CHAPTER 9

From Search Engines to Question Answering Systems – The Problems of World Knowledge, Relevance, Deduction and Precisiation[1]

Lotfi A. Zadeh

Professor in the Graduate School and Director,
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
E-mail: zadeh@eecs.berkeley.edu

Dedicated to Elie Sanchez

Abstract

Existing search engines, with Google at the top, have many truly remarkable capabilities. Furthermore, constant progress is being made in improving their performance. But what is not widely recognized is that there is a basic capability which existing search engines do not have: deduction capability – the capability to synthesize an answer to a query by drawing on bodies of information which reside in various parts of the knowledge base. By definition, a question-answering system, or Q/A system for short, is a system which has deduction capability. Can a search engine be upgraded to a question-answering system through the use of existing tools – tools which are based on bivalent logic and probability theory? A view which is articulated in the following is that the answer is: No.

The first obstacle is world knowledge – the knowledge which humans acquire through experience, communication and education. Simple examples are: “Icy roads are slippery,” “Princeton usually means Princeton University,” “Paris is the capital of France,” and “There are no honest politicians.” World knowledge plays a central role in search, assessment of relevance and deduction. The problem with world knowledge is that much of it is perception-based. Perceptions – and especially perceptions of probabilities – are intrinsically imprecise, reflecting the fact that human sensory organs, and ultimately the brain, have a bounded ability to resolve detail and store information. Imprecision of perceptions stands in the way of using conventional techniques – techniques which are based on bivalent logic and probability theory – to deal with perception-based information. A further complication is that much of world knowledge is negative knowledge in the sense that it

[1] This paper draws in part on *Toward a generalized theory of uncertainty (GTU) – an outline*, Inform. Sci. **172** (2005), 1–40.

FUZZY LOGIC AND THE SEMANTIC WEB
Edited by Elie Sanchez

relates to what is impossible and/or non-existent. For example, “A person cannot have two fathers,” and “Netherlands has no mountains.”

The second obstacle centers on the concept of relevance. There is an extensive literature on relevance, and every search engine deals with relevance in its own way, some at a high level of sophistication. But what is quite obvious is that the problem of assessment of relevance is quite complex and far from solution.

There are two kinds of relevance: (a) question relevance and (b) topic relevance. Both are matters of degree. For example, on a very basic level, if the question is q: Number of cars in California? and the available information is p: Population of California is 37,000,000, then what is the degree of relevance of p to q? Another example: To what degree is a paper entitled “A New Approach to Natural Language Understanding” of relevance to the topic of machine translation.

Basically, there are two ways of approaching assessment of relevance: (a) semantic; and (b) statistical. To illustrate, in the number of cars example, relevance of p to q is a matter of semantics and world knowledge. In existing search engines, relevance is largely a matter of statistics, involving counts of links and words, with little if any consideration of semantics. Assessment of semantic relevance presents difficult problems whose solutions lie beyond the reach of bivalent logic and probability theory. What should be noted is that assessment of topic relevance is more amendable to the use of statistical techniques, which explains why existing search engines are much better at assessment of topic relevance then question relevance.

The third obstacle is deduction from perception-based information. As a basic example, assume that the question is q: What is the average height of Swedes?, and the available information is p: Most adult Swedes are tall. Another example is: Usually Robert returns from work at about 6 pm. What is the probability that Robert is home at about 6:15 pm? Neither bivalent logic nor probability theory provide effective tools for dealing with problems of this type. The difficulty is centered on deduction from premises which are both uncertain and imprecise.

Underlying the problems of world knowledge, relevance and deduction is a very basic problem – the problem of natural language understanding. Much of world knowledge and web knowledge is expressed in a natural language. A natural language is basically a system for describing perceptions. Since perceptions are intrinsically imprecise, so are natural languages, especially in the realm of semantics.

A prerequisite to mechanization of question-answering is mechanization of natural language understanding, and a prerequisite to mechanization of natural language understanding is precisiation of meaning of concepts and proposition drawn from a natural language. To deal effectively with world knowledge, relevance, deduction and precisiation, new tools are needed. The principal new tools are: Precisiated Natural Language (PNL); Protoform Theory (PFT); and the Generalized Theory of Uncertainty (GTU). These tools are drawn from fuzzy logic – a logic in which everything is, or is allowed to be, a matter of degree.

The centerpiece of new tools is the concept of a generalized constraint. The importance of the concept of a generalized constraint derives from the fact that in PNL and GTU it serves as a basis for generalizing the universally accepted view that information is statistical in nature. More specifically, the point of departure in PNL and GTU is the fundamental premise that, in general, information is representable as a system of generalized constraints, with statistical information constituting a special case. This, much more general, view of information is needed to deal effectively with world knowledge, relevance, deduction, precisiation and related problems.

In summary, the principal objectives of this paper are: (a) to make a case for the view that a quantum jump in search engine IQ cannot be achieved through the use of methods based on bivalent logic and probability theory; and (b) to introduce and outline a collection of non-standard concepts, ideas and tools which open the door to addition of deduction capability to search engines.

1. Introduction

If I were asked, "What is the most challenging problem in the realm of information science and technology?" my unequivocal answer would be; conception and design of question-answering systems. And if I were asked what is likely to be the most important application area for fuzzy logic in coming years, my answer would be (a) improvement of performance of search engines; and (b) upgrading search engines to question-answering systems, or Q/A systems, for short.

Existing search engines are not perfect. Nevertheless, they are extremely useful. Without access to a search engine, many of us would be paralyzed. It is evident that existing search engines, with Google at the top but faced with growing competition, have many truly remarkable capabilities. But there is a very important capability which they do not have – deduction capability – the capability to synthesize an answer to a question by drawing on bodies of information which reside in various parts of the knowledge base. What should be noted, however, is that there are many widely used special purpose Q/A systems which have limited deduction capability. Examples of such systems are driving direction systems, reservation systems, diagnostic systems and specialized expert systems, especially in the domain of medicine.

It is of historical interest to note that question-answering systems were an object of considerable attention in the early seventies. The literature abounded with papers dealing with them. A few examples: L.S. Coles [6], M. Nagao and J. Tsujii [18], J.R. McSkimin and J. Minker [16], A.R. Aronson, B.E. Jacobs and J. Minker [1], W.J.H.J. Bronnenberg et al. [5]. There was no discussion of search engines because the needed computer technology was not in existence. Interest in question-answering systems dwindled in the early eighties, when it became obvious that AI was not advanced enough to provide the needed tools and technology.

In recent years, significant progress toward enhancement of web intelligence has been achieved through the use of concepts and techniques related to the Semantic Web [3], OWL [23], CYC [14] and other approaches. But can such approaches, based as they are on bivalent logic and probability theory, be employed to upgrade search engines to question-answering systems? The answer, in my view, is: No. The reason, which is not widely recognized as yet, is that bivalent logic and bivalent-logic-based probability theory have intrinsic limitations. To circumvent these limitations what are needed are new tools based on fuzzy logic and fuzzy-logic-based probability theory. What distinguishes fuzzy logic from standard logical systems is that in fuzzy logic everything is, or is allowed to be graduated, that is, be a matter of degree. Furthermore, in fuzzy logic everything is allowed to be granulated, with a granule being a clump of values drawn together by indistinguishability, similarity or proximity. It is these fundamental features of fuzzy logic that give it a far greater power to deal with problems related to web intelligence than standard tools based on bivalent logic and probability theory. An analogy which will be elaborated at a later point is: In general, a valid model of a nonlinear system cannot be constructed through the use of linear components.

There are three major obstacles to upgrading a search engine to a question-answering system: (a) the problem of world knowledge; (b) the problem of relevance; and (c) the underlying problem of mechanization of natural language understanding and, in particular, the basic problem of precisiation of meaning. A very brief discussion of these problems follows. We will return to a discussion of these problems at a later point. Since the issues to be discussed are not restricted to web-related problems, our discussion will be general in nature.

2. The principal problems

2.1. *The problem of world knowledge*

World knowledge is the knowledge which humans acquire through experience, education and communication. Simple examples are:

- Few professors are rich
- There are no honest politicians
- It is not likely to rain in San Francisco in midsummer
- Most adult Swedes are tall
- There are no mountains in Holland
- Usually Princeton means Princeton University
- Paris is the capital of France
- In Europe, the child-bearing age ranges from about sixteen to about forty-two

The problem with world knowledge is that much of it is perception-based. Examples:

- Most adult Swedes are tall
- Most adult Swedes are much taller than most adult Italians
- Usually a large house costs more than a small house
- There are no honest politicians

Perception-based knowledge is intrinsically imprecise, reflecting the bounded ability of sensory organs, and ultimately the brain, to resolve detail and store information. More specifically, perception-based knowledge is f-granular in the sense that (a) the boundaries of perceived classes are unsharp (fuzzy); and (b) the values of perceived attributes are imprecise (fuzzy). Bivalent-logic-based approaches provide no methods for deduction from perception-based knowledge. For example, given the datum: Most adult Swedes are tall, existing bivalent-logic-based methods cannot be employed to come up with valid answers to the questions q_1: How many adult Swedes are short; and q_2: What is the average height of adult Swedes?

2.2. *The problem of relevance*

The importance of the concept of relevance is hard to exaggerate. Relevance is central to search. Indeed, the initial success of Google is due, in large measure, to its simple but ingenious page ranking algorithm for assessment of relevance [4].

Despite its importance, there are no satisfactory definitions of relevance in the literature, nor can the problem of assessment of relevance be considered to be on the way to solution. In fact, it may be argued that, as in the case of world knowledge, the concept of relevance is much too complex to lend itself to treatment within the limited conceptual framework of bivalent logic and bivalent-logic-based probability theory. An immediate problem is that relevance is not a bivalent concept. Relevance is a matter of degree, that is, it is a fuzzy concept. To define fuzzy concepts, what is needed is the conceptual structure of fuzzy logic. As was stated earlier, in fuzzy logic everything is, or is allowed to be, a matter of degree.

For concreteness, it is convenient to define a relevance function, $R(q/p)$, as a function in which the first argument, q, is a question or a topic; the second argument, p, is a proposition, topic, document, web page or a collection of such objects; and R is the degree to which p is relevant to q.

When q is a question, computation of $R(q/p)$ involves an assessment of the degree of relevance of p to q, with p playing the role of question-relevant information. For example, if q: What is the number of cars in California, and p: Population of California is 37 million, then p is question-relevant to q in the sense that p constrains, albeit imprecisely, the number of cars in California. The constraint is a function of world knowledge.

If q is a topic, e.g., q: Ontology, then a document entitled p: What is ontology?, is of obvious relevance to q, i.e., p is topic-relevant. The problem in both cases is that of assessment of degree of relevance. The problem is far from solution. Basically, what we need is a method of computing the degree of relevance based on the meaning of q and p, that is, we need semantic relevance. Existing search engines have a very limited capability to deal with semantic relevance. Instead, what they use is what may be called statistical relevance. In statistical relevance, what is used is, in the main, statistics of links and counts of words. Performance of statistical methods of assessment of relevance is unreliable, as can be seen from examples which will be considered later in this section.

A major source of difficulty in assessment of relevance relates to non-compositionality of the relevance function. More specifically, assume that we have a question, q, and two propositions p and r. Can the value of $R(q/p, r)$ be composed from the values of $R(q/p)$ and $R(q/r)$? The answer, in general, is: No. As a simple, not web-related, example, suppose that q: How old is Vera; p: Vera's age is the same as Irene's; r: Irene is 65. In this case, $R(q/p) = 0$; $R(q/r) = 0$ and yet $R(q/p, r) = 1$. What this implies is that, in general, relevance cannot be assessed in isolation. This suggests a need for differentiation between relevance and what may be called i-relevance, that is, relevance in isolation. In other words, a proposition, p, is i-relevant if it is relevant by itself, and it is i-irrelevant if it is not of relevance by itself but might be relevant in combination with other propositions. An example is considered in [40,41].

2.3. *The problem of precisiation of meaning – a prerequisite to mechanization of natural language understanding*

Much of world knowledge and web knowledge is expressed in a natural language. This is why issues relating to natural language understanding and natural language reasoning are of direct relevance to search and, even more so, to question-answering.

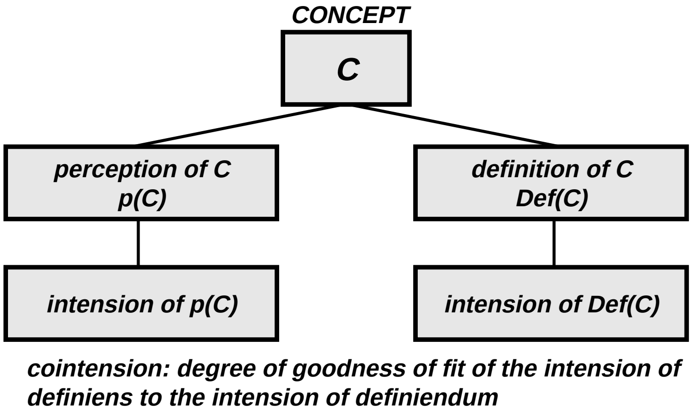


Fig. 1. Cointension of definition.

Humans have no difficulty in understanding natural language, but machines have many. One basic problem is that of imprecision of meaning. A human can understand an instruction such as "Take a few steps," but a machine cannot. To execute this instruction, a machine needs a precisiation of "few."

Precisiation of propositions drawn from a natural language is the province of PNL (Precisiated Natural Language) [40,41]. A forerunner of PNL is PRUF [34]. In PNL, precisiation is interpreted as meaning precisiation rather than value precisiation. A proposition is precisiated through translation into the Generalized Constraint Language (GCL), as will be dismissed in greater detail in Section 3. An element of GCL which precisiates p is referred to as a precisiand of p, $\mathrm{GC}(p)$, with $\mathrm{GC}(p)$ representing a generalized constraint. A precisiand may be viewed as a model of meaning.

A concept which plays a key role in precisiation is cointension, with intension used in its usual logical sense as attribute-based meaning. Thus, p and q are cointensive if the meaning of p is a close approximation to that of q. In this sense, a precisiand, $\mathrm{GC}(p)$, is valid if $\mathrm{GC}(p)$ is cointensive with p.

The concept of cointensive precisiation has an important implication for validity of definitions of concepts. More specifically, if C is a concept and $\mathrm{Def}(C)$ is its definition, then for $\mathrm{Def}(C)$ to be a valid definition, $\mathrm{Def}(C)$ must be cointensive with C (Figure 1).

The concept of cointensive definition leads to an important conclusion: In general, a cointensive definition of a fuzzy concept cannot be formulated within the conceptual structure of bivalent logic and bivalent-logic-based probability theory. This is why cointensive definitions of such concepts as causality, relevance, summary and mountain cannot be found in the literature. The concept of cointensive precisiation will be discussed in greater detail in Section 4.

3. Test queries (Google)

There is no question that existing search engines, with Google at the top, have many remarkable capabilities and are extremely useful. Search engines are extremely useful because the user of a search engine employs human intelligence to compensate for its limitations, especially in relation to world knowledge, assessment of relevance and natural language understanding. A few examples of use of Google are intended to illustrate its strengths and limitations.

Topic 1: Precisiation
Top-ranking results
[UAI] The concept of cointensive **precisiation**
. . . from data expressed in a natural language is **precisiation** of meaning. . . .
In this perspective, the problem of **precisiation** is that of picking a . . .
[PPT] www.cs.berkeley.edu/~nikraves/bisc/Present/Fall0/l . . .
. . . **precisiation** of c, i, and q can be reduced to precisiation of p . . . the issues of definability, **precisiation**, recognition, organization and search loom . . .
AI Magazine: Precisiated natural language
. . . Basically, what PNL provides is a system for **precisiation** of . . . provide a computational framework for **precisiation** of meaning rather than to serve as a . . .
Precisiation of Meaning via Translation into PRUF, Lotfi A. Zadeh, In Vaina, L. and Hintikka, J., Eds., *Cognitive Constraints on Communication*, D. Reidel, 1984: 373–401.

COMMENT. In this test, Google does a remarkably good job. Precisiation is a neologism. I used it for the first time in a paper published in 1984 – a paper which Google finds. The top-ranking result relates to a presentation which I made just a few weeks ago.

Rating of performance: Excellent

Topic 2: Precisiand
Top-ranking results
[UAI] The concept of cointensive precisiation
. . . the meaning of p is referred to as a **precisiand** of p, denoted as $\mathrm{GC}(p)$. . . .
A cointensive **precisiand** would be $\mathrm{GC}(p)$: If speed is less than about 70 mph, . . .
[PDF] Microsoft PowerPoint – Zadeh_presentation
. . . In general, **precisiand** of p is not unique. If $\mathrm{GC}.1.(p), \ldots, .\mathrm{GC}.n.(p)$ are possible precisiands of p, then a basic. question which arises is: which of . . .

COMMENT. Precisiand is a neologism which I introduced in a recent presentation. In this test, Google does a remarkably good job.

Rating of performance: Excellent

Topic 3: Tikhonov regularization
Top-ranking results
Tikhonov Regularization and Total Least Squares
. . . stable solutions to these systems it is necessary to apply **regularization** methods. We show how **Tikhonov's regularization** method, which in its original . . .
Tikhonov Regularization for Finitely and Infinitely Smoothing . . .
. . . The main goal of this paper is to obtain a unified theory of **Tikhonov regularization**, . . . **Tikhonov regularization**, Hilbert scales, spectral measure . . .
Finite Dimensional Approximation and **Tikhonov Regularization** Method
. . . of the **regularization** functional (**Tikhonov** & Arsenin, 1977, Morozov, 1984, . . . and then applying the **Tikhonov regularization** method in the frequency space. . . .

COMMENT. Tikhonov regularization is a specialized topic. Google does extremely well.

Rating of performance: Excellent

Topic 4: Knowledge representation
Top-ranking results
Book results for **knowledge representation** Knowledge Representation – by Arthur B. Markman – 200 pages
What is a **Knowledge Representation**?
An introductory critical paper by R. Davis, H. Shrobe, and P. Szolovits in AI Magazine – 1993.
Knowledge Representation Book
The cover art for John Sowa's book is Ramon Lull's tree of nature and logic. It is reproduced here, along with Sowa's explanation.

COMMENT. This is a much broader topic, and yet Google does extremely well.

Rating of performance: Excellent

Question 1: What is the capital of New York?
Top-ranking results
Web definitions for capital of new york Albany: state capital of New York; located in eastern New York State on the west bank of the Hudson river
News results for what is the capital of New York – View today's top stories After the twin-tower nightmare, New York is back on form, says . . . – Economist –
The New Raiders – BusinessWeek –
Brascan acquires New York-based Hyperion Capital for $50M US

COMMENT. There is some ambiguity in the meaning of New York: Does it refer to the city of New York or the state of New York? Google resolves the ambiguity and comes up with correct answer. The results which follow the correct answer are of little, if any, relevance.

Rating of performance: Very good

Question 2: What is the population of the capital of New York?
Top-ranking results
News results for population of New York – View today's top stories After the twin-tower nightmare, New York is back on form, says . . .
UN: World's population is aging rapidly –
New, deadly threat from AIDS virus

COMMENT. Google fails to come up with the correct answer. However, Google comes up with a correct answer to the query: What is the population of Albany? What this example shows is that Google lacks the ability to synthesize an answer to the question by combining the answers to: What is the capital of New York?, and What is the population of Albany?

Rating of performance: Poor

Question 3: What is the distance between the largest city in Spain and the largest city in Portugal?
Top-ranking results
Porto – Oporto – Portugal Travel Planner
Munich Germany Travel Planner – Hotels Restaurants Languange . . .

COMMENT. This example is a slightly more complicated version of the previous question. Similar problem with deduction incapability.

Rating of performance: Poor

Topic 5: Chirac
Top-ranking results
News results for **Chirac** – View today's top stories **Chirac**, Schroeder meet on EU constitution, cooperation
en.wikipedia.org/wiki/Jacques_Chirac
BBC NEWS | World | Europe | **Chirac** makes case for EU treaty
The French president goes on TV in a bid to boost the faltering Yes campaign on the EU constitution.
Biography – Website of the Office of the French President
. . . Jacques **Chirac** Born on 29 November 1932 in the 5th arrondissement of Paris. Son of François **Chirac**, a company director, and Marie-Louise, née Valette. . . .

COMMENT. Google correctly assumes that Chirac is the president of France. Relevance of many results is low.

Rating of performance: Fairly good

Topic 6: Chirac's son
Top-ranking results
Guardian Unlimited | Special reports | **Chirac's** nemesis
. . . Two events destroyed his position as **Chirac's** spiritual **son**. The first, a little talked about extra-marital romance with the President's daughter Claude . . .
lgf: Moore: a BIG Hit in Britain
. . . 'intellectual' circles, like Mitterand's hidden daughter, or, today, **Chirac's** hidden **son** in Japan. . . . I was not talking of **Chirac's son**, nor mazarine. . . .
International PoliticsandElections.com
. . . **son** of late French president Francois Mitterand. . . . **Chirac's** candidacy appears to be weakend substantially by allegations and investigations of . . .

COMMENT. Most of the results are irrelevant.

Rating of performance: Poor

Question 4: Age of Chirac's son
Top-ranking results
Guardian Unlimited | Special reports | **Chirac's** nemesis
. . . Their **son** plunged himself into politics at a precocious **age**, and was active

in the . . . Two events destroyed his position as **Chirac's** spiritual **son**. . . .
Jacques Chirac
. . . An only child, Chirac is the **son** of a bank clerk and later an executive for
. . . However, **Chirac's** extensive family acquaintances had him ranked back at . . .
Politics and Elections
. . . Jean-Christophe Mitterand, **son** of former President Francois Mitterand, . . .
Chirac's candidacy appears to be weakend substantially by allegations and . . .

COMMENT. No relevant results. No deduction capability.

Rating of performance: Very poor

Question 5: What is the number of cars in California?
Top-ranking results
California Lemon Law – Lemon Check CARFAX Report, used **car** history . . .
California Lemon Law provides FREE CARFAX Report, Kelley Blue Book, VIN **number**, used **car** history, auto insurance, auto warranty, lemon check, . . .
CBS News | States Mull Taxing Drivers By Mile | February 16, 2005 . . .
. . . Officials in **car**-clogged **California** are so worried they may be considering a . . . So Too Does **Number** Of Fakes, • Softer **Cars** To Save Pedestrians . . .
Ev Archive for January 2001
. . . of a decade-long bureaucratic and political fight over electric **cars** in **California**. . . . should be a greater **number** of **cars** that produce no pollution. . . .

COMMENT. No relevant results.

Rating of performance: Poor

Question 6: How many Ph.D. degrees in mathematics were granted by European universities in 1986?
Top-ranking results
A History of the **University of** Podlasie
. . . to grant master's **degrees** in the following programmes: **mathematics**, . . . its inception in **1986**, the Faculty was **granted** rights to confer **Ph.D**. degree in . . .
1999 Gairdner Foundation Winners!
. . . SB **degrees** in **Mathematics** and Economics from MIT in 1968 and his **Ph.D**. . . . of Science **degrees** from the **University of** Manitoba and McGill University. . . .
Enabling technology for distributed multimedia applications . . .
. . . **Ph.D**. degree in computer science from the **University of** Waterloo in **1986**. . . . degree in **mathematics** from the **University of** Regina, and a **Ph.D**. degree in . . .

COMMENT. No relevant results. No deduction capability.

Rating of performance: Very poor

What these tests show is that Google has some remarkable capabilities and serious incapabilities, especially in relation to assessment of semantic relevance and deduction. Google is extremely useful because a skilled human user can exploit Google's strengths and compensate for Google's limitations.

4. The new tools

The principal thesis of the proceeding discussion is that upgrading a search engine to a question-answering system is beyond the reach of methods based on bivalent logic and bivalent-logic-based probability theory. The principal obstacles are the problems of world knowledge, relevance and precisiation.

To deal with these and related problems it is necessary to move from bivalent logic to fuzzy logic – a logic that is far more general than bivalent logic. With fuzzy logic as the base, a complex of new tools can be constructed. The structure of this complex is shown in Figure 2.

In what follows, we discuss the basics of some of the new tools. The centerpiece of the new tools is the concept of a generalized constraint [35].

4.1. *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 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 km/h," are, in reality, not hard. How can such constraints be defined? The concept of a generalized constraint is motivated by questions of this kind.

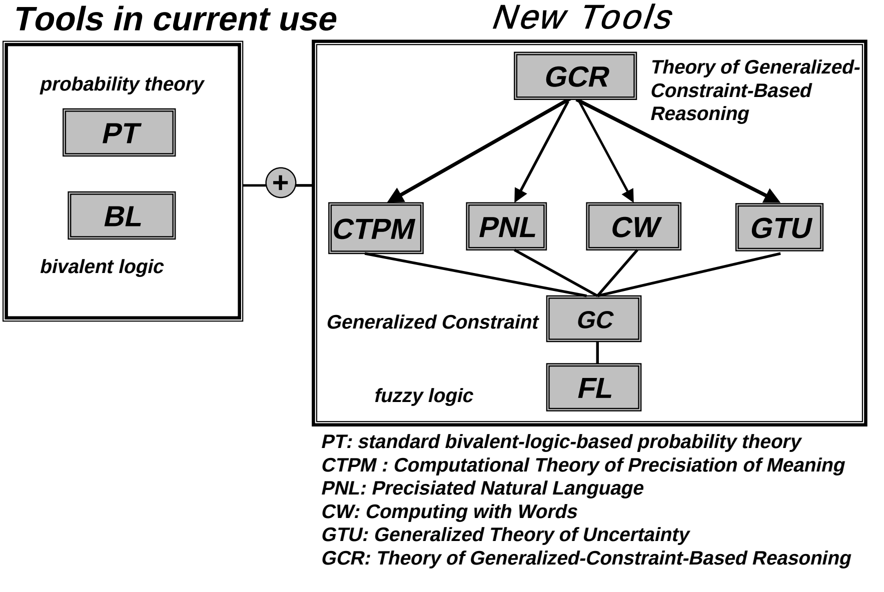


Fig. 2. Structure of new tools.

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, is defined as an expression of the form.

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

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. R will be referred to as a granular value of X.

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

- X is an n-ary variable, $X = (X_1, \ldots, X_n)$
- X is a proposition, e.g., $X =$ 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., $X =$ Location(Residence(Carol))
- X is a group variable. In this case, there is a group, $G[A]$; with each member of the group, Name$_i$, $i = 1, \ldots, n$, associated with an attribute-value, A_i. A_i may be vector-valued. Symbolically

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

  Basically, $G[A]$ is a relation.
- X is a generalized constraint, $X = Y$ is$r$ R.

A generalized constraint, GC, is associated with a test-score function, $ts(u)$ [31] 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 semi-ring [21], an element of a lattice [11] or, more generally, an element of a partially ordered set, or a bimodal distribution – a constraint which will be described later in this section. 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.

4.1.1. *Principal modalities of generalized constraints*

(a) *Possibilistic* ($r =$ 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. If μ_{small} is the membership function of small, then the semantics of "X is small" is defined by

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

where u is a generic value of X.

(b) *Probabilistic* ($r = \mathrm{p}$)

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

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

$$X \text{ isp } N\left(m, \sigma^2\right)$$

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 $\{u_1, \ldots, u_n\}$ with respective probabilities $p_1, \ldots, p_n$, then X may be expressed symbolically as

$$X \text{ isp } (p_1 \backslash u_1 + \cdots + p_n \backslash u_n),$$

with the semantics

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

What is important to note is that in the Generalized Theory of Uncertainty (GTU) [42] 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 [29]. For example.

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

where small is a fuzzy subset of the real line, means that the 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 even "X is small" may be expressed as [25]

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

Hence

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

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

(c) *Veristic* ($r = \mathrm{v}$)

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

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

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

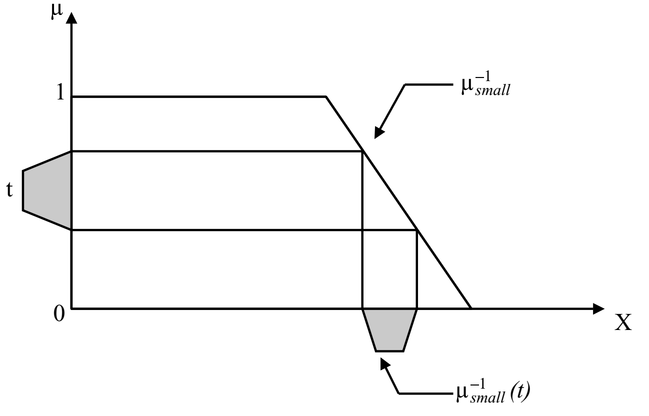


Fig. 3. Truth-qualification: (X is small) is t.

meaning that $\mathrm{Ver}(X = u_i) = t_i$, $i = 1, \ldots, n$.

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

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

When X is a generalized constraint, the expression

$$X \text{ isv } R$$

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

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

should be interpreted as "It is very true that X is small." The semantics of truth qualification is defined in [29,30]

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

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 3.

NOTE. There are two classes of fuzzy sets: (a) possibilistic, and (b) veristic. In the case of a possibilistic fuzzy set, the grade of membership is the degree of possibility. In the case of a veristic fuzzy set, the grade of membership is the degree of verity (truth). Unless stated to the contrary, a fuzzy set is assumed to be possibilistic.

(d) *Usuality* ($r = \mathrm{u}$)

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

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 [13]. For example.

$$X \text{ isu small}$$

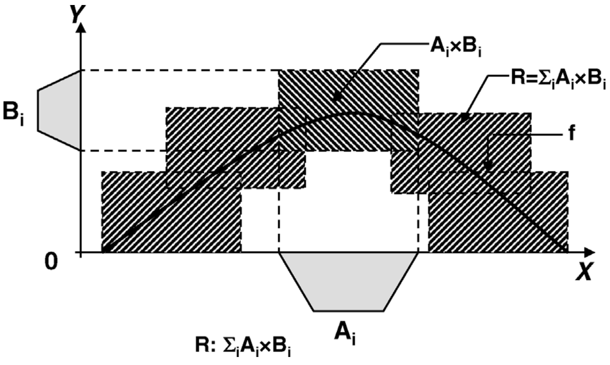


Fig. 4. Fuzzy graph.

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 expected value.

(e) *Random-set* ($r = \text{rs}$)
In

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

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

(f) *Fuzzy-graph* ($r = \text{fg}$)
In

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

X is a function, f, and R is a fuzzy graph [27] which constrains f (Figure 4). A fuzzy graph is a disjunction of Cartesian granules expressed as

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

where the A_i and B_i, $i = 1, \ldots, n$, 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 [26, 36,20,2].

$$R\colon \text{ if } X \text{ is } A_1 \text{ then } Y \text{ is } B_1, \quad i = 1, \ldots, n.$$

The concept of a fuzzy-graph constraint plays an important role in applications of fuzzy logic [2,10,12].

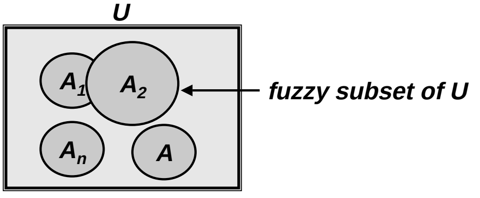


Fig. 5. Bimodal distribution. The A_i are fuzzy subsets of U.

(g) *Bimodal* ($r = \text{bm}$)

In the bimodal constraint,

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

R is a bimodal distribution of the form

$$R\colon \sum_i P_i \backslash A_i, \quad i = 1, \ldots, n,$$

which means that Prob(X is A_i) is P_i [39].

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 [29,30]. Type 1, 2 and 3 bimodal distributions have a common framework but differ in important detail (Figure 5). 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
  $A_1, \ldots, A_n, A$ are events (fuzzy sets)
  $p_i = \text{Prob}(X \text{ is } A_i)$, $\text{Prob}(X \text{ is } A_i)$ is P_i, $i = 1, \ldots, n$,
  $\sum_i p_i$ is unconstrained
  BD: bimodal distribution: $((P_1, A_1), \ldots, (P_n, A_n))$
  or, equivalently,

$$X \text{ isbm } (P_1 \backslash A_1 + \cdots + P_n \backslash A_n)$$

PROBLEM. What is the probability, p, of A? In general, this probability is fuzzy-set-valued.

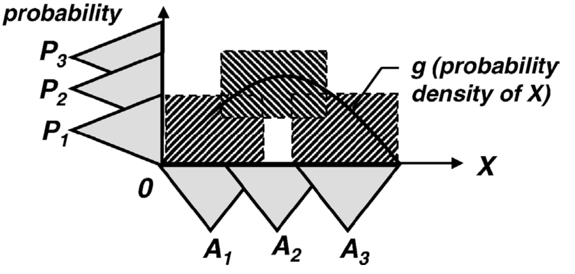


Fig. 6. Basic bimodal distribution.

A special case of bimodal distribution of Type 1 is the basic bimodal distribution (BBD). In BBD, X is a real-valued random variable, and X and P are granular (Figure 6).

- Type 2 (fuzzy random set): X is a fuzzy-set-valued random variable with values $A_1, \ldots, A_n$ (fuzzy sets) $p_i = \text{Prob}(X = A_i)$, $\text{Prob}(X \text{ is } A_i)$ is P_i, $i = 1, \ldots, n$

  BD: X isrs $(P_1 \backslash A_1 + \cdots + P_n \backslash A_n)$

$$\sum_i P_i = 1,$$

  where the P_i are granular probabilities.

PROBLEM. What is the 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) [22]: X is a random variable taking values $X_1, \ldots, X_n$ with probabilities $p_1, \ldots, p_n$.

$$X_i \text{ is a random variable taking values in } A_i, \quad i = 1, \ldots, n$$

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

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.

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.

What is the expected value of a bimodal distribution? This question is considered in Section 4.

(h) *Group* ($r = g$)

In

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

X is a group variable, $G[A]$, and R is a group constraint on $G[A]$. More specifically, if X is a group variable of the form

$$G[A]\colon\ \text{Name}_1/A_1 + \cdots + \text{Name}_n/A_n$$

or

$$G[A]\colon\ \sum_i \text{Name}_i/A_i, \quad \text{for short, } i = 1, \ldots, n,$$

then R is a constraint on the A_i. To illustrate, if we have a group of n Swedes, with Name_i being the name of ith Swede, and A_i being the height of Name_i, then the proposition "most Swedes are tall," is a constraint on the A_i which may be expressed as [40]

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

or, more explicitly,

$$\frac{1}{n}\big(\mu_{\text{tall}}(A_1) + \cdots + \mu_{\text{tall}}(A_n)\big) \text{ is most},$$

where most is a fuzzy quantifier which is interpreted as a fuzzy number.

4.1.2. *Operations on generalized constraints*

There are many ways in which generalized constraints may be operated on. The basic operations – expressed in symbolic form – are the following.

(a) *Conjunction*

$$\frac{\begin{array}{l} X \text{ isr } R \\ Y \text{ iss } S \end{array}}{(X, Y) \text{ ist } T}.$$

EXAMPLE (possibilistic constraints).

$$\frac{\begin{array}{l} X \text{ is } R \\ Y \text{ is } S \end{array}}{(X, Y) \text{ is } R \times S},$$

where $\times$ is the Cartesian product.

EXAMPLE (probabilistic/possibilistic).

$$\frac{\begin{array}{l} X \text{ isp } R \\ (X, Y) \text{ is } S \end{array}}{(X, Y) \text{ isrs } T}.$$

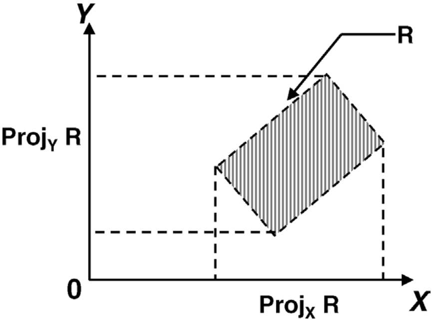


Fig. 7. Projection.

In this example, if S is a fuzzy relation then T is a fuzzy random set. What is involved in this example is a conjunction of a probabilistic constraint and a possibilistic constraint. This type of probabilistic/possibilistic constraint plays a key role in the Dempster–Shafer theory of evidence [22], and in its extension to fuzzy sets and fuzzy probabilities [29,30].

EXAMPLE (possibilistic/probabilistic).

$$\frac{\begin{array}{l} X \text{ is } R \\ (X, Y) \text{ isp } S \end{array}}{Y/X \text{ isp } T}.$$

This example, which is a dual of the proceeding example, is an instance of conditioning.

(b) *Projection* (*possibilistic*) (*Figure* 7)

$$\frac{(X, Y) \text{ is } R}{X \text{ is } S},$$

where X takes values in $U = \{u\}$; Y takes values in $V = \{v\}$; and the projection

$$S = \text{Proj}_X R,$$

is defined as

$$\mu_S(u) = \mu_{\text{Proj}_X R}(u) = \max_v \mu_R(u, v),$$

where μ_R and μ_S are the membership functions of R and S, respectively.

(c) *Projection* (*probabilistic*)

$$\frac{(X, Y) \text{ isp } R}{X \text{ isp } S},$$

where X and Y are real-valued random variables, and R and S are the probability distributions of (X, Y) and X, respectively. The probability density function of S, p_S, is related to that of R, p_R, by the familiar equation

$$p_S(u) = \int p_R(u, v)\, \mathrm{d}v$$

with the integral taken over the real line.

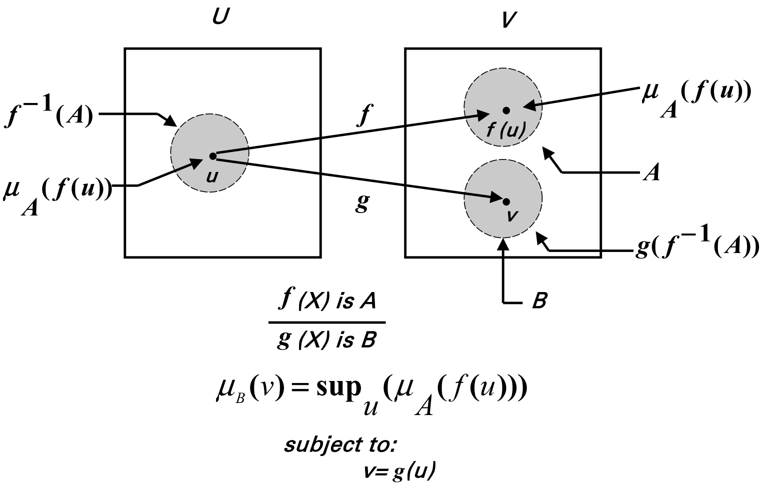


Fig. 8. Extension principle.

(d) *Propagation*

$$\frac{f(X) \text{ isr } R}{g(X) \text{ iss } S},$$

where f and g are functions or functionals.

EXAMPLE (possibilistic constraints) (Figure 8).

$$\frac{f(X) \text{ is } R}{g(X) \text{ is } S},$$

where R and S are fuzzy sets. In terms of the membership function of R, the membership function of S is given by the solution of the variational problem

$$\mu_S(v) = \sup_u \big(\mu_R\big(f(u)\big)\big)$$

subject to

$$v = g(u).$$

NOTE. The constraint propagation rule described in this example is the well-known extension principle of fuzzy logic [24,28]. Basically, this principle provides a way of computing the possibilistic constraint on $g(X)$ given a possibilistic constraint on $f(X)$.

4.1.3. *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:

Possibilistic constraint: X is R

Probabilistic constraint: X isp R

and

Veristic constraint: X 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, GC, 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) [42] embraces all possible mixtures. In this sense the conceptual structure of GTU accommodates most, and perhaps all, of the existing theories of uncertainty.

4.1.4. *The generalized constraint language and standard constraint language*

A concept which has a position of centrality in PNL 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:

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

where $\wedge$ is conjunction.

A very simple example of a semantic rule is:

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

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.

In PNL, 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.

4.2. *The concept of cointensive precisiation*

As was pointed out already, much of world knowledge and web knowledge is expressed in a natural language. For this reason, mechanization of natural language understanding is of direct relevance to enhancement of web intelligence.

In recent years, considerable progress has been made in areas of computational linguistics which relate to mechanization of natural language understanding. But what is widely unrecognized is that there is a fundamental limitation to what can be achieved through the use of commonly-employed methods of meaning representation. The aim of what follows is, first, to highlight this limitation and, second, to suggest ways of removing it.

To understand the nature of the limitation, two facts have to be considered. First, as was pointed out earlier, a natural language, NL, is basically a system for describing perceptions; and second, perceptions are intrinsically imprecise, reflecting the bounded ability of human sensory organs, and ultimately the brain, to resolve detail and store information.

A direct consequence of imprecision of perceptions is semantic imprecision of natural languages. Semantic imprecision of natural languages is not a problem for humans, but is a major problem for machines.

To clarify the issue, let p be a proposition, concept, question or command. For p to be understood by a machine, it must be precisiated, that is, expressed in a mathematically well-defined language. A precisiated form of p, $\text{Pre}(p)$, will be referred to as a precisiand of p and will be denoted as p^*. The object of precisiation, p, will be referred to us precisiend.

To precisiate p we can employ a number of meaning-representation languages, e.g., Prolog, predicate logic, semantic networks, conceptual graphs, LISP, SQL, etc. The commonly-used meaning-representation languages are bivalent, i.e., are based on bivalent logic. Are we moving in the right direction when we employ such languages for mechanization of natural language understanding? The answer is: No. The reason relates to an important issue which we have not addressed: cointension of p^*, with intension used in its logical sense as attribute-based meaning. More specifically, cointension is a measure of the goodness of fit of the intension of a precisiand, p^*, to the intended intension of precisiend, p. Thus, cointension is a desideratum of precisiation. What this implies is that mechanization of natural language understanding requires more than precisiation – it requires cointensive precisiation. Note that definition is a form of precisiation. In plain words, a definition is cointensive if its meaning is a good fit to the intended meaning of the definiendum.

Here is where the fundamental limitation which was alluded to earlier comes into view. In a natural language, NL, most p's are fuzzy, that is, are in one way or another, a matter of degree. Simple examples: propositions "most Swedes are tall" and "overeating causes obesity;" concepts "mountain" and "honest;" question "is Albert honest?" and command "take a few steps."

Employment of commonly-used meaning-representation languages to precisiate a fuzzy p leads to a bivalent (crisp) precisiand p^*. The problem is that, in general, a bi-

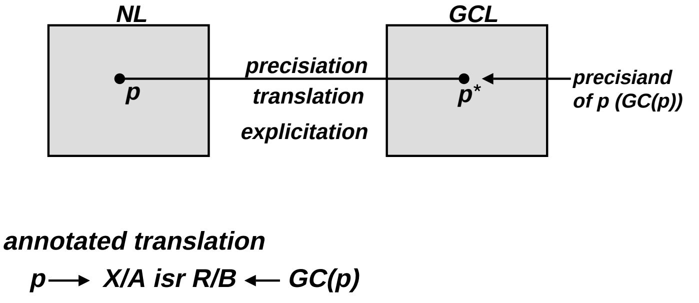


Fig. 9. Precisiation = translation into GCL.

valent p^* is not cointensive. As a simple illustration, consider the concept of recession. The standard definition of recession is: A period of general economic decline; specifically, a decline in GDP for two or more consecutive quarters. Similarly, a 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. (Robert Shuster, Ned Davis Research.) Clearly, neither definition is cointensive. Another example is the classical definition of stability. Consider a ball of diameter D which is placed on an open bottle whose mouth is of diameter d. If D is somewhat larger than d, the configuration is stable: Obviously, as D increases, the configuration becomes less and less stable. But, according to Lyapounov's bivalent definition of stability, the configuration is stable for all values of D greater than d. This contradiction is characteristic of crisp definitions of fuzzy concepts – a well-known example of which is the Greek sorites (heap) paradox. The magnitude of the problem becomes apparent when we consider that many concepts in scientific theories are fuzzy, but are defined and treated as if they are crisp. This is particularly true in fields in which the concepts which are defined are descriptions of perceptions.

To remove the fundamental limitation, bivalence must be abandoned. Furthermore, new concepts, ideas and tools must be developed and deployed to deal with the issues of cointensive precisiation, definability and deduction. The principal tools are Precisiated Natural Language (PNL); Protoform Theory (PFT); and the Generalized Theory of Uncertainty (GTU). These tools form the core of what may be called the Computational Theory of Precisiation of Meaning (CTPM). The centerpiece of CTPM is the concept of a generalized constraint [35].

The concept of a generalized constraint plays a key role in CTPM by providing a basis for precisiation of meaning. More specifically, if p is a proposition or a concept, its precisiand, $\mathrm{Pre}(p)$, is represented as a generalized constraint, GC. Thus, $\mathrm{Pre}(p) = \mathrm{GC}$. In this sense, the concept of a generalized constraint may be viewed as a bridge from natural languages to mathematics (Figure 9).

Representing precisiands of p as elements of GCL is the pivotal idea in CTPM. Each precisiand is associated with the degree to which it is cointensive with p. Given p, the problem is that of finding those precisiands which are cointensive, that is, have a high degree of cointension. If p is a fuzzy proposition or concept, then in general there are no cointensive precisiands in SCL (Figure 10).

In CTPM, a refinement of the concept of precisiation is needed. First, a differentiation is made between v-precision (precision in value) and m-precision (precision in meaning) (Figure 11). For example, proposition p: X is 5, is both v-precise and m-precise; p: X

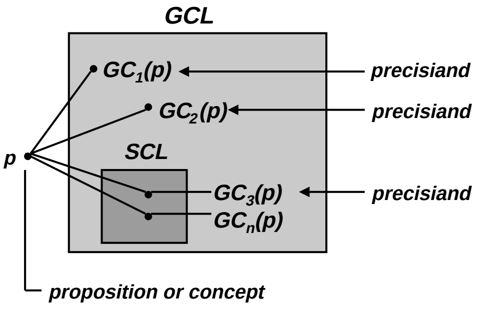


Fig. 10. Precisiands of p.

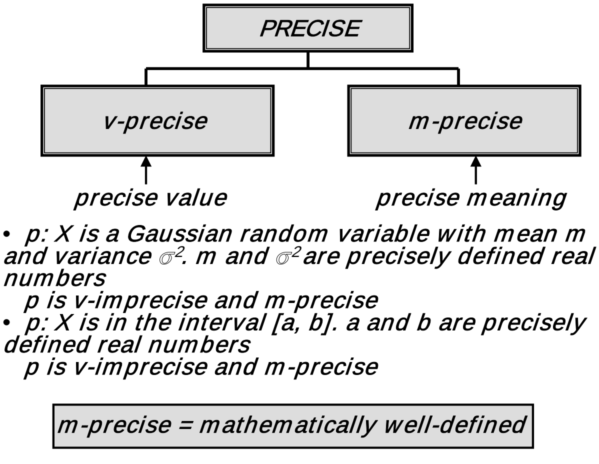


Fig. 11. Two meanings of precise.

is between 5 and 7, is v-imprecise and m-precise; and p: X is small, is both v-imprecise and m-imprecise; however, p can be m-precisiated by defining small as a fuzzy set or a probability distribution. A perception is v-imprecise and its description is m-imprecise. PNL makes it possible to m-precisiate descriptions of perceptions.

Granulation of a variable, e.g., representing the values of age as young, middle-aged and old, may be viewed as a form of v-imprecisiation. Granulation plays an important role in human cognition by serving as a means of (a) exploiting a tolerance for imprecision through omission of irrelevant information; (b) lowering precision and thereby lowering cost; and (c) facilitating understanding and articulation. In fuzzy logic, granulation is m-precisiated through the use of the concept of a linguistic variable.

Further refinement of the concept of precisiation relates to two modalities of m-precisiation: (a) human-oriented, denoted as *mh*-precisiation; and (b) machine-oriented, denoted as *mm*-precisiation (Figure 12). Unless stated to the contrary, in CTPM, precisiation should be understood as *mm*-precisiation.

In a bimodal dictionary or lexicon, the first entry, p, is a concept or proposition; the second entry, p^*, is *mh*-precisiand of p; and the third entry is *mm*-precisiand of p. To illustrate,

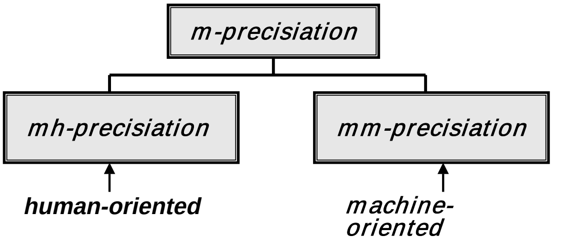


Fig. 12. Modalities of m-precisiation.

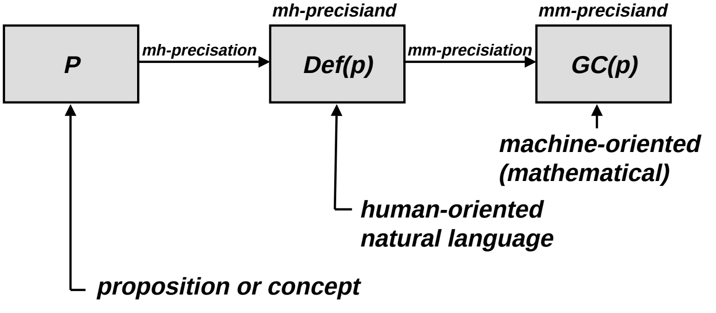


Fig. 13. Bimodal lexicon (PNL).

the entries for recession might read: *mh*-precisiand – a period of general economic decline; and *mm*-precisiand – a decline in GDP for two or more consecutive quarters (Figure 13).

There is a simple analogy which helps to understand the meaning of cointensive precisiation. Specifically, a proposition, p, is analogous to a system, S; precisiation is analogous to modelization; a precisiand, expressed as a generalized constraint, $\mathrm{GC}(p)$, is analogous to a model, $M(S)$, of S; test-score function is analogous to input–output relation; cointensive precisiand is analogous to well-fitting model; GCL is analogous to the class of all fuzzy-logic-based systems; and SCL is analogous to the subclass of all bivalent-logic-based systems (Figure 14). To say that, in general, a cointensive definition of a fuzzy concept cannot be formulated within the conceptual structure of bivalent logic and probability theory, is similar to saying that, in general, a linear system cannot be a well-fitting model of a nonlinear system.

Ramifications of the concept of cointensive precisiation extend well beyond mechanization of natural language understanding. A broader basic issue is validity of definitions in scientific theories, especially in the realms of human-oriented fields such as law, economics, medicine, psychology and linguistics. More specifically, the concept of cointensive precisiation calls into question the validity of many of the existing definitions of basic concepts – among them the concepts of causality, relevance, independence, stability, complexity and optimality.

Translation of p into GCL is made more transparent though annotation. To illustrate,

(a) p: Monika is young $\longrightarrow$ $X/\mathrm{Age(Monika)}$ is R/young
(b) p: It is likely that Monika is young $\longrightarrow$ $\mathrm{Prob}(X/\mathrm{Age(Monika)}$ is $R/\mathrm{young})$ is S/likely

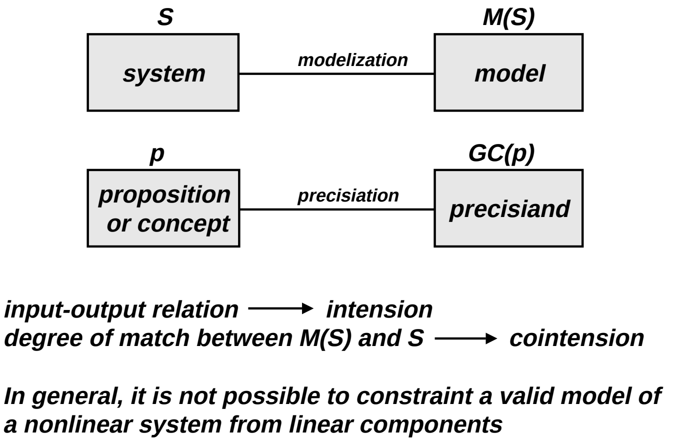


Fig. 14. Analogy between precisiation and modelization.

NOTE. Example (b) is an instance of probability qualification.

More concretely, let $g(u)$ be the probability density function of the random variable, Age(Monika). Then, with reference to our earlier discussion of probability qualification, we have

$$\text{Prob}\big(\text{Age(Monika) is young}\big) \text{ is likely} \longrightarrow \int_0^{100} g(u)\mu_{\text{young}}(u)\,\mathrm{d}u \text{ is likely}$$

or, in annotated form,

$$\text{GC}(g) = X \Big/ \int_0^{100} g(u)\mu_{\text{young}}(u)\,\mathrm{d}u \text{ is } R/\text{likely}.$$

The test-score of this constraint on g is given by

$$ts(g) = \mu_{\text{likely}}\left(\int_0^{100} g(u)\mu_{\text{young}}(u)\,\mathrm{d}u\right).$$

(c) p: Most Swedes are tall

Following (b), let $h(u)$ be the count density function of Swedes, meaning that $h(u)\,\mathrm{d}u =$ fraction of Swedes whose height lies in the interval $[u, u + \mathrm{d}u]$.

Assume that height of Swedes lies in the interval $[a, b]$. Then,

$$\text{fraction of tall Swedes:} \quad \int_a^b h(u)\mu_{\text{tall}}(u)\,\mathrm{d}u \text{ is most.}$$

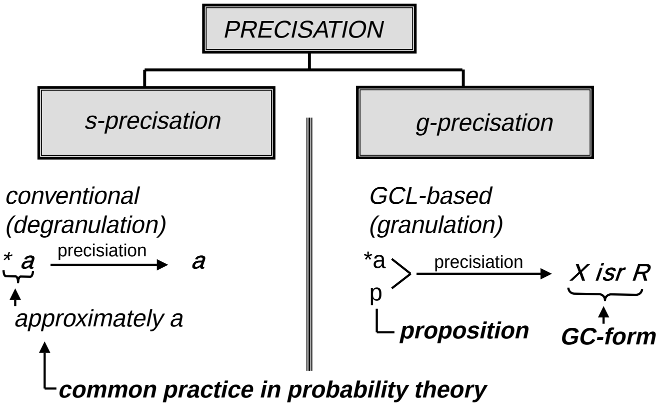


Fig. 15. s-precisiation and g-precisiation.

Interpreting this relation as a generalized constraint on h, the test-score may be expressed as

$$ts(h) = \mu_{\text{most}}\left(\int_0^h h(u)\mu_{\text{tall}}(u)\,\mathrm{d}u\right).$$

In summary, precisiation of "Most Swedes are tall" may be expressed as the generalized constraint.

$$\text{Most Swedes are tall} \longrightarrow \text{GC}(h) = \mu_{\text{most}}\left(\int_a^b h(u)\mu_{\text{tall}}(u)\,\mathrm{d}u\right).$$

An important application of the concept of precisiation relates to precisiation of propositions of the form "X is approximately a," where a is a real number. How can "approximately a," or $^*a$ for short, be precisiated? In other words, how can the uncertainty associated with the value of X which is described as $^*a$, be defined precisely?

There is a hierarchy of ways in which this can be done. The simplest is to define $^*a$ as a. This mode of precisiation will be referred to as singular precisiation, or s-precisiation, for short (Figure 15) s-precisiation is employed very widely, especially in probabilistic computations in which an imprecise probability, $^*a$, is computed with as if it were an exact number, a.

The other ways (Figures 15, 16) will be referred to as granular precisiation, or g-precisiation, for short. In g-precisiation, $^*a$ is treated as a granule. What we see is that various modes of precisiating $^*a$ are instances of the generalized constraint.

The concept of precisiation has an inverse – the concept of imprecisiation, which involves replacing a with $^*a$, with the understanding that $^*a$ is not unique.

Imprecisiation has a negative connotation. In fact, imprecisiation serves an important purposes. More specifically, consider a proposition p of the form

$$p\text{: } X \text{ is } V,$$

where X is a variable and V is its value. X may assume a variety of forms. In particular, X may be a real-valued variable, an n-ary variable, a function or a relation. The value, V, is

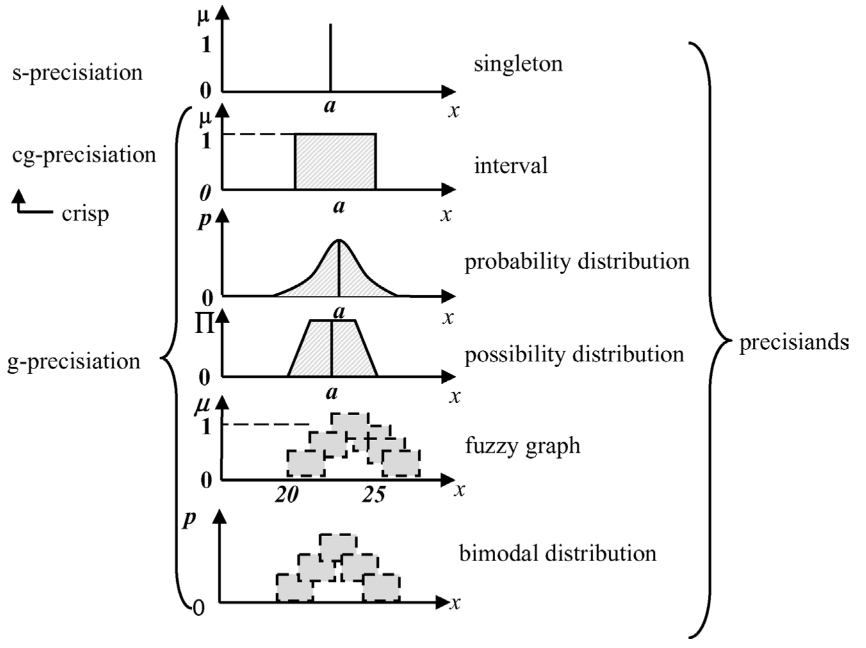


Fig. 16. Precisiands of "X is approximately a."

v-precise if it is singular, that is, V is a singleton. V is v-imprecise if it is granular. In this framework, v-imprecisiation may be interpreted as a transition from singular to granular value of V.

v-imprecisiation is forced (necessary) when the value of V is not known precisely. v-imprecisiation is deliberate (optional) if there is no need for V to be known precisely. In this case, what may be called v-imprecisiation principle comes into play.

v-imprecisiation principle: Precision carries a cost. If there is a tolerance for imprecision, exploit it by employing v-imprecisiation to achieve lower cost, robustness, tractability, decision-relevance and higher level of confidence.

A word about confidence. If V is uncertain, the confidence in p, $\mathrm{Con}(p)$, may be defined as the probability that p is true. Generally, v-imprecisiation of V serves to increase $\mathrm{Con}(p)$. For example, Con(Carol is young) $>$ Con(Carol is 23). Thus, as a rule, confidence increases when specificity decreases.

An important example is granulation. In fuzzy logic, granulation may be interpreted as v-imprecisiation followed by *mm*-precisiation. In this perspective, the concept of granulation – in combination with the associated concept of a linguistic variable – may be viewed as one of the major contributions of fuzzy logic.

A basic problem which relates to imprecisiation is the following. Assume for simplicity that we have two linear equations involving real-valued coefficients and real-valued variables:

$$a_{11}X + a_{12}Y = b_1,$$
$$a_{21}X + a_{22}Y = b_2.$$

Solutions of these equations read,

$$X = \frac{a_{22}b_1 - a_{12}b_2}{a_{11}a_{22} - a_{12}a_{21}},$$
$$Y = \frac{a_{11}b_2 - a_{21}b_1}{a_{11}a_{22} - a_{12}a_{21}}.$$

Now suppose that we imprecisiate the coefficients, replacing, a_{ij} with $^*a_{ij}$, $i, j = 1, 2$, and replacing b_i with $^*b_i$, $i = 1, 2$. How can we solve these equations when imprecisiated coefficients are defined as generalized constraints?

There is no general answer to this question. Assuming that all coefficients are defined in the same way, the method of solution will depend on the modality of the constraint. For example, if the coefficients are interval-valued, the problem falls within the province of interval analysis [17]. If the coefficients are fuzzy-interval-valued, the problem falls within the province of the theory of relational equations [8,7]. And if the coefficients are real-valued random variables, we are dealing with the problem of solution of stochastic equations. In general, solution of a system of equations with imprecisiated coefficients may present complex problems.

One complication is the following. If (a) we solve the original equations, as we have done above; (b) imprecisiate the coefficients in the solution; and (c) employ the extension principle to complete X and Y, will we obtain solutions of imprecisiated equations? The answer, in general, is: No.

Nevertheless, when we are faced with a problem which we do not know how to solve correctly, we proceed as if the answer is: Yes. This common practice may be described as Precisiation/Imprecisiation Principle which is defined in the following.

4.3. *Precisiation/imprecisiation principle (P/I principle)*

Informally, let f be a function or a functional. $Y = f(X)$, where X and Y are assumed to be imprecise, $Pr(X)$ and $Pr(Y)$ are precisiations of X and Y, and $^*Pr(X)$ and $^*Pr(Y)$ are imprecisiations of $Pr(X)$ and $Pr(Y)$, respectively. In symbolic form, the P/I principle may be expressed as

$$f(X) \;{}^*\!\!= {}^*f\big(Pr(X)\big),$$

where $^*=$ denotes "approximately equal," and $^*f$ is imprecisiation of f. In words, to compute $f(X)$ when X is imprecise, (a) precisiate X, (b) compute $f(Pr(X))$; and (c) imprecisiate $f(Pr(X))$. Then, usually, $^*f(Pr(X))$ will be approximately equal to $f(X)$. An underlying assumption is that approximations are commensurate in the sense that the closer $Pr(X)$ is to X, the closer $f(Pr(X))$ is to $f(X)$. This assumption is related to the concept of gradual rules of Dubois and Prade [9].

As an illustration, suppose that X is a real-valued function; f is the operation of differentiation, and $^*X$ is the fuzzy graph of X. Then, using the P/I principle, $^*f(X)$ will have the form shown in Figure 17. It should be underscored that imprecisiation is an imprecise concept.

Use of the P/I principle underlies many computations in science, engineering, economics and other fields. In particular, as was alluded to earlier, this applies to many computations in probability theory which involve imprecise probabilities. It should be emphasized that the

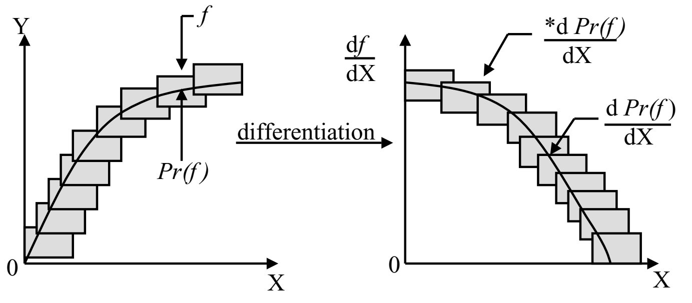


Fig. 17. Illustration of P/I principle.

P/I principle is neither normative (prescriptive) nor precise; it merely describes imprecisely what is common practice – without suggesting that common practice is correct.

4.3.1. *Precisiation of propositions*

In preceding discussion, we focused our attention on precisiation of propositions of the special form "X is $^*a$." In the following, we shall consider precisiation in a more general setting. In this setting, the concept of precisiation in PNL opens the door to a wide-ranging enlargement of the role of natural languages in scientific theories, especially in fields such as economics, law and decision analysis. Our discussion will be brief; details may be found in [40,41].

Within CTPM, precisiation of propositions – and the related issues of precisiation of questions, commands and concepts – falls within the province of PNL. As was stated earlier, the point of departure in PNL is representation of a precisiand of a proposition, p, as a generalized constraint.

$$p \longrightarrow X \text{ isr } R.$$

To illustrate precisiation of propositions and questions, it will be convenient to consider the examples which were discussed earlier in Section 4.

(a) *The Robert example*

p: Usually Robert returns from work at about 6 pm.

Q: What is the probability that Robert is home at about 6:15 pm?

Precisiation of p may be expressed as

p: Prob(Time(Return(Robert)) is *6:00 pm) is usually

where "usually" is a fuzzy probability.

Assuming that Robert stays home after returning from work, precisiation of q may be expressed as

$$q\text{: Prob}\big(\text{Time}\big(\text{Return(Robert)}\big) \text{ is } \leqslant \circ \text{ 6:15 pm}\big) \text{ is } A?$$

where $\circ$ is the operation of composition, and A is a fuzzy probability

(b) *The balls-in-box problem*

p_1: A box contains about 20 black and white balls

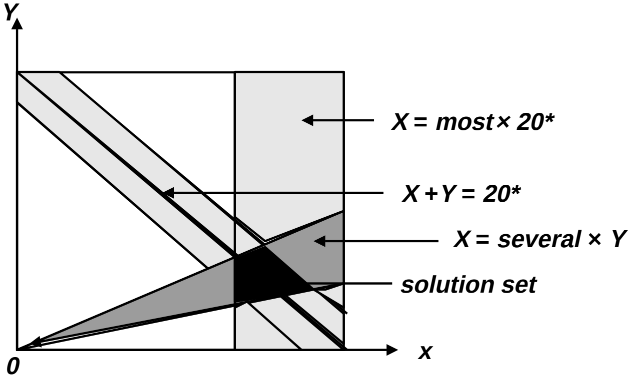


Fig. 18. Fuzzy integer programming.

p_2: Most are black
p_3: There are several times as many black balls as white balls
q_1: What is the number of white balls?
q_2: What is the probability that a ball drawn at random is white?

Let X be the number of black balls and let Y be the number of white balls. Then, in precisiated form, the statement of the problem may be expressed as:

$$\left.\begin{array}{l} p_1\colon\ (X+Y) \text{ is } {}^*20 \\ p_2\colon\ X \text{ is most } \times\ {}^*20 \\ p_3\colon\ X \text{ is several } \times Y \end{array}\right\} \text{data}$$

$$\left.\begin{array}{l} q_1\colon\ Y \text{ is } ?A \\ q_2\colon\ \dfrac{Y}{{}^*20} \text{ is } ?B \end{array}\right\} \text{questions}$$

where $Y/{}^*20$ is the granular probability that a ball drawn at random is white.

Solution of these equations reduces to an application of fuzzy integer programming (Figure 18).

(c) *The tall Swedes problem*

p: Most Swedes are tall.
Q: What is the average height of Swedes?
Q: How many Swedes are short?

As was shown earlier,

$$p\colon \text{ Most Swedes are tall} \longrightarrow \int_a^b h(u)\mu_{\text{tall}}(u)\,\mathrm{d}u \text{ is most,}$$

where h is the count density function.

Precisiations of q_1, and q_2 may be expressed as

$$q_1\colon \int_a^b uh(u)\,\mathrm{d}u \text{ is } ?A,$$

where A is a fuzzy number which represents the average height of Swedes, and

$$q_2\colon \int_a^b h(u)\mu_{\text{short}}(u)\,\mathrm{d}u \text{ is } ?B,$$

where μ_{short} is the membership function of short, and B is the fraction of short Swedes.

(d) *The partial existence problem*

X is a real number. I am uncertain about the value of X. What I know about X is:
p_1: X is much larger than approximately a,
p_2: X is much smaller than approximately b,
where a and b are real numbers, with $a < b$.

What is the value of X?

In this case, precisiations of data may be expressed as
p_1: X is much.larger $\circ\, {}^*a$
p_2: X is much smaller $\circ\, {}^*b$,
where $\circ$ is the operation of composition. Precisiation of the question is:

$$q\colon\ X \text{ is } ?A,$$

where A is a fuzzy number. The solution is immediate:

$$X \text{ is } (\text{much.larger} \circ {}^*a \wedge \text{ much.smaller} \circ {}^*b),$$

when $\wedge$ is min or a t-norm. In this instance, depending on a and b, X may exist to a degree.

These examples point to an important aspect of precisiation. Specifically, to precisiate p we have to precisiate or, equivalently, calibrate its lexical constituents. For example, in the case of "Most Swedes are tall," we have to calibrate "most" and "tall." Likewise, in the case of the Robert example, we have to calibrate "about 6:00 pm," "about 6:15 pm" and "usually." In effect, we are composing the meaning of p from the meaning of its constituents. This process is in the spirit of Frege's principle of compositionality [32,33], Montague grammar [19] and the semantics of programming languages.

An important aspect of precisiation which will not be discussed here relates to precisiation of concepts. It is a deep-seated tradition in science to base definition of concepts on bivalent logic. In probability theory, for example, independence of events is a bivalent concept. But, in reality, independence is a matter of degree, i.e., is a fuzzy concept. PNL, used as a definition language, makes it possible, more realistically, to define independence and other bivalent concepts in probability theory as fuzzy concepts. For this purpose, when PNL is used as a definition language, a concept is first defined in a natural language and then its definition is precisiated through the use of PNL.

4.4. *The concept of a protoform*

Viewed in a broader perspective, what should be noted is that precisiation of meaning is not the ultimate goal – it is an intermediate goal. Once precisiation of meaning is achieved, the next goal is that of deduction from decision-relevant information. The ultimate goal is decision.

In CTPM, a concept which plays a key role in deduction is that of a protoform – an abbreviation for prototypical form. Informally, a protoform of an object is its abstracted summary. More specifically, a protoform is a symbolic expression which defines the deep semantic structure of an object such as a proposition, question, command, concept, scenario, or a system of such objects. In the following, our attention will be focused on protoforms of propositions, with PF(p) denoting a protoform of p (Figure 19). Abstraction has levels, just as summarization does. For this reason, an object may have a multiplicity of protoforms (Figure 20). Conversely, many objects may have the same protoform. Such objects are said to be protoform-equivalent, or PF-equivalent, for short. The set of protoforms of all precisiable propositions in NL, together with rules which govern propagation of generalized constraints, constitute what is called the Protoform Language (PFL).

EXAMPLES.

- instantiation
  Monika is young ⟶ Age(Monika) is young ⟶ $A(B)$ is C
  abstraction
- Monika is much younger than Pat ⟶ $(A(B), A(C))$ is R
  Age Monika Age Pat much younger
- distance between New York and Boston is about 200 mi ⟶ $A(B, C)$ is R
- usually Robert returns from work at about 6 pm ⟶

  Prob$\{A$ is $B\}$ is C
  usually
  about 6 pm
  Time(Robert.returns.from.work)

- Carol lives in a small city near San Francisco ⟶

  $A(B(C)$ is $(D$ and $E)$
  small city
  city near SF
  Carol
  Residence

- Location
  most Swedes are tall ⟶ 1/n ΣCount(G[A] is R) is Q
  Height

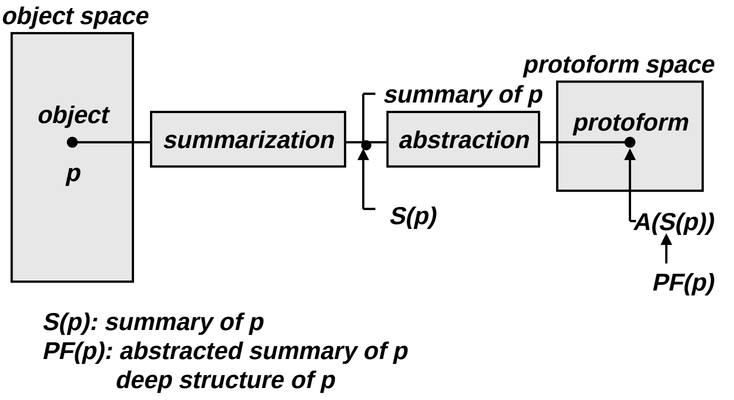


Fig. 19. Definition of protoform of p.

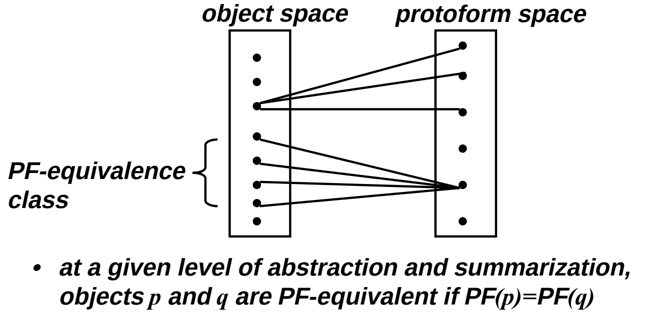
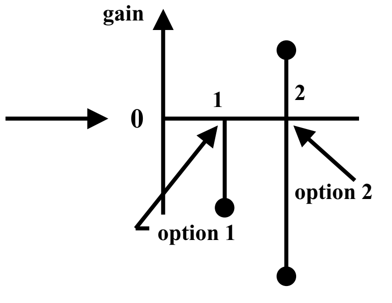


Fig. 20. Protoforms and PF-equivalence.

Alan has severe back pain. He goes to see a doctor. The doctor tells him that there are two options: (1) do nothing; and (2) do surgery. In the case of surgery, there are two possibilities: (a) surgery is successful, in which case Alan will be pain free; and (b) surgery is not successful, in which case Alan will be paralyzed from the neck down.



4.4.1. *Protoformal deduction*

The rules of deduction in CTPM are, basically, the rules which govern constraint propagation. In CTPM, such rules reside in the Deduction Database (DDB), Figure 21. The Deduction Database comprises a collection of agent-controlled modules and submodules, each of which contains rules drawn from various fields and various modalities of generalized constraints (Figure 22). A typical rule has a symbolic part, which is expressed in

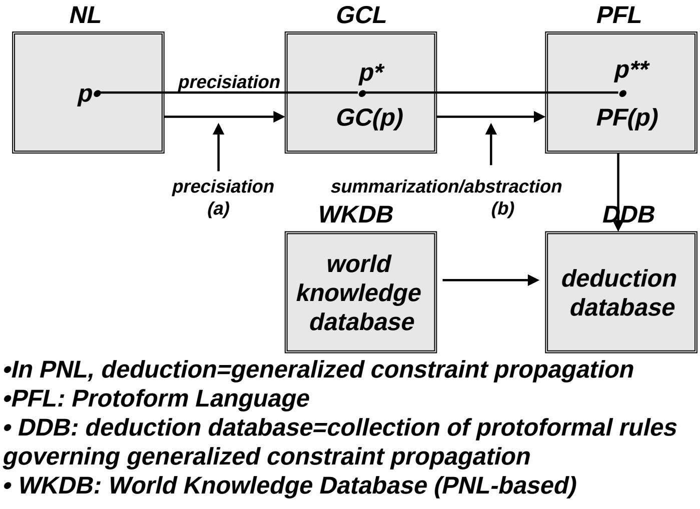


Fig. 21. Basic structure of PNL.

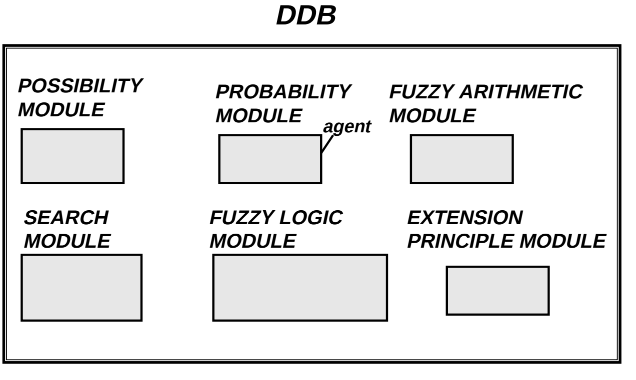


Fig. 22. Structure of deduction database.

terms of protoforms; and a computational part which defines the computation that has to be carried out to arrive at a conclusion. In what follows, we describe briefly some of the basic rules, and list a number of other rules without describing their computational parts. The motivation for doing so is to point to the necessity of developing a set of rules which is much more complete than the few rules which are used as examples in this section.

(a) *Computational rule of inference* [37]

Symbolic part	Computational part
$\begin{array}{l} X \text{ is } A \\ (X, Y) \text{ is } B \\ \hline Y \text{ is } C \end{array}$	$\mu_C(v) = \max_u\big(\mu_A(u) \wedge \mu_B(u, v)\big)$

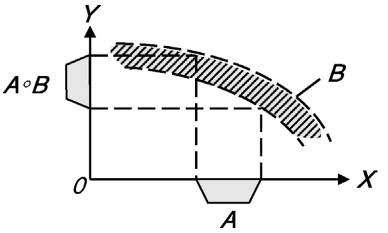


Fig. 23. Compositional rule of inference.

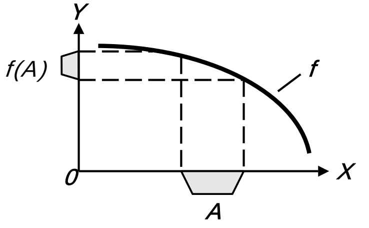


Fig. 24. Basic extension principle.

A, B and C are fuzzy sets with respective membership functions μ_A, μ_B, μ_C. $\wedge$ is min or t-norm (Figure 23).

(b) *Intersection/product syllogism* [32,33]

Symbolic part	Computational part
$\dfrac{\begin{array}{l} Q_1 A\text{'s are } B\text{'s} \\ Q_2(A\&B)\text{'s are } C\text{'s} \end{array}}{Q_3 A\text{'s are } (B\&C)\text{'s}}$	$Q_3 = Q_1 * Q_2$

Q_1 and Q_2 are fuzzy quantifiers; A, B, C are fuzzy sets; $*$ is product in fuzzy arithmetic [13].

(c) *Basic extension principle* [24]

Symbolic part	Computational part
$\dfrac{X \text{ is } A}{f(X) \text{ is } B}$	$\mu_B(v) = \sup_u \big(\mu_A(u)\big)$ subject to $v = f(u)$

g is a given function or functional; A and B are fuzzy sets (Figure 24).

(d) *Extension principle* [38]

This is the principal rule governing possibilistic constraint propagation (Figure 8)

Symbolic part	Computational part
$\dfrac{f(X) \text{ is } A}{g(X) \text{ is } B}$	$\mu_B(v) = \sup_u\big(\mu_B\big(f(u)\big)\big)$ subject to $v = g(u)$

NOTE. The extension principle is a primary deduction rule in the sense that many other deduction rules are derivable from the extension principle. An example is the following rule.

(e) *Basic probability rule*

Symbolic part

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

Computational part

$$\mu_D(v) = \sup_r\left(\mu_B\left(\int_U \mu_A(u)r(u)\,\mathrm{d}u\right)\right)$$

subject to

$$v = \int_U \mu_C(u)r(u)\,\mathrm{d}u$$

$$\int_U r(u)\,\mathrm{d}u = 1.$$

X is a real-valued random variable; A, B, C and D are fuzzy sets: r is the probability density of X; and $U = \{u\}$. To derive this rule, we note that

$$\mathrm{Prob}(X \text{ is } A) \text{ is } B \longrightarrow \int_U r(u)\mu_A(u)\,\mathrm{d}u \text{ is } B$$

$$\mathrm{Prob}(X \text{ is } C) \text{ is } D \longrightarrow \int_U r(u)\mu_C(u)\,\mathrm{d}u \text{ is } D$$

which are generalized constraints of the form

$$f(r) \text{ is } B$$

$$g(r) \text{ is } D.$$

Applying the extension principle to these expressions, we obtain the expression for D which appears in the basic probability rule.

(f) *Bimodal interpolation rule*

The bimodal interpolation rule is a rule which resides in the Probability module of DDB. With reference to Figure 25, the symbolic and computational parts of this rule are:

Symbolic

$$\frac{\mathrm{Prob}(X \text{ is } A_i) \text{ is } P_i}{\mathrm{Prob}(X \text{ is } A) \text{ is } Q}, \quad i = 1, \ldots, n$$

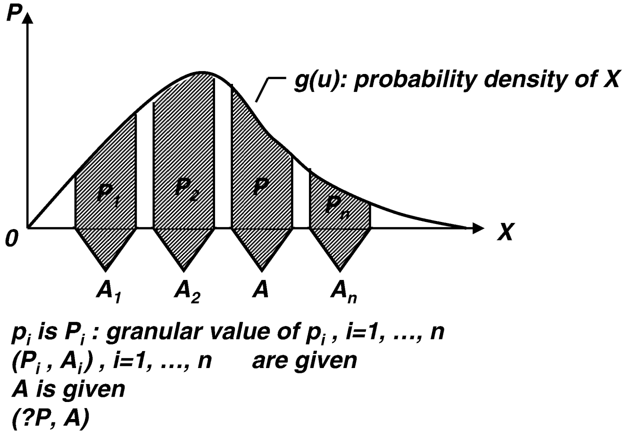


Fig. 25. Interpolation of a bimodal distribution.

Computational

$$\mu_Q(v) = \sup_r \left(\mu_{P_1}\left(\int_U \mu_{A_1}(u) r(u)\,\mathrm{d}u \right) \wedge \cdots \wedge \mu_{P_n}\left(\int_U \mu_{A_n}(u) r(u)\,\mathrm{d}u \right) \right)$$

subject to

$$v = \int_U \mu_A(u) r(u)\,\mathrm{d}u$$

$$\int_U r(u)\,\mathrm{d}u = 1$$

In this rule, X is a real-valued random variable; r is the probability density of X; and U is the domain of X.

NOTE. The probability rule is a special case of the bimodal interpolation rule.

What is the expected value, $E(X)$, of a bimodal distribution? The answer follows through application of the extension principle:

$$\mu_{E(X)}(v) = \sup_r \left(\mu_{P_1}\left(\int_U \mu_{A_1}(u) r(u)\,\mathrm{du} \right) \wedge \cdots \wedge \mu_{P_n}\left(\int_U \mu_{A_n}(u) r(u)\,\mathrm{d}u \right) \right)$$

subject to

$$v = \int_U u r(u)\,\mathrm{d}u$$

$$\int_U r(u)\,\mathrm{d}u = 1$$

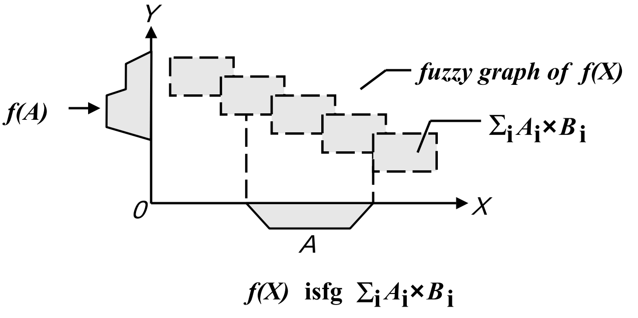


Fig. 26. Fuzzy-graph interpolation.

NOTE. $E(X)$ is a fuzzy subset of U.

(g) *Fuzzy-graph interpolation rule*

This rule is the most widely used rule in applications of fuzzy logic [36]. We have a function, $Y = f(X)$, which is represented as a fuzzy graph (Figure 26). The question is: What is the value of Y when X is A? The A_i, B_i and A are fuzzy sets.

Symbolic part

$$\frac{\begin{array}{l} X \text{ is } A \\ Y = f(X) \\ f(X) \text{ isfg } \sum_i A_i \times B_i \end{array}}{Y \text{ is } C}$$

Computational part

$$C = \sum_i m_i \wedge B_i,$$

where m_i is the degree to which A matches A_i

$$m_i = \sup_u \bigl(\mu_A(u) \wedge \mu_{A_i}(u)\bigr), \quad i = 1, \ldots, n.$$

When A is a singleton, this rule reduces to

$$X = a$$

$$Y = f(X)$$

$$f(X) \text{ isfg } \sum_i A_i \times B_i, \quad i = 1, \ldots, n.$$

$$Y = \sum_i \mu_{A_i}(a) \wedge B.$$

In this form, the fuzzy-graph interpolation rule coincides with the Mamdani rule – a rule which is widely used in control and related applications [15] (Figure 27).

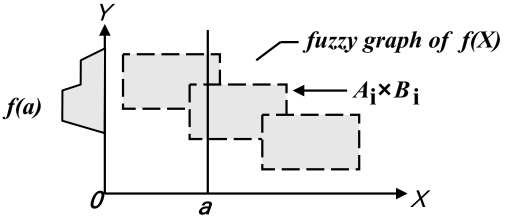


Fig. 27. Mamdani interpolation.

In the foregoing, we have summarized some of the basic rules in DDB which govern generalized constraint propagation. Many more rules will have to be developed and added to DDB. A few examples of such rules are the following.

(a) *Probabilistic extension principle*

$$\frac{f(X) \text{ isp } A}{g(X) \text{ isr } ?B}$$

(b) *Usuality-qualified extension principle*

$$\frac{f(X) \text{ isu } A}{g(X) \text{ isr } ?B}$$

(c) *Usuality-qualified fuzzy-graph interpolation rule*

$$\begin{array}{l} X \text{ is } A \\ Y = f(X) \\ f(X) \text{ isfg } \sum_i \text{ if } X \text{ is } A_i \text{ then } Y \text{ isu } B_i \\ \hline Y \text{ isr } ?B \end{array}$$

(d) *Bimodal extension principle*

$$\begin{array}{l} X \text{ isbm } \sum_i P_i \backslash A_i \\ Y = f(X) \\ \hline Y \text{ isr } ?B \end{array}$$

(e) *Bimodal, binary extension principle*

$$\begin{array}{l} X \text{ isr } R \\ Y \text{ iss } S \\ Z = f(X, Y) \\ \hline Z \text{ ist } T \end{array}$$

In the instance, bimodality means that X and Y have different modalities, and binary means that f is a function of two variables. An interesting special case is one in which X is R and Y isp S.

The deduction rules which were briefly described in the foregoing are intended to serve as examples. How can these rules be applied to reasoning under uncertainty? To illustrate, it will be convenient to return to the examples given in Section 1.

(a) *The Robert example*

p: Usually Robert returns from work at about 6:00 pm. What is the probability that Robert is home at about 6:15 pm?

First, we find the protoforms of the data and the query.

Usually Robert returns from work at about 6:00 pm $\longrightarrow$

$$\longrightarrow \text{Prob}\big(\text{Time(Return(Robert)}\big) \text{ is } {}^{*}6:00 \text{ pm}) \text{ is usually}$$

which in annotated form reads

$$\longrightarrow \text{Prob}\big(X/\text{Time(Return(Robert)}\big) \text{ is } A/{}^{*}6{:}00 \text{ pm}) \text{ is } B/\text{usually}$$

Likewise, for the query, we have

$$\text{Prob}\big(\text{Time(Return(Robert)}\big) \text{ is } \leqslant \circ\, {}^{*}6{:}15 \text{ pm}) \text{ is } ?D$$

which in annotated form reads

$$\longrightarrow \text{Prob}\big(X/\text{Time(Return(Robert)}\big) \text{ is } C/\leqslant \circ\, {}^{*}6{:}15 \text{ pm}) \text{ is } D/\text{usually}$$

Searching the Deduction Database, we find that the basic probability rule matches the protoforms of the data and the query

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

where

$$\mu_D(v) = \sup_g \left(\mu_B\left(\int_U \mu_A(u) g(u)\, \mathrm{d}u\right)\right)$$

subject to

$$v = \int_U \mu_C(u) g(u)\, \mathrm{d}u$$

$$\int_U g(u)\, \mathrm{d}u = 1$$

Instantiating A, B, C and D, we obtain the answer to the query:

Probability that Robert is home at about 6:15 pm is D, where

$$\mu_D(v) = \sup_g \left(\mu_{\text{usually}}\left(\int_U \mu_{{}^{*}6{:}00\text{ pm}}(u) g(u)\, \mathrm{d}u\right)\right)$$

subject to

$$v = \int_U \mu_{\leqslant \circ {}^{*}6{:}15\text{ pm}}(u) g(u)\, \mathrm{d}u$$

and

$$\int_U g(u)\, \mathrm{d}u = 1$$

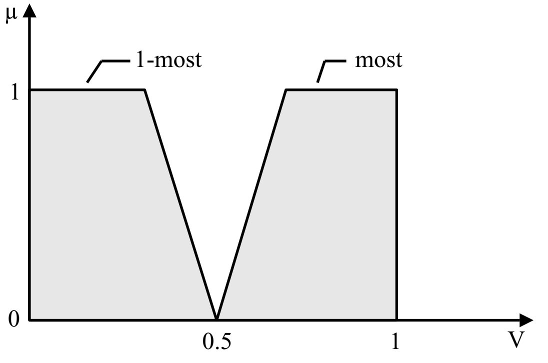


Fig. 28. "Most" and antonym of "most."

(b) *The tall Swedes problem*

We start with the data

p: Most Swedes are tall.

Assume that the queries are:

q_1: How many Swedes are not tall

q_2: How many are short

q_3: What is the average height of Swedes

In our earlier discussion of this example, we found that p translates into a generalized constraint on the count density function, h.

Thus

$$p \longrightarrow \int_a^b h(u)\mu_{\text{tall}}(u)\,\mathrm{d}u \text{ is most}$$

Precisiations of q_1, q_2 and q_3 may be expressed as

$$q_1\colon \longrightarrow \int_a^b h(u)\mu_{\text{not.tall}}(u)\,\mathrm{d}u$$

$$q_2\colon \longrightarrow \int_a^b h(u)\mu_{\text{short}}(u)\,\mathrm{d}u$$

$$q_3\colon \longrightarrow \int_a^b uh(u)\,\mathrm{d}u.$$

Considering q_1, we note that

$$\mu_{\text{not.tall}}(u) = 1 - \mu_{\text{tall}}(u).$$

Consequently

$$q_1 \longrightarrow 1 - \int_a^b h(u)\mu_{\text{tall}}(u)\,\mathrm{d}u$$

which may be rewritten as

$$q_2 \longrightarrow \text{1-most}$$

where 1-most plays the role of the antonym of most (Figure 28).
Considering q_2, we have to compute

$$A\colon \int_a^b h(u)\mu_{\mathrm{short}}(u)\,\mathrm{d}u$$

given that $\int_a^b h(u)\mu_{\mathrm{tall}}(u)\,\mathrm{d}u$ is most.

Applying the extension principle, we arrive at the desired answer to the query:

$$\mu_A(v) = \sup\left(\mu_{\mathrm{most}}\left(\int_a^b h(u)\mu_{\mathrm{tall}}(u)\,\mathrm{d}u\right)\right)$$

subject to

$$v = \int_a^b h(u)\mu_{\mathrm{short}}(u)\,\mathrm{d}u$$

and

$$\int_a^b h(u)\,\mathrm{d}u = 1.$$

Likewise, for q_3 we have as the answer

$$\mu_A(v) = \sup_u\left(\mu_{\mathrm{most}}\left(\int_a^b h(u)\mu_{\mathrm{tall}}(u)\,\mathrm{d}u\right)\right)$$

subject to

$$v = \int_a^b uh(u)\,\mathrm{d}u$$

and

$$\int_a^b h(u)\,\mathrm{d}u = 1.$$

As an illustration of application of protoformal deduction to an instance of this example, consider

p: Most Swedes are tall
q: How many Swedes are short?

We start with the protoforms of p and q (see earlier example):
Most Swedes are tall $\rightarrow 1/n \sum \text{Count}(G[A \text{ is } R])$ is Q
?T Swedes are short $\longrightarrow 1/n \sum \text{Count}(G[A \text{ is } S])$ is T,
where

$$G[A] = \sum_i \text{Name}_i / A_i, \quad i = 1, \ldots, n.$$

An applicable deduction rule in symbolic form is:

$$\frac{1/n \sum \text{Count}(G[A \text{ is } R]) \text{ is } Q}{1/n \sum \text{Count}(G[A \text{ is } S]) \text{ is } T}$$

The computational part of the rule is expressed as

$$\frac{1/n \sum_i \mu_R(A_i) \text{ is } Q}{1/n \sum_i \mu_S(A_i) \text{ is } T}$$

where

$$\mu_T(v) = \sup_{A_i, \ldots, A_n} \mu_Q\left(\sum_i \mu_R(A_i)\right)$$

subject to

$$v = \sum_i \mu_S(A_i).$$

What we see is that computation of the answer to the query, q, reduces to the solution of a variational problem, as it does in the earlier discussion of this example in which protoformal deduction was not employed.

The foregoing examples are merely elementary instances of reasoning through the use of generalized constraint propagation. What should be noted is that the chains of reasoning in these examples are very short. More generally, what is important to recognize is that shortness of chains of reasoning is an intrinsic characteristic of reasoning processes which take place in an environment of substantive imprecision and uncertainty. What this implies is that, in such environments, a conclusion arrived at the end of a long chain of reasoning is likely to be vacuous or of questionable validity.

4.5. *Deduction (extension) principle*

Underlying almost all examples involving computation of an answer to a question, is a basic principle which may be referred to as the Deduction Principle. This principle is closely related to the extension principle of fuzzy logic [24,28].

Assume that we have a database, D, and database variables $X_1, \ldots, X_n$, with u_i being a generic value of X_i, $i = 1, \ldots, n$.

Suppose that q is a given question and that the answer to q, $\text{Ans}(q)$, is a function of the u_i.

$$\text{Ans}(q) = g(u_1, \ldots, u_n), \quad u = (u_1, \ldots, u_n).$$

I do not know the exact values of the u_i. My information about the u_i, $I(u_1, \ldots, u_n)$, is a generalized constraint on the u_i. The constraint is defined by its test-score function $ts(u) = f(u_1, \ldots, u_n)$.

At this point, the problem is that of constraint propagation from $ts(u)$ to $g(u)$. Employing the extension principle, we are led to the membership function of the answer to q. More specifically,

$$\mu_{\mathrm{Ans}(q)}(v) = \sup_u \big(ts(u)\big)$$

subject to

$$v = g(u)$$

This, in brief, is the substance of the Deduction Principle.

As a simple illustration, let us consider an example that was discussed earlier. Suppose that q: What is the average height of Swedes. Assume that D consists of information about the heights of a population of Swedes, $\mathrm{Swede}_1, \ldots, \mathrm{Swede}_n$, with height of ith Swede being h_i, $i = 1, \ldots, n$. Thus, average height may be expressed as

$$\mathrm{Ave}(h) = 1/n(h_1 + \cdots + h_n).$$

Now, I do not know the h_i. What I am given is the datum d: Most Swedes are tall. This datum constrains the h_i. The test-score of this constraint is

$$ts(h) = \mu_{\mathrm{most}}\left(\frac{1}{n}\sum \mu_{\mathrm{tall}}(h_i)\right), \quad h = (h_1, \ldots, h_n).$$

The generalized constraint on the h_i induces a generalized constraint on Ave(h). Thus

$$\mu_{\mathrm{Ave}(h)}(v) = \sup\left(\mu_{\mathrm{most}}\left(\frac{1}{n}\sum_i \mu_{\mathrm{tall}}(h_i)\right)\right), \quad h = (h_1, \ldots, h_n)$$

subject to

$$v = \frac{1}{n}\sum_i h_i.$$

5. Concluding remark

Existing search engines, with Google at the top, have many truly remarkable capabilities. But there is a basic limitation – search engines do not have deduction capability – a capability which a question-answering system is expected to have. Nevertheless, search engines are extremely useful because a skilled human user can get around search engine limitations. In this perspective, a search engine may be viewed as a semi-mechanized question-answering system.

Upgrading a search engine such as Google to a question-answering system is a task whose complexity is hard to exaggerate. To achieve success, new concepts, ideas and tools are needed to address difficult problems which arise when knowledge has to be dealt with in an environment of imprecision, uncertainty and partial truth.

The present paper sketches a conceptual framework which may be viewed as a step in this direction. What lies ahead is the challenging task of applying this framework to web-specific problems.

Acknowledgement

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

References

[1] A.R. Aronson, B.E. Jacobs, J. Minker, *A note on fuzzy deduction*, J. ACM **27** (4) (1980), 599–603.
[2] A. Bardossy, L. Duckstein, *Fuzzy Rule-based Modelling with Application to Geophysical, Biological and Engineering Systems*, CRC Press, 1995.
[3] T. Berners-Lee, J. Hendler, O. Lassila, *The semantic web*, Scientific American **284** (5) (2001), 34–43.
[4] S. Brin, L. Page, *The anatomy of a large-scale hypertextual web search engine*, Computer Networks **30** (1–7) (1998), 107–117.
[5] W.J.H.J. Bronnenberg, M.C. Bunt, S.P.J. Lendsbergen, R.H.J. Scha, W.J. Schoenmakers, E.P.C. van Utteren, *The question answering system PHLIQA1*, in: L. Bola (Ed.), Natural Language Question Answering Systems, Macmillan, 1980.
[6] L.S. Coles, *Techniques for information retrieval using an inferential question-answering system with natural language input*, SRI Report, 1972.
[7] A. Di Nola, S. Sessa, W. Pedrycz, W. Pei-Zhuang, *Fuzzy relation equation under a class of triangular norms: a survey and new results*, in: Fuzzy Sets for Intelligent Systems, Morgan Kaufmann Publishers, San Mateo, CA, 1993, pp. 166–189.
[8] A. Di Nola, S. Sessa, W. Pedrycz, E. Sanchez, *Fuzzy Relation Equations and their Applications to Knowledge Engineering*, Kluwer Academic Publishers, Dordrecht, 1989.
[9] D. Dubois, H. Prade, *Gradual inference rules in approximate reasoning*, Inform. Sci. **61** (1–2) (1992), 103–122.
[10] D. Filev, R.R. Yager, *Essentials of Fuzzy Modeling and Control*, Wiley-Interscience, 1994.
[11] J.A. Goguen, *The logic of inexact concepts*, Synthese **19** (1969), 325–373.
[12] M. Jamshidi, A. Titli, L.A. Zadeh, S. Boverie (Eds.), Applications of Fuzzy Logic – Towards High Machine Intelligence Quotient Systems, Environmental and Intelligent Manufacturing Systems Series, vol. 9, Prentice-Hall, Upper Saddle River, NJ, 1997.
[13] A. Kaufmann, M.M. Gupta, *Introduction to Fuzzy Arithmetic: Theory and Applications*, Van Nostrand, New York, 1985.
[14] D.B. Lenat, *CYC: a large-scale investment in knowledge infrastructure*, Comm. ACM **38** (11) (1995), 32–38.
[15] E.H. Mamdani, S. Assilian, *An experiment in linguistic synthesis with a fuzzy logic controller*, Int. J. Man–Machine Studies **7** (1975), 1–13.
[16] J.R. McSkimin, J. Minker, *The use of a semantic network in a deductive question-answering system*, in: IJCAI, 1977, pp. 50–58.
[17] R.E. Moore, *Interval Analysis*, SIAM Studies in Applied Mathematics, vol. 2, Philadelphia, PA, 1979.
[18] M. Nagao, J. Tsujii, *Mechanism of deduction in a question-answering system with natural language input*, in: ICJAI, 1973, pp. 285–290.
[19] B.H. Partee (Ed.), Montague Grammar, Academic Press, New York, 1976.
[20] W. Pedrycz, F. Gomide, *Introduction to Fuzzy Sets*, MIT Press, Cambridge, MA, 1998.
[21] F. Rossi, P. Codognet (Eds.), Soft Constraints, Special issue on Constraints, vol. 8, N. 1, Kluwer Academic Publishers, 2003.
[22] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, 1976.
[23] M.K. Smith, C. Welty, D. McGuinness (Eds.), OWL Web Ontology Language Guide, W3C Working Draft 31, 2003.
[24] L.A. Zadeh, *Fuzzy sets*, Inform. and Control **8** (1965), 338–353.
[25] L.A. Zadeh, *Probability measures of fuzzy events*, J. Math. Anal. Appl. **23** (1968), 421–427.
[26] L.A. Zadeh, *Outline of a new approach to the analysis of complex systems and decision processes*, IEEE Trans. on Systems Man Cybernet. **3** (1973), 28–44.
[27] L.A. Zadeh, *On the analysis of large scale systems*, in: H. Gottinger (Ed.), Systems Approaches and Environment Problems, Vandenhoeck and Ruprecht, Gottingen, 1974, pp. 23–37.

[28] L.A. Zadeh, *The concept of a linguistic variable and its application to approximate reasoning*, *Part I*, Inform. Sci. **8** (1975), 199–249; *Part II*, Inform. Sci. **8** (1975), 301–357; *Part III*, Inform. Sci. **9** (1975), 43–80.

[29] L.A. Zadeh, *Fuzzy sets and information granularity*, in: M. Gupta, R. Ragade, R. Yager (Eds.), Advances in Fuzzy Set Theory and Applications, North-Holland Publishing Co, Amsterdam, 1979, pp. 3–18.

[30] L.A. Zadeh, *A theory of approximate reasoning*, in: J. Hayes, D. Michie, L.I. Mikulich (Eds.), Machine Intelligence, vol. 9, Halstead Press, New York, 1979, pp. 149–194.

[31] L.A. Zadeh, *Test-score semantics for natural languages and meaning representation via PRUF*, in: B. Rieger (Ed.), Empirical Semantics, Brockmeyer, Bochum, W. Germany, 1982, pp. 281–349. Also Technical Memorandum 246, AI Center, SRI International, Menlo Park, CA, 1981.

[32] L.A. Zadeh, *A computational approach to fuzzy quantifiers in natural languages*, Computers and Mathematics **9** (1983), 149–184.

[33] L.A. Zadeh, *A fuzzy-set-theoretic approach to the compositionality of meaning: propositions, dispositions and canonical forms*, J. Semantics **3** (1983), 253–272.

[34] L.A. Zadeh, *Precisiation of meaning via translation into PRUF*, in: L. Vaina, J. Hintikka (Eds.), Cognitive Constraints on Communication, Reidel, Dordrecht, 1984, pp. 373–402.

[35] L.A. Zadeh, *Outline of a computational approach to meaning and knowledge representation based on a concept of a generalized assignment statement*, in: M. Thoma, A. Wyner (Eds.), Proceedings of the International Seminar on Artificial Intelligence and Man–Machine Systems, Springer-Verlag, Heidelberg, 1986, pp. 198–211.

[36] L.A. Zadeh, *Fuzzy logic and the calculi of fuzzy rules and fuzzy graphs*, Multiple-Valued Logic **1** (1996), 1–38.

[37] L.A. Zadeh, *Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic*, Fuzzy Sets and Systems **90** (1997), 111–127.

[38] L.A. Zadeh, *From computing with numbers to computing with words – from manipulation of measurements to manipulation of perceptions*, IEEE Trans. on Circuits and Systems **45** (1) (1999), 105–119.

[39] L.A. Zadeh, *Toward a perception-based theory of probabilistic reasoning with imprecise probabilities*, J. Statist. Plann. Inference **105** (2002), 233–264.

[40] L.A. Zadeh, *Precisiated natural language (PNL)*, AI Magazine **25** (3) (2004), 74–91.

[41] L.A. Zadeh, *A note on web intelligence, world knowledge and fuzzy logic*, Data and Knowledge Engineering **50** (2004), 291–304.

[42] L.A. Zadeh, *Toward a generalized theory of uncertainty (GTU) – an outline*, Inform. Sci. **172** (2005), 1–40.

Further reading

[43] J. Arjona, R. Corchuelo, J. Pena, D. Ruiz, *Coping with web knowledge*, in: Advances in Web Intelligence, Springer-Verlag, Berlin, 2003, pp. 165–178.

[44] A. Bargiela, W. Pedrycz, *Granular Computing – An Introduction*, Kluwer Academic Publishers, Boston, 2003.

[45] Z. Bubnicki, *Analysis and Decision Making in Uncertain Systems*, Springer-Verlag, 2004.

[46] P.P. Chen, *Entity-relationship Approach to Information Modeling and Analysis*, North-Holland, 1983.

[47] 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.

[48] M.J. Cresswell, *Logic and Languages*, Methuen, London, UK, 1973.

[49] D. Dubois, H. Prade, *On the use of aggregation operations in information fusion processes*, Fuzzy Sets and Systems **142** (1) (2004), 143–161.

[50] T.F. Gamat, *Language, Logic and Linguistics*, University of Chicago Press, 1996.

[51] M. Mares, *Computation over Fuzzy Quantities*, CRC, Boca Raton, FL, 1994.

[52] V. Novak, I. Perfilieva, J. Mockor, *Mathematical Principles of Fuzzy Logic*, Kluwer Academic Publishers, Boston, 1999.

[53] V. Novak, I. Perfilieva (Eds.), Discovering the World with Fuzzy Logic, Studies in Fuzziness and Soft Computing, Physica-Verlag, Heidelberg, 2000.

[54] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht, 1991.

[55] M.K. Smith, C. Welty, *What is ontology*? *Ontology: towards a new synthesis*, in: Proceedings of the Second International Conference on Formal Ontology in Information Systems, 2002.

[56] J.F. Sowa, in: L. Albertazzi (Ed.), Ontological Categories. Shapes of Forms: From Gestalt Psychology and Phenomenology to Ontology and Mathematics, Kluwer Academic Publishers, Dordrecht, 1999, pp. 307–340.
[57] L.A. Zadeh, *A fuzzy-algorithmic approach to the definition of complex or imprecise concepts*, Int. J. Man–Machine Studies **8** (1976), 249–291.
[58] L.A. Zadeh, *Fuzzy logic, neural networks, and soft computing*, Comm. ACM – AI **37** (1994), 77–84.
[59] L.A. Zadeh, *A new direction in AI – toward a computational theory of perceptions*, AI Magazine **22** (1) (2001), 73–84.