ON THE VALIDITY OF DEMPSTER'S RULE OF COMBINATION OF EVIDENCE

by

Lotfi A. Zadeh

Memorandum No. UCB/ERL M79/24

21 March 1979

ON THE VALIDITY OF DEMPSTER'S RULE OF COMBINATION OF EVIDENCE

by

L. A. Zadeh

Memorandum No. UCB/ERL M79/24

21 March, 1979

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

ON THE VALIDITY OF DEMPSTER'S RULE OF COMBINATION OF EVIDENCE

L.A. Zadeh*.

Abstract

In a seminal paper published in 1967, Dempster has described a rule for combining independent sources of information. More recently, Dempster's rule has been employed as a basis for a mathematical theory of evidence. It is suggested in this note that there is a serious flaw in Dempster's rule which restricts its use in many applications.

*Computer Science Division, Department of Electrical Engineering and Computer Sciences and the Electronics Research Laboratory, University of California, Berkeley, CA 94720. Research supported by Naval Electronic Systems Command Contract N00039-78-C0013 and National Science Foundation Grant MCS77-07568.

ON THE VALIDITY OF DEMPSTER'S RULE OF COMBINATION OF EVIDENCE

L.A. Zadeh

1. Introduction

During the past several years, the development of expert systems typified by MYCIN [1], PROSPECTOR [2], IRIS [3] and others [4]--systems which have a sophisticated question-answering capability coupled with the ability to provide the user with an assessment of the degree of credibility of the system's response--has accentuated the need for a better understanding of the issues relating to the concepts of evidence, belief and credibility, and stimulated a critical analysis of some of the earlier work.

In an important paper bearing on these issues [5], Dempster has presented a rule for combining independent sources of information--a rule which yields the degree of belief that the value of a variable X lies in a specified subset of its range. More recently, Dempster's rule of combination of evidence, along with the concepts of upper and lower probabilities which were employed by him, have formed a basis for the construction of a mathematical theory of evidence by Shafer [6], and its applications to medical diagnosis [7] and other fields.

The purpose of this note is to raise a question with regard to the validity of Dempster's rule and suggest a way of correcting what appears to be a serious flaw in its formulation. This flaw becomes apparent when the problem analyzed by Dempster is viewed in the context of a combination of a mixture of probability and possibility distributions.

2. Dempster's Model

The point of departure in Dempster's paper is the following model. Briefly, let U and V be a pair of spaces together with a multivalued mapping Γ which assigns a subset $\Gamma u \subset V$ to every $u \in U$. Let μ be a probability measure defined on U. Then, for a subset A of V, define the subsets A^*, A_* and A_θ of U by

$$A^* \triangleq \{u \mid \Gamma u \cap A \neq \theta\} , \qquad \theta \triangleq \text{empty set} \tag{1}$$

$$A_* \triangleq \{u \mid \Gamma u \subset A \text{ and } \Gamma u \neq \theta\} \tag{2}$$

$$A_\theta \triangleq \{u \mid \Gamma u \neq \theta\} \tag{3}$$

and let $\mu(A^*)$, $\mu(A_*)$ and $\mu(A_\theta)$ be, respectively, the probability measures of A^*, A_* and A_θ. In terms of these measures, the <u>upper</u> and <u>lower</u> probabilities of A are defined, respectively, as [5], [14], [27]:

$$P^*(A) = \mu(A^*)/\mu(A_\theta) \tag{4}$$

and

$$P_*(A) = \mu(A_*)/\mu(A_\theta) \tag{5}$$

where $\mu(A_\theta)$ plays the role of a normalizing factor. Based on these definitions, Shafer in his theory of evidence [6] identifies $P_*(A)$ with the degree of belief and $P^*(A)$ with the degree of plausibility of the proposition $Y \in A$, where Y is a variable taking values in V.

As we shall see in the following section, the crux of the difficulty with Dempster's rule lies in the assumption that, if there are points in U which map into θ and $\mu(A_\theta) < 1$, then it is appropriate to normalize $\mu(A^*)$ and $\mu(A_*)$ in the manner of (4) and (5).

To see the issue in a clearer perspective, we shall reformulate Dempster's model by making use of the concept of a possibility distribution [8].[1] More specifically, if X is a variable taking values in U then the possibility distribution of X is simply the set of possible values that may be assumed by X. If the degree of possibility is allowed to take values in the interval $[0,1]$ --and not just 0 or 1 -- then

$$\pi_X(u) \triangleq \mathrm{Poss}\{X = u\} \tag{6}$$

where $\pi_X\colon U \to [0,1]$ is the possibility distribution function which characterizes Π_X. The possibility measure of a subset A of U is defined as

$$\begin{aligned}\Pi(A) &\triangleq \mathrm{Poss}\{X \in A\} \\ &= \sup_{u \in A} \pi_X(u)\end{aligned} \tag{7}$$

and Π is F-additive in the sense that

$$\Pi(A \cup B) = \Pi(A) \vee \Pi(B) \tag{8}$$

where $\vee$ denotes max (in infix form).

If there are two variables X and Y, ranging over U and V, respectively, then the conditional possibility distribution of Y given X is denoted by $\Pi_{(Y|X)}$ and is characterized by its conditional possibility distribution function

$$\pi_{(Y|X)}(v|u) \triangleq \mathrm{Poss}\{Y = v | X = u\} \ . \tag{9}$$

In terms of $\pi_{(Y|X)}(v|u)$, the joint possibility distribution functions of X and Y may be expressed as

[1]The concept of a possibility distribution may be employed in a similar manner in the characterization of upper and lower probabilities through the use of random relations [9] and random sets [10].

$$\pi_{(X,Y)}(u,v) = \pi_X(u) \wedge \pi_{(Y|X)}(v|u) \tag{10}$$

where $\wedge \triangleq \min$ (in infix form). However, we may also write

$$\pi_{(X,Y)}(u,v) = \pi_{(X|Y)}(u|v) \wedge \pi_{(Y|X)}(v|u) \tag{11}$$

which has no probabilistic analog.

For use in Dempster's model, it is sufficient to assume that the range of possibility distribution functions is the set $\{0,1\}$. Thus, what we have in this model is (a) a pair of variables, (X,Y), ranging over U and V, respectively; and (b) what we know about (X,Y) is the "evidence," E:

$$E = \{P_X, \Pi_{(Y|X)}\} \tag{12}$$

which consists of the probability distribution of X, P_X, and the conditional possibility distribution of Y given X, $\Pi_{(Y|X)}$. The question, then, is: What can be inferred about the distribution of Y from the evidence E?

To begin with, we shall assume that $\mu(A_\theta) = 1$, which in terms of $\Pi_{(Y|X)}$ may be expressed as

$$\text{Prob}\{\Pi_{(Y|X)} = \theta\} = 0 \quad . \tag{13}$$

With this assumption, then, the definitions of $P_*(A)$ and $P^*(A)$ may be expressed as

$$P_*(A) = \text{Prob}\{\Pi_{(Y|X)} \subset A\} \tag{14}$$

and

$$P^*(A) = \text{Prob}\{\Pi_{(Y|X)} \cap A \neq \theta\} \ . \tag{15}$$

To make these definitions more intuitive, let $\text{Cert}\{Y \in A|X\}$ and $\text{Poss}\{Y \in A|X\}$ denote, respectively, the degree of conditional certainty and the degree of conditional possibility that the proposition

$$p \triangleq \{Y \in A | X\} \tag{16}$$

is true. (For Dempster's model, the degrees of certainty and possibility can take only the values 0 and 1.) Then, (14) and (15) may be expressed equivalently in the form

$$P_*(A) = E_X(\mathrm{Cert}\{Y \in A | X\}) \tag{17}$$

and

$$P^*(A) = E_X(\mathrm{Poss}\{Y \in A | X\}) \tag{18}$$

where E_X denotes the expectation with respect to X.

We observe that when no normalization is needed (i.e., $\mathrm{Prob}\{\Pi_{(Y|X)} = \theta\} = 0$), Dempster's upper and lower probabilities, correspond, respectively, to the expectation of conditional certainty and conditional possibility.[2] (In Shafer's work, $P_*(A)$ is identified with the degree of belief and $P^*(A)$ with the degree of plausibility. It should be noted, however, that there is no particular reason why the label "belief" should be associated with P_* rather than with P^*.)

As was remarked earlier, the case where $\mathrm{Prob}\{\Pi_{(Y|X)} = \theta\} > 0$ is handled by Dempster through the normalization of $P_*(A)$ and $P^*(A)$. In the notation of (14) and (15), the normalized expressions for $P_*(A)$ and $P^*(A)$ read

$$P_*(A) = \frac{\mathrm{Prob}\{\Pi_{(Y|X)} \subset A\}}{\mathrm{Prob}\{\Pi_{(Y|X)} \neq \theta\}} \tag{19}$$

and

$$P^*(A) = \frac{\mathrm{Prob}\{\Pi_{(Y|X)} \cap A \neq \theta\}}{\mathrm{Prob}\{\Pi_{(Y|X)} \neq \theta\}} \tag{20}$$

[2]It should be noted that certainty and possibility--in the sense used here--are closely related to the concepts of necessity and possibility in modal logic [11].

The point that we wish to argue is that normalization is not an appropriate way of dealing with the issue in question. For, the crux of the difficulty--which is well recognized in the analysis of presuppositions [12] --is that when $\text{Prob}\{\Pi_{(Y|X)}=\theta\} > 0$, the object Y in the conditional proposition $\{Y \in A|X\}$ does not exist, implying that the proposition $\{Y \in A|X\}$ has no truth-value, rather than the truth-value 0 or 1. As we shall see in the sequel, it is--above all--this issue that casts serious doubt on the validity of Dempster's rule of combination of evidence.[3]

3. Combination of Evidence

Suppose that, in the notation of (12), we have two bodies of evidence

$$E_1 = \{P_{X_1}, \Pi_{(Y|X_1)}\} \tag{21}$$

and

$$E_2 = \{P_{X_2}, \Pi_{(Y|X_2)}\} \tag{22}$$

in which X_1 and X_2 are independent random variables ranging over U_1 and U_2, respectively. The combined body of evidence, then, may be expressed as

$$(E_1, E_2) = \{P_{(X_1,X_2)}, \Pi_{(Y|X_1)} \cap \Pi_{(Y|X_2)}\} \tag{23}$$

in which $P_{(X_1,X_2)}$ is the joint probability distribution of X_1 and X_2, and the intersection $\Pi_{(Y|X_1)} \cap \Pi_{(Y|X_2)}$ is the conditional possibility distribution of Y given X_1 and X_2.

By analogy with (14), let $P_*^{(1,2)}(A)$ denote the lower probability corresponding to (23) for the case where

[3]It should be noted that the validity of Dempster's rule has been questioned on different grounds by P.M. Williams in a review of Shafer's "A Mathematical Theory of Evidence" [13].

$$\mathrm{Prob}\{\Pi_{(Y|X_1)} \cap \Pi_{(Y|X_2)} = \theta\} = 0 \quad . \tag{24}$$

Then, for the combined bodies of evidence, we have

$$P_*^{(1,2)}(A) = \mathrm{Prob}\{\Pi_{(Y|X_1)} \cap \Pi_{(Y|X_2)} \subset A\} \tag{25}$$

or, equivalently,

$$P_*^{(1,2)}(A) = E_{(X_1,X_2)}(\mathrm{Cert}\{Y \in A|(X_1,X_2)\}) \ . \tag{26}$$

As in the case of a single body of evidence, a difficulty arises when (24) does not hold. When the latter is true, Dempster's way of resolving the difficulty leads to the following normalized expression for the combined lower probability:

$$P_*^{(1,2)}(A) = \frac{\mathrm{Prob}\{\Pi_{(Y|X_1)} \cap \Pi_{(Y|X_2)} \subset A\}}{\mathrm{Prob}\{\Pi_{(Y|X_1)} \cap \Pi_{(Y|X_2)} \neq \theta\}} \ . \tag{27}$$

Using this expression (or, more precisely, its equivalent in terms of multivalued mappings), Dempster constructs an explicit rule for computing $P_*^{(1,2)}(A)$ from the knowledge of $P_*^1(A)$ and $P_*^2(A)$ for the case where U and V are finite sets. The basic point that is at issue, however, is not the computation of $P_*^{(1,2)}(A)$ from the knowledge of $P_*^1(A)$ and $P_*^2(A)$, but the legitimacy of normalization of $P_*^{(1,2)}(A)$ as expressed by (27). We believe that this normalization is invalid and that, in general, its use leads to counterintuitive results.

The reason for the invalidity of normalization is actually quite simple. If the intersection of $\Pi_{(Y|X_1)}$ and $\Pi_{(Y|X_2)}$ is empty with positive probability, then for some combinations of values of X_1 and X_2 which have a positive probability, the evidence provided by X_1 concerning the values of

Y is in flat contradiction to that provided by X_2. But, since such contradictions cannot be resolved within the theory, it is not permissible to suppress their existence through the artifice of normalization.

The counterintuitive results yielded by Dempster's rule become clearly apparent when it is applied to the combination of probability distributions (or, in Shafer's terminology, to Bayesian belief functions). Thus, assume that $U = V = \{a,b,c\}$, $\Pi_{(Y|X_1)} = \{X_1\}$, $\Pi_{(Y|X_2)} = \{X_2\}$ and

$$\begin{aligned} \text{Prob}\{X_1 = a\} &= P^1_*(\{a\}) = 0.99 \\ \text{Prob}\{X_1 = b\} &= P^1_*(\{b\}) = 0.01 \\ \text{Prob}\{X_1 = c\} &= P^1_*(\{c\}) = 0 \\ \text{Prob}\{X_2 = a\} &= P^2_*(\{a\}) = 0 \\ \text{Prob}\{X_2 = b\} &= P^2_*(\{b\}) = 0.01 \\ \text{Prob}\{X_2 = c\} &= P^2_*(\{c\}) = 0.99 \end{aligned} \tag{28}$$

In application to this case, Dempster's rule of combination of evidence leads to the conclusion that

$$P^{(1,2)}_*(\{b\}) = 1 \tag{29}$$

i.e., the expectation of the degree of certainty that $Y = b$ given the two bodies of evidence is unity. This is clearly inconsistent with the fact that each of the two bodies of evidence taken separately assigns a low degree of belief (0.01) to the proposition $Y = b$.

In conclusion, it appears that Dempster's rule of combination of evidence is invalid when the probability that the intersection $\Pi_{(Y|X_1)} \cap \Pi_{(Y|X_2)}$ is nonempty is less than unity. In this case, there is a positive probability that Y does not exist--a fact which cannot be suppressed through normalization. Thus, to place the nonexistence of Y in evidence, it is necessary to display the probability

$$\text{Prob}\{\Pi_{(Y|X_1)} \cap \Pi_{(Y|X_2)} = \theta\}$$

as a constituent of the pair

$$(\text{Prob}\{\Pi_{(Y|X_1)} \cap \Pi_{(Y|X_2)} \subset A\},\ \text{Prob}\{\Pi_{(Y|X_1)} \cap \Pi_{(Y|X_2)} = \theta\})$$

and not to combine the two into a single degree of belief in the truth of the proposition $\{Y \in A\}$ given the independent bodies of evidence E_1 and E_2.

It is of interest to note that a somewhat similar situation arises when in announcing the results of a vote we feel that it is necessary to specify separately the yeses, the nays and the abstentions. Viewed in this perspective, the use of normalization in Dempster's rule is somewhat analogous to the suppression of information about the abstentions in a voting process.

Acknowledgment. The author wishes to acknowledge helpful discussions on various issues relating to the concepts of upper and lower probabilities with Pat Suppes, Mario Zanotti, Glenn Shafer and Hung Nguyen.

References and Related Papers

[1] Shortliffe, E. and Buchanan, B.G., A model of inexact reasoning in medicine, Mathematical Biosciences 23: 351-379, 1975.

[2] Duda, R. et al., Development of a computer based consultant for mineral exploration (PROSPECTOR), Annual Report SRI International, Nos. 5821 & 6415, 1977.

[3] Trigoboff, M. and Kulikowski, C., IRIS: a system for the propagation of inferences in a semantic net, Proc. 5th IJACI, M.I.T., 274-280, 1977.

[4] Amarel, S. et al., Panel on applications of artificial intelligence, Proc. 5th IJCAI, M.I.T., 994-1006, 1977.

[5] Dempster, A.P., Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Statist. 38: 325-329, 1967.

[6] Shafer, G., A Mathematical Theory of Evidence, Princeton University Press, 1976.

[7] Smets, P., Un modele mathematico-statistique simulant le processus du diagnostic medicale, Free University of Brussels, 1978.

[8] Zadeh, L.A., Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems 1: 3-28, 1978.

[9] Suppes, P. and Zanotti, M., On using random relations to generate upper and lower probabilities, Synthese 36: 427-440, 1977.

[10] Nguyen, H.T., On random sets and belief functions, J. Math. Anal. Appl. 65: 531-542, 1978.

[11] Hughes, G.E. and Cresswell, M.J., An Introduction to Modal Logic, Methuen, London, 1968.

[12] Kempson, R., Semantic Theory, Cambridge University Press, Cambridge, 1977.

[13] Williams, P.M., On a new theory of epistemic probability (review of G. Shafer: A Mathematical Theory of Evidence), Brit. J. for the Philosophy of Science 29: 375-387, 1978.

[14] Suppes, P., The measurement of belief, J. Roy. Statist. Soc. B 36: 160-175, 1974.

[15] Colby, K.M., Simulations of belief systems, in Computer Models of Thought and Language, R.C. Schank and K.M. Colby, eds., W. Freeman, San Francisco, 1973.

[16] Bruce, B.C., Belief systems and language understanding, N.I.H. Report CBM-TR-41, Rutgers University, 1975.

[17] Abelson, R.P., The structure of belief systems, in Computer Models of Thought and Language, R.C. Schank and K.M. Colby, eds., W. Freeman, San Francisco, 1973.

[18] Rosenkrantz, R.D., Inference, Method and Decision, D. Reidel, Dordrecht, 1977.

[19] Rescher, N., Plausible Reasoning, Van Gorcum, Amsterdam, 1976.

[20] Tversky, A. and Kahneman, D., Judgment under uncertainty: heuristics and biases, Science 185: 1124-1131, 1974.

[21] Zadeh, L.A., PRUF--a meaning representation language for natural languages, Int. J. Man-Machine Studies 10: 395-460, 1978.

[22] Woods, W.A., Kaplan, R.M. and Nash-Webber, B., The lunar sciences natural language information system, BBN Report 2378, Cambridge, Mass., 1972.

[23] Miller, G.A. and Johnson-Laird, P.N., Language and Perception, Harvard University Press, Cambridge, 1976.

[24] Adams, B., A probability model of medical reasoning and the MYCIN model, Mathematical Biosciences 32: 177-186, 1976.

[25] Joshi, A., Inferencing in partial information, in Pattern Directed Inference, F. Hays-Roth and D. Waterman, eds., Academic Press, 1978.

[26] Nguyen, H.T., On conditional possibility distributions, Fuzzy Sets and Systems 1: 299-309, 1978.

[27] Good, I.J., Subjective probability as the measure of a non-measurable set, in Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress, E. Nagel, P. Suppes and A. Tarski, eds., Stanford University Press, 1962.