

Toward Human Level Machine Intelligence— Is It Achievable? The Need for a Paradigm Shift

Abstract: Officially, AI was born in 1956. Since then, very impressive progress has been made in many areas—but not in the realm of human level machine intelligence. Anyone who has been forced to use a dumb automated customer service system will readily agree. The Turing Test lies far beyond. Today, no machine can pass the Turing Test and none is likely to do so in the foreseeable future.

During much of its early history, AI was rife with exaggerated expectations. A headline in an article published in the late forties of last century was headlined, “Electric brain capable of translating foreign languages is being built.” Today, more than half a century later, we do have translation software, but nothing that can approach the quality of human translation. Clearly, achievement of human level machine intelligence is a challenge that is hard to meet.

Humans have many remarkable capabilities; there are two that stand out in importance. First, the capability to reason, converse and make rational decisions in an environment of imprecision, uncertainty, incompleteness of information, partiality of truth and possibility. And second, the capability to perform a wide variety of physical and mental tasks without any measurements and any computations. A prerequisite to achievement of human level machine intelligence is mechanization of these capabilities and, in particular, mechanization of natural language understanding. In my view, mech-



© BRAND X PICTURES & PHOTODISC

anization of these capabilities is beyond the reach of the armamentarium of AI—an armamentarium which in large measure is based on classical, Aristotelian, bivalent logic and bivalent-logic-based probability theory.

To make significant progress toward achievement of human level machine intelligence, a paradigm shift is needed. More specifically, what is needed is an addition to the armamentarium

Digital Object Identifier 10.1109/MCI.2008.926583

of AI of two methodologies: (a) a nontraditional methodology of computing with words (CW) or more generally, NL-Computation; and (b) a countertraditional methodology which involves a progression from computing with numbers to computing with words. The centerpiece of these methodologies is the concept of precisiation of meaning. Addition of these methodologies to AI would be an important step toward the achievement of human level machine intelligence and its applications in decision-making, pattern recognition, analysis of evidence, diagnosis, and assessment of causality. Such applications have a position of centrality in our infocentric society.

1. Introduction

Achievement of human level machine intelligence (HLMI) has profound implications for modern society—a society which is becoming increasingly infocentric in its quest for efficiency, convenience and enhancement of quality of life. Achievement of human level machine intelligence has long been one of the basic objectives of AI. In the fifties of last century, the question “Can machines think?” was an object of many spirited discussions and debates (Dreyfus and Dreyfus 2000). Exaggerated expectations were the norm, with few exceptions. In an article “Thinking machines—a new field in electrical engineering,” published in January 1950, I began with a sample of headlines of articles which appeared in the popular press in the late forties. One of them read “Electric brain capable of translating foreign languages is being built.” Today, more than half a century later, we have translation software, but nothing that approaches the level of human translation. In 1948, on the occasion of inauguration of IBM's Mark I relay computer, Howard Aiken, Director of Harvard's Computation Laboratory, said “There is no problem in applied mathematics that this computer cannot solve.” Today, there is no dearth of problems which cannot be solved by any supercomputer. Exaggerated expectations should be forgiven. As Jules Verne wrote at the turn of last century, “Scientific progress is driven by exaggerated expectations.”

Where do we stand today? What can we expect in the future? Officially, AI was born in 1956. Today, half a century later, there is much that AI can be proud of—but not in the realm of human level machine intelligence. A telling benchmark is summarization. We have software that can passably

summarize a class of documents but nothing that can summarize miscellaneous articles, much less books. We have humanoid robots but nothing that can compare in agility with that of a four year old child. We can automate driving a car in very light city traffic but there is nothing on the horizon that could automate driving in Cairo. Far too often, we have to struggle with a dumb automated customer service system which we are forced to use. Such experiences make us keenly aware that human level machine intelligence is an objective rather than reality. The Turing Test lies far beyond. What should be noted, however, is that authoritative views within the AI community tend to be substantially more optimistic. Representative views can be found in (Guha and Lenat 1994; Hibbard 2002; H. Kurzweil 2005; McCarthy 2007; Minsky et al 2004).

In an article “A new direction in AI—toward a computational theory of perceptions,” *AI Magazine*, 2001, I argued that, in large measure, the lack of significant progress in many realms of human level machine intelligence is attributable to AI's failure to develop a machinery for dealing with perceptions. Underlying human level machine intelligence are two remarkable human capabilities. First, the capability to perform a wide variety of physical and mental tasks, such as driving a car in heavy city traffic, without any measurements and any computations. And second, the capability to reason, converse and make rational decisions in an environment of imprecision, uncertainty, incompleteness of information, partiality of truth and partiality of possibility. A principal objective of human level intelligence is mechanization of these remarkable human capabilities.

What is widely unrecognized within the AI community is that mechanization of these capabilities is beyond the reach of methods based on classical, Aristotelian, bivalent logic and bivalent-logic-based probability theory. In short, if the question is: Can human level machine intelligence be achieved through the use of methods based on bivalent logic and bivalent-logic-based probability theory, then in my view the answer is: No. If the question is: Can human level machine intelligence be achieved sometime in the future, then my answer is: Possibly, but the challenge will be hard to meet. Extensions of existing techniques will not be sufficient. Basically, what is needed is a paradigm shift. More specifically, what is needed is an addition to the armamentarium of AI of two methodologies: (a) a nontraditional methodology of computing with words (CW) or, more generally, NL-Computation; and (b) a countertraditional methodology which involves a progression from computing with numbers to computing with words. The centerpiece of these methodologies is the concept of precisiation of meaning—a concept drawn from fuzzy logic.

What is fuzzy logic? What does it have to offer? There are many misconceptions about fuzzy logic. The following précis of fuzzy logic is intended to correct the misconceptions.

Fuzzy logic is not fuzzy. Basically, fuzzy logic is a precise logic of imprecision and approximate reasoning. In fact, fuzzy logic is much more than a logical system. It has many facets. The principal facets are logical, fuzzy-set-theoretic, epistemic and relational (Figure 1). Most of the applications of fuzzy

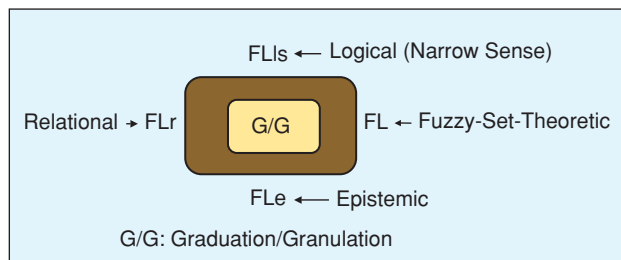


FIGURE 1 Principal facets of fuzzy logic (FL). The core of FL is graduation/granulation (G/G).

logic involve the concept of a linguistic variable and the machinery of fuzzy if-then rules. The formalism of linguistic variables and fuzzy if-then rules is associated with the relational facet. The cornerstones of fuzzy logic are graduation, granulation, precisiation and the concept of a generalized constraint (Figure 2). Graduation should be understood as an association of a concept with grades or degrees.

In fuzzy logic, everything is or is allowed to be a matter of degree or, equivalently, fuzzy. Furthermore, in fuzzy logic everything is or is allowed to be granulated, with a granule being a clump of attribute values drawn together by indistinguishability, equivalence, proximity or functionality. Graduated granulation or, equivalently, fuzzy granulation is inspired by what humans employ to deal with complexity, imprecision and uncertainty. Graduated granulation underlies the concept of a linguistic variable. When Age, for example, is treated as a linguistic variable, its granular values may be young, middle-aged and old. The granular values of Age are labels of fuzzy sets. Informally, a fuzzy set is a class with unsharp boundary. A fuzzy set is defined by its membership function. A trapezoidal membership which defines middle-age is shown in Figure 3.

A concept which plays a pivotal role in fuzzy logic is that of a generalized constraint, (Zadeh 1986, 2008) represented as $X \text{ is } R$, where X is the constrained variable, R is the constraining relation and r is an indexical variable which defines the modality of the constraint, that is, its semantics. The principal generalized constraints are possibilistic, probabilistic and veristic. The fundamental thesis of fuzzy logic is that information may be represented as a generalized constraint. A consequence of the fundamental thesis is that the meaning of a proposition, p , may likewise be represented as a generalized constraint. The concept of a generalized constraint serves as a basis for representation of and computation with propositions drawn from a natural language. This is the province of NL-Computation/Computing with Words—computation with information described in natural language.

NL-Computation opens the door to achievement of human level machine intelligence. The validity of this assertion rests on two basic facts. First, much of human knowledge, and especially world knowledge, is described in natural language. And second, a natural language is basically a system for describing perceptions. What this implies is that NL-Computation serves two major functions: (a) providing a conceptual framework and techniques for precisiation of natural language in the context of human level machine intelligence; and (b) providing a capability to compute with natural language descriptions of perceptions. These capabilities play essential roles in progression toward human level machine intelligence.

Human level machine intelligence has many components. The principal components are shown in Figure 4. Basically, achievement of human level machine intelligence requires a mechanization of the components of HLMI. Among the principal components of HLMI the component which stands out in importance involves mechanization of natural language understanding. A prerequisite to mechanization of natural

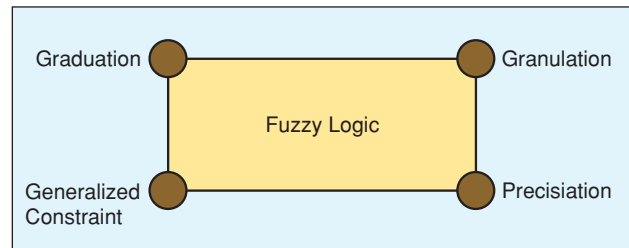


FIGURE 2 The cornerstones of a nontraditional view of fuzzy logic.

- Imprecision of Meaning = Elasticity of Meaning
 - Elasticity of Meaning = Fuzziness of Meaning
- Example: Middle Age

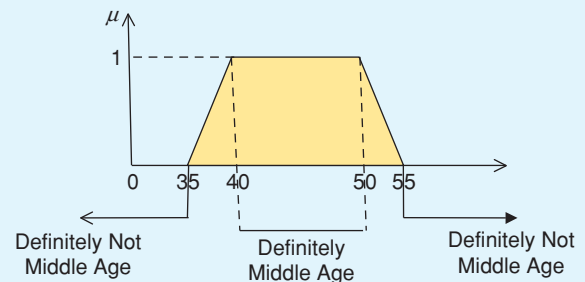


FIGURE 3 Imprecision of meaning.

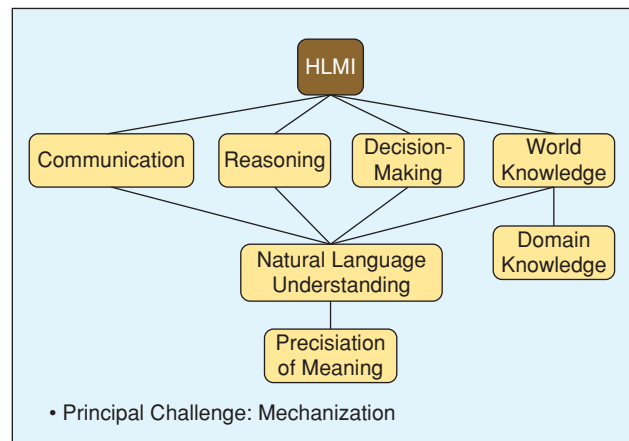


FIGURE 4 HLMI—Principal concepts and ideas.

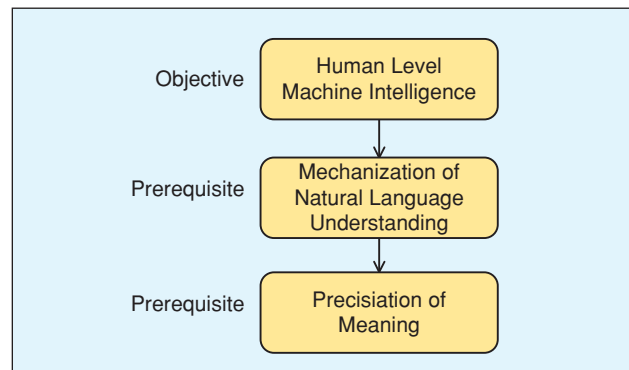


FIGURE 5 Achievement of human level machine intelligence.

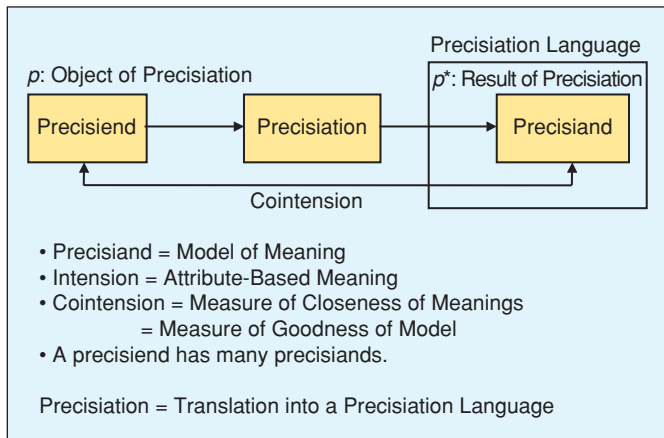


FIGURE 6 Basic concepts.

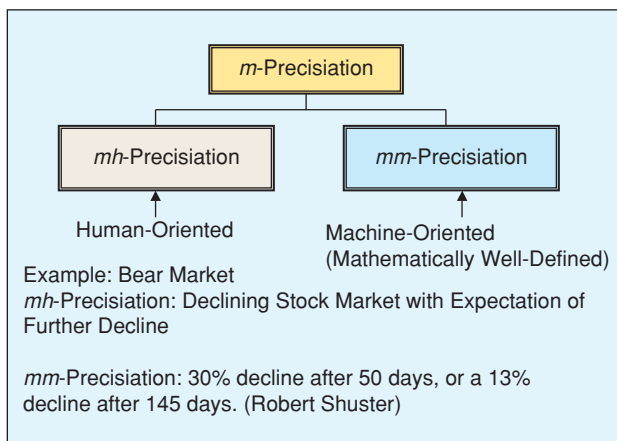


FIGURE 7 Modalities of m -precisiation.

language understanding is precisiation of meaning. Humans can understand unprecisiated natural language but machines cannot (Figure 5).

What has been widely unrecognized is that in the final analysis, progress toward achievement of human level machine intelligence requires a resolution of a critical problem—the problem of precisiation of meaning. The two cornerstones of fuzzy logic—precisiation of meaning and the concept of a generalized constraint—are of direct relevance to human level machine intelligence. The primary purpose of this paper is to bring these concepts to the attention of the Computational Intelligence community. The exposition which follows is based on (Zadeh 2006) and (Zadeh 2008). Additional details may be found in these papers.

2. The Concept of Precisiation

In one form or another, precisiation of meaning has always played an important role in science. Mathematics is a quintessential example of what may be called a meaning precisiation language. Precisiation of meaning has direct relevance to mechanization of natural language understanding. For this reason, precisiation of meaning is an issue that is certain to grow in visibility and importance as we move further into the age of machine intelligence and automated reasoning.

Semantic imprecision of natural languages is a very basic characteristic—a characteristic which is rooted in imprecision of perceptions. Basically, a natural language is a system for describing perceptions. Perceptions are imprecise. Imprecision of perceptions entails semantic imprecision of natural languages.

The concept of precisiation has few precursors in the literature of logic, probability theory and philosophy of languages (Carnap 1950). The reason is that the conceptual structure of bivalent logic—on which the literature is based—is much too limited to allow a full development of the concept of precisiation. In HLMI what is used for this purpose is the conceptual structure of fuzzy logic.

Precisiation and precision have many facets. More specifically, it is expedient to consider what may be labeled λ -precisiation, with λ being an indexical variable whose values identify various modalities of precisiation. In particular, it is important to differentiate between precision in value (v -precision) and precision in meaning (m -precision). For example, proposition $X = 5$ is v -precise and m -precise, but proposition $2 \leq X \leq 6$, is v -imprecise and m -precise. Similarly, proposition “ X is a normally distributed random variable with mean 5 and variance 2,” is v -imprecise and m -precise. Some of the basic concepts relating to precisiation are defined in Figure 6.

A further differentiation applies to m -precisiation. Thus, mh -precisiation is human-oriented meaning precisiation, while mm -precisiation is machine-oriented or, equivalently, mathematically-based meaning precisiation (Figure 7). A dictionary definition may be viewed as a form of mh -precisiation, while a mathematical definition of a concept, e.g., stability, is mm -precisiation whose result is mm -precisiand of stability.

So far as imprecisiation is concerned, it may be forced or deliberate. Imprecisiation is forced when a precise value of a variable is not known. Imprecisiation is deliberate when a precise value is not needed and precision carries a cost. Familiar examples of deliberate imprecisiation are data compression and summarization. For convenience, the precisiand and imprecisiand of p are denoted as p^* and $*p$, respectively.

Note: Deliberate imprecisiation plays an important role in many applications of fuzzy logic especially in the realm of consumer products where cost is an important consideration. In such applications, what is employed is what is referred to as “The fuzzy logic gambit” (Figure 8). In the fuzzy logic gambit deliberate v -imprecisiation is followed by mm -precisiation.

A more general illustration of mm -precisiation relates to representation of a function as a collection of fuzzy if-then rules—a mode of representation which is widely used in practical applications of fuzzy logic (Yen and Langari 1998). More specifically, let f be a function from reals to reals which is represented as (Figure 9).

f : if X is small then Y is small
if X is medium then Y is large
if X is large then Y is small

where small, medium and large are labels of fuzzy sets. In this representation, the collection in question may be viewed as *mh*-precisiand of f . When the collection is interpreted as a fuzzy graph (Zadeh 1974, 1996) representation of f assumes the form.

$$f^* : \text{small} \times \text{small} + \text{medium} \times \text{large} + \text{large} \times \text{small}$$

which is a disjunction of Cartesian products of small, medium and large. This representation is *mm*-precisiand of f .

In general, a precisiand may have many precisiands. As an illustration, consider the proposition “ X is approximately a ,” or “ X is $*a$ ” for short, where a is a real number. How can “ X is $*a$ ” be precisiated?

The simplest precisiand of “ X is $*a$ ” is “ $X = a$,” [Figures 10(a) and 10(b)]. This mode of precisiation is referred to as *s*-precisiation, with *s* standing for singular. This is a mode of precisiation which is widely used in science and especially in probability theory. In the latter case, most real-world probabilities are not known exactly but in practice are frequently computed with as if they are exact numbers. For example, if the probability of an event is stated to be 0.7, then it should be understood that 0.7 is actually $*0.7$, that is, approximately 0.7. A standard practice is to treat $*0.7$ as 0.7000..., that is, as an exact number.

Next in simplicity is representation of $*a$ as an interval centering on a . This mode of precisiation is referred to *cg*-precisiation, with *cg* standing for crisp-granular. Next is *fg*-precisiation of $*a$, with the precisiand being a fuzzy interval centering on a . Next is *p*-precisiation of $*a$, with the precisiand being a probability distribution centering on a . And so on.

An analogy is helpful in understanding the relationship between a precisiand and its precisiands. More specifically, a *mm*-precisiand, p^* , may be viewed as a model of precisiand, p , in the same sense as a differential equation may be viewed as a model of a physical system.

In the context of modeling, an important characteristic of a model is its “closeness of fit.” In the context of NL-Computation, an analogous concept is that of cointension. The concept is discussed in the following.

3. The Concept of Cointensive Precisiation

Precisiation is a prerequisite to computation with information described in natural language. To be useful, precisiation of a precisiand, p , should result in a precisiand, p^* , whose meaning, in some specified sense, should be close to that of p . Basically, cointension of p^* and p is the degree to which the meaning of p^* fits the meaning of p .

In dealing with meaning, it is necessary to differentiate between this intension or, equivalently, the intensional meaning, *i*-meaning, of p , and the extension, or, equivalently, the extensional, *e*-meaning of p . The concepts of extension and intension are drawn from logic and, more particularly, from modal logic and possible world semantics (Cresswell 1973, Lambert 1970, Belohlavek 2006). Basically, *e*-meaning is attribute-free and *i*-meaning is attribute-based. As a simple illustration, if A is a finite set in a universe of discourse, U , then the *e*-meaning of A , that is, its extension is the list of elements of A , $\{u_1, \dots, u_n\}$, u_i being the name of i th element of A , with no attributes associated with the u_i . Let $a(u_i)$ be an attribute-vector associated with each u_i . Then the intension of A is a recognition algorithm which, given $a(u_i)$, recognizes whether u_i is or is not an element of A . If A is a fuzzy set with membership function μ_A then the *e*-meaning and *i*-meaning of A may be expressed compactly as

$$e\text{-meaning of } A : A = \{\mu_A(u_i)/u_i\}$$

where $\mu_A(u)/u$ means that $\mu_A(u)$ is the grade of membership of u_i in A ; and

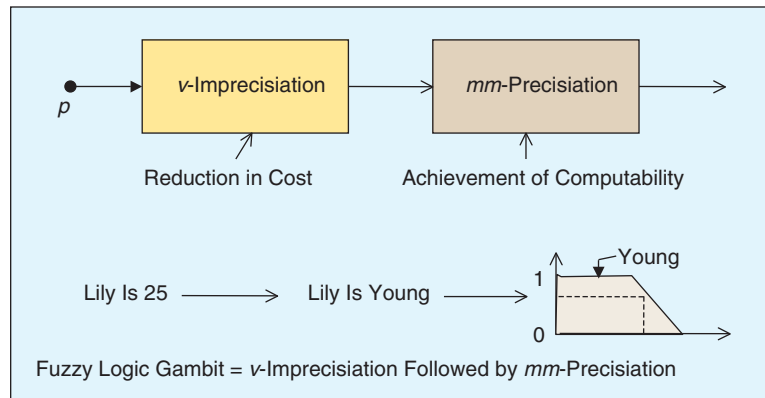


FIGURE 8 The fuzzy logic gambit.

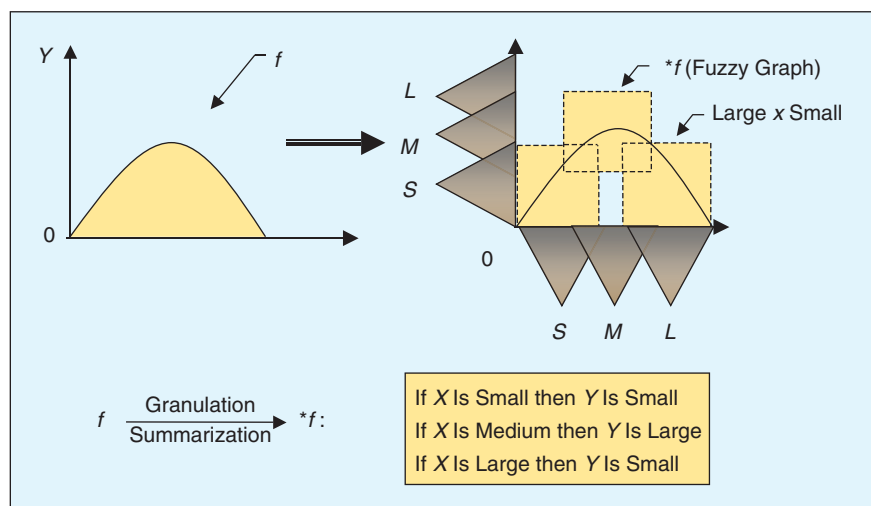


FIGURE 9 Granulation of a function. *S* (small), *M* (medium) and *L* (large) are fuzzy sets. The granuland of f , $*f$, may be viewed as a summary of f .

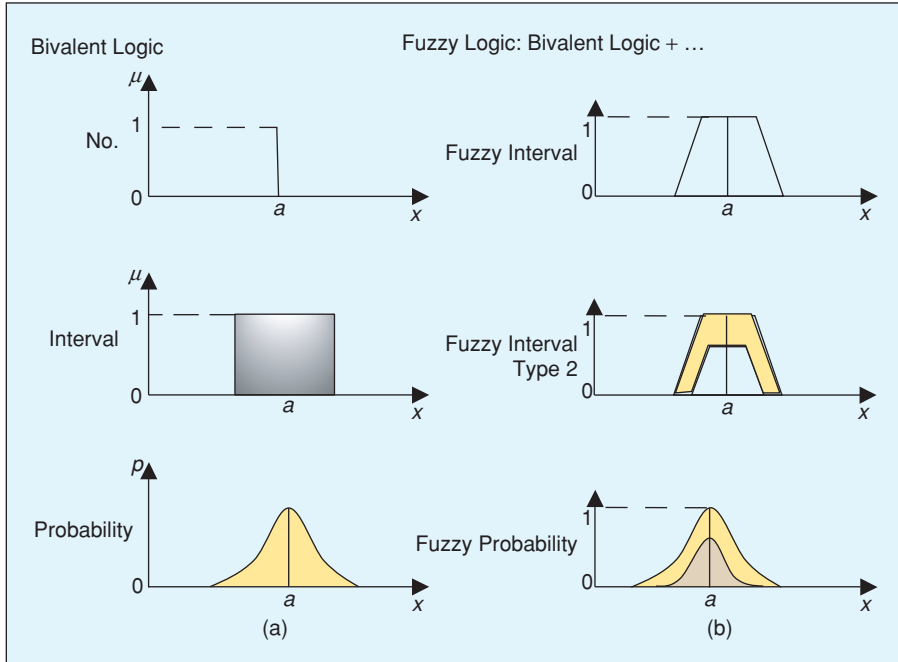


FIGURE 10 (a) Alternative modes of *mm*-precision of “approximately a ,” $*a$, within the framework of bivalent logic. (b) Alternative modes of *mm*-precision of “approximately a ,” $*a$, within the framework of fuzzy logic.

$$i\text{-meaning of } A : \mu_A = \{\mu_A(a(u_i))/u_i\},$$

with the understanding that in the *i*-meaning of A the membership function, μ_A is defined on the attribute space. It should be noted that when A is defined through exemplification, it is said to be defined ostensively. Thus, *o*-meaning of A consists of exemplars of A . An ostensive definition may be viewed as a special case of extensional definition. A neural network may be viewed as a system which derives *i*-meaning from *o*-meaning.

Clearly, *i*-meaning is more informative than *e*-meaning. For this reason, cointension is defined in terms of intensions rather than extensions of precisiend and precisiand. Thus, meaning will be understood to be *i*-meaning, unless stated to the con-

trary. However, when the precisiend is a concept which plays the role of definiendum and we know its extension but not its intension, cointension has to involve the extension of the definiendum (precisiend) and the intension of the definiens (precisiand).

As an illustration, let p be the concept of bear market. A dictionary definition of p —which may be viewed as a *mh*-precisiand of p —reads “A prolonged period in which investment prices fall, accompanied by widespread pessimism.” A widely accepted quantitative definition of bear market is: We classify a bear market as a 30 percent decline after 50 days, or a 13 percent decline after 145 days. (Shuster) This definition may be viewed as a *mm*-precisiand of bear market. Clearly, the quantitative definition, p^* , is not a good fit to the perception of the meaning of bear market which is the basis for

the dictionary definition. In this sense, the quantitative definition of bear market is not cointensive.

Intensions are more informative than extensions in the sense that more can be inferred from propositions whose meaning is expressed intensionally rather than extensionally. The assertion will be precisiated at a later point. For the present, a simple example will suffice.

Consider the proposition p : Most Swedes are tall. Let U be a population of n Swedes, $U = (u_1, \dots, u_n)$, u_1 = name of i th Swede.

A precisiand of p may be represented as

$$\frac{1}{n} \text{Count}(\text{tall.Swedes}) \text{ is most,}$$

where most is a fuzzy quantifier which is defined as a fuzzy subset of the unit interval (Zadeh 1983). Let $\mu_{\text{tall}}(u_i)$, $i = 1, \dots, n$ be the grade of membership of u_i in the fuzzy set of tall Swedes. Then the *e*-meaning of tall Swedes may be expressed in symbolic form as

$$\text{tall.Swedes} = \mu_{\text{tall}}(u_1)/u_1 + \dots + \mu_{\text{tall}}(u_n)/u_n.$$

Accordingly, the *i*-precisiand of p may be expressed as

$$\frac{1}{n}(\mu_{\text{tall}}(u_1)) + \dots + \mu_{\text{tall}}(u_n) \text{ is most.}$$

Similarly, the *i*-precisiand of p may be represented as

$$\frac{1}{n}(\mu_{\text{tall}}(h_1)) + \dots + \mu_{\text{tall}}(h_n) \text{ is most.}$$

As will be seen later, given the *e*-precisiend of p we can compute the answer to the query: How many Swedes are not tall? The answer is 1-most. However, we cannot compute the

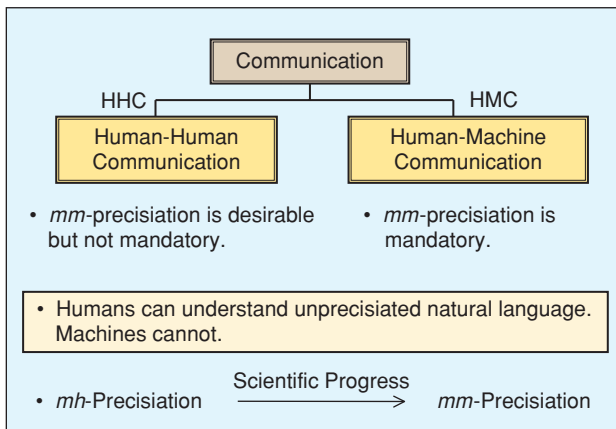


FIGURE 11 Precision in communication.

answer to the query: How many Swedes are short? The same applies to the query: What is the average height of Swedes? As will be shown later, the answers to these queries can be computed given the i -precisiand of p .

The concept of cointensive precision has important implications for the way in which scientific concepts are defined. The standard practice is to define a concept within the conceptual structure of bivalent logic, leading to a bivalent definition under which the universe of discourse is partitioned into two classes: objects which fit the concept and those which do not, with no shades of gray allowed. Such definition is valid when the concept that is defined, the definiendum, is crisp, that is, bivalent. The problem is that in reality most scientific concepts are fuzzy, that is, are a matter of degree. Familiar examples are the concepts of causality, relevance, stability, independence and bear market. In general, when the definiendum (precisiend) is a fuzzy concept, the definiens (precisiand) is not cointensive, which is the case with the bivalent definition of bear market. More generally, bivalent definitions of fuzzy concepts are vulnerable to the Sorites (heap) paradox (Sainsbury 1995).

4. A Key Idea—the Meaning Postulate

In fuzzy logic, a proposition, p , is viewed as an answer to a question, q , of the form “What is the value of X ?” Thus p is a carrier of information about X . In this perspective, the meaning of p , $M(p)$, is the information which p carries about X . An important consequence of the fundamental thesis of fuzzy logic is what is referred to as the meaning postulate. In symbolic form, the postulate is expressed as $M(p) = GC(X(p))$, where $GC(X(p))$ is a generalized constraint on the variable which is constrained by p . In plain words, the meaning postulate asserts that the meaning of a proposition may be represented as a generalized constraint. It is this postulate that makes the concept of a generalized constraint a cornerstone of fuzzy logic. By providing a mechanism for precisiating the meaning of a proposition, the concept of a generalized constraint opens the door to a wide-ranging enlargement of the role of natural languages in scientific theories. What is referred to as PNL, Precisiated Natural Language, serves this purpose (Zadeh 2004).

A point which should be noted is that the question to which p is an answer is not uniquely determined by p ; hence $X(p)$ is not uniquely defined by p . Generally, however, among the possible questions there is one which is most likely. For example, if p is “Monika is young,” then the most likely question is “How old is Monika?” In this example, X is Age(Monika).

The concept of precision has a direct bearing on communication in the context of human level machine intelligence (Figure 11). In communication

between humans, HHC, precision of meaning is desirable but not mandatory. In communication between a human and machine, HMC, precision is mandatory because a machine cannot understand unprecisiated natural language. In the case of HMC, an important issue relates to whether the precisiator is the sender (human, s -precision) or the recipient (machine, r -precision) (Figure 12). In most applications of fuzzy logic, the precisiator is the sender (human), an example is the Honda fuzzy logic transmission (Figure 13). In the case of s -precision, context-dependence is not a problem. As a consequence, precision is a much simpler function than it is in the case of r -precision. It should be noted that mechanization of natural language understanding involves for the most part r -precision.

In HHC, mm-precision is a major application area for generalized-constraint-based semantics (Zadeh 2008). Generalized-constraint-based semantics provides a basis for reformulation of bivalent-logic-based definitions of scientific concepts, associating Richter-like scales with concepts which are traditionally defined as bivalent concepts but in reality are fuzzy concepts. Examples: recession, civil war, arthritis, randomness, causality.

5. The Concept of a Generalized Constraint

Constraints are ubiquitous. A typical constraint is an expression of the form $X \in C$, where X is the constrained variable and

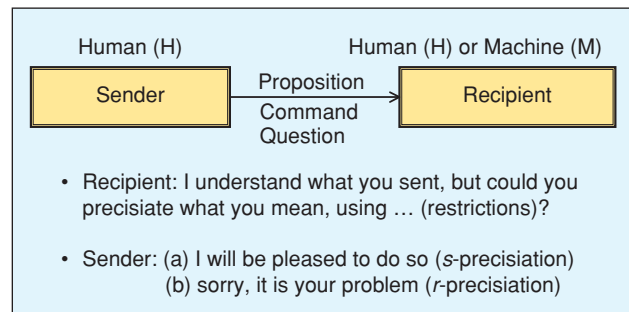


FIGURE 12 Precision in communication—Basic idea.

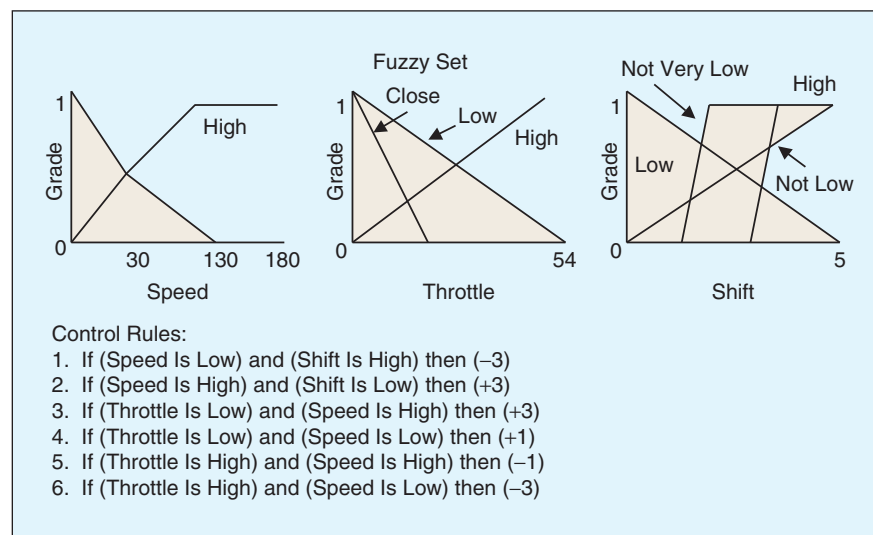


FIGURE 13 Honda fuzzy logic transmission.

C is the set of values which X is allowed to take. A typical constraint is hard (inelastic) in the sense that if u is a value of X then u satisfies the constraint if and only if $u \in C$.

The problem with hard constraints is that most real-world constraints are not hard, meaning that most real-world constraints have some degree of elasticity. For example, the constraints “check-out time is 1 pm,” and “speed limit is 100 kmh,” are, in reality, not hard. How can such constraints be defined? The concept of a generalized constraint is motivated by questions of this kind (Zadeh 2008).

Real-world constraints may assume a variety of forms. They may be simple in appearance and yet have a complex structure. Reflecting this reality, a generalized constraint, $GC(X)$, is defined as an expression of the form.

$$GC(X) : X \text{ isr } R,$$

where X is the constrained variable; R is a constraining relation which, in general, is non-bivalent; and r is an indexing variable which identifies the modality of the constraint, that is, its semantics.

The constrained variable, X , may assume a variety of forms. In particular,

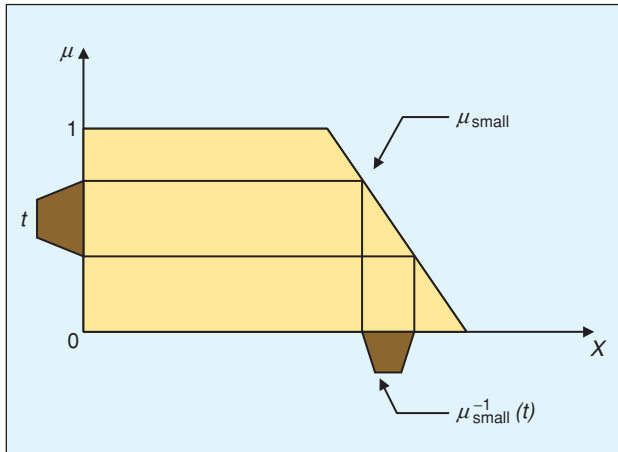


FIGURE 14 Truth-qualification: $(X \text{ is small})$ is t .

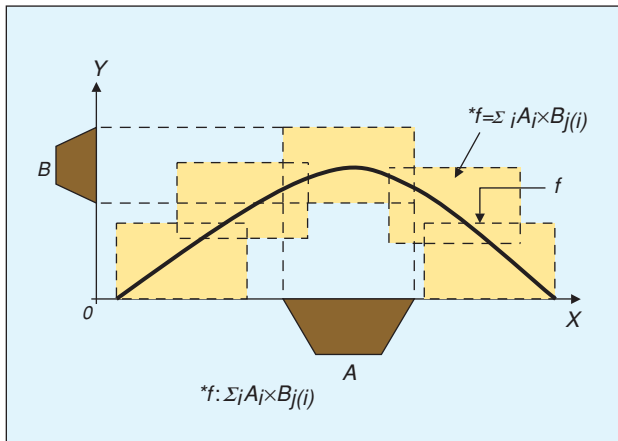


FIGURE 15 Fuzzy-graph extension principle. $B = {}^*f(A)$.

- X is an n -ary variable, $X = (X_1, \dots, X_n)$
- X is a proposition, e.g., $X = \text{Leslie is tall}$
- X is a function
- X is a function of another variable, $X = f(Y)$
- X is conditioned on another variable, X/Y
- X has a structure, e.g., $X = \text{Location}(\text{Residence}(\text{Carol}))$
- X is a group variable. In this case, there is a group, $G[A]$; with each member of the group, $\text{Name}_i, i = 1, \dots, n$, associated with an attribute-value, A_i . A_i may be vector-valued. Symbolically

$$G[A] : \text{Name}_1/A_1 + \dots + \text{Name}_n/A_n.$$

Basically, $G[A]$ is a relation.

- X is a generalized constraint, $X = Y \text{ isr } R$.

A generalized constraint is associated with a test-score function, $ts(u)$, (Zadeh 1982) which associates with each object, u , to which the constraint is applicable, the degree to which u satisfies the constraint. Usually, $ts(u)$ is a point in the unit interval. However, if necessary, the test-score may be a vector, an element of a semiring (Rossi 2003), an element of a lattice (Goguen 1969) or, more generally, an element of a partially ordered set, or a bimodal distribution—a constraint which will be described later. The test-score function defines the semantics of the constraint with which it is associated.

The constraining relation, R , is, or is allowed to be, non-bivalent (fuzzy). The principal modalities of generalized constraints are summarized in the following.

6. Principal Modalities of Generalized Constraints

(a) *Possibilistic* ($r = \text{blank}$)

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

with R playing the role of the possibility distribution of X . For example:

$$X \text{ is } [a, b]$$

means that $[a, b]$ is the set of possible values of X . Another example:

$$X \text{ is small.}$$

In this case, the fuzzy set labeled small is the possibility distribution of X (Zadeh 1978; Dubois and Prade 1988). If μ_{small} is the membership function of small, then the semantics of “ X is small” is defined by

$$\text{Pross}\{X = u\} = \mu_{\text{small}}(u),$$

where u is a generic value of X .

(b) *Probabilistic* ($r = p$)

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

with R playing the role of the probability distribution of X . For example.

$$X \text{ is } N(m, \sigma^2)$$

means that X is a normally distributed random variable with mean m and variance σ^2 .

If X is a random variable which takes values in a finite set $\{u_1, \dots, u_n\}$ with respective probabilities p_1, \dots, p_n , then X may be expressed symbolically as

$$X \text{ isp } (p_1|u_1 + \dots + p_n|u_n),$$

with the semantics

$$\text{Prob}(X = u_i) = p_i, \quad i = 1, \dots, n.$$

What is important to note is that in fuzzy logic a probabilistic constraint is viewed as an instance of a generalized constraint.

When X is a generalized constraint, the expression

$$X \text{ isp } R$$

is interpreted as a probability qualification of X , with R being the probability of X , (Zadeh 1979a, 1981). For example.

$$(X \text{ is small}) \text{ isp likely},$$

where small is a fuzzy subset of the real line, means that probability of the fuzzy event $\{X \text{ is small}\}$ is likely. More specifically, if X takes values in the interval $[a, b]$ and g is the probability density function of X , then the probability of the fuzzy event “ X is small” may be expressed as (Zadeh 1968)

$$\text{Prob}(X \text{ is small}) = \int_a^b \mu_{\text{small}}(u)g(u)du.$$

Hence

$$ts(g) = \mu_{\text{likely}}\left(\int_b^a g(u)\mu_{\text{small}}(u)du\right).$$

This expression for test-score function defines the semantics of probability qualification of a possibilistic constraint.

$$(c) \text{ Veristic} \quad (r = v)$$

$$X \text{ isv } R,$$

where R plays the role of a verity (truth) distribution of X . In particular, if X takes values in a finite set $\{u_1, \dots, u_n\}$ with respective verity (truth) values t_1, \dots, t_n , then X may be expressed as

$$X \text{ isv } (t_1|u_1 + \dots + t_n|u_n),$$

meaning that $\forall i \text{ } r(X = u_i) = t_i, \quad i = 1, \dots, n$

For example, if Robert is half German, quarter French and quarter Italian, then

$$\text{Ethnicity}(\text{Robert}) \text{ isv } 0.5|\text{German} + 0.25|\text{French} + 0.25|\text{Italian}$$

When X is a generalized constraint, the expression

$$X \text{ isv } R$$

is interpreted as verity (truth) qualification of X . For example,

$$(X \text{ is small}) \text{ isv very.true},$$

should be interpreted as “It is very true that X is small.” The semantics of truth qualification is defined by (Zadeh 1979b)

$$\text{Ver}(X \text{ is } R) \text{ is } t \longrightarrow X \text{ is } \mu_R^{-1}(t),$$

where μ_R^{-1} is inverse of the membership function of R , and t is a fuzzy truth value which is a subset of $[0, 1]$, Figure 14.

$$(d) \text{ Usuality} \quad (r = u)$$

$$X \text{ isu } R.$$

The usuality constraint presupposes that X is a random variable, and that probability of the event $\{X \text{ isu } R\}$ is usually, where usually plays the role of a fuzzy probability which is a fuzzy number (Kaufman and Gupta 1985). For example.

$$X \text{ isu small}$$

means that “usually X is small” or, equivalently,

$$\text{Prob}\{X \text{ is small}\} \text{ is usually}.$$

In this expression, small may be interpreted as the usual value of X . The concept of a usual value has the potential of playing a significant role in decision analysis, since it is more informative than the concept of an expected value.

$$(e) \text{ Random-set} \quad (r = vs)$$

In

$$X \text{ isrs } R,$$

X is a fuzzy-set-valued random variable and R is a fuzzy random set

$$(f) \text{ Fuzzy-graph} \quad (r = fq)$$

In

$$X \text{ isfg } R,$$

X is a function, f , and R is a fuzzy graph (Zadeh 1974, 1996) which constrains f (Figure 15). A fuzzy graph is a disjunction of Cartesian granules expressed as

$$R = A_1 \times B_1 + \dots + A_n \times B_n,$$

where the A_i and B_i , $i = 1, \dots, n$, are fuzzy subsets of the real line, and \times is the Cartesian product. A fuzzy graph is frequently described as a collection of fuzzy if-then rules (Zadeh 1973, 1996; Pedrycz and Gomide 1998; Bardossy and Duckstein 1995).

$$R : \text{ if } X \text{ is } A_i \text{ then } Y \text{ is } B_i, \quad i = 1, \dots, n.$$

The concept of a fuzzy-graph constraint plays an important role in applications of fuzzy logic (Bardossy and Duckstein 1995; Filev and Yager 1994; Jamshidi, Titli, Zadeh and Boverie 1997; Yen, Langari and Zadeh 1995).

7. The Concept of Bimodal Constraint/Distribution

In the bimodal constraint,

$$X \text{ isbm } R,$$

R is a bimodal distribution of the form

$$R : \sum_i p_i | A_i, \quad i = 1, \dots, n.$$

with the understanding that $\text{Prob}(X \text{ is } A_i)$ is P_i . (Zadeh 2002), that is, P_i is a granular value of $\text{Prob}(X \text{ is } A_i)$, $i = 1, \dots, n$. (See next section for definition of granular value).

To clarify the meaning of a bimodal distribution it is expedient to start with an example. I am considering buying Ford stock. I ask my stockbroker, “What is your perception of the near-term prospects for Ford stock?” He tells me, “A moderate decline is very likely; a steep decline is unlikely; and a moderate gain is not likely.” My question is: What is the probability of a large gain?

Information provided by my stock broker may be represented as a collection of ordered pairs:

- Price: (unlikely, steep.decline), (very.likely, moderate.decline), (not.likely, moderate.gain))

In this collection, the second element of an ordered pair is a fuzzy event or, generally, a possibility distribution, and the first element is a fuzzy probability. The expression for Price is an example of a bimodal distribution.

The importance of the concept of a bimodal distribution derives from the fact that in the context of human-centric systems, most probability distributions are bimodal. Bimodal distributions can assume a variety of forms. The principal types are Type 1, Type 2 and Type 3 (Zadeh 1979a, 1981). Type 1, 2 and 3 bimodal distributions have a common framework but differ in important detail (Figure 16). A bimodal distribution may be viewed as an important generalization of standard probability distribution. For this reason, bimodal distributions of Type 1, 2, 3 are discussed in greater detail in the following.

- Type 1 (default): X is a random variable taking values in U
 A_1, \dots, A_n, A are events (fuzzy sets)
 $p_i = \text{Prob}(X \text{ is } A_i)$, $i = 1, \dots, n$
 $\sum_i p_i$ is unconstrained
 P_i = granular value of P_i

BD: bimodal distribution: $((P_1, A_1), \dots, (P_n, A_n))$

or, equivalently,

$$X \text{ isbm } (P_1 \setminus A_1 + \dots + P_n \setminus A_n)$$

Problem: What is the granular probability, P , of A ? In general, this probability is fuzzy-set-valued.

- Type 2 (fuzzy random set): X is a fuzzy-set-valued random variable with values

A_1, \dots, A_n (fuzzy sets)

$$P_i = \text{Prob}(X = A_i), \quad i = 1, \dots, n$$

P_i : granular value of P_i

BD: $X \text{ isrs } (P_1 \setminus A_1 + \dots + P_n \setminus A_n)$

$$\sum_i p_i = 1$$

Problem: What is the granular probability, P , of A ? P is not definable. What are definable are (a) the expected value of the conditional possibility of A given BD, and (b) the expected value of the conditional necessity of A given BD

- Type 3 (Dempster-Shafer) (Dempster 1967, Shafer 1976): X is a random variable taking values X_1, \dots, X_n with probabilities p_1, \dots, p_n

X_i is a random variable taking values in A_i , $i = 1, \dots, n$

Probability distribution of X_i in A_i , $i = 1, \dots, n$, is not specified

$X \text{ isp } (p_1 \setminus X_1 + \dots + p_n \setminus X_n)$

Problem: What is the probability, p , that X is in A ? Because probability distributions of the X_i in the A_i are not specified, p is interval-valued. What is important to note is that the concepts of upper and lower probabilities break down when the A_i are fuzzy sets (Zadeh 1979a).

Note: In applying Dempster-Shafer theory it is important to check on whether the data fit Type 3 model. In many cases, the correct model is Type 1 rather than Type 3.

The importance of bimodal distributions derives from the fact that in many realistic settings a bimodal distribution is the best approximation to our state of knowledge. An example is assessment of degree of relevance, since relevance is generally not well defined.

If I am asked to assess the degree of relevance of a book on knowledge representation to summarization, my state of knowledge about the book may not be sufficient to justify an answer such as 0.7. A better approximation to my state of knowledge may be “likely to be high.” Such an answer is an instance of a bimodal distribution.

8. Primary Constraints, Composite Constraints and Standard Constraints

Among the principal generalized constraints there are three that play the role of primary generalized constraints. They are:

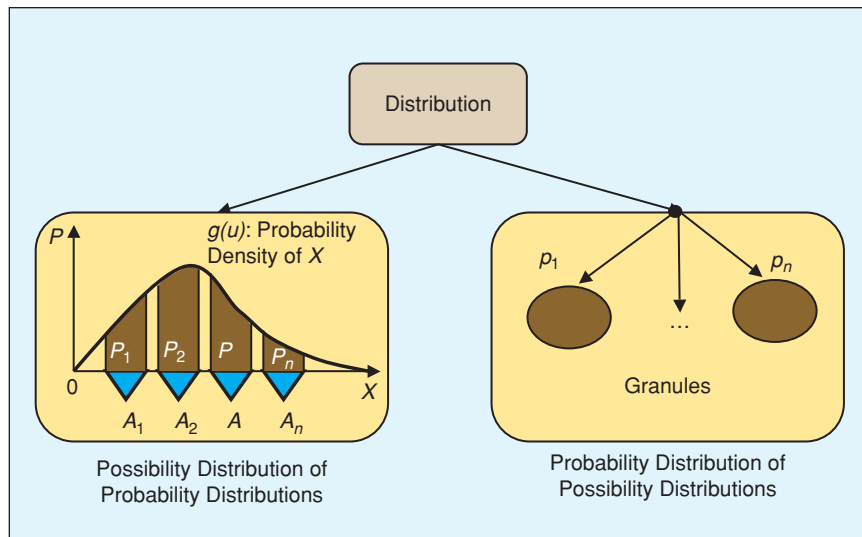


FIGURE 16 Bimodal distributions. Granular versus granule-valued distributions.

Possibilistic constraint: $X \text{ is } R$,

Probabilistic constraint: $X \text{ isp } R$,

and

Veristic constraint: $X \text{ isv } R$.

A special case of primary constraints is what may be called standard constraints: bivalent possibilistic, probabilistic and bivalent veristic. Standard constraints form the basis for the conceptual framework of bivalent logic and probability theory.

A generalized constraint is composite if it can be generated from other generalized constraints through conjunction, and/or projection and/or constraint propagation and/or qualification and/or possibly other operations. For example, a random-set constraint may be viewed as a conjunction of a probabilistic constraint and either a possibilistic or veristic constraint. The Dempster-Shafer theory of evidence is, in effect, a theory of possibilistic random-set constraints. The derivation graph of a composite constraint defines how it can be derived from primary constraints.

The three primary constraints—possibilistic, probabilistic and veristic—are closely related to a concept which has a position of centrality in human cognition—the concept of partiality. In the sense used here, partial means: a matter of degree or, more or less equivalently, fuzzy. In this sense, almost all human concepts are partial (fuzzy). Familiar examples of fuzzy concepts are: knowledge, understanding, friendship, love, beauty, intelligence, belief, causality, relevance, honesty, mountain and, most important, truth, likelihood and possibility. Is a specified concept, C , fuzzy? A simple test is: If C can be hedged, then it is fuzzy. For example, in the case of relevance, we can say: very relevant, quite relevant, slightly relevant, etc. Consequently, relevance is a fuzzy concept.

The three primary constraints may be likened to the three primary colors: red, blue and green. In terms of this analogy, existing theories of uncertainty may be viewed as theories of different mixtures of primary constraints. For example, the Dempster-Shafer theory of evidence is a theory of a mixture of probabilistic and possibilistic constraints. The Generalized Theory of Uncertainty (GTU) (Zadeh 2006) embraces all possible mixtures.

9. The Generalized Constraint Language and Standard Constraint Language

A concept which has a position of centrality in GTU is that of Generalized Constraint Language (GCL). Informally, GCL is the set of all generalized constraints together with the rules governing syntax, semantics and generation. Simple examples of elements of GCL are:

$(X \text{ is small}) \text{ is likely}$
 $((X, Y) \text{ isp } A) \wedge (X \text{ is } B)$
 $(X \text{ isp } A) \wedge ((X, Y) \text{ isv } B)$
 $\text{Proj}_Y((X \text{ is } A) \wedge (X, Y) \text{ isp } B),$

where \wedge is conjunction.

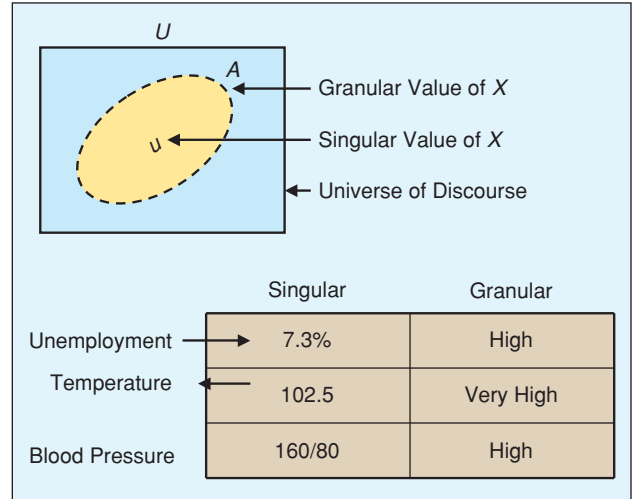


FIGURE 17 Singular and granular values.

A very simple example of a semantic rule is:

$$(X \text{ is } A) \wedge (Y \text{ is } B) \longrightarrow \text{Poss}(X \text{ is } A) \wedge \text{Poss}(Y \text{ is } B), \\ = \mu_A(u) \wedge \mu_B(v),$$

where u and v are generic values of X , Y ; and μ_A and μ_B are the membership functions of A and B , respectively.

In principle, GCL is an infinite set. However, in most applications only a small subset of GCL is likely to be needed.

A key idea which underlies NL-Computation is embodied in the meaning postulate—a postulate which asserts that the meaning of a proposition, p , drawn from a natural language is representable as a generalized constraint. In the context of GCL, the meaning postulate asserts that p may be precisiated through translation into GCL. Transparency of translation may be enhanced through annotation. Simple example of annotation,

$$\text{Monika is young} \longrightarrow X/\text{Age (Monika) is } R/\text{young}.$$

In fuzzy logic, the set of all standard constraints together with the rules governing syntax, semantics and generation constitute the Standard Constraint Language (SCL). SCL is a subset of GCL.

In previous sections, we have employed the concept of a granular value in an informal fashion, without formulating a definition. The concept of a generalized constraint makes it possible to define a granular value more precisely. Let X be a variable taking values in a universe of discourse U , $U = \{u\}$. If a is an element of U , and it is known that the value of X is a , then a is referred to as a singular value of X . If there is some uncertainty about the value of X , the available information induces a restriction on the possible values of X which may be represented as a generalized constraint $GC(X)$, $X \text{ isr } R$. Thus a generalized constraint defines a granule which is referred to as a granular value of X , $G r(X)$ (Figure 17). For example, if it is known to lie in the interval $[a, h]$, then $[a, h]$ is a granular value of X . Similarly, if $X \text{ isp } N(m, \sigma^2)$, then $N(m, \sigma^2)$ is a granular value of X . What is important to note is that defining a granular

value in terms of a generalized constraint makes a granular value *mm*-precise. It is this characteristic of granular values that underlies the concept of a linguistic variable (Zadeh 1973). Symbolically, representing a granular value as a generalized constraint may be expressed as $Gr(X) = GC(X)$. It should be noted that, in general, perception-based information is granular (Figure 17).

The importance of the concept of a granular value derives from the fact that it plays a central role in computation with information described in natural language. More specifically, when a proposition expressed in a natural language is represented as a system of generalized constraints, it is, in effect, a system of granular values. Thus, computation with information described in natural language ultimately reduces to computation with granular values. Such computation is the province of Granular Computing (Zadeh 1979a, 1998; Bargiela and Pedrycz 2002).

10. Concluding Remarks

There are many reasons why achievement of human level machine intelligence is a challenge that is hard to meet. One of the principal reasons is the need for mechanization of two remarkable human capabilities. First, the capability to converse, communicate, reason and make rational decisions in an environment of imprecision, uncertainty, incompleteness of information, partiality of truth and partiality of possibility. And second, the capability to perform a wide variety of physical and mental tasks—such as driving a car in heavy city traffic—without any measurements and any computations. What is well understood is that a prerequisite to mechanization of these capabilities is mechanization of natural language understanding. But what is widely unrecognized is that mechanization of natural language understanding is beyond the reach of methods based on bivalent logic and bivalent-logic-based probability theory. In addition, what is widely unrecognized is that mechanization of natural language understanding is contingent on precisiation of meaning.

Humans can understand unprecisiated natural language but machines cannot. Natural languages are intrinsically imprecise. Basically, a natural language is a system for describing perceptions. Perceptions are intrinsically imprecise. Imprecision of natural languages is rooted in imprecision of perceptions.

The principal thesis of this paper is that to address the problem of precisiation of meaning it is necessary to employ the machinery of fuzzy logic. In addition, the machinery of fuzzy logic is needed for mechanization of human reasoning. In this perspective, fuzzy logic is of direct relevance to achievement of human level machine intelligence. The cornerstones of fuzzy logic are the concepts of graduation, granulation, precisiation and generalized constraint.

Acknowledgment

Research is supported in part by ONR N00014-02-1-0294, BT Grant CT1080028046, Omron Grant, Tekes Grant, Chevron Texaco Grant, and the BISC Program of UC Berkeley.

References

- [1] A. Bargiela and W. Pedrycz, *Granular Computing*. Kluwer Academic Publishers, 2002.
- [2] A. Bardossy and L. Duckstein, *Fuzzy Rule-Based Modelling with Application to Geophysical, Biological and Engineering Systems*. CRC Press, 1995.

- [3] R. Belohlavek and V. Vychodil, *Attribute Implications in a Fuzzy Setting*. B. Ganter and L. Kwuida (Eds.): ICFCA 2006, *Lecture Notes in Artificial Intelligence* 3874, Springer-Verlag, Heidelberg, pp. 45–60, 2006.
- [4] R. Carnap, *The Logical Foundations of Probability*. University of Chicago Press, 1950.
- [5] M.J. Cresswell, *Logic and Languages*. Methuen, London, UK, 1973.
- [6] A.P. Dempster, “Upper and Lower Probabilities Induced by a Multivalued Mapping,” *Ann. Math. Statist.*, vol. 38, pp. 325–329, 1967.
- [7] H. Dreyfus and S. Dreyfus, *Mind and Machine*. Free Press, New York, 2000.
- [8] D. Dubois and H. Prade, “Representation and combination of uncertainty with belief functions and possibility measures,” *Computational Intelligence*, vol. 4, pp. 244–264, 1988.
- [9] D. Filev and R.R. Yager, *Essentials of Fuzzy Modeling and Control*, Wiley-Interscience, 1994.
- [10] J.A. Goguen, “The logic of inexact concepts,” *Synthese*, vol. 19, pp. 325–373, 1969.
- [11] R.V. Guha and D.B. Lenat, “Enabling agents to work together,” *Commun. ACM*, vol. 37, no. 7, pp. 126–142, 1994.
- [12] W. Hibbard, *Super Intelligent Machines*, Springer. New York, 2002.
- [13] M. Jamshidi, A. Tidi, L.A. Zadeh, and S. Boverie (Eds.), *Applications of Fuzzy Logic—Towards High Machine Intelligence Quotient Systems, Environmental and Intelligent Manufacturing Systems Series*. Prentice Hall, Upper Saddle River, NJ, vol. 9, 1997.
- [14] A. Kaufmann and M.M. Gupta, *Introduction to Fuzzy Arithmetic: Theory and Applications*. Von Nostrand, New York, 1985.
- [15] R. Kurzweil, *Singularity is Near*. Viking Press, New York, 2005.
- [16] K. Lambert and B.C. Van Fraassen, “Meaning relations, possible objects and possible worlds,” *Philosophical Problems in Logic*, pp. 1–19, 1970.
- [17] J. McCarthy, “From here to human-level AI,” *Artif. Intell.* vol. 171, no. 18, pp. 1174–1182, 2007.
- [18] M. Minsky, P. Singh, and A. Sloman, “The St. Thomas Common sense symposium: Designing architectures for human-level intelligence,” *AI Magazine*, vol. 25, no. 2, pp. 113–124, 2004.
- [19] W. Pedrycz and F. Gomide, *Introduction to Fuzzy Sets*. MIT Press, Cambridge, MA, 1998.
- [20] F. Rossi and P. Codognet, *Soft Constraints, Special issue on Constraints*. Kluwer, vol. 8, no. 1, 2003.
- [21] R.M. Sainsbury, *Paradoxes*. Cambridge University Press, 1995.
- [22] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ, 1976.
- [23] J. Yen, R. Langari, and L.A. Zadeh (Ed.), *Industrial Applications of Fuzzy Logic and Intelligent Systems*. IEEE, 1995.
- [24] J. Yen and R. Langari, *Fuzzy Logic: Intelligence, Control and Information*. Prentice Hall, 1st edition, 1998.
- [25] L.A. Zadeh, “Probability measures of fuzzy events,” *Jour. Math. Analysis and Appl.*, vol. 23, pp. 421–427, 1968.
- [26] L.A. Zadeh, “Outline of a new approach to the analysis of complex systems and decision processes,” *IEEE Trans. on Systems, Man and Cybernetics SMC-3*, pp. 28–44, 1973.
- [27] L.A. Zadeh, “On the analysis of large scale systems,” *Systems Approaches and Environment Problems*, H. Gottinger (ed.), Vandenhoeck and Ruprecht, Gottingen, pp. 23–37, 1974.
- [28] L.A. Zadeh, “Fuzzy sets and information granularity,” *Advances in Fuzzy Set Theory and Applications*, M. Gupta, R. Ragade and R. Yager (eds.), North-Holland Publishing Co., Amsterdam, pp. 3–18, 1979a.
- [29] L.A. Zadeh, “A theory of approximate reasoning,” *Machine Intelligence 9*, J. Hayes, D. Michie, and L. I. Mikulich (eds.), New York: Halstead Press, pp. 149–194, 1979b.
- [30] L.A. Zadeh, “Possibility theory and soft data analysis,” *Mathematical Frontiers of the Social and Policy Sciences*, L. Cobb and R.M. Thrall (eds.), Boulder, Westview Press, CO, pp. 69–129, 1981.
- [31] L.A. Zadeh, “Test-score semantics for natural languages and meaning representation via PRUF,” *Empirical Semantics*, B. Rieger (ed.), Brockmeyer, Bochum, W. Germany, pp. 281–349, 1982.
- [32] L.A. Zadeh, “A computational approach to fuzzy quantifiers in natural languages,” *Computers and Mathematics*, vol. 9, pp. 149–184, 1983.
- [33] L.A. Zadeh, “Outline of a computational approach to meaning and knowledge representation based on the concept of a generalized assignment statement,” *Proceedings of the International Seminar on Artificial Intelligence and Man-Machine Systems*, M. Thoma and A. Wyner (eds.), Springer-Verlag, Heidelberg, pp. 198–211, 1986.
- [34] L.A. Zadeh, “Fuzzy logic and the calculi of fuzzy rules and fuzzy graphs,” *Multiple-Valued Logic*, vol. 1, pp. 1–38, 1996.
- [35] L.A. Zadeh, “Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information/intelligent systems,” *Soft Computing*, vol. 2, pp. 23–25, 1998.
- [36] L.A. Zadeh, “Toward a perception-based theory of probabilistic reasoning with imprecise probabilities,” *Journal of Statistical Planning and Inference, Elsevier Science*, vol. 105, pp. 233–264, 2002.
- [37] L.A. Zadeh, “Precisiated natural language (PNL),” *AI Magazine*, vol. 25, no. 3, pp. 74–91, 2004.
- [38] L.A. Zadeh, “Generalized theory of uncertainty (GTU)—principal concepts and ideas,” *Computational Statistics & Data Analysis*, vol. 51, pp. 15–46, 2006.
- [39] L.A. Zadeh, “Is there a need for fuzzy logic?,” *Information Sciences*, vol. 178, no. 13, pp. 2751–2779, 2008.