Inflammatory protein variations : medical knowledge representation and approximate reasoning

Robert Bartolin, Vincent Bonniol, Elie Sanchez

INFLAMMATORY PROTEIN VARIATIONS : MEDICAL KNOWLEDGE REPRESENTATION AND APPROXIMATE REASONING

Robert BARTOLIN,
Service de Médecine Interne, Hôpital de l'Hôtel-Dieu,
6 place Daviel, 13 224 Marseille cédex 1, France

Vincent BONNIOL,
Institut de Recherche en Thérapeutique, Hôpital de l'Hôtel-Dieu,
6 place Daviel, 13 224 Marseille cédex 1, France

Elie SANCHEZ,
Service Universitaire de Biomathématiques, Faculté de Médecine,
27 Bd Jean Moulin, 13 385 Marseille cédex 5, France

ABSTRACT

It is proposed a medical knowledge representation of inflammatory protein variations involving complex relationships, frequently encountered in Internal Medicine. A linguistic model has been represented by a fuzzy set pattern expressing relative variations of serum proteins levels. Weights have been introduced into this pattern to translate relative importance among proteins. Three indexes or measures (possibility, necessity and truth-possibility) have been used for pattern matching purposes. Finally, a separating power yielding non fuzzy partitions has allowed to assign the corresponding diagnoses to patients (over 160 cases).

Keywords : Inflammatory proteins, Linguistic model, Medical knowledge representation, Approximate reasoning, Weights of importance, Fuzzy matching, Separating power.

I - INTRODUCTION

In internal medicine, many diseases are associated with an inflammatory syndrome. Given an inflammatory syndrome (I.S.) observed on a patient, it is difficult to assign a right diagnosis, based on specific protein variations. The main reason is that some I.S. are typical but, unfortunately most of the I.S. are non typical. In a typical I.S., proteins are all increased in similar proportions and in the same way, as it is the case for severe Infections and Septicemias. In a non typical I.S., like Vasculitis and Collagen Diseases, protein variations are dissociated, for example two or three protein levels are increased while some others are normal or decreased. Moreover for a given I.S., proteins are not all equally important, so that relative weighting of proteins has to be adjusted.

II - PATTERN OF MEDICAL KNOWLEDGE

By means of nephelemetric methods, seric levels of five proteins involved in biological inflammatory reactions have been measured. These five proteins are : C3 (C3-Complement Fraction), A1AT (Alpha-1-Antitrypsine), Orosomucoid, Haptoglobin, C.R.P. (C-Reactive Protein). Our protein-I.S pattern contains eleven groups : eight inflammatory syndromes (Bacterial Infections, Viral Infections, Vasculitis, Nephrotic syndromes, Acute Glomerular Nephritis, Intravascular Hemolysis with inflammation, Collagen Diseases non-Lupus and without infection, and finally Lupus), Normal condition, Intravascular Hemolysis without inflammation, and Glomerular Renal Insufficiency without inflammation.

In the protein-I.S. model, uncertainty is not of a probabilistic nature, as pointed out in [8]. What really matters, is relative variations of serum level proteins. Moreover, thresholds cannot be defined with good precision to allow a classification of patients in the I.S. pattern. Finally, these variations are easily interpreted in linguistic terms by internists, so that a fuzzy set representation naturally fits such a protein-I.S. pattern .

For example, Vasculitis is characterized by the following rule, involving fuzzy propositions :

IF C3-Complement Fraction is DECREASED OR NORMAL
AND A1- Antitrypsine is DECREASED OR NORMAL
AND Orosomucoid is INCREASED
AND Haptoglobin is VERY INCREASED
AND C-Reactive Protein is VERY INCREASED
THEN VASCULITIS.

Linguistic descriptions have been established [3,4] for the eleven groups (Normal condition, eight I.S. and two syndromes without Inflammation). This linguistic pattern expresses the relative protein variations among the I.S.

III - FUZZY INTERPRETATIONS

A - Pattern

The linguistic informations conveyed by the pattern are treated by conjunctions of fuzzy propositions of the form "X is F", where the variable X represents one of the five proteins and F is a subset of the universe of discourse of the corresponding protein. For example, in the description of Vasculitis, one has :

"Orosomucoid (X) is INCREASED (F)".

The fuzzy set F, which expresses a relative variation compared to the Normal condition (basic levels), is depicted with normalized values in Figure1.

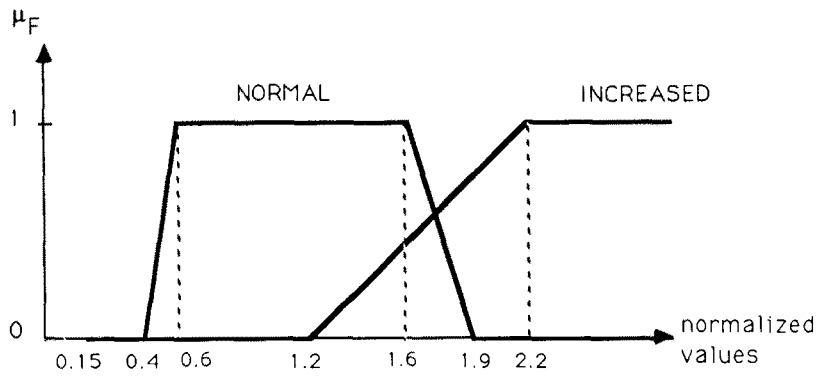

Fig.1 : Illustration of "Orosomucoid is INCREASED".

In order to compare different protein variations among various syndromes, it is customary to represent these variations in a relative scale. For example, NORMAL Orosomucoid is centered on value "0.95 g/l" (mean value evaluated from a reference population of normal subjects with same age and sex), which is here assigned reference value "1" (normalized value). Hence, 1.33 g/l is assigned value 1.40 (i.e. 1.33/0.95). So that fuzzy sets are now expressed according to their relative variations.

B - Relative Weighting

Practically, some proteins are more or less important in the characterization of a group. For a given group, relative importance among proteins can be handled by means of weights (a,b,c, ...) ranging in [0,1]. A value "0" weight assigned to a protein means that this protein is not important at all in the evaluation of the group and hence it can be deleted, whereas a value "1" weight does not modify the importance of the protein. Intermediate grades of importance can be tuned by adjusting values of weights within the interval (0,1).

In the pattern, fuzzy propositions (X is F) characterizing a given group, appear as conjunctions (ANDs), and assignement of a weight "a" to take into account the relative importance of protein variations, assumes the following form [14,16], for F fuzzy subset of a universe of discourse U :

$$F^a = \text{Max}(1\text{-}a, F), \quad \text{i.e.,} \quad \forall x \in U, \ \mu_{F^a}(x) = (1\text{-}a) \vee \mu_F(x).$$

More generally, a t-conorm could replace the max-operator in the above formula [15].

Limit cases

a=0 : $\forall x \in U, \ \mu_{F^0}(x) = 1$, F^0 is neutral for conjunctions and therefore, it can be deleted.

a=1 : $\forall x \in U, \ \mu_{F^1}(x) = \mu_F(x)$, the weight has no effect.

In the case of Vasculitis, the following weights can be assigned, yielding the modified rule:

IF C3-Complement Fraction is (DECREASED OR NORMAL)$^{0.1}$
AND A1- Antitrypsine is DECREASED OR NORMAL
AND Orosomucoid is (INCREASED) $^{0.8}$
AND Haptoglobin is (VERY INCREASED) $^{0.3}$
AND C-Reactive Protein is (VERY INCREASED) $^{0.8}$
THEN VASCULITIS.

Note that C3-Complement Fraction could be neglected (weight close to 0) and that no weight is assigned to DECREASED OR NORMAL for A1- Antitrypsine (weight equal to 1, i.e. no effect of the weight).

The modified fuzzy variations of the proteins, with the above weights, are depicted in Table 1.

C - Fuzzy Interpretations of Data

In order to classify patients according to the pattern of Medical Knowledge, the five protein levels must be measured for each patient. Each measured value (m) is normalized (yielding value d=m/M) and then transformed into a fuzzy number (D) meaning "around d", to take into account imprecision in measurements and fiducial interpretations. M is a normalizing factor (depending on age and sex) and determined by a polynomial regression over a sample of normal patients. For example, if a patient has an Orosomucoid level of 1.20 g/l, this value (i.e. (m)) is normalized yiedling 1.26 (d=1.20/0.95, where M=0.95 is the normalizing factor for a sixty-two years old woman) and finally transformed into the fuzzy number D, as represented in Figure 2. Note that the bandwidth δ of the fuzzy number D is equal to d x 0.10, that is 0.13 here.

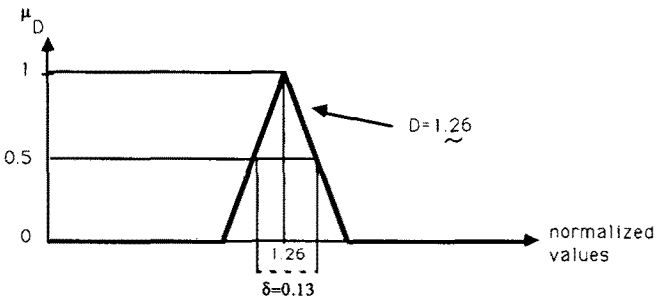

Fig. 2 : Fuzzy number D meaning "Orosomucoid is around 1.26".

V-MATCHING

Fuzzy numbers evaluated from measurements over patients are matched to the corresponding fuzzy sets in the I.S.-Protein pattern, by means of three measures or indexes : possibility measure (π), necessity measure (ν), and truth-possibility index (ρ).

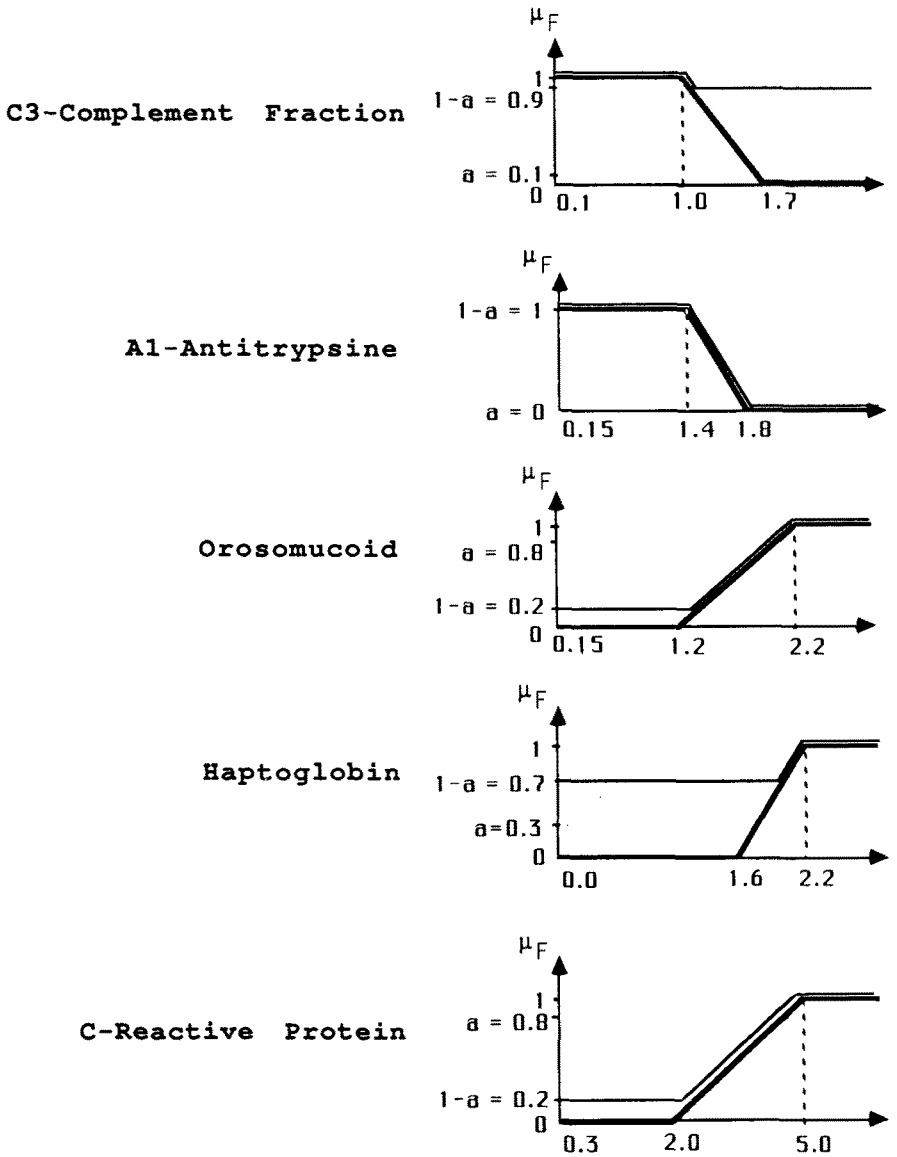

Table 1 : Weighted fuzzy sets (F^a : thiner curves) in the pattern, for Vasculitis.

For each protein (X), let F (or F^a in a weighted form) be a fuzzy set characterizing X in a group, and let D be the fuzzy number issued from the seric level of X on a patient,

i) Possibility measure [17]. By definition, π (F,D) = Sup (F $\cap$ D).

ii) Necessity measure [5]. By definition, $\nu(F,D) = 1 - \pi(F',D)$, where F' denotes the fuzzy complement of F, i.e. F' = 1-F. Note that $\nu(F,D) = 1 - \text{Sup}(F' \cap D) = \text{Inf}(F \cup D')$.

iii) Truth-possibility index [10,11]. By definition , $\rho(F,D) = \pi(\tau_0,\tau_1)$, where τ_0 and τ_1 are related to truth-qualification [17] as follows .

τ_0 is a fuzzy subset of V=[0,1], defined as the compatibility of the fuzzy proposition "X(patient) is D" with the fuzzy proposition in the pattern "X is F" :

$\tau_0 = \mu_F(D)$,

i.e. for all v in V, $\mu_{\tau_0}(v) = \sup \{ \mu_D(x)$ such that $\mu_F(x)=v \}$.

τ_0 is the truth-value of "X is F" relative to "X is D", so that $\mu_D(x) = \mu_{\tau_0}(\mu_F(x))$.

$\tau_1 = (\mu_F(D'))'$,

i.e. for all v in V, $\mu_{\tau_1}(v) = \inf \{ \mu_D(x)$ such that $\mu_F(x)=v \}$.

Given F and D, when no τ exists such that $\mu_D(x) = \mu_\tau(\mu_F(x))$, the following properties [10] are particularly useful for applications :

$$\forall x \in U, \quad \mu_{\tau_1}(\mu_F(x)) \leq \mu_D(x) \leq \mu_{\tau_0}(\mu_F(x)),$$

moreover τ_1 is the greatest τ such that $\mu_\tau(\mu_F(x)) \leq \mu_D(x)$ and τ_0 is the smallest τ such that $\mu_D(x) \leq \mu_\tau(\mu_F(x))$.

One has the semantic entailment :

$$[(X \text{ is } F) \text{ is } \tau_1] \mapsto X \text{ is } D \mapsto [(X \text{ is } F) \text{ is } \tau_0].$$

With the particular fuzzy sets (F) and fuzzy numbers (D) in our model, one can show that the following ranking holds [3] (see figure 3 for an illustration) :

$$\nu \leq \rho \leq \pi$$

so that these indexes can be chosen according to optimistic or pessimistic considerations.

For each of the eleven groups, a patient's condition yields five (one for each protein) triples (ν_i, ρ_i, π_i), $i = 1, \ldots, 5$, which are combined using the min- operator, expressing conjunctions : $(\nu, \rho, \pi) = (\min_i \nu_i, \min_i \rho_i, \min_i \pi_i)$.

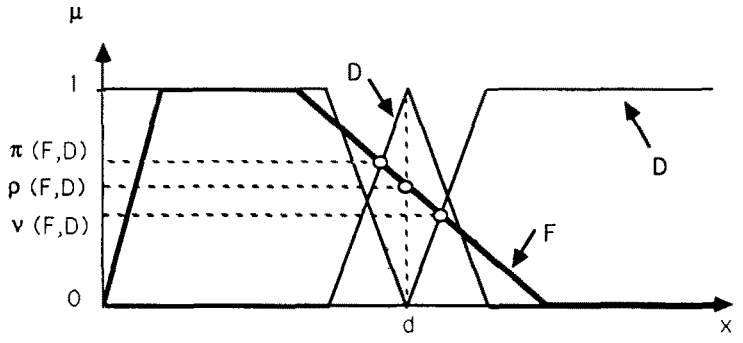

Fig. 3 : Compatibility measures or indexes.

Finally for each patient, one has three fuzzy sets f, g and h derived from ν, ρ, π, respectively, and such that for each group (y), one has $\mu_f(y)=\nu$, $\mu_g(y)=\rho$, $\mu_h(y)=\pi$.

The separating power s(f) [6] allows to evaluate to which extent a fuzzy subset f of a universe of discourse E (E is here the set of the eleven groups), "separates" optimally E into a non fuzzy partition (A,A'), see [7]. The set A is defined as : $s(f) = f*A = \sup \{ f*B$ such that $E \supseteq B, B \neq \varnothing \}$ and $f*B = |\text{card}(f_B) / \text{card}(B) - \text{card}(f_{B'}) / \text{card}(B')|$, where f_B denotes the restriction of μ_f to B, card (B) is the cardinality of B, and card (f_B) is the fuzzy cardinality of f_B ; for example,

$$\text{card}(f_B) = \Sigma_{y \in B} \mu_f(y) .$$

Applying the separating power to f, g and h, it is associated to each fuzzy set the corresponding optimal partition (playing the role of (A, A') above), i.e. (F,F') to f, (G,G') to g, and (H,H') to h. F, G and H are finally the (non fuzzy) sets of diagnostic groups assigned to patients.

V - PATIENTS CASES

The patients cases we present here, have been medically diagnosed as Vasculitis.

* CASE 1 :	C3	A1AT	OrosoM.	Hapto.	CRP
raw data (g/l)	0.53	2.97	3.50	3.67	0.100
normalized data	0.66	1.30	3.98	2.05	16.67

Weighted or Non-Weighted process:

F = Collagen Diseases	($\min_i \nu_i$= 0.69)	s(f)= 0.59
G = Collagen Diseases	($\min_i \rho_i$= 0.84)	s(f)= 0.74
Vasculitis	($\min_i \rho_i$= 0.75)	
H = Collagen Diseases	($\min_i \pi_i$= 0.87)	s(f)= 0.76
Vasculitis	($\min_i \pi_i$= 0.82)	

* CASE 2:	C3	A1AT	OrosoM.	Hapto.	CRP
raw data (g/l)	1.62	1.13	1.75	10.0	0.060
normalized data	2	0.50	1.99	5.59	10.0

Non-Weighted process:

F = Collagen Diseases	($\min_i \nu_i$=0.30)	s(f)= 0.30
G = Collagen Diseases	($\min_i \rho_i$=0.38)	s(g)= 0.38
H = Collagen Diseases	($\min_i \pi_i$=0.42)	s(h)= 0.40

Weighted process:

F = Vasculitis	($\min_i \nu_i$= 0.66)	s(f)= 0.66
G = Vasculitis	($\min_i \rho_i$= 0.77)	s(g)= 0.77
H = Vasculitis	($\min_i \pi_i$= 0.80)	s(h)= 0.78

Let us consider the C3-Complement Fraction (C3). Either the compatibility measures/indexes take value "1" (like in case 1) and hence weighting is of no use (recall the definition of F^a) or, the compatibility measures/indexes take a very low value (this value is

equal to zero in case 2) so that weighting is fully justified for it allows to assign the right diagnosis of Vasculitis. Analogous considerations hold for the other proteins.

Two cases have been reported here. During the conference, detailed results will be presented and discussed (160 patients have been explored without weighting [3], weights are presently being incorporated).

REFERENCES

[1] ADLASSNIG K.P., KOLARZ G., SCHEITHAUER W. : 1985, "Present State of the Medical Expert System CADIAG-2", Meth. Inf. Med., 24, pp. 13-20.

[2] BARTOLIN R., BOUVENOT G., SOULA G., SANCHEZ E.: 1982,"Apport des sous-ensembles flous à l'aide au diagnostic biomédical : à propos de deux applications concrètes", Sem. Hôp. Paris, 58 (22),1631-1635.

[3] BARTOLIN R. :1987, "Aide au diagnostic médical par mesures de comparaisons floues et pouvoir séparateur. Approche linguistique des profils protéiques inflammatoires biologiques". Thèse d'Etat en biologie humaine, Marseille, France.

[4] BARTOLIN R., BONNIOL V., SANCHEZ E.: 1987, "Linguistic characterizations of inflammatory syndromes and fuzzy matching measures". Int. Symp. on Fuzzy Systems and Knowledge Engineering (ISFK), Guangzhou & Guiyang, China, proceedings, pp. 455-464.

[5] CAYROL M., FARRENY H., PRADE H. : 1982, "Fuzzy Pattern Matching", Kybernetes, Vol. 11, pp. 103-116.

[6] DUJET C. : 1980, "Valuation et séparation dans les ensembles flous", Structures de l'information, 18 - Publications du CNRS, pp. 95-105.

[7] DUJET C. , SANCHEZ E., BARTOLIN R. : To appear in 1988, "Fuzzy sets and partitions. A new approach applied to medical decision assistance : inflammatory syndromes", new Int. J. for A.I. in medicine.

[8] FROT J.C., HOFMANN H., MULLER F., BENAZET M.F., GIRAUDET P. :1984, "Le concept du profil protéique", Ann. Biol. Clin., 42, pp. 1-8

[9] GIRAUDET P., FAURE A., FROT J.C. : 1984, "Exploration clinique de la reaction inflammatoire, A: Exploration biochimique du syndrome inflammatoire chez l'adulte", Vigot, Paris, pp. 190-214.

[10] SANCHEZ E. : 1978, "On truth-qualification in natural languages", Int. Conf. on Cybernetics and Society, Tokyo, proceedings, pp.1233-1236.

[11] SANCHEZ E. : 1980, "Mesures de possibilité, Qualifications de vérité et Classification de formes linguistiques en Médecine", Actes Table Ronde CNRS, Lyon.

[12] SANCHEZ E. , GOUVERNET J., BARTOLIN R. , VOVAN L.: 1982, "Linguistic Approach in fuzzy logic of the WHO classification of dislipoproteinemias", in: Recent Developments in Fuzzy Set and Possibility Theory, R.R. Yager (Ed.), Pergamon Press, New-York, pp. 582-588.

[13] SANCHEZ E. : 1986, "Medical applications with fuzzy sets", in : A. JONES et al (Eds), Fuzzy sets theory and applications, D.Reidel, pp. 331-347.

[14] SANCHEZ E. : 1987, "Soft Queries in Knowledge Systems". Second I.F.S.A. World Congress , Tokyo, proceedings, pp. 597-599.

[15] SANCHEZ E. : to appear in 1988, "Handling Requests in Intelligent Retrieval", in: Contributions on Approximate Reasoning and Artificial Intelligence, M. Delgado and J.L. Verdegay (Eds).

[16] SANCHEZ E. : "Importance in Knowledge Systems", submitted to Data & Knowledge Engineering (special isue on Fuzzy data bases).

[17] ZADEH L.A. : 1978, "Fuzzy sets as a basis for a theory of possibility", Fuzzy Sets and Systems, 1, pp. 3-28.