International Journal of Approximate Reasoning 52 (2011) 554–561

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

journal homepage: www.elsevier.com/locate/ijar

Reasoning with doubly uncertain soft constraints

Ronald R. Yager

Machine Intelligence Institute, Iona College, New Rochelle, NY 10801, United States

A R T I C L E I N F O

Article history:
Received 16 November 2009
Received in revised form 24 December 2010
Accepted 5 January 2011
Available online 12 January 2011

Keywords:
Soft information
Fuzzy sets
Inference
Generalized constraints

A B S T R A C T

We describe the basic ideas of the theory of approximate reasoning and indicate how it provides a framework for representing human sourced soft information. We discuss how to translate linguistic knowledge into formal representations using generalized constraints. We consider the inference process within the theory of approximate reasoning and introduce the entailment principle and describe its centrality to this inference process. Next we introduce the idea of doubly uncertain statements such as John's friend is young. In these statements there exists uncertainty both with respect to value of the age, young, and the object associated with the age, John's friend. We suggest a method for representing these complex statements and investigate the problem of making inferences about specific objects.

1. Introduction

Zadeh [1–4] introduced the fuzzy set based theory of approximate reasoning (AR) and the related generalized theory of uncertainty (GTU) to provide a technology for manipulating human sourced linguistically expressed soft knowledge. Useful introductory surveys on the theory of approximate reasoning can be found in [5,6]. Many applications based on these ideas can be found in the literature [7–12]. An important component of this framework is its ability to represent imprecise and uncertain information using the concept of a generalized constraint statement. A typical example of such a statement is Johns is near the river, where we are assigning the variable John's location the value near the river. In this statement uncertainty is associated with the term near-river, the value of the variable. Here we shall extend the capability of this framework by considering an additional source of uncertainty in the generalized constraint statement, uncertainty with respect to variable itself. Formally the generalized constraint statement is of the form Attribute(Object) is Value. In the preceding the attribute is location, the object is John and the value is near-river. In this work we shall consider situations in which there exists some uncertainty with respect to object itself as exemplified by the statement ***a tall men*** *is near the river*. Here the attribute is still location and the value is still near-river but the object is an uncertain object, a tall men. Our goal here is to try to represent these doubly uncertain constraints and infer any information we can about specific objects, such as John's location. We note that closely related issues have considered within the area of probabilistic databases [13].

2. The theory of approximate reasoning

To provide a technology for manipulating human sourced linguistically expressed soft knowledge Zadeh introduced the fuzzy set based theory of approximate reasoning (AR) and the generalized constraint language [1–4]. Fundamental to this approach is the realization that much of our knowledge can be viewed as a constraint on some explicit or implicit variable. The process introduced by Zadeh for manipulating this type of knowledge can be seen to involve four basic steps. The first step is the translation of the available knowledge into the representational language of the theory of approximate reasoning

E-mail address: yager@panix.com

0888-613X/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2011.01.001

suggested by Zadeh. This can be seen as a step of translation. In this step each of the pieces of knowledge are translated into constraints. The second step involves the fusing of the individual pieces of knowledge to form a global constraint that constitutes our knowledge base. The third step involves the making of inferences in response to some objective. The final step is one of retranslation. This step involves taking the inferred knowledge, which is a constraint expressed in mathematical form and turning it into some natural language statement that is easier for a human to understand.

Central to this approach is the task of translation. Here we take knowledge convert it into constraints on some implicit or explicit variable. Numerous recommendations have been suggested for accomplishing this translation [4].

An important class of constraints are possibilistic constraints. These constraints arise from statements such as ***John is tall***. Here we are imposing some constraints on the possible age of John. At a formal level if V is a variable with a domain of discourse X then the formal representation of a possibilistic constraint is

$$V \text{ is } A.$$

Here A is a fuzzy subset of the domain X. Often A is a representation of some linguistic value associated with V. The effect of this constraint is to induce a possibility distribution Π over X where $\Pi(x) = A(x)$ is the possibility that x is the value of V. Here $A(x)$ is the membership grade of x in A.

Two special cases of V is A are worth noting. The first is the case where $A(x) = 1$ for all $x \in X$, here $A = X$. In this case we know nothing about V. Another special case is where $A(x^*) = 1$ and $A(x) = 0$ for $x \neq x^*$. In this case we have that $V = x^*$, it equals the value x^*.

We shall say a fuzzy set is normal if there exists at least one element with membership one. If this is not the case we say it is subnormal. In the framework of Zadeh's theory of approximate reasoning subnormality is some indication of conflict or inconsistency. An extreme case of this occurs when $A = \emptyset$, the null set.

One can consider constraints involving multiple variables such as (V, U) is H. If X is the domain of V and Y the domain of U then H is a fuzzy subset of $X \times Y$. Here we refer to (V,U) as a joint variable. We note that propositions involving joint variables are simultaneously constraining both variables. Thus the statement (V,U) is H is telling us that $H(x, y)$ is the possibility that simultaneously V is x and U is y. More generally we can consider constraints involving joint variables having any number of atomic variables such as $(V_1, V_2, V_3, \ldots, V_n)$ is H.

At times it is useful to extend a proposition involving a single variable to one involving a joint variable. Assume we have the proposition V is A we can naturally extend it to one involving a joint variable (V, U) by simply replacing V is A by V is A and U is Y. The statement U is Y, where Y is the domain of U is essentially saying U can be anything. In this case we get (V, U) is G where $G(x, y) = \text{Min}[A(x), Y(y)] = A(x)$. A significant implication here is that any two propositions can always use expressed in terms of the same joint variables. This can be done by extending each of the propositions with the variables from the other proposition.

Zadeh [1] discusses an approach for projecting the value of a constituent variable from a joint variable. Assume we have (V, U) is H where H is a fuzzy subset of $X \times Y$. From this we can obtain V is F where $F(x) = \text{Max}_y[H(x, y)]$ and U is E where $E(y) = \text{Max}_x[H(x, y)]$. We denote F as the $\text{Proj}_V[(V, U) \text{ is } A]$ and E is $\text{Proj}_U[(V, U) \text{ is } H]$. We say the proposition (V, U) is H is clean in V if $F(x) = 1$ for all x. Here $F = X$, H has no individual information about V.

Given a proposition V is A one may be interested in the truth of some other proposition such as V is B. Consider that we know John is in his twenties. We can clearly answer no to the question is John over 50. We can similarly clearly answer yes to the question is John older than 15. But the question is John over 25 is not directly answerable. It is possibly true but not necessarily true. In the face of uncertain information many questions do not have clear answers.

In the light of this two surrogate measures have been introduced [14]. The first is the measure of possibility defined as $\text{Poss}[V \text{ is } B|V \text{ is } A] = \text{Max}_x[A(x) \wedge B(x)]$. The second is the measure of certainty (or necessity) $\text{Cert}[V \text{ is } B|V \text{ is } A] = 1 - \text{Poss}[V \text{ is } \overline{B}|V \text{ is } A]$. It is well known that if A is normal that $\text{Poss}[V \text{ is } B|V \text{ is } A] \geqslant \text{Cert}[V \text{ is } B|V \text{ is } A]$.

In the following we shall find it at times convenient to use the generic formulation $\mathbf{M}$ is H to indicate a possibilistic constraint in which we have a joint variable.

In Zadeh's approximate reasoning each piece of knowledge (proposition) P induces a constraint C expressed in the language of AR, symbolically we denote this as $P \Rightarrow C$. When we have multiple pieces of knowledge, $\{P_1, \ldots, P_q\}$, the combined constraint induced by these is the conjunction of their individual constraints, if $P_i \Rightarrow C_i$ then $\{P_1, \ldots, P_q\} \Rightarrow C_1 \cap C_2 \cdots \cap C_q$. Thus if we have $P_1 \Rightarrow V$ is A_1, $P_2 \Rightarrow V$ is A_2 and $P_3 \Rightarrow V$ is A_3 then $\{P_1, P_2, P_3\} \Rightarrow V$ is H where $H = A_1 \cap A_2 \cap A_3$, $H(x) = \text{Min}[A_1(x), A_2(x), A_3(x)]$.

In the case where the pieces of knowledge are about different variables $P_1 \Rightarrow V_1$ is B_1, $P_2 \Rightarrow V_2$ is B_2 and $P_3 \Rightarrow V_3$ is B_3 than $\{P_1, P_2, P_3\} \Rightarrow \mathbf{M}$ is H where $\mathbf{M} = (V_1, V_2, V_3)$ and H is a fuzzy subset over $X_1 \times X_2 \times X_3$, the cartesian product of the domains of the V_j such that

$$H(x_1, x_2, x_3) = \text{Min}[A_1(x_1), A_2(x_2), A_3(x_3)].$$

More generally if $P_1, \ldots, P_q$ are a collection of propositions each inducing a constraint $\mathbf{M}_i$ is H_i where $\mathbf{M}_i$ is a collection of variables then the collection $\{P_1, \ldots, P_q\}$ induces the constraint $\mathbf{M}$ is H, where $\mathbf{M} = \cup_i \mathbf{M}_i$ and $H = \cap_i H_i$. Thus the fusion of multiple sources of information results in a constraint on the union of all the variables and the conjunction of their values.

We make the following observation. Assume $\{P_1, \ldots, P_q\}$ is a set of propositions that induces the constraint $\mathbf{M}$ is H. Let $\{P_{\pi(1)}, \ldots, P_{\pi(r)}\}$ be any subset of $\{P_1, \ldots, P_q\}$. Assume $\{P_{\pi(1)}, \ldots, P_{\pi(r)}\}$ induces the constraint, $\mathbf{M}$ is G. Then we can easily show that $H \subseteq G$. We recall $H \subseteq G$ if for each element x in the domain of $\mathbf{M}$ we have $H(x) \leqslant G(x)$.