

Special Contribution

From search engines to question-answering systems—The role of fuzzy logic

Lotfi A. Zadeh

Univ. California, Berkeley

The following brief note is intended to draw attention to an issue whose importance is widely unrecognized. The issue is: Can an existing search engine be upgraded to a question-answering system through the use of methods based on bivalent logic and bivalent-logic-based probability theory?

Existing search engines have many truly remarkable capabilities. But, what is not among them is deduction capability?the capability to answer a query by a synthesis of information drawn from various parts of the knowledge base. A question-answering system is by definition a system which has this capability. One of the principal goals of Web intelligence is that of upgrading search engines to question-answering systems. Achievement of this goal requires a quantum jump in the WIQ (Web IQ) of existing search engines.

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

There are three major problems which relate to: (a) World knowledge; (b) Assessment of relevance; and (c) Deduction. World knowledge is the kind of knowledge that humans acquire through experience, education and communication. Simple examples of frag-

ments of world knowledge are: Usually it does not rain in San Francisco in midsummer; Berkeley is a friendly city; affordable housing is nonexistent in Palo Alto; almost all professors have a Ph.D. degree; Switzerland has no ports; and usually Princeton means Princeton University.

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

Imprecision of perception-based information is a major obstacle to dealing with world knowledge through the use of methods based on bivalent logic and bivalent-logic-based probability theory. What is needed for this purpose is a collection of tools drawn from fuzzy logic?a logic in which everything is, or is allowed to be, a matter of degree. The principal tool is Precisiated Natural Language (PNL).

The point of departure in PNL is the assumption that the meaning of a proposition, p , drawn from a natural language, NL, can be represented as a generalized constraint of the form $X \text{ is } R$, where X is the constrained variable; R is the constraining relation; and r is a modal variable, that is, a variable whose value defines the modality (semantics) of the constraint. The principal modalities are: possibilistic ($r=\text{blank}$); veristic ($r=v$); probabilistic ($r=p$); random set ($r=rs$); fuzzy graph ($r=fg$); usuality ($r=u$); and group ($r=g$). The set of all generalized constraints together with their combinations, qualifications and

rules of constraint propagation, constitutes the Generalized Constraint Language (GCL). By construction, GCL is maximally expressive.

A proposition, p , in NL is precisiable if it is translatable into GCL. In this sense, PNL consists of precisiable propositions, with the understanding that not every proposition in NL is precisiable. The importance of PNL derives from the fact that it has a far greater expressive power than predicate-logic-based synthetic languages like LISP, Prolog, SQL, etc. A concept which plays a key role in PNL is that of a protoform—an abbreviation of “prototypical form.” Informally, the protoform of a lexical entity such as a proposition, command, question, or scenario is its abstracted summary. For example, the protoform of p : Yoko is young, is $A(B)$ is C , where A is abstraction of age, B is abstraction of Yoko, and C is abstraction of young. Similarly, the protoform of p : Most Swedes are tall, is $\text{Count}(B/A)$ is Q , where A is abstraction of Swedes, B is abstraction of tall Swedes, $\text{Count}(B/A)$ is abstraction of the relative count of tall Swedes among Swedes, and Q is abstraction of most.

The importance of the concept of a protoform derives from the fact that it places in evidence the deep semantic structure of the lexical entity to which it applies. In this sense, propositions p and q are PF-equivalent, written as $\text{PFE}(p, q)$, if they have identical protoforms, that is, identical deep semantic structures. As a simple example, p : Most Swedes are tall, and q : Few professors are rich, are PF-equivalent.

The concept of PF-equivalence serves as a basis for what may be called protoform-centered mode of knowledge organization. In this mode, a protoformal module consists of all propositions which have a specified protoform in common, e.g., $A(B)$ is C . Submodules of such a module are generated through instantiation of A , B and C . For example, the partially instantiated protoform: price (B) is low, would represent all objects in a universe of discourse, U , whose price is low.

An important function of PNL is that of serving as a deduction language. For this purpose, PNL contains a Deduction Database, DDB, which consists of so-called protoformal rules of deduction. Basically, such rules govern generalized constraint propagation, with antecedents and consequents expressed as protoforms. Typically, a protoformal rule of deduction has two parts: symbolic and computational. A simple example is the compositional rule of inference in fuzzy logic. In this case, the symbolic part is: if X is A and (X, Y) is B ,

then Y is C ; and the computational part is: $C = A \circ B$, that is, C is the composition of A and B .

The Deduction Database contains a large number of modules and submodules comprising protoformal rules drawn from a wide range of domains. Examples of such modules are: the Search module, the World Knowledge module, the Extension Principle module, the Probability module, the Possibility module, the Usuality module, etc.

The second major problem relates to assessment of relevance. There are two kinds of relevance: (a) semantic relevance; and (b) statistical relevance. To illustrate the differences, consider an example drawn from everyday experience. Suppose that the query, q , is: How old is Motoko; and the datum, d , is: Motoko has a son who is about 25 years old. The question is: What is the degree of relevance of d to q ? Existing methods of assessment of relevance cannot deal with questions of this type.

Statistical relevance involves statistical information about links and word counts. This is what existing search engines use, some at a high level of sophistication. The PageRanking algorithm which is used in Google is an example. As an illustration, assume that the query, q , is: Age of Bush? A search engine will correctly interpret Bush as President George W. Bush, but will have difficulty in coming up with his age.

The third major problem is deduction. For example, given that Motoko has a son who is about 25 years old, how can one estimate Motoko's age? Existing search engines do not have the capability to infer from data that are imprecise. What is needed for this purpose is PNL.

In summary, abandonment of bivalence is a prerequisite to achieving a quantum jump in WIQ. By abandoning bivalence, the door is opened to the use of tools such as PNL for adding to search engines three essential capabilities: (a) capability to operate on perception-based information; (b) capability to deal with semantic relevance; and (c) deduction capability. What should be stressed, however, is that no matter what approach is used, upgrading a search engine to a question-answering system is a challenge that will be hard to meet.

A more detailed discussion of this issue may be found in my paper, “A Note on Web Intelligence, World Knowledge and Fuzzy Logic,” *Data and Knowledge Engineering*, Elsevier Science, Vol. 50, 291-304, 2004.

**Lotfi A. ZADEH**

Lotfi A. Zadeh is a Professor in the Graduate School at the University of California, Berkeley, and serves as Director of the Berkeley Initiative in Soft Computing. His first paper on fuzzy sets was published in 1965, and since then his research has focused on the development of fuzzy logic and its applications. He is a fellow of the American Association for Artificial Intelligence, the ACM, the IEEE and the International Fuzzy Systems Association. He is a member of the National Academy of Engineering and a foreign member of the Russian and Finnish Academies of Science. He has received numerous awards in recognition of his development of fuzzy logic, among them the IEEE Medal of Honor, the ACM 2000 Allen Newell Award, the Honda Prize, the Okawa Prize and twenty-three honorary doctorates.