

Notes on Dempster Shafer Belief Functions

With standard probability approaches, we don't get an idea of any confidence in our beliefs. For instance, consider a coin where we don't know if it is fair or not (or maybe it even has heads on both sides). If we know nothing about this coin, a Bayesian approach would assign the prior probabilities $P(h)=P(t)=.5$. But this is the same value as if we had flipped the coin 1000 times and it had come up heads 500 times!! Clearly, a single probability does not capture how uncertain we are about our probabilities.

One approach to addressing this problem is to consider an interval of probabilities, specifying reasonable upper and lower bounds on what you believe the probability is. For example, if we have never flipped the coin so know nothing about it, our belief about the coin coming up heads might be $[0, 1]$. In contrast, after flipping the coin 1000 times and finding it fair, our belief in the coin coming up heads would be close to $[.5,.5]$.

How might we compute these intervals? We could use a statistical approach, a **confidence interval**, and use statistics to determine, say, what bounds would give us a 95% confidence that the true probability falls between the bounds. We will not have time to explore this option in this course, however.

Rather we will look at interval representation based on more intuitive measures. Say we park our car in our usual spot but we don't know whether a friend might be picking it up during the day. Say Sue comes into the office and says your car is not there. But we know that Sue is absent minded and so only 90% of the time is she looked in the right spot. Thus, 90% of the time we have good evidence that the car is gone, whereas the other 10% we have no information one way or the other. Thus, we might believe that the probability that the car is missing is at least .9, and might be up to 1.0. This we capture with what Brachman and Levesque call a **possibility interval** of $[.9, 1.0]$ - where the lower bound is called the **belief** in the proposition, and the second is the **plausibility** of it being true, say if all the uncertainty were resolved in favor of our hypothesis. We write the first term as $Bel(p)$, and the second as $Pl(p)$, and we can relate them with the formula

$$Bel(p) = 1 - Pl(\neg p)$$

i.e., the first indicates how certain we are about p , where the second indicates how high the probability could go given how certain we are that $\neg P$. In this example, we have no evidence that the car is still there, so the interval is $[.9, 1.0]$.

Say we have another friend Bill, who also says that the car is missing. But we know that Bill only looks in the right spot 80% of the time. Thus, looking at just his evidence, we would have the possibility interval $[.8, 1.0]$. Now consider if both Sue and Bill tell me that the car is missing - how might I combine this evidence?

It is true that the probability that both are reliable is $.8 * .9 = .72$, but this isn't the number we are interested in, since only one of them has to be reliable for us to know that the car is missing. The probability that both Bill and Sue are unreliable is $.1 * .2 = .02$, so 98% of the time one of them is reliable and my car will be missing. Thus, my new degree of belief is .98, and the possibility interval is $[.98, 1.0]$. In this case, all my evidence was consistent, and as long as we believe the two pieces of evidence are independent, our conclusion seems justified.

What happens when evidence is inconsistent? Say Sue says the car is missing, and Bill says the car is there (but might have misrecognized it when he looked in the wrong spot). Thus, for Sue

we have the possibility interval of [.9, 1.0] as before, and for Bill we have [0, .2]. We have to consider four cases:

Sue is reliable, Bill is reliable - impossible since they both can't be right!

Sue is reliable, Bill is not, with probability $.9 * .2 = .18$. In this case, the car is gone.

Bill is reliable, Sue is not, with probability $.1 * .8 = .08$. In this case the car is there.

Both are unreliable, with probability $.2 * .1 = .02$. In this case, we don't know.

How do we convert this into a possibility interval? Essentially we normalize. Since we know at least one is unreliable, the sum of the three non-zero probabilities above should be 1, but the above calculations only give us $.18+.08+.02 = .28$. We fix this by normalizing, and thus we believe that the car is missing with $.18/.28 = .643$ and that the car is not missing as $.08/.28 = .286$. Thus the possibility interval for the car being missing is [.643, 1-.286], which equals [.643, .714].

Formalization

We formalize the above intuitions in terms of a universe of mutually exhaustive distinct hypotheses H . In the above examples, the set might have been {CarThere, CarMissing}. We define a **bpa** (basic probability assignment) as a function m that maps subsets of H to $[0,1]$, such that $m(X)$ is between 0 and 1 for all subsets X of H , and

$$\sum_{X \subseteq H} m(X) = 1$$

We define another function called the **Belief function**, also from subsets of H to $[0, 1]$ such that

$$Bel(A) = \sum_{X \subseteq A} m(X)$$

It is easy to see that $Bel(H) = 1$, and $Bel(\emptyset) = 0$, where \emptyset is the empty set. We also can define a plausibility function **Pl** as before

$$Pl(A) = 1 - Bel(\neg A), \text{ where } \neg A \text{ is the set complement of } A \text{ in } H, \text{ i.e., } H-A.$$

For example, the situation above where Sue and Bill disagree is formalized as follows:

$$H = \{\text{CarThere}, \text{CarMissing}\}$$

$$m_1(\{\text{CarMissing}\}) = .9, m_1(H) = .1 \text{ (Sue's evidence)}$$

$$m_2(\{\text{CarThere}\}) = .8, m_2(H) = .2 \text{ (Bill's evidence)}$$

Note that the bpa for the set H encodes the uncertainty in the evidence.

Combining Evidence

The Dempster-Shafer rule for combining evidence is as follows:

$$m_3(A) = \frac{\sum_{X \cap Y = A} m_1(X) * m_2(Y)}{1 - \sum_{X \cap Y = \emptyset} m_1(X) * m_2(Y)}$$

The denominator performs the normalization of the probabilities as we saw in the example above. The numerator collects the different forms of evidence for each subset. Applying this to the example, we see the following *before normalization*:

$$\text{for } \{\text{CarMissing}\}: m1(\{\text{CarMissing}\}) * m2(H) = .9 * .2 = .18$$

$$\text{for } \{\text{CarThere}\}: m1(H) * m2(\{\text{CarThere}\}) = .1 * .8 = .08$$

$$\text{for } H: m1(H) * m2(H) = .2 * .1 = .02$$

These are exactly the calculations performed in the introduction. If we compute the denominator, we get the normalization constant. The only pair of X and Y that intersect to the empty set is $X = \{\text{CarMissing}\}$ and $Y = \{\text{CarThere}\}$, which produces a value of .72, yielding a denominator of .28 as before.

Thus we have $m3(\{\text{CarMissing}\}) = .643$, $m3(\{\text{CarThere}\}) = .286$, and $m3(H) = .07$. Thus, we can compute $\text{Bel}(\{\text{CarMissing}\}) = .643$, and $\text{Pl}(\{\text{CarMissing}\}) = 1 - \text{Bel}(\{\text{CarThere}\}) = 1 - .286 = .714$, and thus the possibility interval for $\{\text{CarMissing}\}$ is $[\text{.642}, \text{.714}]$.

A More Extensive Example

Consider a medical diagnosis problem in which we have four mutually exclusive hypotheses: C(common cold), F(flu), M(Menengitis) and N(no problem). Thus $H = \{C, F, M, N\}$. Let assume that based on studies of medical cases we know that a fever supports the hypotheses $\{C, F\}$ at level .5, $\{M\}$ at level .2. We capture this as the bpa $m1$, i.e.,

$$m1(\{C, F\}) = .5, m1(\{M\}) = .2, m1(H) = .3$$

Assume also we know that if a patient has nausea then this supports the hypotheses $\{C, F, N\}$ at a level .7. We capture this with the bpa $m2$:

$$m2(\{C, F, N\}) = .7, m2(H) = .3$$

Finally, assume that if a lab test comes back positive, we know it supports the hypotheses $\{M\}$ at level .8. Thus, we have $m3$ defined by

$$m3(\{M\}) = .8, m3(H) = .2$$

Consider if we know that a patient has both a fever and nausea. We can compute the new bpa by consider all possible intersections of $m1$ and $m2$.

	m1	bpa	m2	bpa	intersection	product
$\{C, F\}$.5		$\{C, F, N\}$.7	$\{C, F\}$.35
$\{C, F\}$.5		H	.3	$\{C, F\}$.15
$\{M\}$.2		$\{C, F, N\}$.7	\emptyset	.14
$\{M\}$.2		H	.3	$\{M\}$.06

	H	.3	{C, F, N}	.7	{C, F, N}	.21
H		.3	H	.3	H	.09

Given this table, we can now compute m_3 . Note that there is only one combination that yields the empty set, and its weight is .14. So the denominator will be $1 - .14 = .86$. For each other resulting set, we add all the weights associated with that set:

$$m_3(\{C, F\}) = (.35 + .15) / .86 = .581$$

$$m_3(\{M\}) = .06 / .86 = .07$$

$$m_3(\{C, F, N\}) = .21 / .86 = .244$$

$$m_3(H) = .09 / .86 = .104$$

Consider the values of the Bel function for its non-zero values:

$$\text{Bel}(\{C, F\}) = .581 \quad \text{Pl}(\{C, F\}) = 1 - \text{Bel}(\{N, M\}) = 1 - \text{Bel}(\{M\}) = .93$$

$$\text{Bel}(\{M\}) = .07 \quad \text{Pl}(\{M\}) = 1 - \text{Bel}(\{C, F, N\}) = .175$$

$$\text{Bel}(\{C, F, N\}) = .825 \quad \text{Pl}(\{C, F, N\}) = 1 - \text{Bel}(M) = .93$$

We see quite strong belief that the patient has either C, F or N, indicated by the possibility interval [.825, .93], and a correspondingly weak belief in the meningitis with the interval [.07, .175].

We leave it as an exercise to know see how things change if we get back a positive lab test (m_3).