuzzy membership functions that assign to every finding a degree of membership, which expresses the degree of compatibility of a measured concrete value with the abstract concept under consideration. The degrees of compatibility take their values in $[0, 1]$, where zero means *no* and *unity* full compatibility. Thus, a measured value of 130mg/dl for *serum glucose level* is abstracted to (see also Figure 1):

$$\textit{glucose level is 130mg/dl} \quad \begin{cases} \text{glucose is normal is 0.17;} \\ \text{glucose level is elevated is 0.82;} \\ \text{glucose level is highly elevated is 0.0;} \\ \text{glucose level is reduced is 0.0;} \\ \text{glucose level is highly reduced is 0.0.} \end{cases}$$

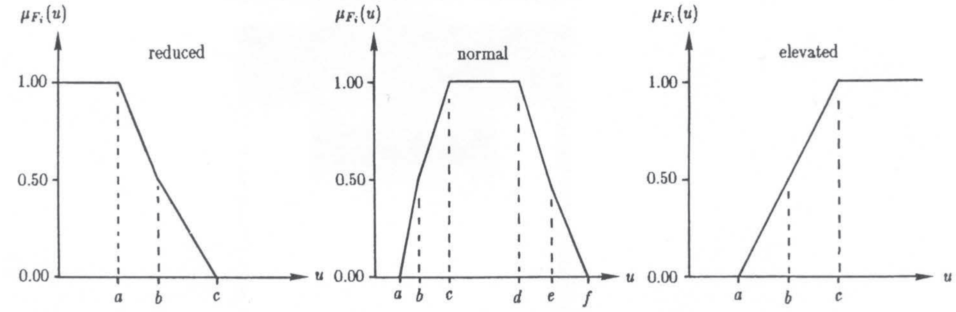

Figure 2: Three standard abstraction functions for *'reduced'*, *'normal'*, and *'elevated'*.

The membership functions are determined by the physician according to the numerical ranges for normal and pathological results, where the physician also indicates the transition zones. Most of them will be additionally adjusted according to sex and age of the patient during the actual diagnostic process.

It is frequently convenient to employ standardized functions with adjustable parameters to define fuzzy membership functions. In CADIAG-2, three types of functions are applied to define these membership functions for abstracting quantitative laboratory test results (see Figure 2).

The first abstraction function, $L_1(x; a, b, c)$, is applied for concepts that express states such as *'reduced'*, *'highly reduced'*, *'smaller than* a', or similar (see Eq. 1):

$$L_1(x; a, b, c) = \begin{cases} 0, & \text{if } x \leq a; \\ 0.5\left(\dfrac{x-a}{b-a}\right), & \text{if } a < x \leq b; \\ 0.5 + 0.5\left(\dfrac{x-b}{c-b}\right), & \text{if } b < x \leq c; \\ 1, & \text{if } x > c. \end{cases} \tag{1}$$

The three parameters a, b, and c are given by the medical expert so that $b_l \leq a \leq b \leq c \leq b_u$, where $b_l \leq x \leq a$ indicates the certain range of x with *full* compatibility of every x with the abstract concept under consideration, $c \leq x \leq b_u$ denotes the certain range of x with *no* compatibility of every x with the respective concept, and $a \leq x \leq c$ signifies the transition zone for borderline test results with an intermediate degree of compatibility of x with the considered concept. Full compatibility is expressed by L_1 taking the value 1, no compatibility is expressed by the value 0, and intermediate degrees between unity and zero indicate partial compatibility.

The second abstraction function, $L_2(x; a, b, c, d, e, f)$, is applied for concepts that express states such as *'normal'*, *'between c and d'*, or similar (see Eq. 2):

$$L_2(x; a, b, c, d, e, f) = \begin{cases} 0, & \text{if } x \leq a; \\ 0.5\left(\dfrac{x-a}{b-a}\right), & \text{if } a < x \leq b; \\ 0.5 + 0.5\left(\dfrac{x-b}{c-b}\right), & \text{if } b < x \leq c; \\ 1, & \text{if } c < x \leq d; \\ 1 - 0.5\left(\dfrac{x-d}{e-d}\right), & \text{if } d < x \leq e; \\ 0.5 - 0.5\left(\dfrac{x-e}{f-e}\right), & \text{if } e < x \leq f; \\ 0, & \text{if } x > f. \end{cases} \tag{2}$$

In case of L_2, we have—similar to function L_1—ranges with full and no compatibility of the measured laboratory test result x with the concept under consideration; there are two gradual transition zones: the first one to a lower adjacent concept, say, *'reduced'*, and the other one to an upper adjacent concept, say, *'elevated'*.

The third abstraction function, $L_3(x; a, b, c)$, is employed for states such as *'elevated'*, *'highly elevated'*, *'larger than c'*, or similar (see Eq. 3):

$$L_3(x; a, b, c) = \begin{cases} 1, & \text{if } x \leq a; \\ 1 - 0.5\left(\dfrac{x-a}{b-a}\right), & \text{if } a < x \leq b; \\ 0.5 - 0.5\left(\dfrac{x-b}{c-b}\right), & \text{if } b < x \leq c; \\ 0, & \text{if } x > c. \end{cases} \tag{3}$$

With L_3, we cover the upper part of the whole range of admissible values x; allowing gradual transition from no to full compatibility with a definable borderline range.

However, because ranges for normality depend very often on patient specific factors, the parameters of the standardized abstraction functions may be altered according to these individual factors. Until now, only sex and age of the patient are taken into account. Different sets of parameters may be defined for male and female patients and for various ranges of age, if necessary.

At present, about 400 of such abstraction functions for about 100 laboratory tests are included into the patient data fuzzy interpreter of CADIAG-2.

2.3. Definitional abstraction of binary data

In WAMIS, medical data from patient's history, physical and psychological examinations, US, X-ray, NMR, endoscopies, histology, etc. are usually either binary in nature or are artificially categorized in such a way that binary data are obtained. These yes/no decisions can easily be marked by the patient or physician on the standard forms used for data collection at the departments and clinics.

Data such as *male*, *female* or *dry skin*, *wet skin*, *warm skin*, *cold skin*, etc. are examples for natural binary data. Data gained, for example, from the evaluation of the mobility of the wrist joints, however, are precategorized into meaningful ranges and documented as binary data in the patient record. The available options in WAMIS, for this example, are those in Table 1:

Table 1: Detailed documentation of the mobility of the wrist joints in WAMIS.

					wrist joints					
		right						left		
		20°	10°	0°	radial-	20°	10°	0°		
10°	20°	30°	40°	50°	ulnar	10°	20°	30°	40°	50°
70°	50°	30°	10°	0°	dorsal-	70°	50°	30°	10°	0°
10°	30°	50°	70°	80°	volar	10°	30°	50°	70°	80°

In case of binary data, the admissible values are *'yes'* and *'no'*, thus, the universe of discourse is $U = \{yes, no\}$, or, we write instead $U = \{1, 0\}$.

Abstraction of binary data is done by evaluating logical combinations of these binary data which are defined by the medical expert. The admissible logical connectives are: *conjunction*, *disjunction*, *negation*, and *at least*. The underlying logic is Boolean logic, because binary data stored in the patient data base are either true (present) or false (absent). The n-ary connective *at least m of n* is introduced to simplify the evaluation of all possible combinations of m findings of an entire set of n findings. An example for the definitional abstraction of binary data is given below:

Table 2: Definitional abstraction of *limited mobility of the wrist joints.*

limited mobility		
of the wrist joint =	at least 1 of 10:	0° radial left;
		10° ulnar left; 20° ulnar left; 30° ulnar left;
		0° dorsal left; 10° dorsal left; 30° dorsal left;
		10° volar left; 30° volar left; 50° volar left;
	$\vee$	
	at least 2 of 4:	10° radial left;
		40° ulnar left;
		50° dorsal left;
		70° volar left;
	$\vee$	
	at least 1 of 10:	0° radial right;
		10° ulnar right; 20° ulnar right; 30° ulnar right;
		0° dorsal right; 10° dorsal right; 30° dorsal right;
		10° volar right; 30° volar right; 50° volar right;
	$\vee$	
	at least 2 of 4:	10° radial right;
		40° ulnar right;
		50° dorsal right;
		70° volar right.

The patient data fuzzy interpreter contains at present about 900 of such definitional abstraction functions.

2.4. Uniform fuzzy representation of patient's medical data

Once the abstracted findings together with their degrees of compatibility are transferred to the reasoning part of CADIAG-2, the degrees of compatibility may be altered by the physician according to his/her subjective perception of the case. By doing this, even findings defined as binary in the documentation system of WAMIS may obtain intermediate values.

In more formal terms, findings F_i take values $\mu_{F_i}(u) \in [0,1] \cup \{\nu\}$. The values $\mu_{F_i}(u)$ indicate degrees of compatibility between F_i and a measured or observed concrete value u, which is a member of the universe of discourse U of the respective finding F_i, i.e., $u \in U$. For the sake of brevity, $\mu_{F_i}(u)$ may also be written as μ_{F_i}. A degree of μ_{F_i} of zero means *no* and unity *full* compatibility with the meaning of the finding F_i. A value ν is assigned to all those findings which were not abstracted from the patient data base and not entered into CADIAG-2 manually before the reasoning cycle starts, i.e., these data are missing.

The degrees of compatibility $\mu_{F_i}(u)$ are interpreted as binary fuzzy relationships $\mu_{PF}(P, F_i)$ between the patient P and the findings F_i, that is $\mu_{F_i}(u) = \mu_{PF}(P, F)$. The relation $R_{PF} \subset \Pi \times \Sigma$ is defined by $\mu_{PF}(P, F_i)$, where $P \in \Pi$ and Π is the set of all patients under consideration, that is $\Pi = \{P_1, \ldots, P_r\}$. If we consider just one patient—as in our case—we have $\Pi = \{P\}$. Furthermore, Σ is the set of all findings F_i contained in the knowledge base of the expert system, thus $\Sigma = \{F_1, \ldots, F_m\}$. Π and Σ are non fuzzy sets.

3. CONCLUSION

The presented method for a uniform representation of patient's medical data—abstracted from measured quantitative medical data and from detailed qualitative pieces of information—provides a means to exploit patient's medical data collected and stored by a medical information system for application in an expert consultation system. During the abstraction process, the inherent vagueness and imprecision of medical terms is taken into account.

References

[1] Adlassnig, K.-P., Kolarz, G., Scheithauer, W., Effenberger, H. & Grabner, G. (1985) 'CADIAG: Approaches to Computer-Assisted Medical Diagnosis'. *Computers in Biology and Medicine* **15**, 315–335.

[2] Adlassnig, K.-P., Kolarz, G. & Scheithauer, W. (1985) 'Present State of the Medical Expert System CADIAG-2'. *Methods of Information in Medicine* **24**, 13–20.

[3] Adlassnig, K.-P. (1986) 'Fuzzy Set Theory in Medical Diagnosis'. *IEEE Transactions on Systems, Man, and Cybernetics* **SMC-16**, 260–265.

[4] Adlassnig, K.-P., Kolarz, G., Scheithauer, W. & Grabner, H. (1986) 'Approach to a Hospital-Based Application of the Medical Expert System CADIAG-2'. *Medical Informatics* **11**, 205–223.

[5] Grabner, H., Marksteiner, A., Dorda, W., Wolf, W. & Grabner, G. (1981) 'A Medical Information System. Conception and Clinical Usage'. *Journal of Clinical Computing* **10**, 154–169.

[6] Clancey, W. J. (1985) 'Heuristic Classification'. *Artificial Intelligence* **27**, 289–350.

[7] Zadeh, L. A. (1965) 'Fuzzy Sets'. *Information and Control* **8**, 338–353.

[8] Zadeh, L. A. (1976) 'A Fuzzy-Algorithmic Approach to the Definition of Complex and Imprecise Concepts'. In Bossel, H., Klaczko, S. & Müller, N. (Eds.) *System Theory in the Social Sciences*. Birkhäuser Verlag, Basel.

[9] Zadeh, L. A. (1981) 'Test Score Semantics for Natural Languages and Meaning Representation via PRUF'. Technical Note 247, AI Center, SRI International, Menlo Park CA. Also in: Rieger, B. B. (Ed.) *Empirical Semantics*. Brockmayer, Bochum, 281–349.

[10] Dubois, D. & Prade, H. (1980) *Fuzzy Sets and Systems—Theory and Applications*. Academic Press, New York.

[11] Zimmermann, H.-J. (1985) *Fuzzy Set Theory—and its Applications*. Kluwer-Nijhoff Publishing, Hingham.

[12] Kandel, A. (1986) *Fuzzy Mathematical Techniques with Applications*. Addison Wesley Publishing Company, Reading/Massachusetts.