WORKING PAPER NO. 164

The Problem of Dependent Evidence

Glenn Shafer

October, 1984

SCHOOL OF BUSINESS
WORKING PAPER SERIES

THE UNIVERSITY OF KANSAS · LAWRENCE

WORKING PAPER NO. 164

The Problem of Dependent Evidence

Glenn Shafer

October, 1984

Research partially supported by NSF Grant MCS-8301282.

The Problem of Dependent Evidence

ABSTRACT: It is appropriate to use Dempster's rule for combining belief functions only if the belief functions combined are based on independent items of evidence. What can be done in the case of dependent evidence? Often the answer is to reframe the problem. Three examples are given: one from everyday experience, one from probabilistic relaxation, and one from expert systems.

The theory of belief functions (Shafer 1976) emphasizes Dempster's rule for combining belief functions based on independent items of evidence. Several authors (e.g., Quinlan 1982) who have explored the use of belief functions in artificial intelligence have pointed out that items of evidence are often dependent. How the theory can be used in this case? Can Dempster's rule be generalized to deal with dependent evidence?

In order to deal with this question, it is necessary to recognize that independence is always relative to a frame of discernment. In practice it is always possible to find common uncertainties affecting the assessment of two items of evidence. If, however, we bring the important common uncertainties explicitly into the frame of discernment, then it may be reasonable to treat the remaining uncertainties as independent with respect to that frame.

This paper discusses some simple examples that show how proper framing can enable us to sort out the uncertainties in our evidence so that Dempster's rule can be used. These examples suggest that more will be gained by seeking methods and insights that will help us with this framing than by seeking generalizations of Dempster's rule.

Example 1: The Missing Cookie

Who took the cookie? There are two items of evidence indicating that Sally took it. First, Billy says he saw Sally take it. Second, there seem to be cookie crumbs in the corner where Sally always hides to eat treats.

The simplest way to assess this evidence is to treat the two items as independent with respect to a frame of discernment that has only two elements:

Sally did.

Sally didn't.

We may represent each item by a simple support function and then combine them by Dempster's rule. Since Billy is an unusually honest child, we assess his testimony by giving a degree of support 0.7 to "Sally did." Since all the other children have ceded the corner to Sally, and those crumbs do look like cookie crumbs, we assess them by giving a degree of support 0.5 to "Sally did." Combination by Dempster's rule yields an overall degree of support of $1-(1-0.7)(1-0.5) = 0.85$.

Perhaps, however, it is not reasonable to regard the two items of evidence as independent with respect to this very coarse frame of discernment. Perhaps Billy ate the cookie himself and deliberately dropped the crumbs in the corner to incriminate Sally.

In order to think about this, we need to refine our frame so as to take the possibility of such deviousness on Billy's part explicitly into account. Here is a refinement with four elements:

1. Sally did, and Billy saw her.

2. Sally did, but Billy didn't see her.

3. Sally didn't, and Billy dropped the crumbs deliberately.

4. Sally didn't, but Billy didn't drop the crumbs deliberately.

In the context of this frame, we can think of our two items of evidence, Billy's testimony and the crumbs, in a slightly different way. Billy's testimony and reputation constitute an argument for element 1, while the crumbs and the puzzle about how they might have gotten into the corner are an argument against element 4. These two arguments do seem to be independent. If we weigh them as before, then we have a simple support function with support 0.7 focused on {1}, and another with support 0.5 focused on {1,2,3}. Combination by Dempster's rule now produces a belief function with mass 0.7 focused on {1}, mass 0.15 focused on {1,2,3}, and mass 0.15 focused on the whole frame. Instead of 0.85, we have a total belief of only 0.7 for {1,2} = "Sally did."

There is more to be said. Billy's reputation gives us reason not only to trust his word but also even stronger reason to doubt that he would be so deviously malicious as to frame Sally. A better analysis would take this into account. One way to do so is to replace the simple support function with belief 0.7 focused on {1} by a consonant belief function, say with mass 0.7 on {1}, mass 0.2 on {1,2,4}, and mass 0.1 on the whole frame. Combination now results in a belief function with mass 0.7 on {1}, mass 0.1 on {1,2}, mass 0.1 on {1,2,4}, mass 0.05 on {1,2,3}, and mass 0.05 on the whole frame. We now have a total belief of 0.8 for {1,2} = "Sally did."

Notice that both of our analyses using the four-element frame essentially reduce to the analysis using the two-element frame if we condition on {1,2,4} -- i.e., on Billy's not having acted maliciously.

We may express this by saying that Billy's not having acted maliciously was one of the implicit assumptions made in the original analysis.

In this example, we could not initially regard our two items of evidence as independent because they involved a common uncertainty: the uncertainty as to whether Billy was being thoroughly devious. Once we brought this common uncertainty explicitly into the frame of discernment, we found that it was reasonable to regard the other uncertainties involved in the items of evidence as independent.

It may be helpful to say that the two items of evidence--Billy's testimony and the crumbs--themselves change as we move from the one frame to the other. Such change is possible because these items of evidence are defined only ostensively. "Billy's testimony" refers not to a single definite fact or proposition, but rather to whatever argument we can construct from Billy's testimony and from our experience with the reliability of Billy and other children. Similarly, "the crumbs" refers to whatever argument we can construct from our experience of things that look like cookie crumbs and from our experience of the consistency with which the other children stay away from Sally's corner. We will construct different arguments depending on the frame with which we are working. The arguments we construct with respect to one frame may involve common uncertainties, while the arguments we construct with respect to another frame may involve independent uncertainties.

This example suggests that the "problem of dependent evidence" is often better described as a trade-off between complexity in our evidence and complexity in our frame. We can usually decompose our evidence into independent items if we are willing to make our frame sufficiently complex. But we seek both independent items of evidence and a simple frame. Probability arguments are most feasible and most convincing when they are built out of probability judgments that relate simple items of evidence to simple questions. The need for simple items of evidence pushes us to decompose our evidence. The need for simple questions pushes us to keep our frame simple.

Example 2: Probabilistic Relaxation

Figure 1 presents a simple example of scene labeling that Rosenfeld et al. (1976) used to illustrate a non-linear probabilistic relaxation method. Part (c) of the figure shows eight possible interpretations for an isolated triangle. The probabilities in Part (d) are obtained by treating these eight interpretations as equally likely. The symbol → has probability 3/8, for example, because 9 of the 24 edges in Part (c) are labeled →. Part (e) similarly gives the conditional probabilities for what symbol will follow a given symbol in the clockwise direction. The symbol - is always followed by →, for example, so $P(\to|-) = 1$.

Suppose we obtain observational evidence about a particular triangle, and this evidence is summarized by probabilities for the types of its three edges. Rosenfeld's non-linear probabilistic relaxation method uses these probabilities together with the prior probabilities in (d) and (e) to select an interpretation for the triangle. If, for example, the observational evidence assigns probabilities 1/2 to → and - for each of the three edges, then this method selects the interpretation I. (See case B of Figure 6 in Rosenfeld et al. 1976.)

The effect of this relaxation method can be mimicked using belief functions. We begin with a frame of discernment that allows unrestricted assignments of types to the three edges; formally,

$$F = \{\rightarrow, \leftarrow, +, -\}^3.$$

(This frame has 64 elements, but only the eight of these that correspond to the interpretations in (e) are actually possible. The element ($\leftarrow$, +, $\leftarrow$), for example, refers to interpretation VI.) Then we construct belief functions as follows:

(i) For each of the three edges, we assign the probabilities given by (d) and minimally extend to a belief function on F. (Doing this for edge E1, for example, gives a belief function that assigns mass 3/8 to $\{\rightarrow\} \times \{\rightarrow, \leftarrow, +, -\}^2$, 3/8 to $\{\leftarrow\} \times \{\rightarrow, \leftarrow, +, -\}^2$ 1/8 to $\{+\} \times \{\rightarrow, \leftarrow, +, -\}^2$, and 1/8 to $\{-\} \times \{\rightarrow, \leftarrow, +, -\}^2$.) This gives three belief functions on F.

(ii) For each pair of edges, and each of the four symbols, we conditionally embed (see Shafer 1982) the probabilities for what the next edge will be given that the first edge is identified by the given symbol. This gives 12 more belief functions on F.

(iii) For each edge, we construct a belief function that represents the observational evidence for its type by assigning probabilities 1/2 to $\rightarrow$ and - for that edge. This gives three more belief functions on F.

(iv) We combine the 18 belief functions by Dempster's rule. The resulting belief function assigns probability 4/7 to I and probability 1/7 to each of III, IV, and V.

If we now mimic the iteration that is used in the relaxation method by repeatedly combining this belief function with itself, we will obtain a probability of one for I.

The preceding analysis is faulty because the belief functions in

(i) and (ii) are not independent; they are all derived from Part (c) of Figure 1. It hardly seems worthwhile, though, to try to formulate a rule of combination that generalizes Dempster's rule in a way appropriate to this particular case of dependent evidence. Instead, we should use the right frame: a frame consisting of the eight possible interpretations.

The belief-function analysis using this frame is straightforward. We begin with a "prior" belief function that assigns probability 1/8 to each of the eight interpretations. We then combine this by Dempster's rule with the three belief functions mentioned in (iii) of the paragraph before last. This results in a belief function that assigns probability 1/4 to each of the four interpretations I, III, IV, and V. There is no priority assigned to interpretation I.

This example makes us wonder how often probabilistic relaxation methods go astray because they treat dependent evidence as if it were independent. Here the dependencies appeared in the "prior" evidence. But the problem also arises for observational evidence; see, e.g., Hanson and Riseman (1978).

<u>Example 3: Electronics Troubleshooting</u>

Cantone et al. (1983) report that IN-ATE, their expert system for electronics troubleshooting, uses Dempster's rule for combining independent test results. They discuss the problem of dependent test results as follows:

> Dempster's rule works fine for combining evidence from sources that are <u>independent</u>. But because of circuit connectivity, test results are not always independent. In IN-ATE, non-independent test results are defined as those

that are taken from a common path in the UUT |Unit Under Test circuitry|. The interpretation assigned by the inference engine given two non-independent test results depends upon whether the tests passed or failed and upon which result is upstream or downstream from the other. For example, if two non-independent results are both bad (good), then only the more useful--the upstream (downstream) result--is retained and interpreted by the inference engine. As another example, when one passed test result later clears away the blame from some component--assigned as a result of an earlier failed test result--that blame is removed from the component and the original failed test result is re-interpreted in light of this new information.

It is worth noting that these examples of dependent evidence can be handled quite reasonably by Dempster's rule if proper frames are used. In the simplest cases, all that is required is conditioning, which is a special case of Dempster's rule.

Consider, for example, the simple circuit depicted in Figure 2. Suppose we begin with the simplifying assumption that at most one of the three components C1, C2, and C3 is bad. Then we can work with a frame of discernment that consists of these two possibilities. If an initial test shows the output of C3 is bad, then we conclude that one of the three components is indeed bad, and following the practice of Cantone et al., we assign equal probabilities to the three possibilities.

If a second test now shows the input to C3 to be bad, then we know that C1 or C2 is bad. This knowledge can be represented by a belief function that assigns unit mass to the subset {C1, C2} of the frame {C1, C2, C3}. Combining this belief function with the initial belief function that assigned equal probabilities to the three possibilities is equivalent to conditioning this initial belief function on {C1, C2}, and hence results in equal probabilities for these two possibilities.

Another way of getting the same result is to test C3 using a different input, one not involving C1 and C2. If we then get a good output from C3, we can remove the blame from C3 simply by conditioning on C3 being good--i.e., by again conditioning on {C1, C2}.

Conclusion

Problems of framing arise in all domains where probability arguments are used. But these problems can be especially acute in artificial intelligence. The need for relatively simple frames is especially acute because initial frames are often already so complex that it is impractical to refine them so as to make common uncertainties explicit. And the need for formal insight into proper framing is especially acute because such framing must be engineered before observational evidence is obtained.

Though the problem of dependent evidence has been discussed here in the context of the theory of belief functions, essentially the same problem arises in other theories of probability judgment. In the Bayesian theory, for example, we find the same trade-off between complexity in the evidence and complexity in the frame. In principle, evidence of arbitrary complexity can be directly assessed to give Bayesian probabilities over simple frames, but in practice simple frames can be used only if reasonable assumptions of independence are possible relative to those frames.

In practice it may often be necessary for artificial intelligence techniques to ignore potential dependencies in evidence. In many cases, what is required is not the most thorough possible analysis of a particular problem but rather a method that will generally work well. Often an immediate answer is required, without immediate attention to the probability or reliability of that answer. It may be useful, nevertheless, to study the relationship between practical techniques that ignore dependencies and the more complex methods that could, in principle, take these dependencies into account.

References

Cantone, Richard R.; Pipitone, Frank J.; Lander, W. Brent; and Marrone, Michael P. (1983). "Model-Based Probabilistic Reasoning for Electronics Troubleshooting," in Proceedings of the Eighth International Joint Conference on Artificial Intelligence.

Hanson, Allen R., and Riseman, Edward M. (1978). "Segmentation of Natural Scenes," pp. 129-163 of Computer Vision Systems, edited by Hanson and Riseman, Academic Press.

Quinlan, J.R. (1982). Inferno: A Cautious Approach to Uncertain Inference, Rand Report N-1898-RC.

Rosenfeld, Azriel; Hummel, Robert A.; and Zucker, Steven W. (1976). "Scene Labeling by Relaxation," IEEE Transactions on Systems, Man, and Cybernetics, SMC-6, 420-433.

Shafer, Glenn (1976). A Mathematical Theory of Evidence, Princeton University Press, Princeton, N.J.

Shafer, Glenn (1982). "Belief Functions and Parametric Models," Journal of the Royal Statistical Society, Series B, 44, 322-352.

(a) The triangle.

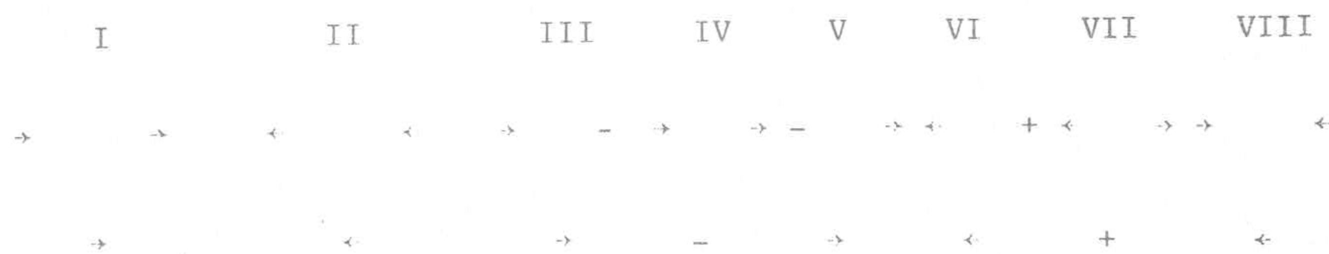
(b) Three types of edge.

+ convex

- concave

→ occluding, with the right-hand region (when one faces along the line in the direction of the arrow) hiding the left-hand region

(c) Eight interpretations of the triangle.



Triangle above background	Triangular hole	Triangular flap folded upward	Triangular flap folded downward

(d) Probabilities.

$P(\rightarrow) = 3/8$

$P(\leftarrow) = 3/8$

$P(+) = 1/8$

$P(-) = 1/8$

(e) Conditional probabilities.

$P(\rightarrow \mid \rightarrow)=2/3$	$P(\rightarrow \mid \leftarrow)=0$	$P(\rightarrow \mid +)=0$	$P(\rightarrow \mid -)=1$
$P(\leftarrow \mid \rightarrow)=0$	$P(\leftarrow \mid \leftarrow)=2/3$	$P(\leftarrow \mid +)=1$	$P(\leftarrow \mid -)=0$
$P(+ \mid \rightarrow)=0$	$P(+ \mid \leftarrow)=1/3$	$P(+ \mid +)=0$	$P(+ \mid -)=0$
$P(- \mid \rightarrow)=1/3$	$P(- \mid \leftarrow)=0$	$P(- \mid +)=0$	$P(- \mid -)=0$

Figure 1.

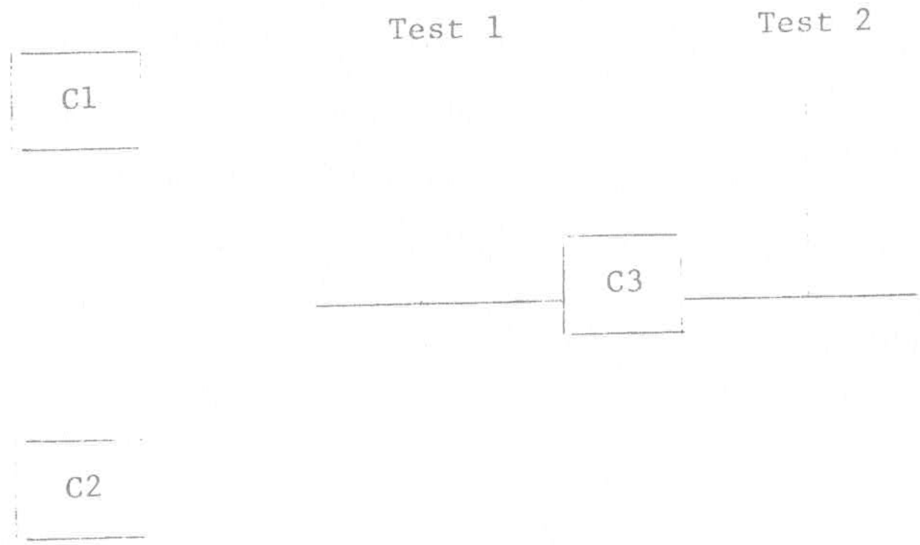


Figure 2.