International Journal of Approximate Reasoning 50 (2009) 812–821

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

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

Comparing approximate reasoning and probabilistic reasoning using the Dempster–Shafer framework

Ronald R. Yager

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

ARTICLE INFO

Article history:
Received 11 August 2008
Received in revised form 3 March 2009
Accepted 4 March 2009
Available online 13 March 2009

Keywords:
Inference
Uncertainty
Fuzzy sets
Modus ponens

ABSTRACT

We investigate the problem of inferring information about the value of a variable *V* from its relationship with another variable *U* and information about *U*. We consider two approaches, one using the fuzzy set based theory of approximate reasoning and the other using probabilistic reasoning. Both of these approaches allow the inclusion of imprecise granular type information. The inferred values from each of these methods are then represented using a Dempster–Shafer belief structure. We then compare these values and show an underling unity between these two approaches.

1. Introduction

Assume *U* and *V* are two variables an important step in many information processing and decision making tasks is inferring information about the value of the variable *V* from knowledge of some relationship between variables *U* and *V* and data about the value of *U*. Two common ways of modeling relationships between variables are with the aid of probabilistic and fuzzy rule-based representations. One shared feature of both these representations is their facility for allowing the types of granular information common to human perception [1–3] in the describing relationships. Thus if *A* is a subset of the domain of some variable *U* both representations allow the use of granular statements such as *U is A* to indicate that the value of *U* lies in the set *A*. The use of this granularization allows for the modeling of imprecision in our knowledge. As discussed by Zadeh [4,5], Bargiela and Pedrycz [6] and Lin et al. [7] granularization and the related idea of granular computing play an important role in the development of computational intelligence.

In this work, we look two methods for performing inference task described above. The first is with the aid Zadeh's framework of approximate reasoning [8,9] in which we use a rule-like representation to describe the relationship between the two variables. The second is with aid of probabilistic reasoning in which we use conditional probability to express our relationship.

Once having determined the inferred value for *V* using each of these two technologies we express these two inferred values in a unified framework using the Dempster–Shafer belief structures. Having a common representational framework we then compare and relate the results of these two approaches to determine the underlying commonality between them. As a result of this comparison we see a unified view if the inference process and in turn get a deeper understanding this process.

E-mail address: yager@panix.com

0888-613X/$ - see front matter
doi:10.1016/j.ijar.2009.03.003

2. Modus ponens inference using approximate reasoning

We here consider the approximate reasoning protoform of modus ponens [10–12]. In this protoform, we have two propositions

P1: If *U is A* then *V is B*
P2: *U is D*

where U and V are variables taking their values in X and Y, respectively. In this protoform A and D are fuzzy subsets of X and B is a fuzzy subset of Y.

Using the representational capability of the theory of approximate reasoning [8,13] we get that P1 is represented as

$$(U, V) \text{ } is \text{ } H.$$

Here (U,V) is a joint variable and H is a fuzzy subset of X x Y with $H(x,y) = \overline{A}(x) \vee B(y)$ and where $\overline{A}$ is the complement of $A, \overline{A}(x) = 1 - A(x)$ [8].

To obtain the inference from P1 and P2 we take their conjunction and get (U,V) *is* M where

$$M(x,y) = H(x,y) \wedge D(x) = (\overline{A}(x) \vee B(y)) \wedge D(x) = (\overline{A}(x) \wedge D(x)) \vee (B(y) \wedge (D(x)).$$

To get the resulting inference *V is E* where E is a fuzzy subset of Y we project onto Y and obtain

$$E(y) = \text{Max}_x[M(x,y)] = \text{Max}_x[(\overline{A}(x) \wedge D(x)) \vee (B(y) \wedge D(x))],$$
$$E(y) = \text{Max}_x[\overline{A}(x) \wedge D(x)] \vee (B(y) \wedge \text{Max}_x[D(x)]).$$

Under the assumption that D is normal[1], has at least one element with membership grade equal 1, we get [14]

$$E(y) = \text{Poss}[\overline{A}/D] \vee B(y),$$

where $\text{Poss}[\overline{A}/D] = \text{Max}_x[\overline{A}(x) \wedge D(x)]$.

Recalling [4] that $\text{Cert}[A/D] = 1 - \text{Poss}[\overline{A}/D]$ we can also express this inference as

$$E(y) = (1 - \text{Cert}[A/D]) \vee B(y).$$

We see that if $\text{Cert}[A/D] = 1$ then we get $E(y) = B(y)$ and hence our inference is *V is B*, the rule has completely fired. At the other extreme if $\text{Cert}[A/D] = 0$ then we get $E(y) = 1$ and we obtain *V is Y*, no information has been obtained.

We recall another protoform from approximate reasoning is related to the translation of certainty qualification [14]. Under this protoform a proposition

$$(V \text{ } is \text{ } G) \text{ } is \text{ } \lambda \text{ Certain}$$

can be translated as *V is H* where $H(y) = G(y) \vee \bar{\lambda}$. $(\bar{\lambda} = 1 - \lambda)$.

Using this if we let $\alpha = \text{Poss}[\overline{A}/D]$ then we can alternatively express the results of the modus ponens protoform as

$$(V \text{ } is \text{ } B) \text{ is } \bar{\alpha} \text{ Certain},$$

this is the same as *V is E* where $E(y) = \alpha \vee B(y)$.

In the special case when our sets are crisp, not fuzzy sets, then $\text{Poss}[\overline{A}/D] \in \{0, 1\}$. In particular

$$\text{Poss}[\overline{A}/D] = 0 \quad \text{if } D \subseteq A,$$
$$\text{Poss}[\overline{A}/D] = 1 \quad \text{if } D \not\subset A.$$

Thus if $D \subseteq A$ then $\alpha = 0$ and we get $E(y) = B(y) \vee \alpha = B(y)$ and we infer *V is B*. On the other hand, if $D \not\subset A$ then $\alpha = 1$ and $E(y) = B(y) \vee \alpha = 1$ and $E = Y$. Here we infer that V can be anything.

We should observe that within P1 there is an implicit third proposition. Essentially the implication P1 tells us when *U is* (**not** *A*) the value of V can be anything. This is formally captured by the proposition

P3 : If *U is* not *A* then *V is Y*.

We can see that this is implicit in the proposition P1 as follows. We first observe that P3 is representable as (U,V) *is* G where $G(x,y) = A(x) \vee Y(y)$. To see that G is implicit in H we note that $G \cap H = R$ where

$$R(x,y) = (\overline{A}(x) \vee B(y)) \wedge (A(x) \vee Y(y)),$$
$$R(x,y) = (\overline{A}(x) \wedge A(x)) \vee (\overline{A}(x) \wedge Y(y)) \vee (B(y) \wedge A(x)) \vee (B(y) \wedge Y(y)),$$
$$R(x,y) = (\overline{A}(x) \vee B(y) = H(x,y).$$

Thus P3 is implicit in P1.

[1] We note normality of D is equivalent to saying D is not null in the crisp case.

3. Inference using probabilistic reasoning

We now consider an alternate approach for making an inference about V using a probabilistic method. Here we assume the same information as in the preceding protoform except it is represented in probabilistic form[2]

S-1: $P(B/A) = 1$
S-2: $P(D) = 1$
S-3: $P(B/\overline{A}) = \text{unknown}$

In this section, our interest is in the obtaining the probability of *V is B*, which we will denote as $P(B)$. In this case, as is well known

$$P(B) = P(B/A)P(A) + P(B/\overline{A})P(\overline{A}).$$

Since $P(B/A) = 1$ we have

$$P(B) = P(A) + P(B/\overline{A})P(\overline{A}).$$

With $P(\overline{A}) = 1 - P(A)$ we get

$$P(B) = P(A) + P(B/\overline{A})(1 - P(A)).$$

Furthermore, we can express

$$P(A) = P(A/D)P(D) + P(A/\overline{D})P(\overline{D}).$$

Since $P(D) = 1$ then $P(\overline{D}) = 1 - P(D) = 0$ hence we have

$$P(A) = P(A/D) = 1 - P(\overline{A}/D).$$

If we let $\lambda = P(\overline{A}/D)$ then $P(A) = (1 - \lambda)$. Using this in our expression for $P(B)$ we get

$$P(B) = \lambda P(B/\overline{A}) + (1 - \lambda)$$

were $P(B/\overline{A})$ is the value unknown. We note that a probability value of unknown can be represented as a set consisting of the unit interval $I = [0,1]$. Using this we have

$$P(B) = (1 - \lambda) + \lambda[0,1] = (1 - \lambda) + [0, \lambda] = [1 - \lambda, 1].$$

Thus $P(B)$ is some value in the interval $1 - \lambda$ to 1. Linguistically this can be expressed as saying that the value of $P(B)$ is "at least $1 - \lambda$".

4. Comparison using Dempster–Shafer representation

Recapitulating, from the first method, using approximate reasoning techniques we have inferred *V is E* where $E(y) = B(y) \vee \alpha$. As we indicated this is semantically equivalent to saying

$$(V \text{ is } B) \quad \text{with } \bar{\alpha} \text{ Certainty},$$

where $\alpha = \text{Poss}[\overline{A}/D] = \text{Max}_x[\overline{A}(x) \wedge D(x)]$. We note that α is the largest membership grade in the intersection of the sets $\overline{A}$ and D.

Using the probabilistic approach we have obtained

$$\text{Prob}(V \text{ is } B) = \lambda \text{Prob}(B/\overline{A}) + (1 - \lambda),$$
$$\text{Prob}(V \text{ is } B) = [1 - \lambda, 1].$$

As we indicated this is semantically to saying that the probability V lies in the set B is *at least* $1 - \lambda$ where $\lambda = \text{Prop}(\overline{A}/D) = \frac{P(\overline{A} \cap D)}{P(D)}$. Since we assumed $\text{Prob}(D) = 1$ then $\lambda = \text{Prob}(\overline{A} \cap D)$.

In order to relate these results we shall use the Dempster–Shafer framework [3,15], it will allow us to represent both of these in a unified framework as belief structures on Y.

First for the inference from the approximate reasoning approach, *V is E*, we can represent this as a D–S belief structure m_1 on Y that has one focal element, the fuzzy subset E, with weight $m_1(E) = 1$. Here $E(y) = B(y) \vee \alpha$ with $\alpha = \text{Poss}[\overline{A}/D] = \text{Max}_x[\overline{A}(x) \wedge D(x)]$.

For the second case, where we used a probabilistic approach and obtained that the Probability that V is in B is at least $1 - \lambda$, we can represent this with a belief structure m_2 on Y with two focal elements B and Y such that

[2] In the following we use the notation $P(B/A)$ and $P(D)$ instead of the more cumbersome notation $P(V \text{ is } B/U \text{ is } A)$ and $P(U \text{ is } D)$.

$$m_2(B) = 1 - \lambda,$$
$$m_2(Y) = \lambda.$$

Here $\lambda = \mathrm{Prob}(\overline{A}/D)$.

We now can calculate the plausibility and belief of some arbitrary set G under these two belief structures. That is the plausibility and belief that V lies in G.

First we recall [16] for any belief structure m with focal elements B_i, i = 1 to K we define the plausibility measure such that for any subset G

$$\mathrm{Pl}(G) = \sum_{i=1}^{K} \mathrm{Poss}[G/B_i]m(B_i).$$

We define the belief measure such that for any subset G

$$\mathrm{Bel}(G) = \sum_{i=1}^{K} \mathrm{Cert}[G/B_i]m(B_i).$$

In the above $\mathrm{Poss}[G/B_i] = \mathrm{Max}_y[G(y) \wedge B_i(y)]$ and $\mathrm{Cert}[G/B_i] = 1 - \mathrm{Poss}[\overline{G}/B_i]$.

Using this we obtain for the belief structure m_1 which has the single focal element F

$$\mathrm{Pl}_1(G) = \mathrm{Poss}[G/F] = \mathrm{Max}_y[G(y) \wedge F(y)],$$
$$\mathrm{Pl}_1(G) = \mathrm{Max}_y[G(y) \wedge (B(y) \vee \alpha)],$$
$$\mathrm{Pl}_1(G) = \mathrm{Max}_y[(G(y) \wedge B(y)) \vee \mathrm{Max}_y[\alpha \wedge G(y)].$$

Assuming G is normal, i.e. has one element with membership grade one, we get

$$\mathrm{Pl}_1(G) = \mathrm{Poss}[G/B] \vee \alpha.$$

For m_2 with two focal elements, B and Y, we get

$$\mathrm{Pl}_2(G) = (1 - \lambda)\mathrm{Poss}[G/B] + \lambda\mathrm{Poss}[G/Y].$$

Under the assumed normality of G we get $\mathrm{Poss}[G/Y] = 1$ and hence we have

$$\mathrm{Pl}_2(G) = (1 - \lambda)\mathrm{Poss}[G/B] + \lambda.$$

A unification of these two results can be useful in providing us a deeper and more unified understanding of the inference process. We now turn to providing this unification. We first recall the t-conorm operator, which provides a generalization the logical'or' operator [17]. A t-conorm S is a mapping S: $[0,1]^2 \to [0,1]$ having the properties

(1) **Symmetry**: $S(a,b) = S(b,a)$
(2) **Monotonicity:** $S(a,b) \geqslant S(c,d)$ if $a \geqslant c$ and $b \geqslant d$
(3) **Associativity:** $S(a,S(b,c)) = S(S(a,b),c)$
(4) **0 as identity**: $S(0,a) = a$

Two important examples of t-conorm are the Max and probabilistic sum

$$S_M(a,b) = \mathrm{Max}(a,b) = a \vee b,$$
$$S_P(a,b) = a + b - ab = b + a(1 - b).$$

We see then that

$$\mathrm{Pl}_1(G) = S_M[\mathrm{Poss}[G/B], \alpha],$$
$$\mathrm{Pl}_2(G) = S_P[\mathrm{Poss}[G/B], \lambda],$$

where we recall $\alpha = \mathrm{Poss}[\overline{A}/D]$ and $\lambda = \mathrm{Prob}[\overline{A}/D]$.

Thus a semantic interpretation of the plausibility of G is degree of possibility of G given B **or** the *feasibility*[3] of not A given D.

We now consider the measure of belief. In the case of the belief structure m_1, we have

$$\mathrm{Bel}_1(G) = \mathrm{Cert}(G/F) = 1 - \mathrm{Poss}[\overline{G}/F] = 1 - \mathrm{Max}_y[\overline{G}(y) \wedge F(y)],$$
$$\mathrm{Bel}_1(G) = 1 - \mathrm{Max}_y[\overline{G}(y) \wedge (B(y) \vee \alpha)],$$
$$\mathrm{Bel}_1(G) = 1 - \mathrm{Max}_y[(\overline{G}(y) \wedge B(y)) \vee (\overline{G}(y) \wedge \alpha)],$$
$$\mathrm{Bel}_1(G) = 1 - (\mathrm{Max}_y[\overline{G}(y) \wedge B(y)] \vee \mathrm{Max}_y[\overline{G}(y) \wedge \alpha]),$$

[3] We are using the term feasibility to generalize the ideas of possibility and probability.

$$\mathrm{Bel}_1(G) = 1 - \mathrm{Poss}(\overline{G}/B) \vee (\alpha \wedge \mathrm{Max}_y[\overline{G}(y)),$$
$$\mathrm{Bel}_1(G) = \mathrm{Cert}[G/B] \wedge (\bar{\alpha} \vee \mathrm{Min}_y[G(y)]),$$
$$\mathrm{Bel}_1(G) = (\bar{\alpha} \wedge \mathrm{Cert}[G/B]) \vee (\mathrm{Min}_y[G(y) \wedge \mathrm{Cert}[G/B]).$$

For the case of the belief structure m_2 we have

$$\mathrm{Bel}_2(G) = (1-\lambda)\mathrm{Cert}[G/B] + \lambda \mathrm{Cert}[G/Y].$$

We observe that $\mathrm{Cert}(G/Y) = 1 - \mathrm{Poss}[\overline{G}/Y]$. We also note that

$$\mathrm{Poss}[\overline{G}/Y] = \mathrm{Max}_y[\overline{G}(y)] = 1 - \mathrm{Min}_y[G(y)].$$

Hence $\mathrm{Cert}(G/Y) = \mathrm{Min}_y[G(y)]$. Thus we

$$\mathrm{Bel}_2(G) = (1-\lambda)\mathrm{Cert}(G/B) + \lambda \mathrm{Min}_y[G(y)].$$

We now consider a unification of these two formulations. To do this we recall the *t*-norm operator which provides a generalization of the 'Anding' operator [17]. A *t*-norm is a mapping $T{:}[0,1]^2 \to [0,1]$ which has the same first three properties as the *t*-conorm, symmetry, monotonicity and associativity, but it has one as is identity, $T(1,a) = a$. Two important examples of *t*-norms are Min and product

$$T_M(a,b) = a \wedge b \quad \textbf{Min},$$
$$T_P(a,b) = ab \quad \textbf{Product}.$$

Consider now the two formulations we just obtained

$$\mathrm{Bel}_1(G) = \mathrm{Cert}[G/B] \wedge (\bar{\alpha} \vee \mathrm{Min}_y[G(y)]),$$
$$\mathrm{Bel}_2(G) = (1-\lambda)\mathrm{Cert}[G/B] + \lambda \mathrm{Min}_y G(y).$$

First we observe that

$$\mathrm{Cert}[G/Y] = \mathrm{Min}_y[G(y) \vee \overline{Y}(y)] = \mathrm{Min}_y[G(y)].$$

Thus $\mathrm{Min}_y[G(y)]$ is the certainty of *V is G* in the case when we know nothing about *V*, that is *V is Y*. Thus we see

$$\mathrm{Bel}_1(G) = \mathrm{Cert}[G/B] \wedge (\bar{\alpha} \vee \mathrm{Cert}[G/Y]),$$
$$\mathrm{Bel}_1(G) = (\bar{\alpha} \wedge \mathrm{Cert}[G/B]) \vee (\mathrm{Cert}[G/Y] \wedge \mathrm{Cert}[G/B]).$$

However, since $\mathrm{Cert}[G/Y] = 1 - \mathrm{Poss}[\overline{G}/Y]$ and $\mathrm{Cert}[G/B] = 1 - \mathrm{Poss}[\overline{G}/B]$ and $B \subseteq Y$ then $\mathrm{Cert}[G/Y] \leqslant \mathrm{Cert}[G/B]$ and hence

$$\mathrm{Bel}_1(G) = (\bar{\alpha} \wedge \mathrm{Cert}[G/B]) \vee \mathrm{Cert}[G/Y].$$

For the case of m_2 we have that

$$\mathrm{Bel}_2(G) = \bar{\lambda}\mathrm{Cert}[G/B] + \lambda \mathrm{Cert}[G/Y].$$

Let us consider the important special case where $\mathrm{Cert}(G/Y) = \mathrm{Min}_y[G(y)] = 0$. Here we are saying *G* is such that if we have no information about *V* we cannot say anything with certainty about *V is G*. In this case we have

$$\mathrm{Bel}_1(G) = \bar{\alpha} \wedge \mathrm{Cert}[G/B],$$
$$\mathrm{Bel}_2(G) = \bar{\lambda}\mathrm{Cert}[G/B].$$

Using the idea of *t*-norm we see that

$$\mathrm{Bel}_1(G) = T_M(\bar{\alpha}, \mathrm{Cert}[G/B]),$$
$$\mathrm{Bel}_2(G) = T_P(\bar{\lambda}, \mathrm{Cert}[G/B]).$$

Furthermore, since $\alpha = \mathrm{Poss}(\overline{A}/D)$ then $1 - \bar{\alpha} = 1 - \mathrm{Poss}[\overline{A}/D] = \mathrm{Cert}[A/D]$ hence

$$\mathrm{Bel}_1(G) = T_M(\mathrm{Cert}[A/D], \mathrm{Cert}[G/B]) = \mathrm{Cert}[A/D] \wedge \mathrm{Cert}[G/B],$$
$$\mathrm{Bel}_1(G) = \mathrm{Cert}[A/D] \quad \textbf{and} \quad \mathrm{Cert}[G/B].$$

Recalling now that $\lambda = \mathrm{Prob}(\overline{A}/D)$ and hence $1 - \lambda = \mathrm{Prob}(\overline{A}/D)$ we get that

$$\mathrm{Bel}_2(G) = T_P[\mathrm{Prob}(A/D), \mathrm{Cert}[G/B]]$$

thus $\mathrm{Bel}_2(G) = \mathrm{Prob}(A/D)$ **and** $\mathrm{Cert}[G/B]$.

Again we have a similar interpreter as the feasibility of *A* given *D* and the certainty of *G* given *B*.

Let us now return to the more general case with

$$\mathrm{Bel}_1(G) = \mathrm{Cert}[G/Y] \vee (\bar{\alpha} \wedge \mathrm{Cert}[G/B]),$$
$$\mathrm{Bel}_2(G) = \bar{\lambda}\mathrm{Cert}[G/B] + \lambda \mathrm{Cert}[G/Y].$$

We observe that we can express $\mathrm{Bel}_2(G)$ as

$$\mathrm{Bel}_2(G) = \bar{\lambda}\mathrm{Cert}[G/B] + \lambda\mathrm{Cert}[G/Y] + \bar{\lambda}\mathrm{Cert}[G/Y] - \bar{\lambda}\mathrm{Cert}[G/Y],$$
$$\mathrm{Bel}_2(G) = \mathrm{Cert}[G/Y] + \bar{\lambda}\mathrm{Cert}[G/B] - \mathrm{Cert}[G/Y]).$$

We see that

$$\mathrm{Bel}_1(G) = T_M(\mathrm{Cert}[G/B]), \bar{\alpha}) \quad \text{or} \quad \mathrm{Cert}[G/Y],$$
$$\mathrm{Bel}_2(G) = T_P(\mathrm{Cert}[G/B], \bar{\lambda}) + T_P(\mathrm{Cert}[G/B], \lambda).$$

We can write

$$\mathrm{Bel}_1(G) = T_M(\mathrm{Cert}[G/B], \bar{\alpha}) \quad \text{or} \quad T_M(\mathrm{Cert}[G/Y], 1).$$

Let us provide an overview and unification of the preceding as we shall see this will give us an interesting protoform of reasoning.

We have two variables U and V and some relationship or association between these variables telling us that knowing that *U is A* allows us to conclude that *V is B*. We refer to this as U is $A \mid\!- V$ is B.

In the current situation we know that *U is D* and our objective is to say something about V. In the following we shall let γ indicate the feasibility of *U is* $\bar{A}$ given *U is D*. Let G be any subset of Y in the following we assume Cert[G/Y] = 0. We can now calculate

$$\mathrm{Pl}(G) = (\mathrm{Poss}[G/B] \ \textbf{or} \ \gamma) = S(\mathrm{Poss}[G/B], \gamma),$$
$$\mathrm{Cert}(G) = (\mathrm{Cert}[G/B] \ \textbf{and} \ \bar{\gamma}) = T(\mathrm{Cert}[G/B], \bar{\gamma})$$

In this situation the choice of S and T depends on the method of measuring the feasibility γ.

5. Calculating the pignistic measure

As discussed in [18] a Dempster–Shafer belief structure can be viewed as a representation of our knowledge about the underlying measure associated with a variable. In this framework the measures of plausibility and belief are two measures compatible with the given belief structure. As a matter of fact they are extreme measures. Plausibility is the measure that assigns the largest value to any subset while the belief always assigns the smallest value to the subset.[4] Another measure is one closely related to Smets' concept of pignistic probability [19,20]. We first briefly describe this measure and then calculate it for the two cases of belief structure we have earlier obtained.

Assume m is a belief structure with focal elements B_i, i = 1 to q, with associated weight $m(B_i)$. The pignistic measure associated with m is defined such that for any set G

$$\mathrm{Pig}(G) = \sum_{i=1}^{q} \frac{\mathrm{Card}(B_i \cap G)}{\mathrm{Card}(B_i)} m(B_i).$$

Here $\mathrm{Card}(B_i) = \sum_{x_j \in X} B_i(x_j)$ and $\mathrm{Card}(B_i \cap G) = \sum_{x_j \in X}(B_i(x_j) \wedge G(x_j))$. The pignistic measure assigns to G a portion of the weight associated with a focal element B_i in proportion to the fraction of the elements lying in B_i that also lie in G. We note in the special case where B and G are crisp sets then Card(B_i) is the number of elements in B_i and Card($B_i \cap G$) is the number of elements in B_i that are also in G.

Consider now the case where our knowledge is m_1 where we have one focal element F where $F(y) = B(y) \vee \alpha$ and $m_1(F) = 1$. In this case

$$\mathrm{Pig}_1(G) = \frac{\mathrm{Card}(F \cap G)}{\mathrm{Card}(F)} = \frac{\sum_y((B(y) \wedge G(y)) \vee (\alpha \wedge G(y)))}{\sum_y(B(y) \vee \alpha)}.$$

If we let Y be divided into two classes Y_1 and Y_2 such that $B(y) \geqslant \alpha$ for all $y \in Y_1$ and $B(y) < \alpha$ for all $y \in Y_2$ then

$$\mathrm{Pig}_1(G) = \frac{\sum_{y \in Y_1}(B(y) \wedge G(y)) + \sum_{y \in Y_2}(\alpha \wedge G(y))}{\sum_{y \in Y_1}(B(y) + \alpha|Y_2|}.$$

$|Y_2|$ being the cardinality of Y_2. We can alternatively express $\mathrm{Pig}_1(G)$ as

$$\mathrm{Pig}_1(G) = \frac{\sum_{y \in Y}(B(y) \wedge G(y)) + \sum_{y \in Y_2}((\alpha \wedge G(y) - (B(y) \wedge G(y)))}{\sum_{Y}(B(y) + \sum_{y \in Y_2}(\alpha - B(y))}.$$

[4] Intuitively we can view plausibility as the measure compatible with the D–S structure that assigns the largest expectation to finding the value of a variable in a subset. The belief assigns the smallest. Plausibility is an optimistic type measure while belief is a pessimistic one.

For the special case where G is a crisp set then

$$\mathrm{Pig}_1(G) = \frac{\sum_{y \in G}(B(y) \vee \alpha)}{\sum_{y \in Y}(B(y) \vee \alpha)}.$$

In the further special case where G is a singleton, $G = \{y^*\}$, then

$$\mathrm{Pig}_1(\{y^*\}) = \frac{B(y^*) \vee \alpha}{\sum_Y (B(y) \vee \alpha)}.$$

If $\alpha = 1$ then we have $\mathrm{Pig}_1(\{y^*\}) = \frac{1}{|Y|}$, it is simply unknown. If $\alpha = 0$ then

$$\mathrm{Pig}_1(\{y^*\}) = \frac{B(y^*)}{\sum_y B(y)}.$$

Let us consider the calculation of $\mathrm{Pig}_2(G)$ for m_2 where

$$m_2(B) = 1 - \lambda,$$
$$m_2(Y) = \lambda.$$

In this case

$$\mathrm{Pig}_2(G) = \frac{\mathrm{Card}(B \cap G)}{\mathrm{Card}(B)} \cdot \bar{\lambda} + \frac{\mathrm{Card}(G \cap Y)}{\mathrm{Card}(Y)} \lambda,$$
$$\mathrm{Pig}_2(G) = \frac{\sum_y B(y) \wedge G(y)}{\sum_y B(y)} \cdot \bar{\lambda} + \frac{\mathrm{Card} G}{|Y|} \lambda.$$

We can express this as

$$\mathrm{Pig}_2(G) = \frac{1}{\mathrm{Card}(B)} \left(\bar{\lambda} \mathrm{Card}(B \cap G) + \lambda \mathrm{Card}(B) \frac{\mathrm{Card}(G)}{\mathrm{Card}(Y)} \right),$$
$$\mathrm{Pig}_2(G) = \frac{1}{\mathrm{Card}(B)} (\mathrm{Card}(B \cap G) + \lambda \left(\frac{\mathrm{Card}(B)\mathrm{Card}(G)}{\mathrm{Card}(Y)} - \mathrm{Card}(B \cap G) \right).$$

For the case where G is a singleton, $G = \{y^*\}$ then

$$\mathrm{Pig}_2(\{y^*\}) = \bar{\lambda} \frac{B(y^*)}{\mathrm{Card}(B)} + \lambda \frac{1}{\mathrm{Card}(Y)}.$$

We recall that for the case of m_1 we had

$$\mathrm{Pig}_1(\{y^*\}) = \frac{B(y^*) \vee \alpha}{\sum_y B(y) \vee \alpha} = \frac{F(y^*)}{\mathrm{Card}(F)}.$$

6. Reasoning with partially certain relationships

We now generalize the above situation by associating with our relationship between *U is* A and *V is B* a degree of validity, certainty in the case of approximate reasoning and probability is the case of probabilistic reasoning.

We consider first the modus ponens protoform from approximate reasoning. Here we start with

P1′ : If *U is A* then *V is B* <u>with ρ certainty</u>.

In addition, we still have the statement

P2 : *U is D*.

We can represent **P1′**, using the representational capability of approximate reasoning as

$$(U, V) \textit{ is } H,$$

where $H(x, y) = \overline{A}(x) \vee B(y) \vee \bar{\rho}$.

Conjuncting this with our knowledge from **P2** we get

$$(U, V) \textit{ is } M,$$

where $M(x, y) = (\overline{A}(x) \vee B(y) \vee \bar{\rho}) \wedge D(x)$.

To get the resulting inference, *V is E*, we calculate

$$E(y) = \mathrm{Max}_x[(\overline{A}(x) \wedge D(x)) \vee (\bar{\rho} \wedge D(x)) \vee (B(y) \wedge D(x))].$$

Under the assumption that $D(x)$ is normal, there exists some x^* for which $D(x^*) = 1$, that we get

$$E(y) = \text{Poss}[\overline{A}/D] \vee \bar{\rho} \vee B(y).$$

Denoting $\text{Poss}[\overline{A}/D] = \alpha$ we get $E(y) = \alpha \vee \bar{\rho} \vee B(y)$. Since $\text{Poss}[\overline{A}/D]$ is the same as $(1 - \text{Cert}[A/D])$ we can also express this as

$$E(y) = (1 - \text{Cert}[A/D]) \vee (1 - \rho) \vee B(y).$$

We see that if $\rho = 1$, then this reduces to the earlier case, $E(y) = \alpha \vee B(y)$. On the other hand, if $\rho = 0$, no confidence in the rule **P1**, then $E(y) = \alpha \vee 1 \vee B(y) = 1$. Here $E = Y$ and we infer nothing about the value of V. We further observe that if $\text{Cert}[A/D] = 0$, $\text{Poss}[\overline{A}/D] = 1$, we also get $E = Y$.

Consider now the probabilistic representation of our knowledge. Here we assume

S-1′: $P(B/A) = \rho$
S-2: $P(D) = 1$
S-3: $P(B/\overline{A})$ – unknown

In this case we have

$$P(B) = P(B/A)P(A) + P(B/\overline{A})P(\overline{A}) = \rho P(A) + P(B/\overline{A})P(\overline{A}),$$
$$P(B) = \rho P(A) + P(B/\overline{A})(1 - P(\overline{A})).$$

As in the preceding we have

$$P(A) = P(A/D)P(D) + P(A/\overline{D})P(\overline{D})$$

with $P(D) = 1$ we get $P(A) = P(A/D) = 1 - P(A/\overline{D})$. Denoting $\lambda = P(\overline{A}/D)$ then $P(A) = 1 - \lambda$ and hence we have

$$P(B) = \bar{\lambda}\rho + \lambda P(B/A)$$

with $P(B/\overline{A}) = [0, 1]$, unknown, we have

$$P(B) = \bar{\lambda}\rho + \lambda[0, 1] = \bar{\lambda}\rho + [0, \lambda],$$
$$P(B) = [\bar{\lambda}\rho, \bar{\lambda}\rho + \lambda].$$

We see in this case if $\rho = 1$ then $P(B) = [\bar{\lambda}, 1]$. This is the earlier value. In this case where $\rho = 0$ we get $P(B) = [0, \lambda]$.

Consider $P(B) = [\bar{\lambda}\rho, \bar{\lambda}\rho + \lambda]$ it is worth noting that

$$\bar{\lambda}\rho = T_P(\bar{\lambda}, \rho),$$
$$\bar{\lambda}\rho + \lambda = (1 - \lambda)\rho + \lambda = \lambda + \rho - \lambda\rho = S_p(\lambda, \rho).$$

Thus we have that $P(B) = [\rho$ and not $\bar{\lambda}, \rho$ or $\lambda]$.

The Dempster–Shafer representation of the first case the one using approximate reasoning is a belief structure m_3 with one focal element F where $F(y) = \alpha \vee \bar{\rho} \vee B(y)$ and $m_3(F) = 1$.

In this case for any subset G we get the plausibility

$$\text{Pl}_3(G) = \text{Poss}[G/B] \vee \alpha \vee \bar{\rho},$$
$$\text{Pl}_3(G) = S_M[\text{Poss}(G/B), \alpha, \bar{\rho}].$$

For the belief of G under m_3 we get

$$\text{Bel}_3(G) = \text{Cert}[G/F] = \text{Cert}[G/B] \wedge ((\overline{\alpha \vee \bar{\rho}}) \vee \text{Min}_y[G(y)]),$$
$$\text{Bel}_3(G) = \text{Cert}[G/B] \wedge (\bar{\alpha} \wedge \rho \vee \text{Min}_y(G(y)).$$

If $\text{Min}_y[G(y)] = \text{Cert}[G/Y] = 0$ then we get

$$\text{Bel}_3(G) = \text{Cert}[G/B] \wedge \bar{\alpha} \wedge \rho,$$
$$\text{Bel}_3(G) = T_M(\text{Cert}[G/B], \bar{\alpha}, \rho).$$

At this point it is worth noting the duality relationship between t-norms and t-conorms. In particular for any t-norm there exists a dual t-conorm such that $T(a, b) = 1 - S(\bar{a}, \bar{b})$ [21]. We further recall that Min and Max are duals, $T_M(a, b) = 1 - S_M(\bar{a}, \bar{b})$. Making use of this duality we see that

$$\text{Bel}_3(G) - T_M(\text{Cert}[G/P], \bar{\alpha}, \rho) = 1 - S_M(\overline{\text{Cert}[G/B]}, \alpha, \bar{\rho}).$$

Further recalling that $\text{Cert}[G/B] = 1 - \text{Poss}[\overline{G}/B]$ then we have

$$\overline{\text{Cert}[G/B]} = 1 - (1 - \text{Poss}[\overline{G}/B]) = \text{Poss}[\overline{G}/B].$$

Thus we have

$$\mathrm{Bel}_3(G) = 1 - S_M(\mathrm{Poss}[\overline{G}/B], \alpha, \bar{\rho}),$$
$$\mathrm{Bel}_3(G) = 1 - \mathrm{Pl}_3(\overline{G}).$$

The belief of G is the complement of the plausibility of the complement of G.

We now consider the D–S representation of the probabilistic inference

$$P(B) = (\bar{\lambda}\rho, \lambda + \bar{\lambda}\rho).$$

We first observe that

$$1 - \bar{\rho}\bar{\lambda} = 1 - (1-\lambda)(1-\rho) = 1 - (1 - \rho - \lambda + \rho\lambda) = \rho + \lambda - \rho\lambda = \lambda + \rho\bar{\lambda}.$$

Using this we see that

$$P(B) = [\bar{\lambda}\rho, 1 - \bar{\rho}\bar{\lambda}].$$

From this we can represent this as a D–S belief structure m_4 with three focal elements

$$m_4(B) = \bar{\lambda}\rho,$$
$$m_4(\overline{B}) = \bar{\rho}\bar{\lambda},$$
$$m_4(Y) = \lambda.$$

With this we can calculate for any subset G

$$\mathrm{Pl}_4(G) = \mathrm{Poss}[G/Y]\lambda + \mathrm{Poss}[G/B]\bar{\lambda}\rho + \mathrm{Poss}[G/\overline{B}]\bar{\rho}\bar{\lambda}.$$

With G normal, Poss[G/Y] = 1, then

$$\mathrm{Pl}_4(G) = \lambda + \bar{\lambda}(\rho\mathrm{Poss}[G/B] + \bar{\rho}\mathrm{Poss}[G/\overline{B}]),$$
$$\mathrm{Pl}_4(G) = S_P(\lambda, (\rho\mathrm{Poss}[G/B] + \bar{\rho}\mathrm{Poss}[G/\overline{B}])).$$

Here we see

$$\mathrm{Pl}_4(G) = \lambda \quad \textbf{or} \quad (\rho\mathrm{Poss}[G/B] + \bar{\rho}\mathrm{Poss}[G/\overline{B}]).$$

An interesting special case is when $\mathrm{Poss}[G/\overline{B}] = 1$, in this case

$$\rho\ \mathrm{Poss}[G/B] + \bar{\rho}\ \mathrm{Poss}[G/\overline{B}]) = \bar{\rho} + \rho\ \mathrm{Poss}[G/B] = S_P(\bar{\rho}, \mathrm{Poss}[G/B]).$$

Hence in this case we get

$$\mathrm{Pl}_4[G] = S_P(\lambda, \bar{\rho}, \mathrm{Poss}[G/B]).$$

We now turn to the expression for the belief in this case of m_4.

$$\mathrm{Bel}_4[G] = \lambda\mathrm{Cert}[G/Y] + \bar{\lambda}\rho\mathrm{Cert}[G/B] + \bar{\rho}\bar{\lambda}\mathrm{Cert}[G/\overline{B}],$$
$$\mathrm{Bel}_4[G] = \lambda\mathrm{Cert}[G/Y] + \bar{\lambda}(\rho\mathrm{Cert}[G/B] + \overline{\rho}\mathrm{Cert}[G/B]).$$

In the case where Cert[G/Y] = Min$_y$[$G(y)$] = 0 we get then

$$\mathrm{Bel}_4[G] = \bar{\lambda}(\rho\mathrm{Cert}[G/B] + \overline{\rho}\mathrm{Cert}[G/\overline{B}]),$$
$$\mathrm{Bel}_4[G] = T_P(\bar{\lambda}, (\mathrm{Cert}[G/B] + \overline{\rho}\mathrm{Cert}[G/\overline{B}])).$$

If we further assume here that $\mathrm{Poss}[\overline{G}/B] = 1$, then $\mathrm{Cert}[G/\overline{B}] = 1 - \mathrm{Poss}[\overline{G}/\overline{B}] = 1$ hence

$$\mathrm{Bel}_4[G] = T_P(\bar{\lambda}, \rho\mathrm{Cert}[G/B]).$$

7. Conclusion

We looked at the problem of inferring information about the value of a variable V using information about the value of some other variable U and knowledge about a relationship between values of the variables U and V. Two approaches were used. One used the Zadeh's theory of approximate reasoning and the other used probabilistic reasoning. We pointed out that both of these methods allowed for the inclusion of granular type information. This feature makes these methods compatible with human type reasoning which strongly relies on the use of granular information. The inferred values from each of these methods were expressed in the framework of Dempster–Shafer theory. We then showed an underling unity between these two approaches. This underlying unity can help provide a deeper understanding of the inference process in the case of granular imprecise information.

References

[1] L.A. Zadeh, Outline of a computational theory of perceptions based on computing with words, in: N.K. Sinha, M.M. Gupta (Eds.), Soft Computing and Intelligent Systems, Academic Press, Boston, 1999, pp. 3–22.
[2] L.A. Zadeh, Toward a logic of perceptions based on fuzzy logic, in: W. Novak, I. Perfilieva (Eds.), Discovering the World with Fuzzy Logic, Physica-Verlag, Heidelberg, 2000, pp. 4–28.
[3] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, NJ, 1976.
[4] L.A. Zadeh, Fuzzy sets and information granularity, in: M.M. Gupta, R.K. Ragade, R.R. Yager (Eds.), Advances in Fuzzy Set Theory and Applications, North-Holland, Amsterdam, 1979, pp. 3–18.
[5] L.A. Zadeh, From imprecise to granular probabilities, Fuzzy Sets and Systems (2005) 370–374.
[6] A. Bargiela, W. Pedrycz, Granular Computing: An Introduction, Kluwer Academic Publishers, Amsterdam, 2003.
[7] T.S. Lin, Y.Y. Yao, L.A. Zadeh, Data Mining, Rough Sets and Granular Computing, Physica-Verlag, Heidelberg, 2002.
[8] L.A. Zadeh, A theory of approximate reasoning, in: J. Hayes, D. Michie, L.I. Mikulich (Eds.), Machine Intelligence, vol. 9, Halstead Press, New York, 1979, pp. 149–194.
[9] L.A. Zadeh, Toward a generalized theory of uncertainty (GTU) – an outline, Information Sciences 172 (2005) 1–40.
[10] R.R. Yager, Knowledge trees and protoforms in question answering systems, Journal of the American Society for Information Science and Technology 57 (2006) 550–563.
[11] C.G. Wagner, Modus tollens probabilized, British Journal for the Philosophy of Science 55 (2004) 747–753.
[12] N. Pfeifer, and G. Kleiter, Human reasoning with imprecise probabilities: modus ponens and denying the antecedent, in: Proceedings Fifth International Symposium on Imprecise Probability: Theories and Applications (ISIPTA'07), Prague, 2007.
[13] L. Zadeh, The concept of a linguistic variable and its application to approximate reasoning: Part 1, Information Sciences 8 (1975) 199–249.
[14] R.R. Yager, Approximate reasoning as a basis for computing with words, in: L.A. Zadeh, J. Kacprzyk (Eds.), Computing with Words in Information/Intelligent Systems, vol. 1, Springer-Verlag, Heidelberg, 1999, pp. 50–77.
[15] R.R. Yager, L. Liu, in: A.P. Dempster, G. Shafer (Advisory Eds.), Classic Works of the Dempster–Shafer Theory of Belief Functions, Springer, Heidelberg, 2008.
[16] R.R. Yager, On the Dempster–Shafer framework and new combination rules, Information Sciences 41 (1987) 93–137.
[17] C. Alsina, E. Trillas, L. Valverde, On some logical connectives for fuzzy set theory, Journal of Mathematical Analysis and Applications 93 (1983) 15–26.
[18] R.R. Yager, A class of fuzzy measures generated from a Dempster–Shafer belief structure, International Journal of Intelligent Systems 14 (1999) 1239–1247.
[19] P. Smets, R. Kennes, The transferable belief model, Artificial Intelligence 66 (1994) 191–234.
[20] T. Denoeux, L.M. Zouhal, Handling possibilistic labels in pattern classification using evidential reasoning, Fuzzy Sets and Systems 122 (2001) 47–62.
[21] E.P. Klement, R. Mesiar, E. Pap, Triangular Norms, Kluwer Academic Publishers, Dordrecht, 2000.