FROM COMPUTING WITH NUMBERS TO COMPUTING WITH WORDS — FROM MANIPULATION OF MEASUREMENTS TO MANIPULATION OF PERCEPTIONS*

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@cs.berkeley.edu
http: //http.cs.berkeley.edu/People/Faculty/Homepages/zadeh.html

Computing, in its usual sense, is centered on manipulation of numbers and symbols. In contrast, computing with words, or CW for short, is a methodology in which the objects of computation are words and propositions drawn from a natural language, e.g., *small, large, far, heavy, not very likely, the price of gas is low and declining, Berkeley is near San Francisco, it is very unlikely that there will be a significant increase in the price of oil in the near future*, etc. Computing with words is inspired by the remarkable human capability to perform a wide variety of physical and mental tasks without any measurements and any computations. Familiar examples of such tasks are parking a car, driving in heavy traffic, playing golf, riding a bicycle, understanding speech and summarizing a story. Underlying this remarkable capability is the brain's crucial ability to manipulate perceptions – perceptions of distance, size, weight, color, speed, time, direction, force, number, truth, likelihood and other characteristics of physical and mental objects. Manipulation of perceptions plays a key role in human recognition, decision and execution processes. As a methodology, computing with words provides a foundation for a computational theory of perceptions – a theory which may have an important bearing on how humans make – and machines might make – perception-based rational decisions in an environment of imprecision, uncertainty and partial truth.

A basic difference between perceptions and measurements is that, in general, measurements are crisp whereas perceptions are fuzzy. One of the fundamental aims of science has been and continues to be that of progressing from perceptions to measurements. Pursuit of this aim has led to brilliant successes. We have sent men to the moon; we can build computers that are capable of performing billions of computations per second; we have constructed telescopes that can explore the far reaches of the universe; and we can date the age of rocks that are millions of years old. But alongside the brilliant successes stand conspicuous underachievements and outright failures. We cannot build robots which can move with the agility of animals or humans; we cannot automate driving in heavy traffic; we cannot translate from one language to another at the level of a human interpreter; we cannot create programs which can summarize non-trivial stories; our ability to model the behavior of economic systems leaves much to be desired; and we cannot build machines that can compete with children in the performance of a wide variety of physical and cognitive tasks.

It may be argued that underlying the underachievements and failures is the unavailability of a methodology for reasoning and computing with perceptions rather than measurements. An outline of such a methodology – referred to as a computational theory of perceptions – is presented in this paper. The computational theory of perceptions, or CTP for short, is based on the methodology of computing with words (CW). In CTP, words play the role of labels of perceptions and, more generally, perceptions are expressed as propositions in a natural language. CW-based techniques are employed to translate propositions expressed in a natural language into what is called the Generalized Constraint Language (GCL). In this language, the meaning of a proposition is expressed as a generalized constraint, X isr R, where X is the constrained variable, R is the constraining relation and isr is a variable copula in which r is a variable whose value defines the way in which R constrains X. Among the basic types of constraints are: possibilistic, veristic, probabilistic, random set, Pawlak set, fuzzy graph and usuality. The wide variety of constraints in GCL makes GCL a much more expressive language than the language of predicate logic.

In CW, the initial and terminal data sets, IDS and TDS, are assumed to consist of propositions expressed in a natural language. These propositions are translated, respectively, into antecedent and consequent constraints. Consequent constraints are derived from antecedent constraints through the use of rules of constraint propagation. The principal constraint propagation rule is the generalized extension principle. The derived constraints are retranslated into a natural language, yielding the terminal data set (TDS). The rules of constraint propagation in CW coincide with the rules of inference in fuzzy logic. A basic problem in CW is that of explicitation of X, R and r in a generalized constraint, X isr R, which represents the meaning of a proposition, p, in a natural language.

There are two major imperatives for computing with words. First, computing with words is a necessity when the available information is too imprecise to justify the use of numbers; and second, when there is a tolerance for imprecision which can be exploited to achieve tractability, robustness, low solution cost and better rapport with reality. Exploitation of the tolerance for imprecision is an issue of central importance in CW and CTP. At this juncture, the computational theory of perceptions – which is based on CW – is in its initial stages of development. In time, it may come to play an important role in the conception, design and utilization of information/intelligent systems. The role model for CW and CTP is the human mind.

1. Introduction

In the fifties, and especially late fifties, circuit theory was at the height of importance and visibility. It played a pivotal role in the conception and design of electronic circuits and was enriched by basic contributions of Darlington, Bode, McMillan, Guillemin, Carlin, Youla, Kuh, Desoer, Sandberg and other pioneers.

However, what could be discerned at that time was that circuit theory was evolving into a more general theory – system theory – a theory in which the physical identity of the elements of a system is subordinated to a mathematical characterication of their input/output relations. This evolution was a step in the direction of greater generality and, like most generalizations, it was driven by a quest for models which make it possible to reduce the distance between an object that is modeled – the modelizand – and its model in a specified class of systems.

In a paper published in 1961 entitled “From Circuit Theory to System Theory,” (Zadeh, 1961) I discussed the evolution of circuit theory into system theory and observed that the high effectiveness of system theory in dealing with mechanistic systems stood in sharp contrast to its low effectiveness in the realm of humanistic systems – systems exemplified by economic systems, biological systems, social systems, political systems and, more generally, manmachine systems of various types. In more specific terms, I wrote:

> There is a fairly wide gap between what might be regarded as “animate” system theorists and ‘inanimate’ system theorists at the present time, and it is not at all certain that this gap will be narrowed, much less closed, in the near future. There are some who feel that this gap reflects the fundamental inadequacy of conventional mathematics – the mathematics of precisely-defined points, functions, sets, probability measures, etc. – for coping with the analysis of biological systems, and that to deal effectively with such systems, which are generally orders of magnitude more complex than man-made systems, we need a radically different kind of mathematics, the mathematics of fuzzy or cloudy quantities which are not describable in terms of probability distributions. Indeed, the need for such mathematicsis becoming increasingly apparent even in the realm of inanimate systems, for in most practical cases the *a priori* data as well as the criteria by which the performance of a man-made system are judged are far from being precisely specified or having accuratelyknown probability distributions.

It was this observation that motivated my development of the theory of fuzzy sets, starting with the 1965 paper “Fuzzy Sets” (Zadeh, 1965), which was published in *Information and Control*.

Subsequently, in a paper published in 1973, “Outline of a New Approach to the Analysis of Complex Systems and Decision Processes,” (Zadeh, 1973) I introduced the concept of a linguistic variable, that is, a variable whose values are words rather than numbers. The concept of a linguistic variable has played and is continuing to play a pivotal role in the development of fuzzy logic and its applications.

The initial reception of the concept of a linguistic variable was far from positive, largely because my advocacy of the use of words in systems and decision analysis clashed with the deep-seated tradition of respect for numbers and disrespect for words. The essence of this tradition was succinctly stated in 1883 by Lord Kelvin:

> In physical science the first essential step in the direction of learning any subject is to find principles of numerical reckoning and practicable methods for measuring some quality connected with it. I often say that when you can measure what you are speaking about and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a

> meagre and unsatisfactory kind: it may be the beginning of knowledge but you have scarcely, in your thoughts, advanced to the state of science, whatever the matter may be.

The depth of scientific tradition of respect for numbers and derision for words was reflected in the intensity of hostile reaction to my ideas by some of the prominent members of the scientific elite. In commenting on my first exposition of the concept of a linguistic variable in 1972, Rudolph Kalman had this to say:

> I would like to comment briefly on Professor Zadeh's presentation. His proposals could be severely, ferociously, even brutally criticized from a technical point of view. This would be out of place here. But a blunt question remains: Is Professor Zadeh presenting important ideas or is he indulging in wishful thinking? No doubt Professor Zadeh's enthusiasm for fuzziness has been reinforced by the prevailing climate in the U.S. one of unprecedented permissiveness. 'Fuzzification' is a kind of scientific permissiveness; it tends to result in socially appealing slogans unaccompanied by the discipline of hard scientific work and patient observation.

In a similar vein, my esteemed colleague Professor William Kahan – a man with a brilliant mind – offered this assessment in 1975:

> "Fuzzy theory is wrong, wrong, and pernicious." says William Kahan, a professor of computer sciences and mathematics at Cal whose Evans Hall office is a few doors from Zadeh's. "I can not think of any problem that could not be solved better by ordinary logic." What Zadeh is saying is the same sort of things 'Technology got us into this mess and now it can't get us out.' Well, technology did not get us into this mess. Greed and weakness and ambivalence got us into this mess. What we need is more logical thinking, not less. The danger of fuzzy theory is that it will encourage the sort of imprecise thinking that has brought us so much trouble."

What Lord Kelvin, Rudolph Kalman, William Kahan and many other brilliant minds did not appreciate is the fundamental importance of the remarkable human capability to perform a wide variety of physical and mental tasks without any measurements and any computations. Familiar examples of such tasks are parking a car; driving in heavy traffic; playing golf; understanding speech and summarizing a story.

Underlying this remarkable ability is the brain's crucial ability to manipulate perceptions – perceptions of size, distance, weight, speed, time, direction, smell, color, shape, force, likelihood, truth and intent, among others. A fundamental difference between measurements and perceptions is that, in general, measurements are crisp numbers whereas perceptions are fuzzy numbers or, more generally, fuzzy granules, that is, clumps of objects in which the transition from membership to nonmembership is gradual rather than abrupt.

The fuzziness of perceptions reflects finite ability of sensory organs and the brain to resolve detail and store information. A concomitant of fuzziness of perceptions is the preponderant partiality of human concepts in the sense that the validity of most human concepts is a matter of degree. For example, we have partial knowledge, partial understanding, partial certainty, partial belief and accept partial solutions, partial truth and partial causality. Furthermore, most human concepts have a granular structure and are context-dependent.

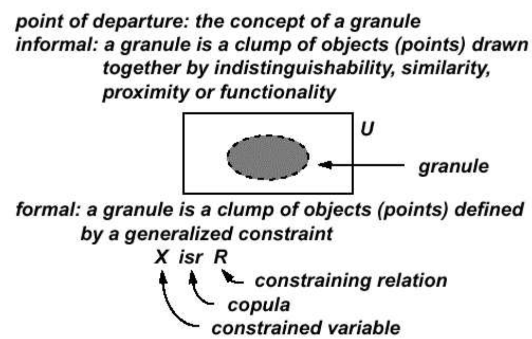

Fig. 1. Informal and formal definitions of a granule.

In essence, a granule is a clump of physical or mental objects (points) drawn together by indistinguishability, similarity, proximity or functionality (Fig. 1). A granule may be crisp or fuzzy, depending on whether its boundaries are or are not sharply defined. For example, age may be granulated crisply into years and granulated fuzzily into fuzzy intervals labeled very young, young, middle-aged, old and very old (Fig. 2). A partial taxonomy of granulation is shown in Figs. 3(a) and 3(b).

Fig. 2. Examples of crisp and fuzzy granulation.

In a very broad sense, granulation involves a partitioning of whole into parts. Modes of information gran-

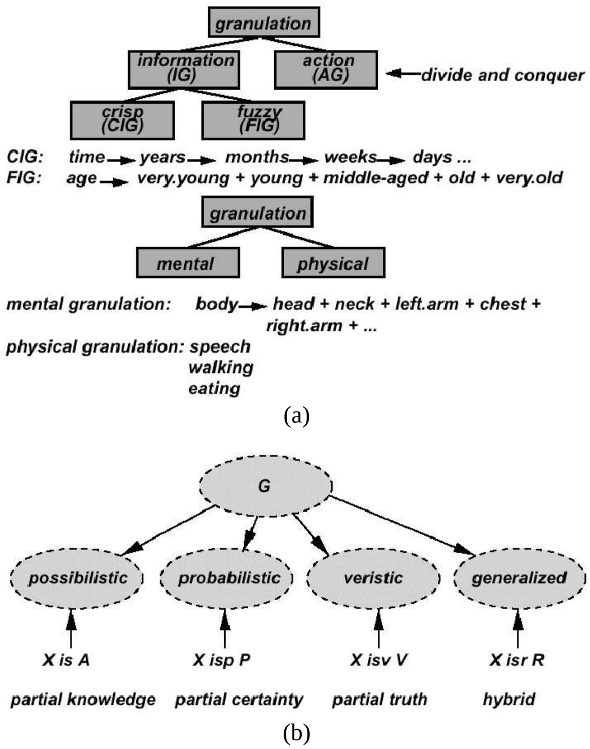

Fig. 3. (a) Partial taxonomy of granulation; (b) Principal types of granules.

ulation (IG) in which granules are crisp play important roles in a wide variety of methods, approaches and techniques. Among them are: interval analysis, quantization, chunking, rough set theory, diakoptics, divide and conquer, Dempster-Shafer theory, machine learning from examples, qualitative process theory, decision trees, semantic networks, analog-to-digital conversion, constraint programming, image segmentation, cluster analysis and many others.

Important though it is, crisp IG has a major blind spot. More specifically, it fails to reflect the fact that most human perceptions are fuzzy rather than crisp. For example, when we mentally granulate the human body into fuzzy granules labeled head, neck, chest, arms, legs, etc., the length of neck is a fuzzy attribute whose value is a fuzzy number. Fuzziness of granules, their attributes and their values is characteristic of ways in which human concepts are formed, organized and manipulated. In effect, fuzzy information granulation (fuzzy IG) may be viewed as a human way of employing data compression for reasoning and, more particularly, making rational decisions in an environment of imprecision, uncertainty and partial truth.

The tradition of pursuit of crispness and precision in scientific theories can be credited with brilliant successes. We have sent men to the moon; we can build computers that are capable of performing billions of computations per second; we have constructed telescopes that can explore the far reaches of the universe; and we can date the age of rocks that are millions of years old. But alongside the brilliant successes stand conspicuous underachievements and outright failures. We cannot build robots which can move with the agility of animals or humans; we cannot automate driving in heavy traffic; we cannot translate from one language to another at the level of a human interpreter; we cannot create programs which can summarize non-trivial stories; our ability to model the behavior of economic systems leaves much to be desired; and we cannot build machines that can compete with children in the performance of a wide variety of physical and cognitive tasks.

What is the explanation for the disparity between the successes and failures? What can be done to advance the frontiers of science and technology beyond where they are today, especially in the realms of machine intelligence and automation of decision processes? In my view, the failures are conspicuous in those areas in which the objects of manipulation are, in the main, perceptions rather than measurements. Thus, what we need are ways of dealing with perceptions, in addition to the many tools which we have for dealing with measurements. In essence, it is this need that motivated the development of the methodology of computing with words (CW) – a methodology in which words play the role of labels of perceptions.

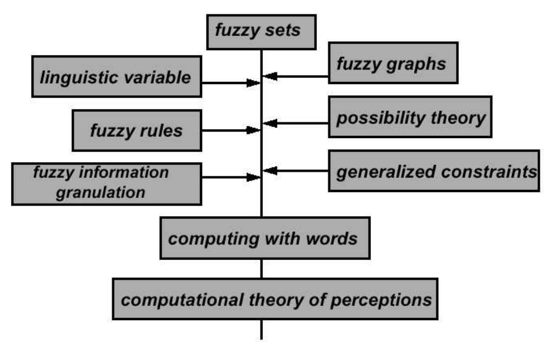

Fig. 4. Conceptual structure of computational theory of perceptions.

Computing with words provides a methodology for what may be called a *computational theory of perceptions* (CTP) (Fig. 4). However, the potential impact of the methodology of computing with words is much broader. Basically, there are four principal rationales for the use of CW:

1) *The don't know rationale*. In this case, the values of variables and/or parameters are not known with sufficient precision to justify the use of conventional methods of numerical computing. An example is decision-making with poorly defined probabilities and utilities.

2) *The don't need rationale*. In this case, there is a tolerance for imprecision which can be exploited to achieve tractability, robustness, low solution cost and better rapport with reality. An example is the problem of parking a car.

3) *The can't solve rationale*. In this case, the problem cannot be solved through the use of numerical computing. An example is the problem of automation of driving in city traffic.

4) *The can't define rationale*. In this case, a concept that we wish to define is too complex to admit of definition in terms of a set of numerical criteria. A case in point is concept of causality. Causality is an instance of what may be called an amorphic concept.

The basic idea underlying the relationship between CW and CTP is conceptually simple. More specifically, in CTP perceptions and queries are expressed as propositions in a natural language. Then, propositions and queries are processed by CW-based methods to yield answers to queries. Simple examples of linguistic characterization of perceptions drawn from everyday experiences are:

Robert is highly intelligent
Carol is very attractive
Hans loves wine
Overeating causes obesity
Most Swedes are tall
Berkeley is more lively than Palo Alto
It is likely to rain tomorrow
It is very unlikely that there will be a significant increase in the price of oil in the near future

Examples of correct conclusions drawn from perceptions through the use of CW-based methods are shown in Fig. 5(a). Examples of incorrect conclusions are shown in Fig. 5(b).

Perceptions have long been an object of study in psychology. However, the idea of linking perceptions to computing with words is in a different spirit. An interesting system-theoretic approach to perceptions is described in a recent work of R. Vallée (1995). A logic of perceptions has been described by H. Rasiowa (1989). These approaches are not related to the approach described in our paper.

An important point that should be noted is that classical logical systems such as propositional logic, predical logic and modal logic, as well as AI-based techniques for natural language processing and knowledge representation, are concerned in a fundamental way with propositions expressed in a natural language. The main difference between such approaches and CW is that the methodology of CW – which is based on fuzzy logic – provides a much more expressive language for knowledge representation and much more versatile machinery for reasoning and computation.

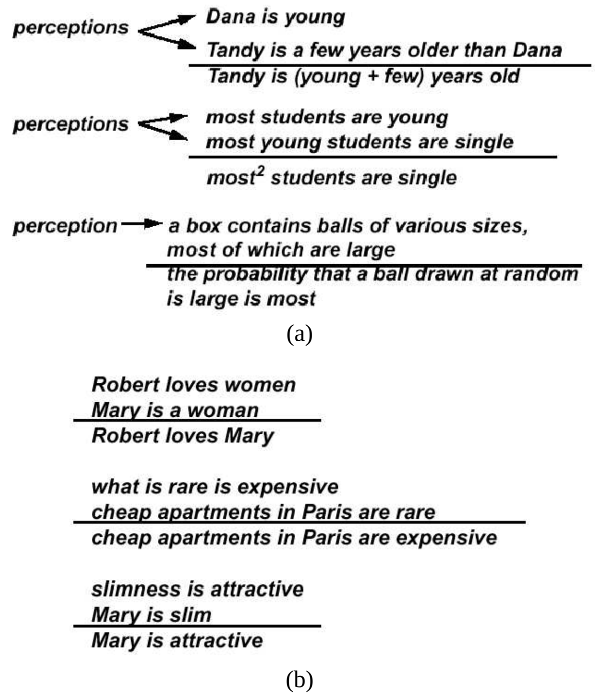

Fig. 5. (a) Examples of reasoning with perceptions; (b) Examples of incorrect reasoning.

In the final analysis, the role model for computing with words is the human mind and its remarkable ability to manipulate both measurements and perceptions. What should be stressed, however, is that although words are less precise than numbers, the methodology of computing with words rests on a mathematical foundation. An exposition of the basic concepts and techniques of computing with words is presented in the following sections. The linkage of CW and CTP is discussed very briefly because the computational theory of perceptions is still in its early stages of development.

2. What is CW?

In its traditional sense, computing involves for the most part manipulation of numbers and symbols. By contrast, humans employ mostly words in computing and reasoning, arriving at conclusions expressed as words from premises expressed in a natural language or having the form of mental perceptions. As used by humans, words have fuzzy denotations. The same applies to the role played by words in CW.

The concept of CW is rooted in several papers starting with my 1973 paper "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes," (Zadeh, 1973) in which the concepts of a linguistic variable and granulation were introduced. The concepts of a fuzzy constraint and fuzzy constraint propagation were introduced in "Calculus of Fuzzy Restrictions," (Zadeh, 1975a), and developed more fully in "A Theory of Approximate Reasoning," (Zadeh, 1979b) and "Outline of a Computational Approach to Meaning and Knowledge Representation Based on a Concept of a Generalized Assignment Statement," (Zadeh, 1986). Application of fuzzy logic to meaning representation and its role in testscore semantics are discussed in "PRUF – A Meaning Representation Language for Natural Languages," (Zadeh, 1978b), and "Test-Score Semantics for Natural Languages and MeaningRepresentation via PRUF," (Zadeh, 1981). The close relationship between CW and fuzzy information granulation is discussed in "Toward a Theory of Fuzzy Information Granulation and its Centrality in Human Reasoning and Fuzzy Logic (Zadeh, 1997)."

Although the foundations of computing with words were laid some time ago, its evolution into a distinct methodology in its own right reflects many advances in our understanding of fuzzy logic and soft computing – advances which took place within the past few years. (See References and Related Papers.) A key aspect of CW is that it involves a fusion of natural languages and computation with fuzzy variables. It is this fusion that is likely to result in an evolution of CW into a basic methodology in its own right, with wide-ranging ramifications and applications.

We begin our exposition of CW with a few definitions. It should be understood that the definitions are dispositional, that is, admit of exceptions.

As was stated earlier, a concept which plays a pivotal role in CW is that of a granule. Typically, a granule is a fuzzy set of points drawn together by similarity (Fig. 1). A word may be atomic, as in *young*, or composite, as in *not very young* (Fig. 6). Unless stated to the contrary, a word will be assumed to be composite. The denotation of a word may be a higher order predicate, as in Montague grammar (Hobbs, 1978; Partee, 1976).

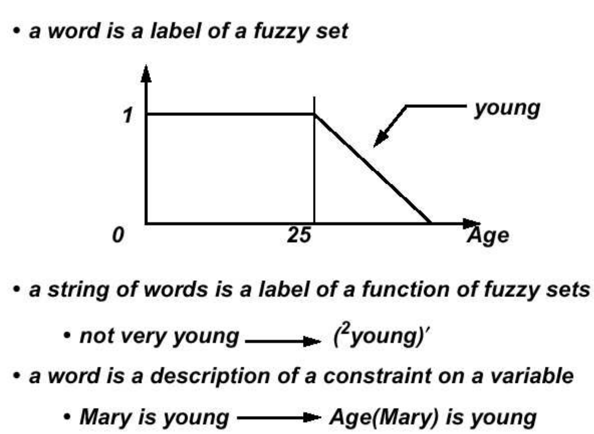

Fig. 6. Words as labels of fuzzy sets.

In CW, a granule, g, which is the denotation of a word, w, is viewed as a fuzzy constraint on a variable. A pivotal role in CW is played by fuzzy constraint propagation from premises to conclusions. It should be noted that, as a basic technique, constraint propagation plays important roles in many methodologies, especially in mathematical programming, constraint programming and logic programming. (See References and Related Papers.)

As a simple illustration, consider the proposition *Mary is young*, which may be a linguistic characterization of a perception. In this case, young is the label of a granule young. (Note that for simplicity the same symbol is used both for a word and its denotation.) The fuzzy set *young* plays the role of a fuzzy constraint on the age of Mary (Fig. 6).

As a further example consider the propositions

$$p_1 = \textit{Carol lives near Mary}$$

and

$$p_2 = \textit{Mary lives near Pat.}$$

In this case, the words *lives near* in p_1 and p_2 play the role of fuzzy constraints on the distances between the residences of Carol and Mary, and Mary and Pat, respectively. If the query is: How far is Carol from Pat?, an answer yielded by fuzzy constraint propagation might be expressed as p_3, where

$$p_3 = \textit{Carol lives not far from Pat.}$$

More about fuzzy constraint propagation will be said at a later point.

A basic assumption in CW is that information is conveyed by constraining the values of variables. Furthermore, information is assumed to consist of a collection of propositions expressed in natural or synthetic language. Typically, such propositions play the role of linguistic characterization of perceptions.

A basic generic problem in CW is the following.

We are given a collection of propositions expressed in a natural language which constitute the *initial data set*, or IDS for short.

From the initial data set we wish to infer an answer to a query expressed in a natural language. The answer, also expressed in a natural language, is referred to as the *terminal data set*, or TDS for short. The problem is to derive TDS from IDS (Fig. 7).

A few problems will serve to illustrate these concepts. At this juncture, the problems will be formulated by not solved.

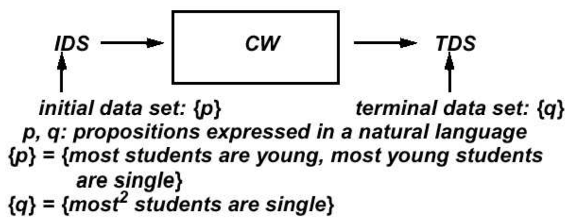

Fig. 7. Computing with words as a transformation of an initial data set (IDS) into a terminal data set (TDS).

1) Assume that a function f, $f : U \to V$, $X \in U$, $Y \in V$, is described in words by the fuzzy if-then rules

f : if X is *small* then Y is *small*

if X is *medium* then Y is *large*

if X is *large* then Y is *small*

What this implies is that f is approximated to by a fuzzy graph f^* (Fig. 8), where

$$f^* = \textit{small} \times \textit{small} + \textit{medium} \times \textit{large} + \textit{large} \times \textit{small}$$

In f^*, $+$ and $\times$ denote respectively, the disjunction and cartesian product. An expression of the form $A \times B$, where A and B are words, will be referred to as a *Cartesian product, Cartesian granule*. In this sense, a fuzzy graph may be viewed as a disjunction of cartesian granules. In essence, a fuzzy graph serves as an approximation to a function or a relation (Zadeh, 1974; 1996a). Equivalently, it may be viewed as a linguistic characterization of a perception of f (Fig. 9).

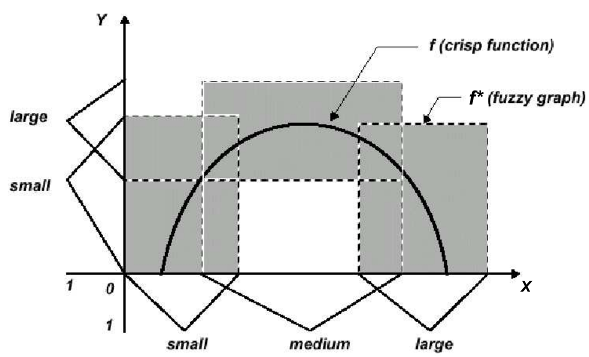

Fig. 8. Fuzzy graph of a function.

In the example under consideration, the IDS consists of the fuzzy rule-set f. The query is: What is the maximum value of f (Fig. 10)? More broadly, the problem is: How can one compute an attribute of a function, f, e.g., its maximum value or its area or its roots if is described in words as a collection of fuzzy if-then rules? Determination of the maximum value will be discussed in greater detail at a later point.

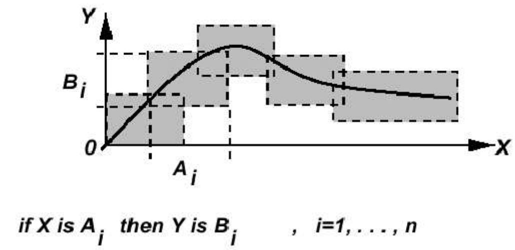

Fig. 9. A fuzzy graph of a function represented by a rule-set.

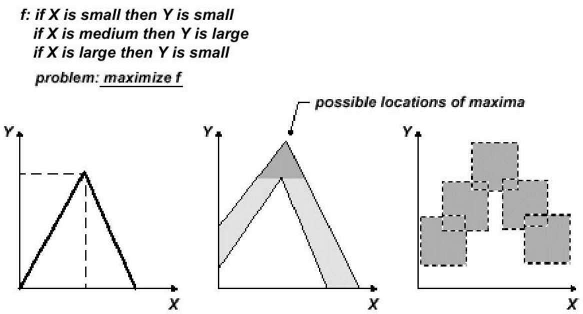

Fig. 10. Fuzzy graph of a function defined by a fuzzy rule-set.

2) A box contains ten balls of various sizes of which several are large and a few are small. What is the probability that a ball drawn at random is neither large nor small? In this case, the IDS is a verbal description of the contents of the box; the TDS is the desired probability.

3) A less simple example of computing with words is the following.

Let X and Y be independent random variables taking values in a finite set $V = \{v_1, \dots, v_n\}$ with probabilities $p_1, \dots, p_n$ and $q_1, \dots, q_n$, respectively. For simplicity of notation, the same symbols will be used to denote X and Y and their generic values, with p and q denoting the probabilities of X and Y, respectively.

Assume that the probability distributions of X and Y are described in words through the fuzzy if-then rules (Fig. 11):

P : if X is *small* then p is *small*

if X is *medium* then p is *large*

if X is *large* then p is *small*

and

Q : if Y is *small* then q is *large*

if Y is *medium* then q is *small*

if Y is *large* then q is *large*

where the granules *small*, *medium* and large are values of linguistic variables X and Y in their respective universes of discourse. In the example under consideration, these rule-sets constitute the IDS. Note that *small* in P need not have the same meaning as *small* in Q, and likewise for *medium* and *large*.

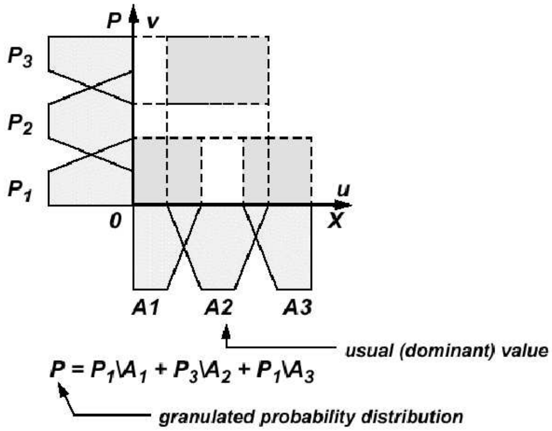

Fig. 11. A fuzzy graph representation of a granulated probability distribution.

The query is: How can we describe in words the joint probability distribution of X and Y? This probability distribution is the TDS.

For convenience, the probability distributions of X and Y may be represented as fuzzy graphs:

$$P: \quad \textit{small} \times \textit{small} + \textit{medium} \times \textit{large} + \textit{large} \times \textit{small}$$

$$Q: \quad \textit{small} \times \textit{large} + \textit{medium} \times \textit{small} + \textit{large} \times \textit{large}$$

with the understanding that the underlying numerical probabilities must add up to unity.

Since X and Y are independent random variables, their joint probability distribution (P, Q) is the product of P and Q. In words, the product may be expressed as (Zadeh, 1996a):

$$\begin{aligned}(P,Q): \quad & \textit{small} \times \textit{small} \times (\textit{small} * \textit{large}) \\ & + \textit{small} \times \textit{medium} \times (\textit{small} * \textit{small}) \\ & + \textit{small} \times \textit{large} \times (\textit{small} * \textit{large}) \\ & + \cdots + \textit{large} \times \textit{large} \times (\textit{small} * \textit{large}),\end{aligned}$$

where $*$ is the arithmetic product in fuzzy arithmetic (Kaufmann and Gupta, 1985). In this example, what we have done, in effect, amounts to a derivation of a linguistic characterization of the joint probability distribution of X and Y starting with linguistic characterizations of the probability distribution of X and the probability distribution of Y.

A few comments are in order. In linguistic characterizations of variables and their dependencies, words serve as values of variables and play the role of fuzzy constraints. In this perspective, the use of words may be viewed as a form of granulation, which in turn may be regarded as a form of fuzzy quantization.

Granulation plays a key role in human cognition. For humans, it serves as a way of achieving data compression. This is one of the pivotal advantages accruing through the use of words in human, machine and man-machine communication.

The point of departure in CW is the premise that the meaning of a proposition, p, in a natural language may be represented as an implicit constraint on an implicit variable. Such a representation is referred to as a *canonical form* of p, denoted as $CF(p)$ (Fig. 12). Thus, a canonical form serves to make explicit the implicit constraint which resides in p. The concept of a canonical form is described in greater detail in the following section.

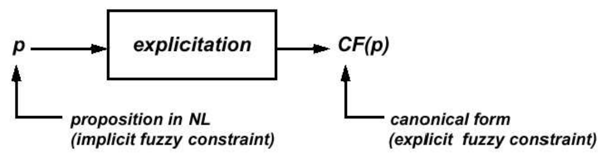

Fig. 12. Canonical form of a proposition.

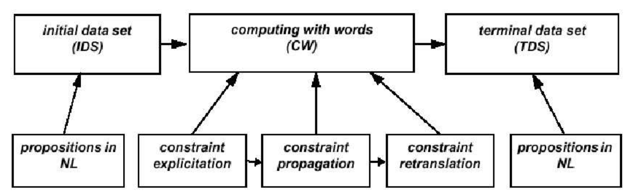

Fig. 13. Conceptual structure of computing with words.

As a first step in the derivation of TDS from IDS, propositions in IDS are translated into their canonical forms, which collectively represent *antecedent* constraints. Through the use of rules for constraint propagation, antecedent constraints are transformed into *consequent* constraints. Finally, consequent constraints are translated into a natural language through the use of *linguistic approximation* (Freuder and Snow, 1990; Mamdani and Gaines, 1981), yielding the terminal data set TDS. This process is schematized in Fig. 13.

In essence, the rationale for computing with words rests on two major imperatives. First, computing with words is a necessity when the available information is too imprecise to justify the use of numbers. And second, when there is a tolerance for imprecision which can be exploited to achieve tractability, robustness, low solution cost and better rapport with reality.

In computing with words, there are two core issues that arise. First is the issue of representation of fuzzy constraints. More specifically, the question is: How can the fuzzy constraints which are implicit in propositions expressed in a natural language be made explicit. And second is the issue of fuzzy constraint propagation, that is, the question of how can fuzzy constraints in premises, i.e., antecedent constraints, be propagated to conclusions, i.e., consequent constraints.

These are the issues which are addressed in the following.

3. Representation of Fuzzy Constraints, Canonical Forms and Generalized Constraints

Our approach to the representation of fuzzy constraints is based on test-score semantics (Zadeh, 1981; 1982). In outline, in this semantics, a proposition, p, in a natural language is viewed as a network of fuzzy (elastic) constraints. Upon aggregation, the constraints which are embodied in p result in an overall fuzzy constraint which can be represented as an expression of the form

$$X \text{ is } R$$

where R is a constraining fuzzy relation and X is the constrained variable. The expression in question is the canonical form of p. Basically, the function of a canonical form is to place in evidence the fuzzy constraint which is implicit in p. This is represented schematically as

$$P \rightarrow X \text{ is } R$$

in which the arrow $\rightarrow$ denotes explicitation. The variable X may be vector-valued and/or conditioned.

In this perspective, the meaning of p is defined by two procedures. The first procedure acts on a so-called explanatory database, ED, and returns the constrained variable, X. The second procedure acts on ED and returns the constraining relation, R.

An explanatory database is a collection of relations in terms of which the meaning of p is defined. The relations are empty, that is, they consist of relation names, relations attributes and attribute domains, with no entries in the relations. When there are entries in ED, ED is said to be *instantiated* and is denoted EDI. EDI may be viewed as a description of a possible world in possible world semantics (Cresswell, 1973), while ED defines a collection of possible worlds, with each possible world in the collection corresponding to a particular instantiation of ED (Zadeh, 1982).

As a simple illustration, consider the proposition

$$p = \textit{Mary is not young}.$$

Assume that the explanatory database is chosen to be

$$\text{ED} = \text{POPULATION [Name; Age]} + \text{ YOUNG [Age; } \mu]$$

in which POPULATION is a relation with arguments Name and Age; YOUNG is a relation with arguments Age and μ; and $+$ is the disjunction. In this case, the constrained variable is the age of Mary, which in terms of ED may be expressed as

$$X = \text{Age (Mary)} = {}_{\text{Age}}\text{POPULATION [Name} = \text{Mary]}.$$

This expression specifies the procedure which acts on ED and returns X. More specifically, in this procedure, Name is instantiated to Mary and the resulting relation is projected on Age, yielding the age of Mary. The constraining relation, R, is given by

$$R = ({}^2\text{YOUNG})'$$

which implies that the intensifier *very* is interpreted as a squaring operation, and the negation *not* as the operation of complementation (Zadeh, 1972).

Equivalently, R may be expressed as

$$R = \text{YOUNG [Age; } 1 - \mu^2].$$

As a further example, consider the proposition

$$p = \textit{Carol lives in a small city near San Francisco}$$

and assume that the explanatory database is:

$$\begin{aligned} ED = {} & \text{POPULATION [Name; Residence]} \\ & + \text{SMALL [City; } \mu] + \text{NEAR [City1; City2; } \mu] \end{aligned}$$

In this case,

$$\begin{aligned} X &= \text{Residence (Carol)} \\ &= {}_{\text{Residence}}\text{POPULATION [Name} = \text{Carol]} \end{aligned}$$

and

$$\begin{aligned} R = {} & \textit{SMALL } [\textit{City}, \mu] \\ & \cap_{\text{City1}} \text{NEAR } [\textit{City2} = \text{San\_Francisco}] \end{aligned}$$

In R, the first constituent is the fuzzy set of small cities; the second constituent is the fuzzy set of cities which are near San Francisco; and $\cap$ denotes the intersection of these sets.

So far we have confined our attention to constraints of the form

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

In fact, constraints can have a variety of forms. In particular, a constraint – expressed as a canonical form – may be conditional, that is, of the form

$$\text{if } X \text{ is } R \text{ then } Y \text{ is } S$$

which may also be written as

$$Y \text{ is } S \text{ if } X \text{ is } R.$$

The constraints in question will be referred to as *basic*.

For purposes of meaning representation, the richness of natural languages necessitates a wide variety of constraints in relation to which the basic constraints form an important though special class. The so-called generalized constraints (Zadeh, 1986) contain the basic constraints as a special case and are defined as follows. The need for generalized constraints becomes obvious when one attempts to represent the meaning of simple propositions such as

Robert loves women
John is very honest
checkout time is 11 am
slimness is attractive

in the language of standard logical systems.

A generalized constraint is represented as

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

where isr, pronounced "ezar", is a variable copula which defines the way in which R constrains X. More specifically, the role of R in relation to X is defined by the value of the discrete variable r. The values of r and their interpretations are defined below:

e : equal (abbreviated to $=$);
d : disjunctive (possibilistic) (abbreviated to blank);
ν : veristic;
p : probabilistic;
γ : probability value;
u : usuality;
rs : random set;
rfs : random fuzzy set;
fg : fuzzy graph;
ps : rough set (Pawlak set);
$\ldots$ ....

As an illustration, when $r = e$, the constraint is an equality constraint and is abbreviated to $=$. When r takes the value d, the constraint is *disjunctive* (possibilistic) and isd abbreviated to is, leading to the expression

$$X \text{ is } R$$

in which R is a fuzzy relation which constrains X by playing the role of the possibility distribution of X. More specifically, if X takes values in a universe of discourse, $U = \{u\}$, then $Poss\{X = u\} = \mu_R(u)$, where μ_R is the membership function of R, and Π_X is the possibility distribution of X, that is, the fuzzy set of its possible values (Zadeh, 1978a). In schematic form:

$$X \text{ is } R \left\{ \begin{array}{c} \Pi_X = R \\ Poss\,\{X = u\} = \mu_R(u) \end{array} \right.$$

Similarly, when r takes the value ν, the constraint is *veristic*. In the case,

$$X \text{ isv } R$$

means that if the grade of membership of u in R is μ, then $X = u$ has truth value μ. For example, a canonical form of the proposition

$p =$ *John is proficient in English, French and German*

may be expressed as

$$\text{Proficiency (John) isv } (1|\text{English} + 0.7|\text{French} + 0.6|\text{German})$$

in which 1.0, 0.7 and 0.6 represent, respectively, the truth values of the propositions *John is proficient in English, John is proficient in French and John is proficient in German*. In a similar vein, the veristic constraint

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

represents the meaning of the proposition *John is half German, quarter French and quarter Italian*.

When $r = p$, the constraint is *probabilistic*. In this case,

$$X \text{ isp } R$$

means that R is the probability distribution of X. For example

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

means that X is normally distributed with mean m and variance σ^2. Similarly,

$$X \text{ isp } \left(0.2\backslash a + 0.5\backslash b + 0.3\backslash c\right)$$

means that X is a random variable which takes the values, a, b and c with respective probabilities 0.2, 0.5 and 0.3.

The constraint

$$X \text{ isu } R$$

is an abbreviation for

$$\textit{usually}(X \text{ is } R)$$

which in turn means that

$$\textit{Prob}\{X \text{ is } R\} \text{ is } \textit{usually}.$$

In this expression X is R is a fuzzy event and *usually* is its fuzzy probability, that is, the possibility distribution of its crisp probability.

The constraint

$$X \text{ isrs } P$$

is a random set constraint. This constraint is a combination of probabilistic and possibilistic constraints. More specifically, in a schematic form, it is expressed as

$$X \text{ isp } P$$
$$(X, Y) \text{ is } Q$$
$$\overline{Y \text{ isrs } R},$$

where Q is a joint possibilitistic constraint on X and Y, and R is a random set. It is of interest to note that the Dempster-Shafer theory of evidence (Shafer, 1976) is, in essence, a theory of random set constraints.

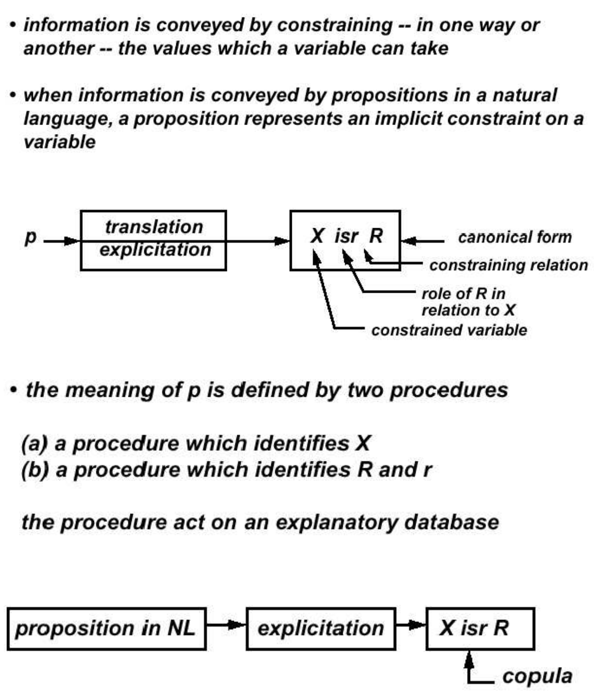

Fig. 14. Representation of meaning in test-score semantics.

In computing with words, the starting point is a collection of propositions which play the role of premises. In many cases, the canonical forms of these propositions are constraints of the basic, possibilistic type. In a more general setting, the constraints are of the generalized type, implying that explicitation of a proposition, p, may be represented as

$$p \rightarrow X \text{ isr } R,$$

where X isr R is the canonical form of p (Fig. 14).

As in the case of basic constraints, the canonical form of a proposition may be derived through the use of testscore semantics. In this context, the depth of p is, roughly, a measure of the effort that is needed to explicitate p, that is, to translate p into its canonical form. In this sense, the proposition X isr R is a surface constraint (depth=zero), with the depth of explicitation increasing in the downward direction (Fig. 15). Thus a proposition such as *Mary is young* is shallow, whereas *it is unlikely that there will be a substantial increase in the price of oil in the near future,* is not.

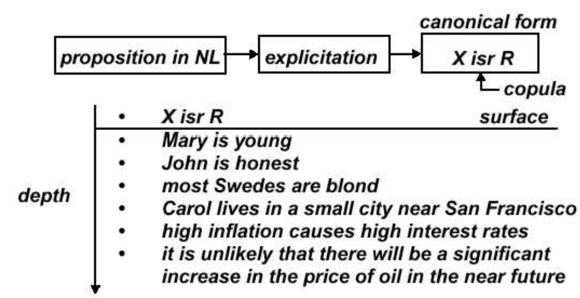

Fig. 15. Depth of explicitation.

Once the propositions in the initial data set are expressed in their canonical forms, the groundwork is laid for fuzzy constraint propagation. This is a basic part of CW which is discussed in the following section.

4. Fuzzy Constraint Propagation and the Rules of Inference in Fuzzy Logic

The rules governing fuzzy constraint propagation are, in effect, the rules of inference in fuzzy logic. In addition to these rules, it is helpful to have rules governing fuzzy constraint modification. The latter rules will be discussed at a later point in this section.

In a summarized form, the rules governing fuzzy constraint propagation are the following (Zadeh, 1996a). (A and B are fuzzy relations. Disjunction and conjunction are defined, respectively, as max and min, with the understanding that, more generally, they could be defined via t-norms and s-norms (Klir and Yuan, 1995; Pedrycz and Gomide, 1998). The antecedent and consequent constraints are separated by a horizontal line.)

Conjunctive Rule 1:

$$\frac{X \text{ is } A \qquad X \text{ is } B}{X \text{ is } A \cap B}$$

Conjunctive Rule 2: $(X \in U,\ Y \in B,\ A \subset U,\ B \subset V)$

$$\frac{X \text{ is } A \qquad Y \text{ is } B}{(X, Y) \text{ is } A \times B}$$

Disjunctive Rule 1:

$$\frac{X \text{ is } A \quad \text{or} \quad X \text{ is } B}{X \text{ is } A \cup B}$$

Disjunctive Rule 2: $(A \subset U,\ B \subset V)$

$$\frac{A \text{ is } A \qquad Y \text{ is } B}{(X, Y) \text{ is } A \times V \cup U \times B,}$$

where $A \times V$ and $U \times B$ are cylindrical extensions of A and B, respectively.

Conjunctive Rule for isv:

$$\frac{X \text{ isv } A \qquad X \text{ isv } B}{X \text{ isv } A \cup B}$$

Projective Rule:

$$\frac{(X, Y) \text{ is } A}{Y \text{ is } \text{proj}_V A,}$$

where $\text{proj}_V A = \sup_u A$.

Surjective Rule:

$$\frac{X \text{ is } A}{(X, Y) \text{ is } A \times V}$$

Derived Rules:

Compositional Rule:

$$\frac{X \text{ is } A \qquad (X, Y) \text{ is } B}{Y \text{ is } A \circ B,}$$

where $A \circ B$ denotes the composition of A and B.

Extension Principle (mapping rule) (Zadeh, 1965; 1975b):

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

where $f : U \to V$, and $f(A)$ is defined by $\mu_{f(A)}(\nu) = \sup_{u|\nu=f(u)} \mu_A(u)$.

Inverse Mapping Rule:

$$\frac{f(X) \text{ is } A}{X \text{ is } f^{-1}(A),}$$

where $\mu_{f^{-1}(A)}(u) = \mu_A(f(u))$.

Generalized modus ponens:

$$\frac{X \text{ is } A \qquad \text{if } X \text{ is } B \text{ then } Y \text{ is } C}{Y \text{ is } A \circ \big((\neg B) \oplus C\big),}$$

where the bounded sum $\neg B \oplus C$ represents Lukasiewicz's definition of implication.

Generalized Extension Principle:

$$\frac{f(X) \text{ is } A}{q(X) \text{ is } q\big(f^{-1}(A)\big),}$$

where $\mu_q(\nu) = \sup_{u|\nu=q(u)} \mu_A(f(u))$.

The generalized extension principle plays a pivotal role in fuzzy constraint propagation. However, what is used most frequently in practical applications of fuzzy logic is the *basic interpolative rule*, which is a special case of the compositional rule of inference applied to a function which is defined by a fuzzy graph (Zadeh, 1974; 1996a). More specifically, if f is defined by a fuzzy rule set

$$f : \textit{if } X \textit{ is } A_i \textit{ then } X \textit{ is } B_i, \quad i = 1, \dots, n$$

or equivalently, by a fuzzy graph

$$f \textit{ is } \sum_i A_i \times B_i$$

and its argument, X, is defined by the antecedent constraint

$$X \text{ is } A$$

then the consequent constraint on Y may be expressed as

$$Y \textit{ is } \sum_i m_i \wedge B_i,$$

where m_i is a matching coefficient,

$$m_i = \sup(A_i \cap A)$$

which serves as a measure of the degree to which A matches A_i.

Syllogistic Rule: (Zadeh, 1984)

$$\frac{\begin{array}{l} Q_1 A\text{'s are } B\text{'s} \\ Q_2(A \text{ and } B)\text{'s are } C\text{'s} \end{array}}{(Q_1 \otimes Q_2)A\text{'s are } (B \text{ and } C)\text{'s},}$$

where Q_1 and Q_2 are fuzzy quantifiers; A, B and C are fuzzy relations; and $Q_1 \otimes Q_2$ is the product of Q_1 and Q_2 in fuzzy arithmetic.

Constraint Modification Rules: (Zadeh, 1972; 1978b)

$$X \text{ is } mA \rightarrow X \text{ is } f(A),$$

where m is a modifier such as *not, very, more or less,* and $f(A)$ defines the way in which m modifies A. Specifically,

$$\text{if } m = \textit{not} \text{ then } f(A) = A' \text{ (complement)}$$

$$\text{if } m = \textit{very} \text{ then } f(A) = {}^2A \text{ (left square)},$$

where $\mu_{^2A}(u) = (\mu_A(u))^2$. This rule is a convention and should not be constructed as a realistic approximation to the way in which the modifier *very* functions in a natural language.

Probability Qualification Rule: (Zadeh, 1979b)

$$(X \text{ is } A) \text{ is } \Lambda \rightarrow P \text{ is } \Lambda,$$

where X is a random variable taking values in U with probability density $p(u)$; Λ is a linguistic probability expressed in words like *likely, not very likely,* etc.; and P is the probability of the fuzzy event X, expressed as

$$P = \int_U \mu_A(u) p(u)\, \mathrm{d}u.$$

The primary purpose of this summary is to underscore the coincidence of the principal rules governing fuzzy constraint propagation with the principal rules of inference in fuzzy logic. Of necessity, the summary is not complete and there are many specialized rules which are not included. Furthermore, most of the rules in the summary apply to constraints which are of the basic, possibilistic type. Further development of the rules governing fuzzy constraint propagation will require an extension of the rules of inference to generalized constraints.

As was alluded to in the summary, the principal rule governing constraint propagation is the generalized extension principle which in a schematic form may be represented as

$$\frac{f(X_1, \dots, X_n) \text{ is } A}{q(X_1, \dots, X_n) \text{ is } q\big(f^{-1}(A)\big).}$$

In this expression, $X_1, \dots, X_n$ are database variables; the term above the line represents the constraint induced by the IDS; and the term below the line is the TDS expressed as a constraint on the query $q(X_1, \dots, X_n)$. In the latter constraint, $f^{-1}(A)$ denotes the preimage of the fuzzy relation A under the mapping $f : U \rightarrow V$, where A is a fuzzy subset of V and U is the domain of $f(X_1, \dots, X_n)$.

Expressed in terms of the membership functions of A and $q(f^{-1}(A))$, the generalized extension principle reduces the derivation of the TDS to the solution of the constrained maximization problem

$$\mu_q(X_1, \dots, X_n)(\nu) = \sup_{(u_1, \dots, u_n)} \big(\mu_A\big(f(u_1, \dots, u_n)\big)\big)$$

in which $u_1, \dots, u_n$ are constrained by

$$\nu = q(u_1, \dots, u_n).$$

The generalized extension principle is simpler than it appears. An illustration of its use is provided by the following example.

The IDS is:

most Swedes are *tall*

The query is: *What is the average height of Swedes?*

The explanatory database consists of a population of N Swedes, $Name_1, \dots, Name_N$. The database variables are $h_1, \dots, h_N$, where h_i is the height of $Name_i$, and the grade of membership of $Name_i$ in *tall* is $\mu_{tall}(h_i)$, $i = 1, \dots, n$.

The proportion of Swedes who are tall is given by the sigma-count (Zadeh, 1978b)

$$\sum \textit{Count (tall – Swedes / Swedes)} = \frac{1}{N} \sum_i \mu_{tall}(h_i)$$

from which it follows that the constraint on the database variables induced by the IDS is

$$\frac{1}{N} \sum_i \mu_{tall}(h_i) \text{ is } \textit{most}.$$

In terms of the database variables $h_1, \ldots, h_N$, the average height of Swedes is given by

$$h_{ave} = \frac{1}{N} \sum_i h_i.$$

Since the IDS is a fuzzy proposition, h_{ave} is a fuzzy set whose determination reduces to the constrained maximization problem

$$\mu_{h_{ave}}(\nu) = \sup_{h_1,\ldots,h_N} \left(\mu_{most}\Big(\frac{1}{N} \sum_i \mu_{tall}(h_i) \Big) \right)$$

subject to the constraint

$$\nu = \frac{1}{N} \sum_i h_i.$$

It is possible that approximate solutions to problems of this type might be obtainable through the use of neurocomputing or evolutionary-computing-based methods.

As a further example, we will return to a problem stated in an earlier section, namely, maximization of a function, f, which is described in words by its fuzzy graph, f^* (Fig. 10). More specifically, consider the standard problem of maximization of an objective function in decision analysis. Let us assume – as is frequently the case in real-world problems – that the objective function, f, is not well-defined and that what we know about can be expressed as a fuzzy rule-set

$f:$ *if* X *is* A_1 *then* Y *is* B_1

if X *is* A_2 *then* Y *is* B_2

....

if X *is* A_n *then* Y *is* B_n

or, equivalently, as a fuzzy graph

$$f \text{ is } \sum_i A_i \times B_i.$$

The question is: What is the point or, more generally, the maximizing set (Zadeh, 1998) at which f is maximized, and what is the maximum value of f?

The problem can be solved by employing the technique of α-*cuts* (Zadeh, 1965; 1975b). With refererence to Fig. 16, if A_{i_α} and B_{i_α} are α-*cuts* of A_i and B_i, respectively, then the corresponding α-*cut* of f^* is given by

$$f^*_\alpha = \sum_i A_{i_\alpha} \times B_{i_\alpha}.$$

From this expression, the maximizing fuzzy set, the maximum fuzzy set and maximum value fuzzy set can readily be derived, as shown in Figs. 16 and 17.

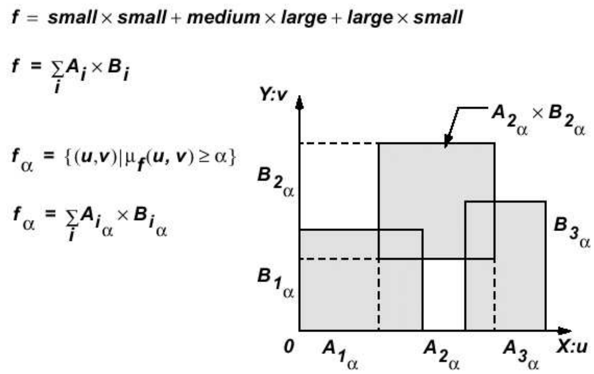

Fig. 16. α-cuts of a function described by a fuzzy graph.

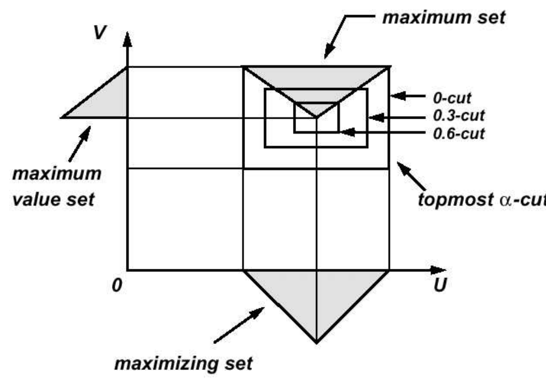

Fig. 17. Computation of maximizing set, maximum set and maximum value set.

A key point which is brought out by these examples and the preceding discussion is that explicitation and constraint propagation play pivotal roles in CW. This role can be concretized by viewing explicitation and constraint propagation as translation of propositions expressed in a natural language into what might be called the *generalized constraint language* (GCL) and applying rules of constraint propagation to expressions in this language – expressions which are typically canonical forms of propositions expressed in a natural language. This process is schematized in Fig. 18.

The conceptual framework of GCL is substantively differently from that of conventional logical systems, e.g., predicate logic. But what matters most is that the expressive power of GCL – which is based on fuzzy logic – is much greater than that of standard logical calculi. As an illustration of this point, consider the following problem.

A box contains ten balls of various sizes of which several are large and a few are small. What is the probability that a ball drawn at random is neither large nor small?

To be able to answer this question it is necessary to be able to define the meanings of *large, small, several large balls, few small balls* and *neither large nor small*. This is

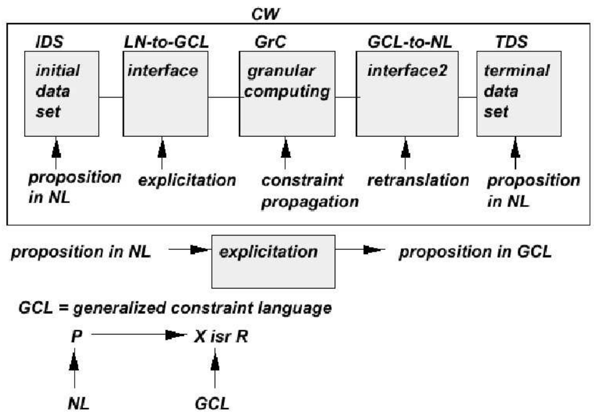

Fig. 18. Conceptual structure of computing with words.

a problem in semantics which falls outside of probability theory, neurocomputing and other methodologies.

An important application area for computing with words and manipulation of perceptions is decision analysis since in most realistic settings the underlying probabilities and utilities are not known with sufficient precision to justify the use of numerical valuations. There exists an extensive literature on the use of fuzzy probabilities and fuzzy utilities in decision analysis. In what follows, we shall restrict our discussion to two very simple examples which illustrate the use of perceptions.

First, consider a box which contains black balls and white balls (Fig. 19). If we could count the number of black balls and white balls, the probability of picking a black ball at random would be equal to the proportion, r, of black balls in the box.

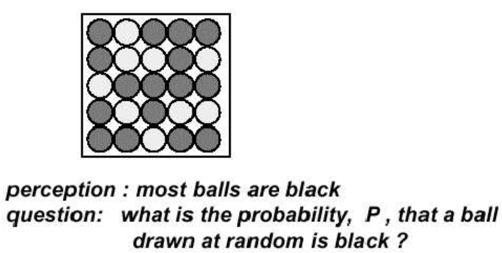

Fig. 19. A box with black and white balls.

Now suppose that we cannot count the number of black balls in the box but our perception is that most of the balls are black. What, then, is the probability, p, that a ball drawn at random is black?

Assume that *most* is characterized by its possibility distribution (Fig. 20). In this case, p is a fuzzy number whose possibility distribution is *most*, that is,

$$p \text{ is } most.$$

Next, assume that there is a reward of a dollars if the ball drawn at random is black and a penalty of b dollars

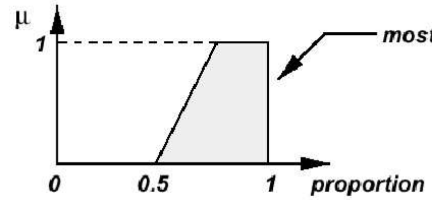

Fig. 20. Membership function of *most*.

if the ball is white. In this case, if p were known as a number, the expected value of the gain would be:

$$e = ap - b(1 - p).$$

Since we know not p but its possibility distribution, the problem is to compute the value of e when p is *most*. For this purpose, we can employ the extension principle (Zadeh, 1965; 1975b), which implies that the possibility distribution, E, of e is a fuzzy number which may be expressed as

$$E = a\ most\ - b(1 - most).$$

For simplicity, assume that *most* has a trapezoidal possibility distribution (Fig. 20). In this case, the trapezoidal possibility distribution of E can be computed as shown in Fig. 21.

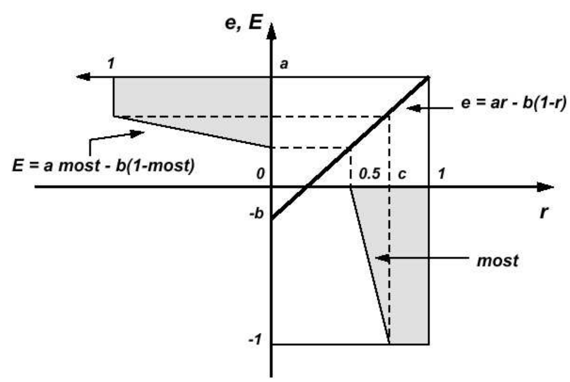

Fig. 21. Computation of expectation through use of the extension principle.

It is of interest to observe that if the support of E is an interval $[\alpha, \beta]$ which straddles O (Fig. 22), then there is no non-controversial decision principle which can be employed to answer the question: Would it be advantageous to play a game in which a ball is picked at random from a box in which most balls are black, and a and b are such that the support of E contains O.

Next, consider a box in which the balls $b_1, \dots, b_n$ have the same color but vary in size, with b_i, $i = 1, \dots, n$ having the grade of membership μ_i in the fuzzy set of large balls (Fig. 23). The question is: What is the

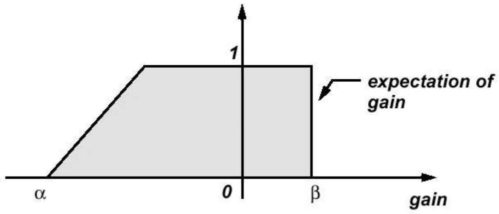

Fig. 22. Expectation of gain.

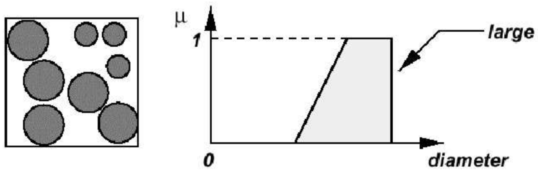

Fig. 23. A box with balls of various sizes and a definition of large ball.

probability that a ball drawn at random is large, given the perception that most balls are large?

The difference between this example and the preceding one is that the event *the ball drawn at random is large* is a fuzzy event, in contrast to the crisp event *the ball drawn at random is black*.

The probability of drawing b_i is $1/n$. Since the grade of membership of b_i in the fuzzy set of large balls is μ_i, the probability of the fuzzy event *the ball drawn at random is large* is given by (Zadeh, 1968)

$$P = \frac{1}{n} \sum \mu_i .$$

On the other hand, the proportion of large balls in the box is given by the relative sigma-count (Zadeh, 1975b; 1978b)

$$\sum \text{Count } (\textit{large.balls / balls.in.box}) = \frac{1}{n} \sum \mu_i .$$

Consequently, the canonical form of the perception *most balls are large* may be expressed as

$$\frac{1}{n} \sum \mu_i \text{ is } \textit{most}$$

which leads to the conclusion that

$$P \text{ is } \textit{most}.$$

It is of interest to observe that the possibility distribution of P is the same as in the preceding example.

If the question were: What is the probability that a ball drawn at random is *small*, the answer would be

$$P \text{ is } \frac{1}{n} \sum \nu_i$$

where $\nu_i, \quad i = 1, \dots, n$, is the grade of membership of b_i in the fuzzy set of small balls, given that

$$\frac{1}{n} \sum \mu_i \text{ is } \textit{most}.$$

What is involved in this case is constraint propagation from the antecedent constraint on the μ_i to a consequent constraint on the ν_i. This problem reduces to the solution of a nonlinear program.

What this example points to is that in using fuzzy constraint propagation rules, application of the extension principle reduces, in general, to the solution of a nonlinear program. What we need – and do not have at present – are approximate methods of solving such programs which are capable of exploiting the tolerance for imprecision. Without such methods, the cost of solutions may be excessive in relation to the imprecision which is intrinsic in the use of words. In this connection, an intriguing possibility is to use neurocomputing and evolutionary computing techniques to arrive at approximate solutions to constrained maximization problems. The use of such techniques may provide a closer approximation to the ways in which human manipulate perceptions.

5. Concluding Remarks

In our quest for machines which have a high degree of machine intelligence (high MIQ), we are developing a better understanding of the fundamental importance of the remarkable human capacity to perform a wide variety of physical and mental tasks without any measurements and any computations. Underlying this remarkable capability is the brain's crucial ability to manipulate perceptions – perceptions of distance, size, weight, force, color, numbers, likelihood, truth and other characteristics of physical and mental objects. A basic difference between perceptions and measurements is that, in general, measurements are crisp whereas perceptions are fuzzy. In a fundamental way, this is the reason why to deal with perceptions it is necessary to employ a logical system that is fuzzy rather than crisp.

Humans employ words to describe perceptions. It is this obvious observation that is the point of departure for the theory outlined in the preceding sections.

When perceptions are described in words, manipulation of perceptions is reduced to computing with words (CW). In CW, the objects of computation are words or, more generally, propositions drawn from a natural language. A basic premise in CW is that the meaning of a proposition, p, may be expressed as a generalized constraint in which the constrained variable and the constraining relation are, in general, implicit in p.

In coming years, computing with words and perceptions is likely to emerge as an important direction in science and technology. In a reversal of long-standing attitudes, manipulation of perceptions and words which describe them is destined to gain in respectability. This is certain to happen because it is becoming increasingly clear that in dealing with real-world problems there is much to be gained by exploiting the tolerance for imprecision, uncertainty and partial truth. This is the primary motivation for the methodology of computing with words (CW) and the computational theory of perceptions (CTP) which are outlined in this paper.

Acknowledgement

The author acknowledges Prof. Michio Sugeno, who has contributed so much and in so many ways to the development of fuzzy logic and its applications.

References

Berenji H.R. (1994): *Fuzzy reinforcement learning and dynamic programming*, In: Fuzzy Logic in Artificial Intelligence (A.L. Ralescu, Ed.). — Proc. IJCAI'93 Workshop, Berlin: Springer-Verlag, pp. 1–9.

Black M. (1963): *Reasoning with loose concepts*. — Dialog 2, pp. 1–12.

Bosch P. (1978): *Vagueness, ambiguity and all the rest*, In: Sprachstruktur, Individuum und Gesselschaft (M. Van de Velde and W. Vandeweghe, Eds.). — Tubingen: Niemeyer.

Bowen J., Lai R. and Bahler D. (1992a): *Fuzzy semantics and fuzzy constraint networks*. — Proc. 1st IEEE Conf. *Fuzzy Systems*, San Francisco, pp. 1009–1016.

Bowen J., Lai R. and Bahler D. (1992b): *Lexical imprecision in fuzzy constraint networks*. — Proc. Nat. Conf. *Artificial Intelligence*, pp. 616–620.

Cresswell M.J. (1973): *Logic and Languages*. — London: Methuen.

Dubois D., Fargier H. and Prade H. (1993): *The calculus of fuzzy restrictions as a basis for flexible constraint satisfaction*. — Proc. 2nd IEEE Int. Conf. *Fuzzy Systems*, San Francisco, pp. 1131–1136.

Dubois D., Fargier H. and Prade H. (1994): *Propagation and satisfaction of flexible constraints*, In: Fuzzy Sets, Neural Networks, and Soft Computing (R.R. Yager, L.A. Zadeh, Eds.). — New York: Von Nostrand Reinhold, pp. 166–187.

Dubois D., Fargier H. and Prade H. (1996): *Possibility theory in constraint satisfaction problems: Handling priority, preference and uncertainty*. — Appl. Intell. J., Vol. 6, No. 4, pp. 287–309.

Freuder E.C. and Snow P. (1990): *Improved relaxation and search methods for approximate constraint satisfaction with a maximin criterion*. — Proc. 8th Biennial Conf. *Canadian Society for Computational Studies of Intelligence*, Ontario, pp. 227–230.

Goguen J.A. (1969): *The logic of inexact concepts*. — Synthese 19, pp. 325–373.

Hobbs J.R. (1978): *Making computation sense of Montague's intensional logic*. — Artif. Intell., Vol. 9, pp. 287–306.

Katai O., Matsubara S., Masuichi H., Ida M., *et al.* (1992): *Synergetic computation for constraint satisfaction problems involving continuous and fuzzy variables by using Occam*, In: Transputer/Occam, (S. Noguchi and H. Umeo, Eds.). — Proc. 4th Transputer/Occam Int. Conf., Amsterdam: IOS Press, pp. 146–160.

Kaufmann A. and Gupta M.M. (1985): *Introduction to Fuzzy Arithmetic: Theory and Applications*. — New York: Von Nostrand.

Klir G. and Yuan B. (1995): *Fuzzy Sets and Fuzzy Logic*. — New Jersey: Prentice Hall.

Lano K. (1991): *A constraint-based fuzzy inference system*. — Proc. 5th Portuguese Conf. *Artificial Intelligence*, EPIA'91 (P. Barahona, L.M. Pereira, and A. Porto, Eds.), Berlin: Springer-Verlag, pp. 45–59.

Lodwick W.A. (1990): *Analysis of structure in fuzzy linear programs*. — Fuzzy Sets Syst., Vol. 38, No. 1, pp. 15–26.

Mamdani E.H. and Gaines B.R., Eds. (1981): *Fuzzy Reasoning and its Applications*. — London.

Mares M. (1994): *Computation Over Fuzzy Quantities*. — Boca Raton: CRC Press.

Novak V. (1991): *Fuzzy logic, fuzzy sets, and natural languages*. — Int. J. General Syst., Vol. 20, No. 1, pp. 83–97.

Novak V., Ramik M., Cerny M. and Nekola J., Eds. (1992): *Fuzzy Approach to Reasoning and Decision-Making*. — Boston: Kluwer.

Oshan M.S., Saad O.M. and Hassan A.G. (1995): *On the solution of fuzzy multiobjective integer linear programming problems with a parametric study*. — Adv. Modell. Anal. A, Vol. 24, No. 2, pp. 49–64.

Partee B. (1976): *Montague Grammar*. — New York: Academic Press.

Pedrycz W. and Gomide F. (1998): *Introduction to Fuzzy Sets*. — Cambridge: MIT Press, pp. 38–40.

Qi G. and Friedrich G. (1992): *Extending constraint satisfaction problem solving in structural design*. — 5th Int. Conf. *Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, IEA/AIE-92* (F. Belli and F.J. Radermacher, Eds.), Berlin: Springer-Verlag, pp. 341–350.

Rasiowa H. and Marek M. (1989): *On reaching consensus by groups of intelligent agents*, In: Methodologies for Intelligent Systems (Z.W. Ras, Ed.). — Amsterdam: North-Holland, pp. 234–243.

Rosenfeld A., Hummel R.A. and Zucker S.W. (1976): *Scene labeling by relaxation operations*. — IEEE Trans. Syst. Man Cybern., Vol. 6, pp. 420–433.

Sakawa M., Sawada K. and Inuiguchi M. (1995): *A fuzzy satisficing method for large-scale linear programming problems with block angular structure*. — Europ. J. Oper. Res., Vol. 81, No. 2, pp. 399–409.

Shafer G. (1976): *A Mathematical Theory of Evidence*. — Princeton: Princeton University Press.

Tong S.C. (1994): *Interval number and fuzzy number linear programming*. — Adv. Modell. Anal. A, Vol. 20, No. 2, pp. 51–56.

Vallée R. (1995): *Cognition et Systeme*. — Paris: l'Interdisciplinaire Systeme(s).

Yager R.R. (1989): *Some extensions of constraint propagation of label sets*. — Int. J. Approx. Reas., Vol. 3, pp. 417–435.

Zadeh L.A. (1961): *From circuit theory to system theory*. — Proc. IRE, Vol. 50, pp. 856–865.

Zadeh L.A. (1965): *Fuzzy sets*. — Inf. Control, Vol. 8, pp. 338–353.

Zadeh L.A. (1968): *Probability measures of fuzzy events*. — J. Math. Anal. Appl., Vol. 23, pp. 421–427.

Zadeh L.A. (1972): *A fuzzy-set-theoretic interpretation of linguistic hedges*. — J. Cybern., Vol. 2, pp. 4–34.

Zadeh L.A. (1973): *Outline of a new approach to the analysis of complex system and decision processes*. — IEEE Trans. Syst. Man Cybern., Vol. SMC–3, pp. 28–44.

Zadeh L.A. (1974): *On the analysis of large scale systems*. — Systems Approaches and Environment Problems (H. Gottinger, Ed.), Gottingen: Vandenhoeck and Ruprecht, pp. 23–37.

Zadeh L.A. (1975a): *Calculus of fuzzy restrictions*, In: Fuzzy Sets and Their Applications to Cognitive and Decision Processes, (L.A. Zadeh, K.S. Fu, M. Shimura, Eds.). — New York: Academic Press, pp. 1–39.

Zadeh L.A. (1975b): *The concept of a linguistic variable and its application to approximate reasoning*. — Part I: Inf. Sci., Vol. 8, pp. 199–249; Part II: Inf. Sci., Vol. 8, pp. 301–357; Part III: Inf. Sci., Vol. 9, pp. 43–80.

Zadeh L.A. (1976): *A fuzzy-algorithmic approach to the definition of complex or imprecise concepts*. — Int. J. Man-Mach. Stud., Vol. 8, pp. 249–291.

Zadeh L.A. (1978a): *Fuzzy sets as a basis for a theory of possibility*. — Fuzzy Sets Syst., Vol. 1, pp. 3–28.

Zadeh L.A. (1978b): *PRUF – A meaning representation language for natural languages*. — Int. J. Man-Mach. Stud., Vol. 10, pp. 395–460.

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

Zadeh L.A. (1979b): *A theory of approximate reasoning*. — Mach. Intell., Vol. 9 (J. Hayes, D. Michie and L.I. Mikulich, Eds.), New York: Halstead Press, pp. 149–194.

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

Zadeh L.A. (1982): *Test-score semantics for natural languages*. — Proc. 9-th Int. Conf. *Computational Linguistics*, Prague, pp. 425–430.

Zadeh L.A. (1984): *Syllogistic reasoning in fuzzy logic and its application to reasoning with dispositions*. — Proc. Int. Symp. *Multiple-Valued Logic*, Winnipeg, Canada, pp. 148–153.

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

Zadeh L.A. (1994): *Fuzzy logic, neural networks and soft computing*. — Comm. ACM, Vol. 37, No. 3, pp. 77–84.

Zadeh L.A. (1996a): *Fuzzy logic and the calculi of fuzzy rules and fuzzy graphs: A precis*. — Multiple Valued Logic 1, Gordon and Breach Science Publishers, pp. 1–38.

Zadeh L.A. (1996b): *Fuzzy logic = Computing with words*. — IEEE Trans. Fuzzy Syst., Vol. 4, pp. 103–111.

Zadeh L.A. (1997): *Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic*. — Fuzzy Sets Syst., Vol. 90, pp. 111–127.

Zadeh L.A. (1998): *Maximizing Sets and Fuzzy Markoff Algorithms*. — IEEE Trans. Syst. Man Cybern., Part C: Appli. Rev., Vol. 28, pp. 9–15.