OUTLINE OF A NAIVE SEMANTICS FOR REASONING WITH QUALITATIVE LINGUISTIC INFORMATION *

Daniel G. Schwartz
Department of Computer Science and the
Center for Artificial Intelligence
Florida State University
Tallahassee, Florida 32306-4019, U.S.A.
schwaxtz@nu.cs.fsu.edu

Abstract

This paper describes the mathematical basis for a computer language which can be used for representing natural human reasoning with imprecise linguistic information. The approach to doing this employs a collection of abstraction mechanisms which are based on the concept of a *linguistic variable* first introduced by Zadeh [1975]. The present semantics differs from that of Zadeh, however, in that (i) it does not require the use of fuzzy sets for the interpretation of linguistic terms, and (ii) the meanings of logical inferences are given as algorithms which act directly on linguistic terms themselves, rather than on their underlying interpretations. Two distinct types of deduction algorithm are proposed. The overall objective is to devise a reasoning system having sufficient generality that it can conveniently employ these plus others in a unified frame.

1 Introduction

One of the central problems in the theory of approximate reasoning is how to model natural human reasoning with imprecise linguistic information. To illustrate, one would like to have an effective means of representing inferences like "MOST professional basketball players are VERY TALL; Bill is a professional basketball player; therefore it is VERY LIKELY that Bill is AT LEAST TALL."

A wide variety of approaches to this problem have been proposed. Primary among these is the semantics based on fuzzy sets, known as possibility theory, developed by Zadeh [1965, 1975, 1978, 1986] Dubois and Prade [1979, 1987], and many others. As a result of more than two decades of research, possibility theory has now reached a respectable level of conceptual sophistication and has encompassed a rich collection of natural language phenomena. A second approach has focused on the probabalistic concept of uncertainty as a means of representing linguistic imprecision. Works in this genre include those of Shafer [1976] and Pearl [19S8]. A third approach, explored separately by Baldwin [1987] and by Dubois and Prade [1987, 1989], employs a notion of possibility and necessity measures. Such measures are an extension of possibility theory that evidently are inspired by the literature on modal logics.

All such reasoning systems have their relative advantages and disadvantages. Typically there is an inherent tradeoff between semantic richness and computational tractability. In Zadeh's work semantic richness is achieved at the expense of almost overwhelming computational complexity. The approaches developed by Baldwin and Dubois and Prade, on the other hand, are computationally manageable but treat only limited aspects of the overall problem. Shafer's and Pearl's systems similarly encompass only a portion of the desired set of linguistic ideas.

The present work represents an ongoing effort to resolve these difficulties through a somewhat different conception of natural language reasoning. A major difference between this conception and those that have gone before is that here no attempt is made to ground the various deduction procedures on an underlying semantics for the linguistic terms. For example, in the inference rule given above, Zadeh's approach would be to interpret each linguistic term as a possibility distribution over a universe of discourse (e.g., TALL is interpreted as a distribution over a set of heights), and the inference itself is then modeled as an operation on those distributions. A primary motivation for this and all similar approaches has been to establish a coherent model of natural language reasoning on an intuitively plausible foundation.

By contrast, the approach taken here yields what may be regarded as a *naive* semantics in that it makes no similar attempt. Rather, logical inferences are defined as operations performed directly on the linguistic

* This work was supported by the Office of Naval Research, Grant Number N00014-87-G-0219.

terms—i.e., without appeal to their underlying interpretations. In this respect, the present system is akin to classical logic in that it employs direct manipulation of symbols but at the same time implicitly captures the *linguistic invariance* of the logical connectives. In what follows, one obtains invariant aspects of the logical OR and AND, together with various *linguistic hedges* (like VERY and AT LEAST) and several varieties of linguistic negation. Moreover, two distinct variants of logical inference are portrayed, and what forms of linguistic qualification are allowed will in part depend on the type of inference scheme that is to be employed.

Thus the present research accomplishes at least a part of the overall goal. In reference to the example inference given above, however, it still falls short of dealing with linguistic quantifiers (like MOST and FEW) and the concept of linguistic likelihood. Nonetheless, it is believed that the present system can be extended to capture these concepts as well, and a future work is planned which deals with these, together with a concept of linguistic temporality (with modifiers like SELDOM and USUALLY). In this manner one should arrive at a system which is simple to implement, yet which at the same time is sufficiently comprehensive to be useful in many real-world applications. The initial aim shall be to develop deduction techniques for use in rule-based expert systems. Once this is accomplished, then these likely will be adaptable to associative nets and inheritance systems. It also seems feasible that such techniques could be implemented on a type of neural computer.

2 Linguistic Variables

A *linguistic variable* A will be represented as a triple (T, U, M), where T is a set of *linguistic terms,* U is a *universe of discourse,* and M is a *meaning assignment,* each described as follows.

The set T of linguistic terms is of the form E u S, where E is a set of *elementary linguistic terms* and S is a (possibly empty) set of *synonyms* for elementary terms. For each A, the corresponding E is assumed to contain a unique *primary term* A. To illustrate, for the linguistic variable Height, a natural choice of primary term would be TALL. Where ant(A) represents the concept of an antonym, and med(A) represents the concept of an intermediate term, we shall allow that E may have any of six possible forms. For notational convenience, r, v, and e will be used as abbreviations for the *linguistic hedges* RATHER, VERY, and EXTREMELY. The six forms are

a) {ant(A), A}

b) {ant(A), med(A), A}

c) {ant(A), r-ant(A), med(A), r-A, A)

d) {v-ant(λ), ant(λ), med(λ), λ, v-λ}

e) {v-ant(λ), ant(λ), r-ant(λ), med(λ), r-λ, λ, v-λ}

f) {e-ant(λ), v-ant(λ), ant(λ), r-ant(λ), med(λ), r-λ, λ, v-λ, e-λ}

It will be assumed that the terms in each version of E are ordered by a relation < in the manner shown.

If nonempty, the set S contains alternative linguistic equivalents for members of E. For example, if A = Height, with primary term λ = TALL, then natural choices of synonyms for med(λ) and ant(λ) would be MEDIUM and SHORT. In addition, S might contain phrases considered as being equivalent with an elementary term, e.g., the synonyms for med(TALL) might include NEITHER TALL NOR SHORT. The introduction of synonyms into the system is not only for the convenience of the user, but also to enrich the system's overall expressive power. They play no essential role, however, in any of the deduction schemes. Rather, deductions are defined exclusively on elementary terms. For this reason, whenever a member of S appears in a deduction, it is implicitly assumed as representing the corresponding elementary term.

In some instances there will be more that one natural choice of primary term for a given linguistic variable. For example, if A = Age, then one might choose λ = OLD or A = YOUNG. Under the current definition, one obtains a distinct linguistic variable for each such choice of primary term.

The universe U is a set of objects that is used for providing meanings for the terms in T. For example, if A = Age is intended as a linguistic variable for ages of people, then a choice for U might be the ages in years from 0 to 150. It is allowed that U be empty. Such would be appropriate for a linguistic variable like Kindness, for which there is no presumed measurement scale.

The meaning assignment M is defined only if U # 0. There will be three permissible interpretation schemes. For r € T, M(r) is either

1) a subinterval of U,

2) a possibility distribution over U, or

3) a probability distribution over U.

The mapping M is used to determine which term in T should be ascribed to an individual *A,* given some measurement *u* for *A* along U. For example, suppose M is defined for a version of Age using subintervals, and M(OLD) = [70, 125]. Then, if *A* is known to be 84 years old, one would ascribe *A* the term OLD. If possibility distributions are used, then *A* is ascribed the term in T with which *A's* age has the highest degree of possibility. If probability distributions are used, then one ascribes the term for which A's age has the highest probability. Note that in the latter it would be natural to use probability distributions determined by a statistical sampling. Which interpretation method is used, or whether any interpretations are used at all, turns out to be irrelevant for the purposes of the deduction methods described below.

3 Inference Method A

An earlier version of this method appeared as [Schwartz, 1987]. There the problem was simplified by limiting the consideration to elementary decision rules of the form

$$\tau_1, \dots, \tau_n \Rightarrow \tau,$$

where $\tau_1, \dots, \tau_n$, and τ are given as linguistic terms considered as unary relations all of the same individual variable. To illustrate, where T=TALL, C=COORDINATED, M=MOTIVATED, and S=SUITABLE, the inference

T(X), C(X), M(X)=>S(X),

might be used to determine the suitability of an individual *X* for a basketball team. The deduction algorithm is defined in such a way that an individual's having a strong rating along one hypothesis will counterbalance that individual's having a weak rating along another. For example, even though individual *A* is only RATHER COORDINATED, if *A* is MOTIVATED and VERY TALL, then *A* should be SUITABLE. The deduction scheme in this way mimics the type of reasoning employed in multi-criteria decision making.

Here this method is extended to include terms that represent n-ary relations and to permit a few additional variations. One will now be able to express inferences such as

SIMILAR(X,Y) => SIMILAR(Y,X)

and, where PREF denotes a preference relation,

PREF(X, Y), PREF(y, *Z)* => PREF(A, *Z),*

and to make more specialized requirements on what such inferences should mean.

Let A = (T,U,M) be a linguistic variable with primary term λ. A *rank g* is assigned to each term $\tau \in \mathbf{T}$ according to: (i) if T consists of only the two terms $\mathrm{ant}(\lambda)$ and A, then $\varrho(\mathrm{ant}(\lambda))$ = —1 and $\varrho(\lambda)$ = 1, (ii) in all of the five other cases, $\varrho(\mathrm{med}(\lambda))$ = 0, the terms greater than med(λ) with respect to the ordering < are assigned in increasing order the successive positive integers 1,2,..., and the terms less than med(λ) are assigned in decreasing order the negative integers —1, —2, ... Then a *distance measure ó* may be defined on the terms r in T by

$$\delta(\tau, \tau') = \varrho(\tau') - \varrho(\tau).$$

This measure forms the basis for the inference algorithm. Where the sequence $X'_1, \dots, X'_m$ is a subsequence of $X_{1,1}, \dots, X_{1,m_1}, \dots, X_{n,1}, \dots, X_{n,m_n}$, let

$$\tau_1(X_{1,1}, \dots, X_{1,m_1}), \dots, \tau_n(X_{n,1}, \dots, X_{n,m_n}) \Rightarrow \tau(X'_1, \dots, X'_m)$$

be an inference composed of linguistic terms from the term sets for some linguistic variables $\Lambda_1, \dots, \Lambda_n, \Lambda$. These Λ's need not be distinct. Suppose that, for individuals $A_{1,1}, \dots, A_{n,m_n}$, it has been determined that all of

$$\tau'_1(A_{1,1}, \dots, A_{1,m_1}), \dots, \tau'_n(A_{n,1}, \dots, A_{n,m_n})$$

hold true, where $\tau'_1, \dots, \tau'_n$ are from the term sets for $\Lambda_1, \dots, \Lambda_n$. Then, the simplest scheme to be included under Inference Method A allows one to conclude

$$\tau'(A_1, \dots, A_m)$$

where τ' is the term from the term set for Λ for which the distance $\delta(\tau, \tau')$ is closest to

$$\sigma = \sum_{i=1}^{n} \delta(\tau_i, \tau'_i).$$

In practice it is likely that one will want to tailor this scheme to different situations. Such will in fact be necessary in case the term set for the inference's conclusion contains only the two terms A and ant(A). Here one must specify which of these to choose as τ' if $\sigma = 0$. Other modifications may also be appropriate under certain conditions, e.g., one might want special provisions in case some of the τ's in the inference are strictly less than rned(λ) in their respective term sets. An extreme modification would be to individually specify τ' for each possible choice of $\tau'_1, \dots, \tau'_n$, i.e., completely by-passing the use of the summation. It is envisioned that, in an implementation, each inference $\tau_1, \dots, \tau_n \Rightarrow \tau$ would be represented within a more complex data structure (akin to a *frame)* which includes an indication of the associated inference computation.

A variant of this method was discussed by Lee and Schwartz and Lee [1988]. There linguistic terms were reinterpreted via some "generic" possibility distributions, and the distance measure *6* was given as a horizontal distance between such distributions. This has the advantage that the distance measure to a certain extent reflects the shape of the distributions. It has the disadvantage, however, that the measure is not additive: for terms τ_1, τ_2, and T3 from the same linguistic variable, one does not in general have that $\delta(\tau_1, \tau_3) = \delta(\tau_1, \tau_2) + \delta(\tau_2, \tau_3)$. For the purposes of application in expert systems, it may turn out that this disadvantage outweighs the advantages.

Another variation, also considered in [Lee and Schwartz, 1988], is to apply weighting factors to the distances in the summation, reflecting that some of the hypotheses are more *important* than others. This further exploits well-known methods of multi-criteria decision analysis.

4 Inference Method B

Earlier versions of the ideas in this section have appeared as [Schwartz 1987 and 1988b]. For the present purposes, the concepts of Section 2 must be expanded to include a set of operators defined generally for all linguistic variables. In addition to proposing some intuitively plausible renditions of these operators, the definitions themselves serve to illustrate a general definition method. It should be clear that many further operators can similarly be introduced by these means.

Let A be a linguistic variable with term set T. The *expressions* of A are defined as follows: (i) terms in T are expressions of A, (ii) if *e* and *ε'* are expressions of A, then (ε OR *ε'), (ε* AND *ε'),* and NOT_t *ε* are expressions of A, (iii) if r is a term in T, then all of NOT_0 r, NOT_8 r, NOT_a r, NOT_v r, AT LEAST r, and NO MORE THAN r are expressions of A.

The expressions *e* of A are provided with a *relative meaning p(ε),* given as a subset of T. Here "relativity" is with respect to the set T. Let *r* € T, and let < be the ordering defined on members of T. For the special case that T does not contain the term med(λ), assume that statements such as r < med(A) are replaced by the analogous statements involving ranks, *p(r)* < 0.

For arbitrary r €T, set

p(r) = *{T}.*

Thus the relative meaning of any term is just the singleton composed of itself. For arbitrary expressions *ε,ε'* of A set

p(e OR ε') =: *p(ε)Up(ε'),*

p(ε AND *ε')* = *p(ε)np(ε'),*

p(KOT$_t$ε) = T-p(ε),

where U, ∩, and — are ordinary set operations.

By virtue of these interpretations, any subset T' of T is taken as representing the logical disjunction of the terms in T'. The special case that *p(ε)* = 0 is taken as saying that *e* is *contradictory* or *impossible.* It follows that the conjunction of any two distinct elementary terms is contradictory. The particular form of NOT defined above is called *total negation* and is interpreted as expressing "anything except *ε."* We now consider the other forms of NOT and the two adverbial phrases.

Ordered Negation. A very common form of negation is one wherein NOT r means "something less than *T"* if r is above med(A) and means "something more than r" if *r* is below med(A). This may be represented here as an operator NOT_o defined by

$$\rho(\mathrm{NOT_o}\ \tau) = \begin{cases} \{\tau' | \tau' < \tau\}, & \text{if } \tau > \mathrm{med}(\lambda); \\ \{\tau' | \tau' > \tau\}, & \text{if } \tau < \mathrm{med}(\lambda); \\ \text{undefined}, & \text{if } \tau = \mathrm{med}(\lambda). \end{cases}$$

Strong Ordered Negation. A form that is closely related to ordered negation is one in the sense of "not at all r", which would naturally apply only to either A or ant(A). This may be defined by

$$\rho(\mathrm{NOT_s}\ \tau) = \begin{cases} \{\tau' | \tau' < \mathrm{med}(\lambda)\}, & \text{if } \tau = \lambda; \\ \{\tau' | \tau' > \mathrm{med}(\lambda)\}, & \text{if } \tau = \mathrm{ant}(\lambda); \\ \text{undefined}, & \text{otherwise}. \end{cases}$$

A possible variant of this might additionally include a clause for "not at all med(A)," represented as a subset of T with a gap in the middle. How this is defined, however, may depend on which form of T is being employed.

Antonymical Negation. Another frequently used form of negation is the reference to the antonym of a term. This also will ordinarily be applied only to either A or ant(A). We have

$$\rho(\mathrm{NOT_a}\ \tau) = \begin{cases} \mathrm{ant}(\lambda), & \text{if } \tau = \lambda; \\ \lambda, & \text{if } \tau = \mathrm{ant}(\lambda); \\ \text{undefined}, & \text{otherwise}. \end{cases}$$

The "Not Very" Negation. In the literature on fuzzy logic, one frequently finds a term like NOT VERY TALL being interpreted as meaning the same thing as VERY SHORT. In ordinary English usage, however, this expression more often means something like RATHER TALL. Thus we may define

$$\rho(\mathrm{NOT_v}\ \tau) = \begin{cases} \text{r-}\lambda, & \text{if } \tau = \text{v-}\lambda; \\ \text{r-ant}(\lambda), & \text{if } \tau = \text{v-ant}(\lambda); \\ \text{undefined}, & \text{otherwise}. \end{cases}$$

The Adverbial Modifier "At Least." This operator may be delied by

$$\rho(\text{AT LEAST } \tau) = \begin{cases} \{\tau' | \tau' \geq \tau\}, & \text{if } \tau \geq \mathrm{med}(\lambda); \\ \{\tau' | \tau' \leq \tau\}, & \text{if } \tau < \mathrm{med}(\lambda). \end{cases}$$

An abbreviation for AT LEAST will be AL.

The Adverbial Modifier "No More Than." This may be defined by

$$\text{NO MORE THAN } \tau = \begin{cases} \{\tau' | \tau' \leq \tau\}, & \text{if } \tau \geq \text{med}(\lambda); \\ \{\tau' | \tau' \geq \tau\}, & \text{if } \tau < \text{med}(\lambda). \end{cases}$$

An abbreviation for NO MORE THSN will be NMT.

We may now consider the inferences for Method B. These have the same general form as for Method A, with the exception that the hypotheses and conclusion may be expressions. Thus, where T=TALL, S=SHORT, and A=ACCEPTABLE, one may have inferences such as

NOT_0 T AND NOT_0 S(X) => AT LEAST *A(X),*

and, where P denotes a preference relation,

$$\text{AL P}(X, Y), \text{ AL P}(Y, Z) \Rightarrow \text{AL P}(X, Z).$$

The associated deduction algorithm is as follows. Where the collection of variables $X_1', \ldots, X_m'$ is a subset of $X_{1,1}, \ldots, X_{1,m1}, X_{n1}, \ldots, X_{n,mn}$, let

$$\varepsilon_1(X_{1,1}, \ldots, X_{1,m_1}), \ldots, \varepsilon_n(X_{n,1}, \ldots, X_{n,m_n}) \Rightarrow \varepsilon(X_1', \ldots, X_m')$$

be an inference composed of expressions from some (not necessarily distinct) linguistic variables $\Lambda_1, \ldots, \Lambda_n, \Lambda$. A hypothesis $\varepsilon_i(X_{i,1}, \ldots, X_{i,m_i})$ is said to be *satisfied* for individuals $A_{i,1}, \ldots, A_{i,m_i}$ if we have that the instantiated expression $\varepsilon_i'(A_{i,1}, \ldots, A_{i,m_i})$ holds and

$$\rho(\varepsilon_i') \subset \rho(\varepsilon_i).$$

Then the algorithm for Scheme B has that, if all of the hypotheses are satisfied in this way by individuals $A_{1,1}, \ldots, A_{n,m_n}$, then one may conclude

$$\varepsilon(A_1', \ldots, A_m'),$$

where $A_1', \ldots, A_m'$ are the individuals corresponding to $X_1', \ldots, X_m'$. It is worth noting that, even though Prolog does not provide all the abstraction mechanisms described here, the general effect of this deduction method can be replicated in that language.

5 The Synthesis

Integrating the two inference methods into a unified system requires dealing with three separate but interrelated problems: forward chaining, backward chaining, and evidence combination. Due to lack of space, only the first of these will be discussed. Suppose we have two inferences

$$\tau_1, \ldots, \tau_n \Rightarrow \tau \text{ and } \tau_1', \ldots, \tau_n' \Rightarrow \tau'$$

where the conclusion *r* of the former is composed of term(s) from the same linguistic variable as one of the hypotheses τ_i' of the latter. First consider the case that the former inference is of type A and the latter is of type B. Then the conclusion of the former is a term, and, since the relative meaning of terms is defined, there is no problem in determining whether τ_i' is satisfied. Thus, for this situation, the mechanism for forward chaining is already provided.

Second suppose that the former inference is of type B and the latter of type A. Then τ will in general be represented as a subset of the associated term set T, while the deduction algorithm for the latter inference involves a distance measure which is defined only between terms. This shortcoming is easily remedied by simply extending the distance measure *6* to a measure δ_s defined on the subsets of T: where T_1, T_2 C T, set

$$\delta_s(T_1, T_2) = \min_{\tau_1 \in T_1, \tau_2 \in T_2} \delta(\tau_1, \tau_2)$$

or, alternatively, replace min with max. This distance measure may be applied to the relative meanings of expressions. In order for this to make intuitive sense, it will typically be necessary to restrict such meanings to be only those which consist of unbroken sequences of linguistic terms, i.e., sets of terms which do not contain gaps.

Backward chaining will be more complex than in Prolog, but should nonetheless be tractable. Evidence combination will have several different schemes, with the choice depending partly on the particular linguistic variable and partly on the particular inference scheme being employed.

References

[Baldwin, 1987] J. Baldwin, Evidential support logic programming. *Fuzzy Sets and Systems,* 24 (1987) 1-26.

[Dubois and Prade, 1979] D. Dubois and H. Prade, *Fuzzy sets and systems: theory and applications,* New York: Academic Press, 1979.

[Dubois and Prade, 1987] D. Dubois and H. Prade, Necessity measures and the resolution principle. *IEEE Transactions on Systems, Man and Cybernetics,* 17 (1987) 474-478.

[Dubios and Prade, 1988] D. Dubios and H. Prade, *Possibility theory: an approach to computerized processing of uncertainty,* E.F. Harding (trans.), New York: Plenum Press, 1988.

[Dubios and Prade, 1989] D. Dubois and H. Prade, Resolution principles in possibilistic logic. *International Journal of Approximate Reasoning,* to appear.

[Lee and Schwartz, 1988] M.-J. Lee and D. G. Schwartz, A scheme for logical inference with linguistic attributes, *Proceedings of the Florida Artificial Intelligence Research Symposium, FLAIRS-88,* Orlando, Florida, March 4-6, 1988, pp. 208-212.

[Pearl, 1988] J. Pearl, *Probabalistic reasoning in intelligent systems: networks of plausible inference,* San Mateo, CA: Morgan Kaufmann, 1988.

[Schwartz, 1987] D. G. Schwartz, A model of expert decision making with fuzzy criteria. *Second Congress of the International Fuzzy Systems Association, IFSA-87,* Tokyo, Japan, July 20-25, 1987, pp. 622-624.

[Schwartz, 1988a] D. G. Schwartz, A computationally simple interpretation of linguistic decision rules. *Proceedings of the North American Fuzzy Information Processing Society, NAFIPS-88,* San Francisco, California, June 8-10, 1988, pp. 216-220.

[Schwartz, 1988b] D. G. Schwartz, An alternative semantics for linguistic variables. *Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge Based Systems, IPMU-88,* Urbino, Italy, July 3-8, 1988, pp. 87-92.

[Shafer, 1976] G. Shafer, *A mathematical theory of evidence,* Princeton, New Jersey: Princeton University Press, 1976.

[Zadeh, 1965] L. A. Zadeh, Fuzzy sets. *Information and Control,* 8 (1965) 338-353.

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

[Zadeh, 1978] L. A. Zadeh, PRUF—a meaning representation language for natural languages, *International Journal of Man-Machine Studies* 10 (1978) 395-460.

[Zadeh, 1986] L. A. Zadeh, Outline of a theory of usuality based on fuzzy logic. In A. Jones, A. Kaufmann, and H.-J. Zimmerman (eds.), *Fuzzy Sets Theory and Applications,* Dordrecht: Reidel, 1986, pp. 79-97.