Available online at www.sciencedirect.com
SCIENCE DIRECT®
Data & Knowledge Engineering 50 (2004) 291–304
DATA & KNOWLEDGE ENGINEERING
www.elsevier.com/locate/datak

A note on web intelligence, world knowledge and fuzzy logic

Lotfi A. Zadeh *

Berkeley Initiative in Soft Computing (BISC), Computer Science Division and the Electronics Research Laboratory, Department of EECS, University of California, Berkeley, CA 94720-1776, USA

Available online 22 April 2004

To Vilem Novak and Irina Perfilieva

Abstract

Existing search engines—with Google at the top—have many remarkable capabilities; but what is not among them is deduction capability—the capability to synthesize an answer to a query from bodies of information which reside in various parts of the knowledge base.

In recent years, impressive progress has been made in enhancing performance of search engines through the use of methods based on bivalent logic and bivalent-logic-based probability theory. But can such methods be used to add nontrivial deduction capability to search engines, that is, to upgrade search engines to question-answering systems? A view which is articulated in this note is that the answer is "No." The problem is rooted in the nature of world knowledge, the kind of knowledge that humans acquire through experience and education.

It is widely recognized that world knowledge plays an essential role in assessment of relevance, summarization, search and deduction. But a basic issue which is not addressed is that much of world knowledge is perception-based, e.g., "it is hard to find parking in Paris," "most professors are not rich," and "it is unlikely to rain in midsummer in San Francisco." The problem is that (a) perception-based information is intrinsically fuzzy; and (b) bivalent logic is intrinsically unsuited to deal with fuzziness and partial truth.

To come to grips with fuzziness of world knowledge, new tools are needed. The principal new tool—a tool which is briefly described in this note—is Precisiated Natural Language (PNL). PNL is based on fuzzy logic and has the capability to deal with partiality of certainty, partiality of possibility and partiality of truth. These are the capabilities that are needed to be able to draw on world knowledge for assessment of relevance, and for summarization, search and deduction.

Keywords: Web intelligence; World knowledge; Fuzzy logic; Precisiated Natural Language

* Tel.: +1-510-642-4959; fax: +1-510-642-1712.
E-mail address: zadeh@cs.berkeley.edu (L.A. Zadeh).

0169-023X/$ - see front matter
doi:10.1016/j.datak.2004.04.001

1. Introduction

In moving further into the age of machine intelligence and automated reasoning, we have reached a point where we can speak, without exaggeration, of systems which have a high machine IQ (MIQ) [15]. The Web and especially search engines—with Google at the top—fall into this category. In the context of the Web, MIQ becomes Web IQ, or WIQ, for short.

Existing search engines have many remarkable capabilities. However, what is not among them is deduction capability—the capability to answer a query by a synthesis of information which resides in various parts of the knowledge base. A question-answering system is by definition a system which has this capability. One of the principal goals of Web intelligence is that of upgrading search engines to question-answering systems. Achievement of this goal requires a quantum jump in the WIQ of existing search engines [1].

Can this be done with existing tools such as the Semantic Web [3], Cyc [7], OWL [11] and other ontology-centered systems [10,12]—tools which are based on bivalent logic and bivalent-logic-based probability theory? It is beyond question that, in recent years, very impressive progress has been made through the use of such tools. But can we achieve a quantum jump in WIQ? A view which is advanced in the following is that bivalent-logic-based methods have intrinsically limited capability to address complex problems which arise in deduction from information which is pervasively ill-structured, uncertain and imprecise.

The major problem is world knowledge—the kind of knowledge that humans acquire through experience and education [4]. Simple examples of fragments of world knowledge are: Usually it is hard to find parking near the campus in early morning and late afternoon; Berkeley is a friendly city; affordable housing is nonexistent in Palo Alto; almost all professors have a Ph.D. degree; most adult Swedes are tall; and Switzerland has no ports.

Much of the information which relates to world knowledge—and especially to underlying probabilities—is perception-based [8,14] (Fig. 1). Reflecting the bounded ability of sensory organs, and ultimately the brain, to resolve detail and store information, perceptions are intrinsically imprecise. More specifically, perceptions are f-granular in the sense that (a) the boundaries of perceived classes are unsharp; and (b) the values of perceived attributes are granular, with a granule being a clump of values drawn together by indistinguishability, similarity, proximity or functionality [2].

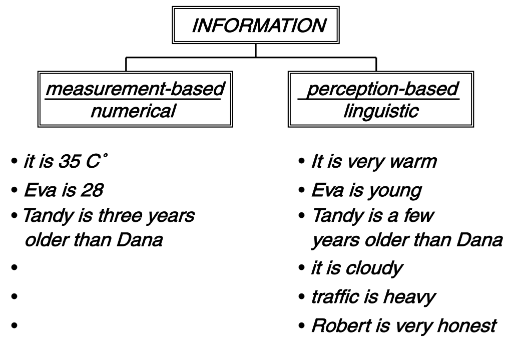

Fig. 1. Measurement-based and perception-based information.

Imprecision of perception-based information is a major obstacle to dealing with world knowledge through the use of methods based on bivalent logic and bivalent-logic-based probability theory—both of which are intolerant of imprecision and partial truth. What is the basis for this contention? A very simple example offers an explanation.

Suppose that I have to come up with a rough estimate of Pat's age. The information which I have is: (a) Pat is about 10 years older than Carol; and (b) Carol has two children: a son, in mid-twenties; and a daughter, in mid-thirties.

How would I come up with an answer to the query q: How old is Pat? First, using my world knowledge, I would estimate Carol's age, given (b); then I would add "about 10 years," to the estimate of Carol's age to arrive at an estimate of Pat's age. I would be able to describe my estimate in a natural language, but I would not be able to express it as a number, interval or a probability distribution.

How can I estimate Carol's age given (b)? Humans have an innate ability to process perception-based information—an ability that bivalent-logic-based methods do not have; nor would such methods allow me to add "about 10 years" to my estimate of Carol's age.

There is another basic problem—the problem of relevance. Suppose that instead of being given (a) and (b), I am given a collection of data which includes (a) and (b), and it is my problem to search for and identify the data that are relevant to the query. I come across (a). Is it relevant to q? By itself, it is not. Then I come across (b). Is it relevant to q? By itself, it is not. Thus, what I have to recognize is that, in isolation, (a) and (b) are irrelevant, that is, uninformative, but, in combination, they are relevant i.e., are informative. It is not difficult to recognize this in the simple example under consideration, but when we have a large database, the problem of identifying the data which in combination are relevant to the query, is very complex.

Suppose that a proposition, p, is, in isolation, relevant to q, or, for short, is i-relevant to q. What is the degree to which p is relevant to q? For example, to what degree is the proposition: Carol has two children: a son, in mid-twenties, and a daughter, in mid-thirties, relevant to the query: How old is Carol? To answer this question, we need a definition of measure of relevance. The problem is that there is no quantitative definition of relevance within existing bivalent-logic-based theories. We will return to the issue of relevance at a later point.

The example which we have considered is intended to point to the difficulty of dealing with world knowledge, especially in the contexts of assessment of relevance and deduction, even in very simple cases. The principal reason is that much of world knowledge is perception-based, and existing theories of knowledge representation and deduction provide no tools for this purpose. Here is a test problem involving deduction form perception-based information.

The Tall Swedes Problem

Perception: Most adult Swedes are tall, with adult defined as over about 20 years in age.

Query: What is the average height of Swedes?

A fuzzy-logic-based solution of the problem will be given at a later point.

A concept which provides a basis for computation and reasoning with perception-based information is that of Precisiated Natural Language (PNL) [13]. The capability of PNL to deal with perception-based information suggests that it may play a significant role in dealing with world knowledge. A quick illustration is the Carol example. In this example, an estimate of

Carol's age, arrived at through the use of PNL would be expressed as a bimodal distribution of the form

$$\text{Age(Carol) is } ((P_1, V_1) + \cdots + (P_n, V_n)), \quad i = 1, \ldots, n$$

where the V_i are granular values of Age, e.g., less than about 20, between about 20 and about 30, etc.; P_i is a granular probability of the event (Age(Carol) is V_i), $i = 1, \ldots, n$; and + should be read as "and." A brief description of the basics of PNL is presentee in the following.

2. Precisiated Natural Language (PNL)

PNL deals with perceptions indirectly, through their description in a natural language, Zadeh [13]. In other words, in PNL a perception is equated to its description in a natural language. PNL is based on fuzzy logic—a logic in which everything is, or is allowed to be, a matter of degree. It should be noted that a natural language is, in effect, a system for describing perceptions.

The point of departure in PNL is the assumption that the meaning of a proposition, p, in a natural language, NL, may be represented as a generalized constraint of the form (Fig. 2)

$$X \text{ is} r\ R,$$

where X is the constrained variable, R is a constraining relation which, in general, is not crisp (bivalent); and r is an indexing variable whose values define the modality of the constraint. The principal modalities are: possibilistic ($r =$ blank); veristic ($r = v$); probabilistic ($r = p$); random set ($r = rs$); fuzzy graph ($r = fg$); usuality ($r = u$); and Pawlak set ($r = ps$). The set of all generalized constraints, together with their combinations, qualifications and rules of constraint propagation, constitutes the Generalized Constraint Language (GCL). By construction, GCL is maximally expressive. In general, X, R and r are implicit in p. Thus, in PNL the meaning of p is precisiated through explicitation of the generalized constraint which is implicit in p, that is, through translation into GCL (Fig. 4). Translation of p into GCL is exemplified by the following.

(a) Monika is young → Age (Monika) is young, where young is a fuzzy relation which is characterized by its membership function μ_{young}, with $\mu_{\text{young}}(u)$ representing the degree to which a numerical value of age, u, fits the description of age as "young."

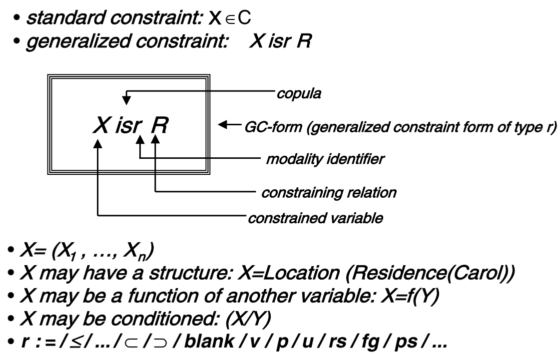

Fig. 2. Generalized constraint.

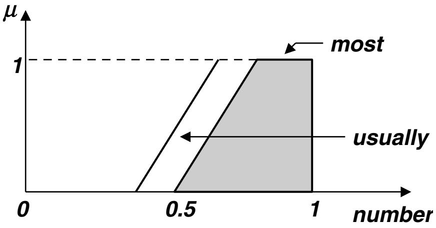

Fig. 3. Calibration of *most* and *usually* represented as trapezoidal fuzzy numbers.

(b) Carol lives in a small city near San Francisco $\rightarrow$ Location (Residence(Carol)) is SMALL [City; μ] $\cap$ NEAR [City; μ], where SMALL [City; μ] is the fuzzy set of small cities; NEAR [City; μ] is the fuzzy set of cities near San Francisco; and $\cap$ denotes fuzzy set intersection (conjunction).

(c) Most Swedes are tall $\rightarrow$ $\sum$Count(tall·Swedes/Swedes) is most. In this example, the constrained variable is the relative count of tall Swedes among Swedes, and the constraining relation is the fuzzy quantifier "most," with "most" represented as a fuzzy number (Fig. 3). The relative $\sum$count is defined as follows. If A and B are fuzzy sets in a universe of discourse $U = \{u_i, \ldots, u_n\}$, with the grades of membership of u_i in A and B being μ_i and v_i, respectively, then by definition the relative $\sum$Count of A in B is expressed as

$$\sum \mathrm{Count}(A/B) = \frac{\sum_i \mu_i \wedge v_i}{\sum_i v_i},$$

where the conjunction, $\wedge$, is taken to be min. More generally, conjunction may be taken to be a t-norm [9].

What is the rationale for introducing the concept of a generalized constraint? Conventional constraints are crisp (bivalent) and have the form $X \varepsilon C$, where X is the constrained variable and C is a crisp set. On the other hand, perceptions are f-granular, as was noted earlier. Thus, there is a mismatch between f-granularity of perceptions and crisp constraints. What this implies is that the meaning of a perception does not lend itself to representation as a collection of crisp constraints; what is needed for this purpose are generalized constraints or, more generally, an element of the Generalized Constraint Language (GCL) [13].

Representation of the meaning of a perception as an element of GCL is a first step toward computation with perceptions—computation which is needed for deduction form data which are resident in world knowledge. In PNL, a concept which plays a key role in deduction is that of a protoform—an abbreviation of "prototypical form," [13].

If p is a proposition in a natural language, NL, its protoform, PF(p), is an abstraction of p which places in evidence the deep semantic structure of p. For example, the protoform of "Monika is young," is "$A(B)$ is C," where A is abstraction of "Age," B is abstraction of "Monika" and C is abstraction of "young." Similarly,

Most Swedes are tall $\rightarrow$ Count(A/B) is Q,

where A is abstraction of "tall Swedes," B is abstraction of "Swedes," and Q is abstraction of "most." Two propositions, p and q, are protoform-equivalent, or PF-equivalent for short, if they have identical protoforms. For example, p: Most Swedes are tall, and q: Few professors are rich, are PF-equivalent.

The concept of PF-equivalence suggests an important mode of organization of world knowledge, namely protoform-based organization. In this mode of organization, items of knowledge are grouped into PF-equivalent classes. For example, one such class may be the set of all propositions whose protoform is $A(B)$ is C, e.g., Monika is young. The partially instantiated class Price(B) is low, would be the set of all objects whose price is low. As will be seen in the following, protoform-based organization of world knowledge plays a key role in deduction form perception-based information.

Basically, abstraction is a means of generalization. Abstraction is a familiar and widely used concept. In the case of PNL, abstraction plays an especially important role because PNL abandons bivalence. Thus, in PNL, the concept of a protoform is not limited to propositions whose meaning can be represented within the conceptual structure of bivalent logic.

In relation to a natural language, NL, the elements of GCL may be viewed as precisiations of elements of NL. Abstraction of elements of GCL gives rise to what is referred to as Protoform Language, PFL (Fig. 4). A consequence of the concept of PF-equivalence is that the cardinality of PFL is orders of magnitude smaller than that of GCL, or, equivalently, the set of precisiable propositions in NL. The small cardinality of PFL plays an essential role in deduction.

The principal components of the structure of PNL (Fig. 5) are: (1) a dictionary from NL to GCL; (2) a dictionary from GCL to PFL (Fig. 6); (3) a multiagent, modular deduction database, DDB; and (4) a world knowledge database, WKDB. The constituents of DDE are modules, with a module consisting of a group of protoformal rules of deduction, expressed in PFL (Fig. 7), which are drawn from a particular domain, e.g., probability, possibility, usuality, fuzzy arithmetic, fuzzy logic, search, etc. For example, a rule drawn from fuzzy logic is the compositional rule of inference [9], expressed as

symbolic part →

$$\frac{\begin{array}{c} X \text{ is } A \\ (X, Y) \text{ is } B \end{array}}{Y \text{ is } A \circ B}$$

$$\mu_{A \circ B}(v) = \max_u (\mu_A(u) \wedge \mu_B(u, v))$$

← computational part

where $A \circ B$ is the composition of A and B, defined in the computational part, in which μ_A, μ_B and $\mu_{A \circ B}$ are the membership functions of A, B and $A \circ B$, respectively. Similarly, a rule drawn from probability is

symbolic part →

$$\frac{\text{Prob } (X \text{ is } A) \text{ is } B}{\text{Prob } (X \text{ is } C) \text{ is } D}$$

$$\mu_D(u) = \max_g \left(\mu_B\left(\int_U \mu_A(u) g(u) du\right)\right)$$

subject to: $v = \int_U \mu_C(u) g(u) du$

$$\int_U g(u) du = 1$$

← computational part

where D is defined in the computational part and g is the probability density function.

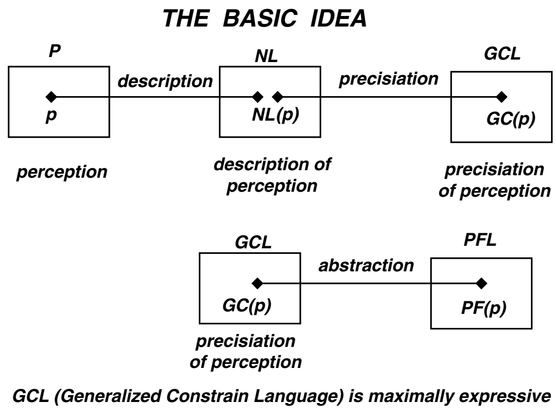

Fig. 4. Precisiation and abstraction.

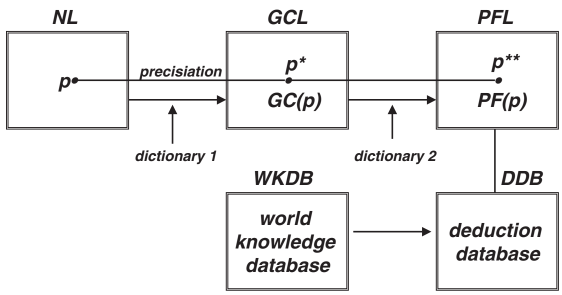

Fig. 5. Basic structure of PNL.

1:

proposition in NL	precisiation
p	p* (GC-form)
most Swedes are tall	$\sum$ Count (tall.Swedes/Swedes) is most

2:

precisiation	protoform
p* (GC-form)	PF(p*)
$\sum$ Count (tall.Swedes/Swedes) is most	Q A's are B's

Fig. 6. Structure of PNL: dictionaries.

The rules of deduction in DDB are, basically, the rules which govern propagation of generalized constraints. Each module is associated with an agent whose function is that of controlling

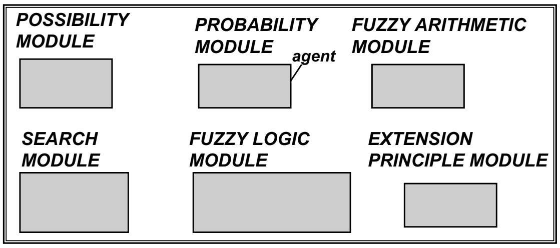

Fig. 7. Basic structure of PNL: modular deduction database.

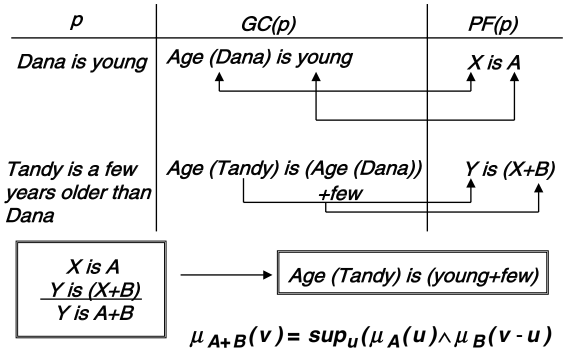

Fig. 8. Example of protoformal reasoning.

execution of rules and performing embedded computations. The top-level agent controls the passing of results of computation from a module to other modules. The structure of protoformal, i.e., protoform-based, deduction is shown in Fig. 5. A simple example of protoformal deduction is shown in Fig. 8.

The principal deduction rule in PNL is the extension principle [17]. Its symbolic part is expressed as

$$\frac{f(X) \text{ is } A}{g(X) \text{ is } B}$$

in which the antecedent is a constraint on X through a function of X, $f(X)$; and the consequent is the induced constraint on a function of X, $g(X)$.

The computational part of the rule is expressed as

$$\mu_B(v) = \sup_u(\mu_A(f(u)))$$

subject to

$$v = g(u)$$

To illustrate the use of the extension principle, we will consider the Tall Swedes problem

p: Most adult Swedes are tall
q: What is the average height of Swedes?

Let $P_1(u_1,\dots,u_N)$ be a population of Swedes, with the height of u_i being h_i, $i=1,\dots,N$, and $\mu_a(u_i)$ representing the degree to which u_i is an adult. The average height of adult Swedes is denoted as h_{ave}. The first step is precisiation of p, using the concept of relative $\sum$Count:

$$p \to \sum \text{Count}(\text{tall} \wedge \text{adult} \cdot \text{Swedes}/\text{adult} \cdot \text{Swedes}) \text{ is most; more explicitly,}$$

$$p \to \frac{\sum_i \mu_{\text{tall}}(h_i) \wedge \mu_a(u_i)}{\sum_i \mu_a(u_i)} \text{ is most.}$$

The next step is precisiation of q:

$$q \to h_{\text{ave}} = \frac{1}{N}\sum_i h_i$$

$$h_{\text{ave}} \text{ is } ?B,$$

where B is a fuzzy number.

Using the extension principle, computation of h_{ave} as a fuzzy number is reduced to the solution of the variational problem.

$$\mu_B(v) = \sup_{\underline{h}}\left(\mu_{\text{most}}\left(\frac{\sum_i \mu_{\text{tall}}(h_i) \wedge \mu_a(u_i)}{\sum_i \mu_a(u_i)}\right)\right), \quad \underline{h} = (\underline{h}_1,\dots,\underline{h}_N)$$

subject to

$$v = \frac{1}{N}\left(\sum_i h_i\right)$$

Note that we assume that the problem is solved once it is reduced to the solution of a well-defined computational problem.

3. PNL as a definition language

One of the important functions of PNL is that of serving as a high level definition language. More concretely, suppose that I have a concept, C, which I wish to define. For example, the concept of distance between two real-valued functions, f and g, defined on the real line.

The standard approach is to use a metric such as L_1 or L_2. But a standard metric may not capture my perception of the distance between f and g. The PNL-based approach would be to describe my perception of distance in a natural language and then precisiate the description.

This basic idea opens the door to (a) definition of concepts for which no satisfactory definitions exist, e.g., the concepts of causality, relevance and rationality, among many others; and (b) redefinition of concepts whose existing definitions do not provide a good fit to reality. Among such concepts are the concepts of similarity, stability, independence, stationarity and risk.

How can we concretize the meaning of "good fit?" In what follows, we do this through the concept of cointension.

More specifically, let U be a universe of discourse and let C be a concept which I wish to define, with C relating to elements of U. For example, U is a set of buildings and C is the concept of tall building. Let $p(C)$ and $d(C)$ be, respectively, my perception and my definition of C. Let $I(p(C))$ and $I(d(C))$ be the intensions of $p(C)$ and $d(C)$, respectively, with "intension" used in its logical sense [5,6], that is, as a criterion or procedure which identifies those elements of U which fit $p(C)$ or $d(C)$. For example, in the case of tall buildings, the criterion may involve the height of a building.

Informally, a definition, $d(C)$, is a good fit or, more precisely, is cointensive, if its intension coincides with the intension of $p(C)$. A measure of goodness of fit is the degree to which the intension of $d(C)$ coincides with that of $p(C)$. In this sense, cointension is a fuzzy concept. As a high level definition language, PNL makes it possible to formulate definitions whose degree of cointensiveness is higher than that of definitions formulated through the use of languages based on bivalent logic.

A substantive exposition of PNL as a definition language is beyond the scope of this note. In what follows, we shall consider as an illustration a relatively simple version of the concept of relevance.

4. Relevance

We shall examine the concept of relevance in the context of a relational model such as shown in Fig. 9. For concreteness, the attributes $A_1, \ldots, A_n$ may be interpreted as symptoms and D as diagnosis. For convenience, rows which involve the same value of D are grouped together. The entries are assumed to be labels of fuzzy sets. For example, A_5 may be blood pressure and a_{53} may be "high."

An entry represented as $*$ means that the entry in question is conditionally redundant in the sense that its value has no influence on the value of D (Fig. 10). An attribute, A_j, is redundant, and hence deletable, if it is conditionally redundant for all values of Name. An algorithm, termed

RELEVANCE, REDUNDANCE AND DELETABILITY

DECISION TABLE

Name	A_1	A_j	A_n	D
$Name_1$	a_{11}	a_{1j}	a_{in}	d_1
.	.	.	.	.
$Name_k$	a_{k1}	a_{kj}	a_{kn}	d_1
$Name_{k+1}$	$a_{k+1,1}$	$a_{k+1,j}$	$a_{k+1,n}$	d_2
.	.	.	.	.
$Name_l$	a_{l1}	a_{lj}	a_{ln}	d_l
.	.	.	.	.
$Name_n$	a_{m1}	a_{mj}	a_{mn}	d_r

A_j: jth symptom

a_{ij}: value of jth symptom of Name

D: diagnosis

Fig. 9. A relational model of decision.

REDUNDANCE ⟶ DELETABILITY

Name	A_1	A_j	A_n	D
.	.	.	.	.
$Name_r$	a_{r1}	*	a_{rn}	d_2
.	.	.	.	.

A_j is conditionally redundant for $Name_r$, A_1 is a_{r1}, A_n is a_{rn}
*If D is d_s for all possible values of A_j in **

A_j is redundant if it is conditionally redundant for all values of Name

- *compactification algorithm (Zadeh, 1976); Quine-McCluskey algorithm*

Fig. 10. Conditional redundance and redundance.

compactification, which identifies all deletable attributes is described in Zadeh [16]. Compactification algorithm is a generalization of the Quine-McCluskey algorithm for minimization of switching circuits. The reduct algorithm in the theory of rough sets is closely related to the compactification algorithm.

The concept of relevance (informativeness) is weaker and more complex than that of redundance (deletability). As was noted earlier, it is necessary to differentiate between relevance in isolation (*i*-relevance) and relevance as a group. In the following, relevance should be interpreted as *i*-relevance.

A value of A_1 say a_{rj}, is irrelevant (uninformative) if the proposition A_j is a_{rj} does not constrain D (Fig. 11). For example, the knowledge that blood pressure is high may convey no information about the diagnosis (Fig. 12).

An attribute, A_j, is irrelevant (uninformative) if, for all a_{rj}, the proposition A_j is a_{rj} does not constrain D. What is important to note is that irrelevance does not imply deletability, as redundance does. The reason is that A_j may be *i*-irrelevant but not irrelevant in combination with other attributes. An example is shown in Fig. 13.

As defined above, relevance and redundance are bivalent concepts, with no degrees of relevance or redundance allowed. But if the definitions in question are interpreted as idealizations, then a measure of the departure from the ideal could be used as a measure of the degree of relevance or

Name	A_1	A_j	A_n	D
$Name_r$	.	a_{ij}	.	d_1 . d_1
$Name_{i+s}$	.	a_{ij}	.	d_2 . d_2

(A_j is a_{ij}) is irrelevant (uninformative)

Fig. 11. Irrelevance.

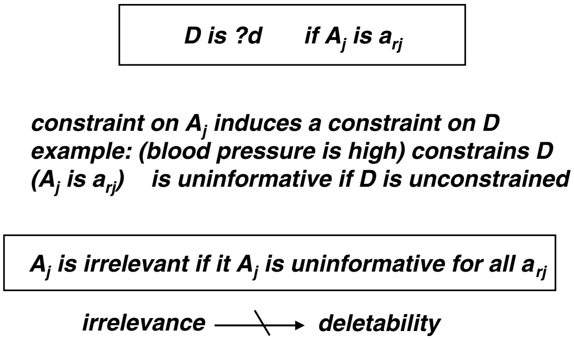

Fig. 12. Relevance and irrelevance.

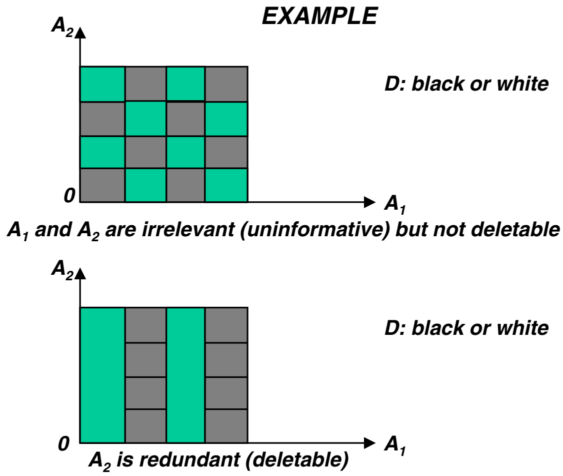

Fig. 13. Redundance and irrelevance.

redundance. Such a measure could be defined through the use of PNL. In this way, PNL may provide a basis for defining relevance and redundance as matters of degree, as they are in realistic settings. However, what should be stressed is that our analysis is limited to relational models. Formalizing the concepts of relevance and redundance in the context of the Web is a far more complex problem—a problem for which no cointensive solution is in sight.

5. Concluding remark

Much of the information which resides in the Web—and especially in the domain of world knowledge—is imprecise, uncertain and partially true. Existing bivalent-logic-based methods of knowledge representation and deduction are of limited effectiveness in dealing with information which is imprecise or partially true. To deal with such information, bivalence must be abandoned and new tools, such as PNL, should be employed. What is quite obvious is that, given the deeply entrenched tradition of basing scientific theories on bivalent logic, a call for abandonment of bivalence is not likely to meet a warm response. Abandonment of bivalence will eventually become a reality but it will be a gradual process.

Acknowledgements

Research supported in part by ONR N00014-02-1-0294, ONR N00014-00-1-0621 and the BISC Program of UC Berkeley.

References

[1] J. Arjona, R. Corchuelo, J. Pena, D. Ruiz, Coping with web knowledge, in: Advances in Web Intelligence, Springer-Verlag, Berlin, Heidelberg, 2003, pp. 165–178.
[2] A. Bargiela, W. Pedrycz, Granular Computing—An Introduction, Kluwer Academic Publishers, Boston, Dordrecht, London, 2003.
[3] T. Berners-Lee, J. Hendler, O. Lassila, The Semantic Web, Scientific American, 2001, pp. 34–49.
[4] M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam, S. Slattery, Learning to construct knowledge bases from the World Wide Web, Artificial Intelligence 118 (1–2) (2000) 69–113.
[5] M.J. Cresswell, Logic and Languages, Methuen, London, UK, 1973.
[6] T.F. Gamat, Language, Logic and Linguistics, University of Chicago Press, 1996.
[7] D.B. Lenat, A large-scale investment in knowledge infrastructure, Communications of the ACM 38 (11) (1995) 32–38.
[8] V. Novak, I. Perfilieva (Eds.), Discovering the World with Fuzzy Logic. Studies in Fuzziness and Soft Computing, Physica-Verlag, New York, Heidelberg, 2000.
[9] W. Pedrycz, F. Gomide, Introduction to Fuzzy Sets, MIT Press, Cambridge, MA, 1998.
[10] Smith B, Welty C. What is Ontology? Ontology: Towards a new synthesis, in: Proceedings of the Second International Conference on Formal Ontology in Information Systems, 2002.
[11] M.K. Smith, C. Welty, D. McGuinness (Eds.), 2003. OWL Web Ontology Language Guide. W3C Working Draft 31.
[12] J.F. Sowa, Ontological categories, in: L. Albertazzi (Ed.), Shapes of Forms: From Gestalt Psychology and Phenomenology to Ontology and Mathematics, Kluwer Academic Publishers, Dordrecht, 1999, pp. 307–340.
[13] L.A. Zadeh, A new direction in AI—toward a computational theory of perceptions, AI Magazine 22 (1) (2001) 73–84.
[14] L.A. Zadeh, From computing with numbers to computing with words from manipulation of measurements to manipulation of perceptions, IEEE Transactions on Circuits and Systems 45 (1) (1999) 105–119.
[15] L.A. Zadeh, Fuzzy logic, neural networks, and soft computing, Communications of the ACM—AI 37 (1994) 77–84.
[16] L.A. Zadeh, A fuzzy-algorithmic approach to the definition of complex or imprecise concepts, International Journal of Man-Machine Studies 8 (1976) 249–291.
[17] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, Part I, Information Sciences 8 (1975) 199–249;
L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, Part II, Information Sciences 8 (1975) 301–357;
L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, Part III, Information Sciences 9 (1975) 43–80.

Lotfi A. Zadeh is a Professor in the Graduate School, Computer Science Division, Department of EECS, University of California, Berkeley. In addition, he is serving as the Director of BISC (Berkeley Initiative in Soft Computing). He is an alumnus of the University of Teheran, MIT and Columbia University. He held visiting appointments at the Institute for Advanced Study, Princeton, NJ; MIT; IBM Research Laboratory, San Jose, CA; SRI International, Menlo Park, CA; and the Center for the Study of Language and Information, Stanford University. His earlier work was concerned in the main with systems analysis, decision analysis and information systems. His current research is focused on fuzzy logic, computing with words and soft computing, which is a coalition of fuzzy logic, neurocomputing, evolutionary computing, probabilistic computing and parts of machine learning. He is a Fellow of the IEEE, AAAS, ACM, AAAI, and IFSA. He is a member of the National Academy of Engineering and a Foreign Member of the Russian Academy of Natural Sciences and the Finnish Academy of Sciences. He is a recipient of the IEEE Education Medal, the IEEE Richard W. Hamming Medal, the IEEE Medal of Honor, the ASME Rufus Oldenburger Medal, the B. Bolzano Medal of the Czech Academy of Sciences, the Kampe de Feriet Medal, the AACC Richard E. Bellman Control Heritage Award, the Grigore Moisil Prize, the Honda Prize, the Okawa Prize, the AIM Information Science Award, the IEEE-SMC J.P. Wohl Career Achievement Award, the SOFT Scientific Contribution Memorial Award of the Japan Society for Fuzzy Theory, the IEEE Millennium Medal, the ACM 2001 Allen Newell Award, the Norbert Wiener Award of the Systems, Man and Cybernetics Society, other awards and twenty honorary doctorates. He has published extensively on a wide variety of subjects relating to the conception, design and analysis of information/intelligent systems, and is serving on the editorial boards of over 50 journals.