Computers Math. Applic. Vol. 19, No. 11, pp. 105–119, 1990
Printed in Great Britain.

0097-4943/90 $3.00 + 0.00

COMPUTATIONAL MODELS OF UNCERTAINTY REASONING IN EXPERT SYSTEMS

J. F. BALDWIN
Department of Engineering Mathematics, University of Bristol, Bristol BS8 1TR, England

Abstract—The use of support pairs associated with the facts and rules of a knowledge base of an expert system to capture various aspects of inductive reasoning is discussed. The concept of semantic unification is introduced with reference to fuzzy sets theory. In this respect a probabilistic interpretation for this semantic unification is described using a population voting model. Examples are discussed including default reasoning using support logic.

INTRODUCTION

In this paper the use of support pairs associated with Prolog type clauses to capture various aspects of inductive reasoning will be discussed. This represents a support logic programming system and it has been implemented in the form of the language Fril [1]. Fril can operate as a pure Prolog system if no uncertainties are involved. False facts, equivalent statements and a true logic negation can be used.

Fril is being used on a variety of applications involving reasoning with uncertainty. These include AI applied to scene analysis, medical and fault diagnosis, expert systems in command and control, analogical reasoning, probabilistic grammars, program evaluation etc.

Support logic programming [2–5] is an evidential reasoning system which, rather than proving theorems, collects evidence to support an hypothesis and also to support the negation of the hypothesis. These supports do not have to add up to unity. If no evidence is available then a support pair (0 1) is returned corresponding to zero support for and zero support against, i.e. total uncertainty. A logic of support has been studied under various names by different people. Koopman called it a logic of intuitive probability and Carnap a theory of confirmation [6]. These and other theories are based on a single number representing the supports and some are comparitive only.

Zadeh's fuzzy sets theory forms a basis for a possibility theory [7]. Bellman made contributions to the analytical formalism of the theory of fuzzy sets [8]. Support logic programming uses this theory. Fuzzy sets can be used as the referents of concepts and semantic unification is used to match two fuzzy terms which are syntactically different but which semantically have something in common. An interpretation of fuzzy sets which shows how semantic unification can be given a probabilistic interpretation, necessary for support logic programming, is given below. The calculus of Fril is consistent with probability theory and this supposes that all propositions are true or false, although it may not be possible to acquire evidence which will allow a probability of 1 or 0 to be obtained.

Expert systems and other knowledge engineering systems such as vision understanding and speech recognition programs must be able to cope with knowledge bases with incomplete information. Incomplete information can be of two types: one type concerns lack of data and the other, lack of concept definition.

For example, in a vision system only part of the object may be in view because of occlusion. This gives rise to uncertainty concerning what the object may be since the part that can be seen could belong to several different objects. Possible extensions of the part of the object will give rise to different interpretations and each extension will have a probability associated with it. This probability is assessed from other evidence picked up from other parts of the picture and may not be able to be assessed with complete accuracy. It may be possible to assess its value as being contained within a certain interval.

On the other hand, the whole object may be in complete view but it is still difficult to say exactly what it is. For example, if the object in total view was a bushy tree like object, no complete support could be given for it being a tree or for it being a bush, simply because of a lack of precise definition for bush and tree. Most concepts which we use in our daily lives are of this nature. For example we cannot prescribe sufficient and/or necessary conditions for classifying what we mean by a "humane society", "a good business venture", "a comfortable seat", "a well-structured program", "a stable system", "a reliable system" etc. Even accepting that these definitions are context dependent, we will still have difficulty in giving exact definitions within a given context. Furthermore, the relevant context may again be a border line decision.

In law, cases are often looked at by considering similar cases from the past where the judgements are known. The present case may not precisely fit any of the historical ones but only have similarities with each. The judgement on the present case will then be influenced in some form by the combination of those judgements of similar past cases. What form this so called combination should take is not at all obvious.

Heuristics used by experts are probabilistic in nature. Truth is not guaranteed when certain conditions hold. Difficulties of entailment when true propositions are replaced by highly probable propositions are well-known. Contraposition is valid for deductive entailments but does not hold for high probabilities. Deductive entailment is transitive but strong inductive support is not. More importantly the following valid argument of deductive logic does not carry over into the inductive case. If A entails B then A AND C entails B. In fact if $\Pr(B \mid A)$ is high, this provides no constraint on the value of $\Pr(B \mid A, C)$ since

$$\Pr(B \mid A, C) = \frac{\Pr(C \mid A, B) \cdot \Pr(B \mid A)}{\Pr(C \mid A)}.$$

This, of course, is obvious from sample space considerations. All relevant criteria must be considered when giving supports to predicates as suggested by Hempel's maximum specificity conditions [9].

Causal connections are important in expert systems. Any sensible theory of causation is probabilistic. Frequent conjunctions often occur, constant conjunctions rarely [10]. The modelling of causality is also discussed in Ref. [11].

Probabilistic reasoning can be viewed as constraint reasoning in which the various probabilistic statements given provide evidence to constrain the probability of another statement to be contained within a certain interval.

If we know that:

$$\Pr(P \to Q) = 2/3,$$
$$\Pr(P) = 4/5,$$

what can we conclude about $\Pr(Q)$? A point value probability cannot be determined since the above two probabilistic statements gives insufficient information for this. The statements constrain the interval which contains $\Pr(Q)$. Three possible cases must be analysed, since the case corresponding to NOT($P \to Q$) AND (NOT P) is not possible because of the inconsistency of the two statements.

Case 1	Case 2	Case 3
$P \to Q$ P	NOT($P \to Q$) P	$P \to Q$ NOT P
Q	NOT P	World 1: Q World 2: NOT Q
Q: {1, 1}	Q: {0, 0}	Q: {0, 1}
x_1	x_2	x_3

where x_i is Pr(case i) and $\{a, b\}$ means that NEC(Q) = a, POS(Q) = b, where $a = 0$ if NEC(Q) is false and 1 if it is true, and $b = 0$ if POS(Q) is false and 1 if it is true. NEC and POS are modal logic operators. Then since

$$x_1 + x_2 + x_3 = 1,$$
$$x_1 + x_2 = 4/5,$$

since $\Pr(P) = \Pr(P \text{ AND } P \rightarrow Q) + \Pr(P \text{ AND NOT } (P \rightarrow Q))$

$$x_1 + x_3 = 2/3,$$

since $\Pr(P \rightarrow Q) = \Pr(P \rightarrow Q \text{ AND } P) + \Pr(P \rightarrow Q \text{ AND NOT } P)$ so that

$$x_1 = 7/15, \quad x_2 = 1/3, \quad x_3 = 1/5.$$

Hence $\Pr(\text{NEC } Q) = 7/15$ and $\Pr(\text{POS } Q) = 7/15 + 1/5 = 2/3$. From $\Pr(\text{NEC } Q) \leqslant \Pr(Q) \leqslant \Pr(\text{POS } Q)$ it follows that $\Pr(Q)$ lies in the interval $[7/15, 2/3]$.

This interval can be determined using linear programming formulations and this is discussed in Ref. [5].

SUPPORT PAIRS

The theory of uncertainty which forms the basis of support logic programming is based on the association of support pairs with Horn clauses as used in Prolog.

Any proposition P is assumed to be true or false. A two valued logic is assumed. There is no mention of truth values lying between 0 and 1. Furthermore any valid formula of first order logic, F say, will be such that there is support of 1 for and 0 against.

Evidence, E, is used to assign a necessary support, $\text{Sn}(P \mid E)$ for, and a necessary support, $\text{Sn}(\text{NOT } P \mid E)$ against any proposition P being true. Possible supports $\text{Sp}(P \mid E)$ and $\text{Sp}(\text{NOT } P \mid E)$ are defined as

$$\text{Sp}(P \mid E) = 1 - \text{Sn}(P \mid E); \quad \text{Sp}(\text{NOT } P \mid E) = 1 - \text{Sn}(\text{NOT } P \mid E).$$

These can be further interpreted in terms of the modal logic necessity and possibility operators, namely

$$\text{Sn}(P \mid E) = \Pr(\text{NEC } P \mid E); \quad \text{Sp}(P \mid E) = \Pr(\text{POS } P \mid E),$$

where modal operators are to be understood in the context of possible world semantics.

The belief that the truth value 1 can be assigned to P using evidence E, $\Pr(P \mid E)$, lies in an interval determined by the necessary and possible supports for P:

$$\Pr(P \mid E) \text{ lies in } [\text{Sn}(P \mid E), \quad \text{Sp}(P \mid E)].$$

It is necessarily true that

$$\text{Sn}(P \text{ AND NOT } P) = 0 \quad \text{and} \quad \text{Sp}(P \text{ AND NOT } P) = 0,$$

$$\text{Sn}(P \text{ OR NOT } P) = 1 \quad \text{and} \quad \text{Sp}(P \text{ OR NOT } P) = 1.$$

SUPPORT LOGIC PROGRAMMING

A *support logic program* consists of a sequence of *support clauses.*

A *support clause* is a *clause* with an associated *support structure.*

A *clause* is a *list* of one or more *atoms.*

An *atom* is an *atomic formula* which is a *list* whose first element is a *predicate* symbol or a *relation* and the remaining elements are *terms.*

A *term* is a *number, constant, variable* or *list.*

The elements of a *list* are *terms.*

A *support structure* can be a single *support pair* or a list of two *support pairs.*

A *support pair* is a list of two elements; the first element being called the *necessary support* and the second element the *possible support.*

Variables, constants, numbers and lists have their usual meaning.

Support clauses can be further divided into *simple support clauses* and *compound support clauses.*

An example of a simple support clause is:

((com12 is a large senate committee)):(0.6 0.8),

which could mean that the degree of belief that com12 is a large senate committee is some number in the interval [0.6 0.8]. The doubt expressed by using this interval arises because of the imprecise definition of large. This support pair could be determined by asking a large representative sample of university members to vote whether they accepted that senate committees of various sizes were large. The vote could be "yes", "no" or "abstain" for each size presented. The proportion who vote "yes" for a given size would represent the necessary support for it being a large committee. This number plus the number of abstentions would give the possible support. The number of "no's" would give the necessary support against and this with the abstentions would give the possible support against. Of course, the doubt could also arise because of the uncertainty in the actual numbers on a committee. The final support pair used must take account of both these cases of uncertainty and this could be done using more rules to determine the support pair.

An example of a compound statement with a single support pair is:

((committee *C* contains a professor)
(*C* is a large senate committee)):(0.9 1),

which says that at least 90% of large senate committees contain a professor. In other words the conditional probability Pr(committee *C* contains a professor | *C* is a large committee) lies in the interval [0.9 1]. In Prolog terms this corresponds to a fact.

The simple clause

$$((p)):(0\ 0)$$

says that p is false.

An example of a compound statement with two support pairs is:

((performance *X* good)
(engineers_report *X* ok)
(efficiency *X* near_optimal)):((0.9 1)(0 0.2)),

which says that if the body of the rule, in this case the conjunction of the two atoms (engineers_report *X* ok) and (efficiency *X* near_optimal), is true then the probability that the performance of *X* is good lies in the interval [0.9 1], while if the body is false this probability lies in the interval [0 0.2].

A special case of this rule, namely

$$((p)(q)):((1\ 1)(0\ 0))$$

says that p is equivalent to q.

THE CALCULUS OF SUPPORT LOGIC

The calculus used in support logic programming is fully described in Ref. [5]. We will not repeat this here but discuss a simple example to illustrate the main points of the calculus.

Since propositions are assumed to be either true or false, assuming one accepts the scoring argument of De Finetti and Lindley [12] then if the support pairs correspond to single numbers, a probability calculus must be used. The calculus for the support pairs is then easily determined since probabilities are contained within intervals determined by the support pairs. A unique probability is not determined and a simple constrained optimisation problem gives the required support pair for any compound statement in terms of the support pairs of its parts.

Consider the following example in which we know that

$$\Pr(a \mid q) = 0.5, \quad \Pr(a \mid \text{NOT } q) = 0.4,$$

$$\Pr(a \mid s) = 0.8, \quad \Pr(a \mid \text{NOT } s) = 0.4,$$

$$\Pr(q) = 0.7,$$

$$\Pr(s) = 0.175,$$

then we can determine Pr(a) in two ways, namely

$$\Pr(a) = \Pr(a \mid q)\cdot\Pr(q) + \Pr(a \mid \text{NOT } q)\cdot\Pr(\text{NOT } q) = 0.47,$$

$$\Pr(a) = \Pr(a \mid s)\cdot\Pr(s) + \Pr(a \mid \text{NOT } s)\cdot\Pr(\text{NOT } s) = 0.47.$$

If the answers in each case had been different, we would have concluded that the knowledge base was inconsistent.

The problem expressed in Fril is:

```
((a) (q)):((0.5 0.5)(0.4 0.4)),
((a) (s)):((0.8 0.8)(0.4 0.4)),
((q)):(0.7 0.7),
((s)):(0.175 0.175),
```

and the query

```
qs((a)),
```

yields the solution

(0.47 0.47).

Fril uses the two proof paths to provide the answer to the query and gives the answer (0.47 0.47) in each case. These intervals are intersected to give (0.47 0.47) as the final solution.

We will now consider a modified problem in which the point probabilities are not precisely known.

Pr($a \mid q$) is in [0.45, 0.55], Pr($a \mid$ NOT q) is in [0.35, 0.45],
Pr($a \mid s$) is in [0.75, 0.85], Pr($a \mid$ NOT s) is in [0.35, 0.45],
Pr(q) is in [0.65, 0.75]
Pr(s) is in [0.1, 0.2]

We can use the theorem of total probability as before to obtain $P(a)$ but we must use interval arithmetic. The two methods give [0.38, 0.57] and [0.355, 0.545], respectively for Pr(a). Any point in the final interval containing the point probability Pr(a) must lie in both these intervals using a consistency argument. Therefore we must intersect the intervals to obtain the final interval. This defines the rule of how solutions are combined from different proof paths in Fril. For this case the final answer for $Pr(a)$ is that it is contained in the interval [0.38, 0.545].

The Fril program for this case is

```
((a)) (q)):((0.45 0.55)(0.35 0.45)),
((a) (s)):((0.75 0.85)(0.35 0.45)),
((q)):(0.65 0.75),
((s)):(0.1 0.2),
```

and the query

```
qs((a)),
```

yields the solution

(0.38 0.545).

The basic rule used to combine support pairs from different proof paths is the intersection rule. An alternative method of combining proof paths is available in Fril and corresponds to using a Dempster type rule [13]. This should only be used when the proof paths correspond to independent viewpoints. In this case conflicts can occur and the Dempster rule is one way of resolving the conflicts. If the user has some other way he wishes to combine solutions from different viewpoints he can express this as a rule in Fril.

Nothing has been said about finding the support pairs of a conjunction or disjunction when given support pairs for each atom. The rules used for Fril are consistent with probability theory.

DEFAULT REASONING

Consider the Fril program:

```
((live_another_five_years X)
      (english X)
      (age X 30)
      (not suffers_from_lung_cancer X)):(0.9 1)
((live_another_five_years X)
      (english X)
      (age X 30)
      (suffers_from_lung_cancer X)):(0 0.1).
```

If we do not know anything about the health of a 30-year old English person, these two rules will use ((suffers_from_lung_cancer person)):(0 1) and conclude (0 1) as the support pair for (live another_five_years person). Intuitively we may feel that an answer something like (0.85 1) should have been given, since we know that most 30-year old Englishmen do live another five years. We could therefore add the additional rule to our program:

```
((live_another_five_years X)
      (english X)
      (age X 30)):(x y),
```

where x and y are chosen appropriately. What is appropriate? Strictly $(x\ \ y) = (0\ \ 1)$ to be consistent with the other two rules. But this would not satisfy the reason we are introducing this rule. We will choose $(x\ \ y) = 0.85\ \ 1)$.

If nothing is known about the health of the person then Fril will use each of the rules, obtaining (0 1) from the first two and (0.85 1) from the last giving the answer (0.85 1). If the second rule is applicable then rules 2 and 3 will give inconsistent answers. Fril recognizes this and because the body of rule 3 is contained in the body of rule 2, ignores rule 3 and uses the first two rules only. This is a consequence of the maximum specificity requirement.

Non-monotonic logic is used to avoid problems like this but these logics have inconsistencies [14]. By using Fril there is no reason to introduce these various additional logics and default reasoning. In the case of the standard problem that "all birds can fly", "a penguin is a bird" "a penguin cannot fly" it is, of course, false to say that all birds can fly. Most birds can fly so that if all that is known is that X is a bird there is a high probability that it can fly. This will not be the case if X is a penguin and is treated as the above problem. Details of this and similar problems can be found in Refs [3, 5].

A recursive definition of a concept "tall", for example, can be written in Fril. This uses the fact that if you remove a little height from a tall man there is still a high support, but not a certain support, for him still being tall. Details can be found in Refs [1, 4].

VOTING MODEL INTERPRETATION OF A FUZZY SET

A fuzzy subset f with respect to the set F is defined by means of a membership function $\mathrm{Mf}: F \rightarrow [0,\ 1]$. In other words an element, e, of the set F belongs to the fuzzy subset f with a degree of membership $\mathrm{Mf}(e)$. How can we interpret this membership level in more specific terms which will give some justification to its actual value and also its existence and use?

One possible interpretation is in terms of the voting behaviour of a population P, say, of persons, all of whom have their own understanding of the meaning of f. We all use the term "tall" in relation to a person's height without having a precise understanding of what it means. When we use the word in ordinary conversation, we assume others will be able to interpret it in more or less the same way as ourselves. It is certainly true that there is a set of heights which everyone would accept as satisfying the concept of "tall height" and there is a set of heights which every one would accept as not satisfying this concept. The difficulty arises for the set of heights in between these two sets. If this intermediate set is null then we have an exact definition for "tall" but otherwise we do not. Each member of this intermediate set can have a degree of membership in the set of heights representing "tall", but how do we choose the actual degree?

Consider a set F and let f be a fuzzy subset of this set. Let each person belonging to a population P vote on whether to accept or reject the membership of a given element e of set F as belonging to f. Each person must accept or reject, agree or not agree to the elements membership. Abstentions and partial agreements are not allowed. Mf(e) is equated to the proportion of persons who vote for accepting e as a member of f. Similarly (1 − Mf(e)) is the proportion of persons who reject e as belonging to f so that this is the membership level for e not belonging to f. Each individual will have a threshold level such that if his doubt in an element e belonging to f increases above this level he will reject its membership, otherwise he will accept it. It is similar to a member of a jury having to say guilty or innocent for the person on trial. The evidence presented at the trial may not be conclusive but the person must still make a final judgement. If people are alowed to abstain we return to an analogue with support pairs.

Example

Let F be the set of positive integers {50 55 60 65 70 75} and f be defined by the following membership function

$$\mathrm{Mf}(55) = 0.2,\quad \mathrm{Mf}(60) = 0.5,\quad \mathrm{Mf}(65) = 0.8,\quad \mathrm{Mf}(70) = 1,\quad \mathrm{Mf}(50) = \mathrm{Mf}(75) = 0.$$

The fuzzy set f can therefore be represented as

$$f = 55|0.2 + 60|0.5 + 65|0.8 + 70|1.$$

If we take P as made up of 10 people then the voting pattern to give consistency with this definition could be

Person									
1	2	3	4	5	6	7	8	9	10
70	70	70	70	70	70	70	70	70	70
65	65	65	65	65	65	65	65		
60	60	60	60	60					
55	55								

The integers given are those integers which the person accepted as satisfying f. Therefore person 1 accepts {70, 65, 60, 55} as satisfying f while person 7 only accepts {70, 65}.

This interpretation assumes that persons who vote yes for 55 also vote yes for 60 and for 65. Similarly it assumes persons who vote yes for 60 vote yes for 65. The assumption that this interpretation uses is that a person who votes for an element h in the set F with membership value Mf(h) as belonging to f will also vote for any other element of F satisfying f if it has a higher membership value than Mf(h). We will call this the constant threshold model since it corresponds to each person having a threshold level for acceptance of an element of F in f which does not vary with the element of F chosen.

An alternative interpretation could be:

Person									
1	2	3	4	5	6	7	8	9	10
70	70	70	70	70	70	70	70	70	70
65	65	65	65	65	65	65	65		
60	60	60	60	60					
								55	55

Other possible interpretations can be given but the one which intuitively seems more reasonable is the constant threshold model.

INTERSECTION AND UNION OF FUZZY SETS

Consider two fuzzy sets $f1$, $f2$, with membership functions Mf1, Mf2, both defined as fuzzy subsets of the set F.

Consider an element h of F. The proportion of persons of population P who vote for h satisfying both the concepts defined by $f1$ and $f2$ is contained in the interval

$$\text{Iconj} = [\max\{\text{Mf1}(h) + \text{Mf2}(h) - 1, 0\}, \quad \min\{\text{Mf1}(h), \quad \text{Mf2(h)}\}].$$

If we use the constant threshold model then one assumes that the threshold levels of the persons P stay constant for judging different concepts. This means that if a person votes yes for one concept when having a certain degree of doubt, that person will also vote yes for another concept if faced with the same degree of doubt. In this case we assume that those people who voted yes for the concept with the lower membership value will also have voted yes for the other. Thus for this assumption the membership value for element h for the intersection of $f1$ and $f2$ will be $\min\{\text{Mf1}(h), \text{Mf2}(h)\}$.

Similarly for the union of $f1$ and $f2$ the membership value for element h will lie in the

$$\text{Idisj} = [\max\{\text{Mf1}(h), \quad \text{Mf2}(h)\}, \quad \min\{\text{Mf1}(h) + \text{Mf2}(h), \quad 1\}].$$

With the same assumption as above, the minimum number of persons would vote yes for the union, so that the membership level for element h belonging to the union of $f1$ and $f2$ is $\max\{\text{Mf1}(h), \text{Mf2}(h)\}$.

This is the assumption we make in Fril for fuzzy sets and is the usual definition for fuzzy conjunction and disjunction.

More generally we can define a mapping T

$$T\colon [0, 1] * [0, 1] \rightarrow [0, 1]$$

which satisfies the axioms

(1) $T(a, 1) = a$,
(2) $T(a, b) = T(b, a)$,
(3) $T(a, b) \geqslant T(c, d)$ if $a \geqslant c$ and $b \geqslant d$,
(4) $T(a, T(b, c)) = T(T(a, b), c)$.

T is called a T-norm and generalizes the AND corresponding to conjunction

Examples of instances of T are

(1) $T(a, b) = \min\{a, b\}$,
(2) $T(a, b) = a \cdot b$,
(3) $T(a, b) = \max\{a + b - 1, 0\}$.

A dual norm, called the T-conorm, S, exists which generalizes disjunction. For any T-norm T there exists a dual norm S such that

$$S(a, b) = 1 - T((1 - a), (1 - b)).$$

The mapping

$$S\colon [0, 1] * [0, 1] \rightarrow [0, 1].$$

satisfies the same axioms as for T except that (1) is replaced by (1′) where

(1′) $S(a \ 0) = a$.

Examples of instances of S conorms corresponding to the T norms (1)–(3) above are respectively

(1) $S(a \ b) = \max\{a \ b\}$,
(2) $S(a \ b) = a + b - a \cdot b$,
(3) $S(a \ b) = \min\{a + b \ 1\}$.

Assumptions can be made about the voting model to obtain each of these answers. For example, if it is assumed that no preference can be made for any possible voting pattern for P in relation to $f1$ or $f2$, then all possible distributions must be allowed. For each pair of distributions the proportion of those persons voting for both and the proportion of those persons voting for at least one can be determined. This gives the values of the conjunction and disjunction, respectively for this pair of distributions. This is repeated for all possible pairs of distributions and the values for the conjunction and disjunction determined in each case. If it is assumed that any pair of distributions is as likely as any other, then the expected values for the conjunction and disjunction will be equal to $\mathrm{Mf1}(h)\cdot\mathrm{Mf2}(h)$ and $\mathrm{Mf1}(h)+\mathrm{Mf2}(h)-\mathrm{Mf1}(h)\cdot\mathrm{Mf2}(h)$, respectively.

SEMANTIC UNIFICATION

Let $f1, f2$ be two fuzzy subsets of the set F and suppose that each of these can be associated with some object X. Then we can ask the question, what is the probability of "X is $f1$" given that we know that "X is $f2$".

Consider the more specific case in which we know that John is between 5 ft 10 in. and 6 ft. Then the probability that John is over 5 ft 10 in. is 1. The probability that John is below 5 ft 9 in. is 0. The probability that John is between 5 ft 9 in. and 5 ft 11 in. lies between 0 and 1. This is so since John's actual height could belong to the interval [5 ft 10 in., 5 ft 11 in.] which would give a probability of 1 of John being between 5 ft 9 in. and 5 ft 11 in., but it could also belong to [5 ft 11 in., 6 ft] which would give zero probability. The first can occur with a probability x_1 and the second case with a probability x_2. If all that we know about x_1 and x_2 is that both are non-negative and they sum to one, then the probability of John being between 5 ft 9 in. and 5 ft 11 in. lies anywhere in the interval [0, 1]. If we can estimate x_1 and x_2 then

$$\Pr(\text{John is between 5 ft 9 in. and 5 ft 11 in.}) = x_1.$$

If an equally likely distribution is assumed over the interval

$$[5\text{ ft 10 in., 6 ft}] \quad \text{then} \quad x_1 = 1/2.$$

This example illustrates the non-fuzzy version of the situation posed above, with respect to the fuzzy subsets $f1$ and $f2$. We can ask a similar question for the fuzzy case. What is the probability that John is tall given that we know that John is a little above average height? We should be able to arrive at an answer using a similar approach to that used for the non-fuzzy case but taking into account that not every height has membership level of 1 or 0 in the sets "tall" and "a little above average height".

SEMANTIC UNIFICATION AND POPULATION VOTING MODEL

We will now return to the example above of determining $\Pr(X \text{ is } f1 \mid X \text{ is } f2)$ when we are given actual definitions for $f1, f2$ and F.

The $\Pr(X \text{ is } f1 \mid X \text{ is } f2)$ can be interpreted in this voting model as $\Pr(P \text{ accepts } X \text{ is } f1 \mid P \text{ is told } X \text{ is } f2)$. Now

$$\Pr(P \text{ accepts } X \text{ is } f1 \mid P \text{ is told } X \text{ is } f2)$$

$$= \underset{h}{\mathrm{SUM}}\ \{\Pr(P \text{ accepts } X \text{ is } f1 \mid P \text{ accepts } X \text{ is } h,\ P \text{ is told } X \text{ is } f2).$$

$$\Pr(P \text{ accepts } X \text{ is } h \mid P \text{ is told } X \text{ is } f2)\}$$

$$= \underset{h}{\mathrm{SUM}}\ \{\Pr(P \text{ accepts } X \text{ is } f1 \mid P \text{ accepts } X \text{ is } h).$$

$$\Pr(P \text{ accepts } X \text{ is } h \mid P \text{ is told } X \text{ is } f2)\}.$$

Now

$$\Pr(P \text{ accepts } X \text{ is } f1 \mid P \text{ accepts } X \text{ is } h) = \Pr(h \text{ is accepted as } f1) = \mathrm{Mf1}(h).$$

Therefore

$$\Pr(P \text{ accepts } X \text{ is } f1 \mid P \text{ is told } X \text{ is } f2)$$

$$= \underset{h}{\mathrm{SUM}}\ \mathrm{Mf1}(h)\cdot\Pr(P \text{ accepts } X \text{ is } h \mid P \text{ is told } X \text{ is } f2).$$

$$\Pr(P \text{ accepts } X \text{ is } h \mid P \text{ is told } X \text{ is } f2)$$

$$= \underset{i}{\mathrm{SUM}}\ \Pr(\text{person } i \text{ chooses } X \text{ is } h \mid X \text{ is } f2).$$

If person i is told that X is $f2$, then this person has an interpretation for the label $f2$ in the form of a set of acceptable values. If h is not one of these values Pr(person i chooses X is $h \mid X$ is $f2) = 0$. If the set of values consists only of h then Pr(person i chooses X is $h \mid X$ is $f2) = 1$. If the set contains more than one value including h then probabilities must be assigned to each value. All that is known about these probabilities is that they sum to 1.

Example

Consider

$$f1 = 55|0.2 + 60|0.5 + 65|0.8 + 70|1,$$

$$f2 = 55|1 + 60|0.2,$$

$$F = \{50, 55, 60, 65, 70, 75\},$$

then P interprets $f2$ as

Person									
1	2	3	4	5	6	7	8	9	10
55 60	55 60	55	55	55	55	55	55	55	55

so that

$$\Pr(P \text{ accepts } X \text{ is } 55 \mid P \text{ is told } X \text{ is } f2) = 4/5 + x \cdot 1/5,$$

$$\Pr(P \text{ accepts } X \text{ is } 60 \mid P \text{ is told } X \text{ is } f2) = (1 - x) \cdot 1/5$$

where $0 \leqslant x \leqslant 1$.

This follows since persons 3–10 accept X is 55 as this is the only value they can choose. Persons 1 and 2 have a choice and x represents their probability of choosing 55.

Therefore

$$\Pr(P \text{ accepts } X \text{ is } f1 \mid P \text{ is told } X \text{ is } f2) = 0.2(0.8 + 0.2x) + 0.5 * 0.2(1 - x); \quad 0 \leqslant x \leqslant 1,$$

so that Pr(P accepts X is $f1 \mid P$ is told X is $f2$) lies in the interval [0.2, 0.26].

We thus conclude that Pr(X is $f1 \mid X$ is $f2$) is in [0.2, 0.26]. This is equivalent to solving the following optimization problem

$$\max/\min z = 0.2x_1 + 0.5x_2 + 0.8x_3 + 1x_4,$$

subject to

$$x_1 + x_2 \leqslant 1,$$
$$x_2 \leqslant 0.2,$$
$$x_3 = 0,$$
$$x_4 = 0,$$
$$x_1 + x_2 + x_3 + x_4 = 1,$$

for the constant threshold model used for interpreting $f2$.

If all possible distributions corresponding to interpretations of $f2$ are considered then the following optimization model results

$$\max/\min z = 0.2x_1 + 0.5x_2 + 0.8x_3 + 1x_4,$$

subject to

$$x_1 \leqslant 1,$$
$$x_2 \leqslant 0.2,$$
$$x_3 = 0,$$
$$x_4 = 0,$$
$$x_1 + x_2 + x_3 + x_4 = 1.$$

In both cases: min z gives lower bound and max z gives upper bound for Pr(X is $f1 \mid X$ is $f2$).

In this example the support pair is the same whatever interpretation is used for $f2$. The next example will yield different results for different interpretations.

A MORE COMPLEX EXAMPLE

$$F = \{e1, e2, e3, e4, e5\, e6\},$$

$$f1 = e1 \mid 0.1 + e2 \mid 0.3 + e3 \mid 0.5 + e4 \mid 0.7\, e5 \mid 1,$$

$$f2 = e1 \mid 0.2 + e2 \mid 1 + e3 \mid 0.7 + e4 \mid 0.1,$$

Pr(X is $f1 \mid X$ is $f2$) lies in [z min, z max], where z min and z max are determined by solving one of the following optimization models.

(1) Using constant threshold model:

$$\text{min/max } z = 0.1x_1 + 0.3x_2 + 0.5x_3 + 0.7x_4 + x_5,$$

subject to

$$x_4 \leqslant 0.1,$$

$$x_1 + x_4 \leqslant 0.2,$$

$$x_1 + x_3 + x_4 \leqslant 0.7,$$

$$x_1 + x_2 + x_3 + x_4 = 1.$$

Therefore

$$z \text{ min} = 0.1 * 0.2 + 0.3 * 0.8 = 0.26,$$

$$z \text{ max} = 0.7 * 0.1 + 0.5 * 0.6 + 0.3 * 0.3 = 0.46.$$

Thus

Pr(X is $f1 \mid X$ is $f2$) lies in [0.26, 0.46].

(2) Allowing for all possible interpretations of $f2$

$$\text{min/max } z = 0.1x_1 + 0.3x_2 + 0.5x_3 + 0.7x_4 + x_5,$$

subject to

$$x_4 \leqslant 0.1$$

$$x_1 \leqslant 0.2$$

$$x_3 \leqslant 0.7$$

$$x_2 \leqslant 1$$

$$x_1 + x_2 + x_3 + x_4 = 1.$$

Therefore

$$z \text{ min} = 0.1 * 0.2 + 0.3 * 0.8 = 0.26,$$

$$z \text{ max} = 0.7 * 0.1 + 0.5 * 0.7 + 0.3 * 0.2 = 0.48.$$

Thus

Pr(X is $f1 \mid X$ is $f2$) lies in [0.26, 0.48]

RESTRICTION MODEL

The generalization of the linear programming solutions for general fuzzy subsets $f1$ and $f2$, defined on F, when X is $f2$ is given and Pr(X is $f1 \mid X$ is $f2$) is to be determined, is as follows:

Let

$$F = \{e_i\};\quad i = 1 \ldots n,$$

$$f1 = \underset{e_i}{\mathrm{SUM}}\ \{e_i \mid \mathrm{Mf1}(e_i)\},$$

$$f2 = \underset{e_i}{\mathrm{SUM}}\ \{e_i \mid \mathrm{Mf2}(e_i)\}.$$

(1) Using the constant threshold model for $f2$:

$$\max/\min z = \underset{e_i}{\mathrm{SUM}}\ \mathrm{Mf1}(e_i)\cdot x_{e_i},$$

subject to

$$\underset{\{k:\mathrm{Mf2}(k) \leqslant \mathrm{Mf2}(e_i)\}}{\mathrm{SUM}} x_k \leqslant \mathrm{Mf2}(e_i);\quad \text{all}\quad e_i,\quad \underset{e_i}{\mathrm{SUM}}\, x_{e_i} = 1.$$

(2) Using all possible interpretations for $f2$:

$$\max/\min z = \underset{e_i}{\mathrm{SUM}}\ \mathrm{Mf1}(e_i)\cdot x_{e_i},$$

subject to

$$x_{e_i} \leqslant \mathrm{Mf2}(e_i);\quad \text{all}\quad e_i,\quad \underset{e_i}{\mathrm{SUM}}\, x_{e_i} = 1.$$

If z min, z max correspond to the min z and max z, subject to the given constraints, respectively then

$$\Pr(X \text{ is } f1 \mid !X \text{ is } f2) \text{ lies in } [z \min, z \max].$$

This corresponds to interpreting membership functions of fuzzy sets as possibility restrictions. For some element e_i in F, $\mathrm{Mf1}(e_i)$ gives the upper bound to the possibility that e_i belongs to $f1$. Mf1 restricts the possibility that h belongs to $f1$ to $\mathrm{Mf1}(e_i)$. If we know the probability distribution over the elements of F consistent with the statement that X is $f2$, i.e. we know $\Pr(X \text{ is } e_i \mid X \text{ is } f2)$ for all e_i in F, then

$$\Pr(X \text{ is } f1 \mid X \text{ is } f2) = \underset{e_i}{\mathrm{SUM}}\ \mathrm{Mf1}(e_i)\cdot \Pr(X \text{ is } e_i \mid X \text{ is } f2).$$

This uses a weighted sum of the conditional probabilities where the weights are the membership values of the fuzzy set $f1$. For the non-fuzzy case the characteristic values of the characteristic function for $f1$ would be used.

In fact we do not know what the values for $\{\Pr(X \text{ is } e_i \mid X \text{ is } f2)\}$ are. All we know is that the value of $\Pr(X \text{ is } e_i \mid X \text{ is } f2)$ will be constrained by the fact that X is $f2$. We will use one of the following assumptions for the form that this constraint can take:

(1) Using the constant threshold model for $f2$:

$$\Pr(X \text{ is one of } Fs \mid X \text{ is } f2) \leqslant \max\{\mathrm{Mf2}(e_i) : e_i \text{ in } Fs\} \text{ for any subset } Fs \text{ of } F.$$

(2) Using all possible interpretations for $f2$:

$$\Pr(X \text{ is } e_i \mid X \text{ is } f2) \leqslant \mathrm{Mf2}(e_i);\quad \text{all}\quad e_i \text{ in } F.$$

If we put $\Pr(X \text{ is } e_i \mid X \text{ is } f2) = x_{e_i}$, then this gives the constraints given above for each of the models.

There is now no unique solution for $\Pr(X \text{ is } f1 \mid X \text{ is } f2)$ so that we can find upper and lower bounds for this by maximizing and minimizing $\mathrm{SUM}\{\mathrm{Mf1}(e_i)\cdot x_{e_i}\}$ subject to the constraints on x_{e_i} given above.

Furthermore, as given below, this can be generalized once again to handle continuous membership functions.

SPECIAL CASE

We will consider the special cases of determining Pr(X is $f|X$ is f). This often causes difficulty since the value of this is not necessarily 1 as many seem to expect. We will consider an example and then justify this answer by using the population model for interpreting the results.

Let

$$F = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8, e_9, e_{10}\}$$

and

$$f = e_1|0.1 + e_2|0.2 + e_3|0.3 + e_4|0.4 + e_5|0.5 + e_6|0.6 + e_7|0.7 + e_8|0.8 + e_9|0.9 + e_{10}|1.0,$$

then

(1) Using the constant threshold model:

$$\max/\min z = 0.1x_1 + 0.2x_2 + 0.3x_3 + 0.4x_4 + 0.5x_5 + 0.6x_6 + 0.7x_7 + 0.8x_8 + 0.9x_9 + x_{10},$$

subject to

$$x_1 \leqslant 0.1,$$
$$x_1 + x_2 \leqslant 0.2,$$
$$x_1 + x_2 + x_3 \leqslant 0.3,$$
$$x_1 + x_2 + x_3 + x_4 \leqslant 0.4,$$
$$x_1 + x_2 + x_3 + x_4 + x_5 \leqslant 0.5,$$
$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 \leqslant 0.6,$$
$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 \leqslant 0.7$$
$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 \leqslant 0.8,$$
$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 \leqslant 0.9,$$
$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_y + x_8 + x_9 + x_{10} \leqslant 1,$$

so that

$$z \min = 0.1 * (0.1 + 0.2 + 0.3 + 0.4 + 0.5 + 0.6 + 0.7 + 0.8 + 0.9 + 1) = 0.1 * 5.5 = 0.55,$$

$$z \max = 1 * 1 = 1.$$

Therefore using the constant threshold model

$$\Pr(X \text{ is } f|X \text{ is } f) = [0.55, 1].$$

We might note that if we had chosen f to be

$$f = e_1|0 + e_2|0.1 + e_3|0.2| + e_4|0.3 + e_5|0.4 + e_6|0.5 + e_7|0.6 + e_8|0.7 + e_9|0.8 + e_{10}|0.9,$$

then Pr(X is $f|X$ is f) would lie in [0.45, 1].

(2) Using all possible interpretations for f:

$$\max/\min z = 0.1x_1 + 0.2x_2 + 0.3x_3 + 0.4x_4 + 0.5x_5 + 0.6x_6 + 0.7x_7 + 0.8x_8 + 0.9x_9 + x_{10},$$

subject to

$$x_1 \leqslant 0.1,$$
$$x_2 \leqslant 0.2,$$
$$x_3 \leqslant 0.3,$$
$$x_4 \leqslant 0.4,$$
$$x_5 \leqslant 0.5,$$
$$x_6 \leqslant 0.6,$$
$$x_7 \leqslant 0.7,$$
$$x_8 \leqslant 0.8,$$
$$x_9 \leqslant 0.9,$$
$$x_{10} \leqslant 1,$$

so that

$$z \text{ min} = 0.1*0.1 + 0.2*0.2 + 0.3*0.3 + 0.4*0.4 = 0.3,$$

$$z \text{ max} = 1,$$

$$\Pr(X \text{ is } f \mid X \text{ is } f) \text{ lies in } [0.3, 1].$$

We might note that if we had chosen f to be

$$f = e_1|0 + e_2|0.1 + e_3|0.2| + e_4|0.3 + e_5|0.4 + e_6|0.5 + e_7|0.6 + e_8|0.7 + e_9|0.8 + e_{10}|0.9,$$

then

$$\Pr(X \text{ is } f \mid X \text{ is } f) \text{ would lie in } [0.4, 1].$$

It is quite easy to justify this result using the population voting model. If the population P is told that X is f then P will interpret this. There will be some elements of the set F which some members of P, but not all, will accept as possible values for X. When asked if X is f these members of P will choose these values with a certain probability as possible values for X, but not all members of P will choose these values.

Consider the situation that the population is told that John is tall. Some members of the population know that other members accept heights corresponding to tall which they do not. Therefore these members know that there is a probability that John's height will be a value which is unacceptable to them as representing tall. Therefore the probability that John is tall when the population is told that John is tall cannot be 1 since some members of the population will not vote yes for certain.

CONTINUOUS CASE

The above is easily generalized to the continuous case.

CONCLUSIONS

A theory of reasoning with uncertainties applicable to expert systems and other AI applications has been described. It is being applied to many applications and evidence to date indicates that it is relatively easy to apply. Fril is a powerful AI systems language and ideal for writing expert system shells.

REFERENCES

1. J. F. Baldwin, T. P. Martin and B. W. Pilsworth, *Fril Manual*. Equipu AIR Ltd, Bristol (1987).
2. J. F. Baldwin, Support logic programming. *Proc. NATO Advanced Study Institute on Fuzzy Sets Theory and Applications*. Louvain-La-Neuve, Belgium (1985).

3. J. F. Baldwin, Support logic programming. In *Fuzzy Sets Theory and Applications* (Eds A. Jones and H. J. Zimmermann), pp. 133–170. Reidel, Dordrecht, Holland (1986).
4. J. F. Baldwin, Evidential support logic programming. *Fuzzy Sets Systems* **10,** 1-26 (1987).
5. J. F. Baldwin, A theory of support pairs (in press).
6. I. Hacking, *Logic of Statistical Inference*. Cambridge Univ. Press (1965).
7. L. A. Zadeh, Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Systems* **1,** 3–28 (1978).
8. R. Bellman and M. Giertz, On the analytic formalism of the theory of fuzzy sets. *Inform. Sci.* **5,** 149–156 (1973).
9. C. G. Hempel, Maximum specificity and lawlikeness in probabilistic explanation. *Phil. Sci.* **35,** 116–133 (1968).
10. P. Suppes, *Probabilistic Metaphysics*. Blackwell, Oxford (1984).
11. W. C. Salmon, *Scientific Explanation and the Causal Structure of the World.* Princeton Univ. Press, N.J. (1984).
12. D. V. Lindley, Scoring rules and the inevitability of probability. *Int. Stat. Rev.* **50,** 1-26 (1982).
13. G. Shafer, *A Mathematical Theory of Evidence*. Princeton Univ. Press, N.J. (1976).
14. V. O. Homolka, The role of nonmonotonic reasoning in an intelligent maintenance system, ITRC73. Information Technology Research Centre, Bristol Univ. (1985).