A Simple View of the Dempster-Shafer Theory of Evidence and its Implication for the Rule of Combination

Lotfi A. Zadeh

Computer Science Division, University of California, Berkeley, California 94720

Abstract

During the past two years, the Dempster-Shafer theory of evidence has attracted considerable attention within the AI community as a promising method of dealing with uncertainty in expert systems. As presented in the literature, the theory is hard to master. In a simple approach that is outlined in this paper, the Dempster-Shafer theory is viewed in the context of relational databases as the application of familiar retrieval techniques to second-order relations, that is, relations in which the data entries are relations in first normal form. The relational viewpoint clarifies some of the controversial issues in the Dempster-Shafer theory and facilitates its use in AI-oriented applications

The emergence of expert systems as one of the major areas of activity within AI has resulted in a rapid growth of interest within the AI community in issues relating to the management of uncertainty and evidential reasoning. During the past two years, in particular, the Dempster-Shafer theory of evidence has attracted considerable attention as a promising method of dealing with some of the basic problems arising in combination of evidence and data fusion. To develop an adequate understanding of this theory requires considerable effort and a good background in probability theory. There is, however, a simple way of approaching the Dempster-Shafer theory that only requires a minimal familiarity with relational models of data. For someone with a background in AI or database management, this approach has the advantage of relating in a natural way to the familiar framework of AI and databases. Furthermore, it clarifies some of the controversial issues in the Dempster-Shafer theory and points to ways in which it can be extended and made useful in AI-oriented applications.[1]

The Basic Idea

The basic idea underlying the approach in question is that in the context of relational databases the Dempster-Shafer theory can be viewed as an instance of inference from second-order relations, that is, relations in which the entries are first-order relations.[2] To clarify this point, let us first consider a standard example of retrieval from a first-order relation, such as the relation EMPLOYEE1 (or EMP1, for short) that is tabulated in the following:

EMP1	Name	Age
	1	23
	2	28
	3	21
	4	27
	5	30

As a point of departure, consider a simple example of a range query: What fraction of employees are between 20 and 25 years old, inclusively? In other words,

Research sponsored by the NASA Grant NCC-2-275, NESC Contract NOOO39-84-C-0243, and NSF Grant IST-8420416

[1]The approach described in this article is derived from the application of the concepts of possibility and certainty (or necessity) to information granularity and the Dempster-Shafer model of uncertainty (Zadeh, 1979a, 1981) Extensive treatments of the concepts of possibility and necessity and their application to retrieval from incomplete databases can be found in recent papers by Dubois and Prade (1982, 1984)

[2]In the terminology of relational databases, a first-order relation is a relation which is in first normal form, that is, a relation whose elements are atomic rather than set-valued.

what fraction of employees satisfy the condition Age(i) ε Q, i = 1,...,5, where Q is the query set Q = [20,25]. Counting those i's which satisfy the condition, the answer is 2/5.

Next, let us assume that the age of i is not known with certainty. For example, the age of 1 might be known to be in the interval [22,26]. In this case, the EMP1 relation becomes a second-order relation, for example:

EMP2	Name	Age
	1	[22,26]
	2	[20,22]
	3	[30,35]
	4	[20,22]
	5	[28,30]

Thus, in the case of 1, for example, the interval-valued attribute [22, 26] means that the age of 1 is known to be an element of the set $\{22, 23, 24, 25, 26\}$. In effect, this set is the set of possible values of the variable Age(1) or, equivalently, the *possibility distribution* of Age(1). Viewed in this perspective, the data entries in the column labeled Age are the possibility distributions of the values of Age. Similarly, the query set Q can also be regarded as a possibility distribution. In this sense, the information resident in the database and the queries about it can be described as *granular* (Zadeh, 1979a, 1981), with the data and the queries playing the roles of *granules.*

When the attribute values are not known with certainty, tests of set membership such as Age(i) ε Q cease to be applicable. In place of such tests then, it is natural to consider the *possibility* of Q given the possibility distribution of Age(i). For example, if Q = [20, 25] and Age(1) ε [22, 26], it is *possible* that Age(1) ε Q; in the case of 3, it is *not possible* that Age(3) ε Q; and in the case of 4, it is *certain (or necessary)* that Age(4) ε Q; more generally:

(a) *Age(i)εQ* is *possible*, if the possibility distribution of Age(i) intersects Q; that is, $D_i \cap Q \neq \Theta$ where D_i denotes the possibility distribution of Age(i) and Θ is the empty set.

(b) Q is *certain (or necessary)* if the possibility distribution of Age(i) is contained in Q, that is, $D_i \subset Q$.

(c) Q is *not possible* if the possibility distribution of Age(i) does not intersect Q or, equivalently, is contained in the complement of Q. This implies that—as in modal logic—possibility and necessity are related by

necessity of Q = not (possibility of complement of Q).

In the case of EMP2, the application of these tests to each row of the relation yields the following results for $Q = [20, 25]$:

EMP2	Name	Age	Test
	1	[22,26]	poss
	2	[20,22]	cert
	3	[30,35]	¬ poss
	4	[20,22]	cert
	5	[28,30]	¬ poss

(In the Test column, poss, cert, and ¬ poss, are abbreviations for possible, certain, and not possible, respectively.)

We are now in a position to construct a surrogate answer to the original question: What fraction of employees are between 20 and 25 years old, inclusively? Clearly, the answer will have to be in two parts, one relating to the certainty (or necessity) of Q and the other to its possibility; in symbols:

$$Resp(Q) = (N(Q); \Pi(Q)), \qquad (1)$$

where Resp(Q), $N(Q)$, and $\Pi(Q)$ denote, respectively, the response to Q, the certainty (or necessity) of Q, and the possibility of Q. For the example under consideration, counting the test results in EMP2 leads to the response:

$$\text{Resp}[20, 25] = (N([20, 25]) = 2/5; \Pi([20, 25]) = 3/5),$$

with the understanding that *cert* counts also as *poss* because certainty implies possibility. Basically, a two-part response of this form, that is, *certainly α and possibly β*, where α and β are absolute or relative counts of objects with a specified property, is characteristic of responses based on incomplete information; for example, *certainly 10% and possibly 30%* in response to: How many households in Palo Alto own a VCR?

The first constituent in Resp(Q) is what is referred to as the measure of *belief* in the Dempster-Shafer theory, and the second constituent is the measure of *plausibility.* Seen in this perspective then, the measures of belief and plausibility in the Dempster-Shafer theory are, respectively, the certainty (or necessity) and possibility of the query set Q in the context of retrieval from a second-order relation in which the data entries are possibility distributions.

There are two important observations that can be made at this point. First, assume that EMP is a relation in which the values of Age are singletons chosen from the possibility distributions in EMP2. For such a relation, the response to Q would be a number, say, alpha. Then, it is evident that the values of $N(Q)$ and $\Pi(Q)$ obtained for Q (that is, 2/5 and 3/5) are the *lower* and *upper bounds,* respectively, on the values of alpha. This explains why in the Dempster-Shafer theory the measures of belief and plausibility are interpreted, respectively, as the *lower* and *upper probabilities* of Q.

Second, because the values of $N(Q)$ and $\Pi(Q)$ represent the result of averaging of test results in EMP2, what matters is the *distribution* of test results and not their association with particular employees. Viewing this distribution as a summary of EMP2, this implies that $N(Q)$ and $\Pi(Q)$ are computable from a summary of EMP2 which specifies the fraction of employees whose ages fall in each of the interval-valued entries in the Age column.

More specifically, assume that in a general setting EMP2 has n rows, with the entry in row $i, i = 1, \ldots, n$, under Age being D_i. Furthermore, assume that the D_i are comprised of k distinct sets $A_1, \ldots A_K$ so that (a) each D is one of the $A_s, s = 1, \ldots, k$. For example, in the case of EMP2,

$$\begin{array}{lllllll} n & = & 5, & & k & = & 4 \\ D_1 & = & [22,26] & & A_1 & = & [22,26] \\ D_2 & = & [20,22] & & A_2 & = & [20,22] \\ D_3 & = & [30,35] & & A_3 & = & [30,35] \\ D_4 & = & [20,22] & & A_4 & = & [28,30] \\ D_5 & = & [28,30] & & & & \end{array}$$

Viewing EMP2 as a *parent relation*, its summary can be expressed as a *granular distribution*, Δ, of the form

$$\Delta = \{(A_1, p_1), (A_2, p_2), \ldots, (A_k, p_k)\},$$

in which $p_s, s = 1, \ldots k$,[3]is the fraction of D's that are A_s Thus, in the case of EMP2, we have

$$\begin{aligned} \Delta = \{&([22,26], 1/5), \\ &([20,22], 2/5), \\ &([30,35], 1/5), \\ &([28,30], 1/5)\}. \end{aligned} \tag{2}$$

As is true of any summary, a granular distribution can have a multiplicity of parents, because Δ is invariant under permutations of the values of Name. At a later point, we see that this observation has an important bearing on the so-called Dempster-Shafer rule of combination of evidence.

In summary, given a query set Q, the response to Q has two components, $N(Q)$ and $\Pi(Q)$. In terms of the granular distribution $\Delta, N(Q)$ and $\Pi(Q)$ can be expressed as

$$N(Q) = \sum_s p_s \text{ such that } (A_s \subset Q, \text{ s} = 1, \ldots, \text{k})$$
$$\Pi(Q) = \sum_s p_s \text{ such that } (A_s \cap Q \neq \Theta, \text{ s} = 1, \ldots, \text{k.}) \tag{3}$$

[3]The relative counts $p_1, \ldots p_k$ are referred to as the *basic probability numbers* in the Dempster-Shafer theory

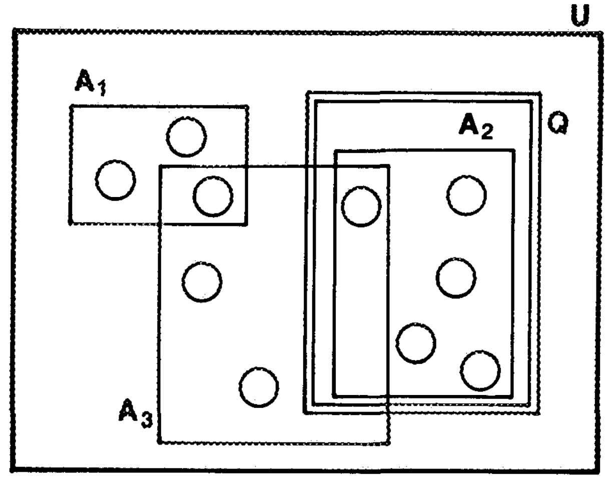

Figure 1

These expressions for the necessity and possibility of Q are identical with the expressions for belief and plausibility in the Dempster-Shafer theory.

The Ball-Box Analogy

The relational model of the Dempster-Shafer theory has a simple interpretation in terms of what might be called the *ball-box analogy*.

Specifically, assume that, as shown in Figure 1, we have n unmarked steel balls which are distributed among k boxes $A_1, \ldots, A_k$, with p_i representing the fraction of balls put in A_i. The boxes are placed in a box U and are allowed to overlap. The position of each ball within the box in which it is placed is unspecified. In this model, the granular distribution Δ describes the distribution of the balls among the boxes. (Note that the number of balls put in A_i, is unrelated to that put in A_j. Thus, if $A_i \subset A_j$, the number of balls put in A_i can be larger than the number of balls put in A_j. It is important to differentiate between the number of balls put in A_i and the number of balls in A_i. The need for differentiation arises because the A_i might overlap, and the boundary of each box is penetrable, except that a ball put in A_i is constrained to stay in A_i.)

Now, given a region Q in U, we can ask the question: How many balls are in Q? To simplify visualization, we assume that, as in Figure 1, the boxes as well as Q are rectangular.

Because the information regarding the position of each ball is incomplete, the answer to the question will, in general, be interval-valued. The upper bound can readily be found by visualizing Q as an attractor, for example, a magnet. Under this assumption, it is evident that the proportion of balls drawn into Q is given by

$$\Pi(Q) = \textstyle\sum_s p_s, \qquad A_s \cap Q \neq \Theta, \qquad s = l, \ldots, k,$$

which is the expression for plausibility in the Dempster-Shafer theory. Similarly, the lower bound results from visualizing Q as a repeller. In this case, the lower bound is given by

$$N(Q) = \textstyle\sum_s p_s, \qquad A_s \subset Q, \qquad s = 1, \ldots, k,$$

which coincides with the expression for belief in the Dempster-Shafer theory. Note that making Q an attractor is equivalent to making Q' (the complement of Q) a repeller. From this it follows at once that

$$\Pi(Q) = 1 - N(Q'),$$

which has already been cited as one of the basic identities in the Dempster-Shafer theory.

The ball-box analogy has the advantage of providing a pictorial—and, thus, easy to grasp–interpretation of the Dempster-Shafer model. As a simple illustration of its use, consider the following problem. There are 20 employees in a department. Five are known to be under 20, 3 are known to be over 40 and the rest are known to be between 25 and 45. How many are over 30? The answer that is yielded at once by the analogy is between 3 and 15.

The Issue of Normalization

A controversial issue in the Dempster-Shafer theory relates to the normalization of upper and lower probabilities and its role in the Dempster-Shafer rule of combination of evidence.

To view this issue in the context of relational databases, assume that the attribute tabulated in the EMP2 relation is not the employee's age but the age of the employee's car, Age(Car(i)), i = 1,..., 5, with the understanding that Age(Car(i)) = 0 means the car is brand new and that Age(Car(i)) = Θ, where Θ is the empty set (or, equivalently, a null value) means i does not have a car. For convenience in reference, an attribute is said to be *definite* if it cannot take a null value and *indefinite* if it can. In these examples, Age is definite, whereas Age (Car) is not.

The question that arises is: How should the null values be counted? Questions of this type arise, generally, when the referent in a proposition does not exist. In the theory of presuppositions, for example, a case in point is the proposition "The King of France is bald," with the question being: What is the truth-value of this proposition if the King of France does not exist? Closer to AI, similar issues arise in the literature on cooperative responses to database queries (Joshi, 1982; Joshi & Webber, 1982; Kaplan, 1982) and the treatment of null values in relational models of data (Biskup, 1980).

In the Dempster-Shafer theory, the null values are not counted, giving rise to what is referred to as *normalization.*[4] However, it is easy to see that normalization can lead to a misleading response to a query. Consider, for example, the relation EMP3 shown in the following:

EMP3	Name	Age(Car)
	1	[3,4]
	2	Θ
	3	[2,3]
	4	Θ
	5	Θ

For the query set $Q \triangleq [2,4]$, normalization would lead to the unqualified conclusion that *all* employees have a car that is two to four years old.

Such misleading responses can be avoided, of course, by not allowing normalization or, better, by providing a relative count of all the null values. As an illustration, in the example under consideration, avoiding normalization would lead to the response $N(Q) = \Pi(Q) = 2/5$. Adding the information about the null values would result in a response with three components: $N(Q) = 2/5; \Pi(Q) = 2/5; RC\Theta = 3/5$, where RC denotes the relative count of the null values.

As pointed out in Zadeh (1979b), normalization can lead to serious problems in the case of what has come to be known as the *Dempster-Shafer rule of combination.* As is seen in the following section, this rule has a simple interpretation in the context of retrieval from relational databases—an interpretation that serves to clarify the implications of normalization and points to ways in which the rule can be made useful.

The Dempster-Shafer Rule

In the examples considered so far, we have assumed that there is just one source of information concerning the attribute Age. What happens when there are two or more sources, as in the relation EMP4 tabulated below?

[4] In Shafer's theory (Shafer, 1976), null values are not allowed in the definition of belief functions but enter the picture in the rule of combination of evidence

EMP4	Name	Age 1	Age 2
	1	[22,23]	[22,24]
	2	[19,21]	[20,21]
	3	[20,21]	[19,20]
	4	[21,22]	[19,20]
	5	[22,23]	[19,21]

Because the entries in Age 1 and Age 2 are possibility distributions, it is natural to combine the sources of information by forming the intersection (or, equivalently, the conjunction) of the respective possibility distributions for each i, resulting in the relation EMP5

EMP5	Name	Age 1 * Age 2
	1	[22,23]
	2	[20,21]
	3	20
	4	Θ
	5	Θ

in which the aggregation operator * has the meaning of intersection.

Using the combined relation to compute the nonnormalized response to the query set $Q = [20, 25]$, leads to

$$\text{Resp}(Q) = (N(Q) = 3/5; \Pi(Q) = 3/5; \text{RC}\Theta = 2/5).$$

With normalization, the response is given by

$$\text{Resp}(Q) = (N(Q) = 1; \Pi(Q) = 1). \tag{4}$$

Note the normalized response suppresses the fact that in the case of 4 and 5 the two sources are flatly contradictory.

Next, consider the case where we know the distribution of the possibility distributions associated with the two sources but not their association with particular employees. Thus, in the case of Age 1, the information conveyed by source 1 is that the possibility distributions of the Age variable and their relative counts in Age1 are given by the granular distribution

$$\Delta_1 = \{(A_1^1, p_1), \ldots, (A_k^1, p_k)\}$$

and in the case of Age 2, the corresponding granular distribution is

$$\Delta_2 = \{(A_1^2, q_1), \ldots, (A_m^2, q_m)\}.$$

Because we do not know the association of A's with particular employees, to combine the two sources we have to form all possible intersections of A^1's and A^2's. As a result, in the combined column Age 1 * Age 2, the data entries will be of the form

$$A_s^1 \cap A_t^2, s = 1, \ldots, k, t = 1, \ldots, m,$$

and the relative count of $A_s^1 \cap A_t^2$'s will be $p_s q_t$.

The result of the combination then is the following granular distribution:

$$\Delta_{1,2} = \{(A_s^1 \cap A_t^2, p_s q_t);\ \text{s} = 1, \ldots, k,\ \text{t}=1,\ \ldots,\ m\} \tag{5}$$

Knowing $\Delta_{1,2}$, we can compute the responses to Q using (3) and (4) with or without normalization. It is the first choice that leads to the Dempster-Shafer rule.

As a simple illustration, assume that we wish to combine the following granular distributions:

$$\Delta_1 = \{([20,21], 0.8), ([22,24], 0.2)\}$$

$$\Delta_2 = \{([19,20], 0.6), ([20,23], 0.4)\}.$$

In this case, (5) becomes

$$\Delta_{1,2} = \{(20, 0.48), ([20,21], 0.32), ([22,23], 0.08), (\Theta, 0.12)\},$$

and if Q is assumed to be given by $Q = [20,22]$, the nonnormalized and normalized responses can be expressed as

$$\text{Resp}(Q) = (N(Q) = 0.8; \Pi(Q) = 0.88; RC\Theta = 0.12)$$

$$\text{Norm. Resp}(Q) = (N(Q) = 0.8/0.88; \Pi(Q) = 1).$$

If we are dealing with a definite attribute, that is, an attribute which is not allowed to take null values, then it is reasonable to reject the null values in the combined distribution. However, if the attribute is indefinite, such rejection can lead to counterintuitive results.

The relational point of view leads to an important conclusion regarding the validity of the Dempster-Shafer rule. Specifically, if we assume that the attribute is definite, then the intersection of the attributes associated with any entry cannot be empty, that is, the relation must be *conflict-free.* Now, if we are given two granular distributions Δ_1 and Δ_2, then there must be at least one parent relation for Δ_1 and Δ_2 that is conflict-free. In this case, we say that Δ_1 and Δ_2 are *combinable.*

What this implies is that in the case of a definite attribute one cannot, in general, combine two arbitrarily specified granular distributions. In more specific terms, this conclusion can be stated as the following conjecture:

In the case of definite attributes, the Dempster-Shafer rule of combination of evidence is not applicable unless the underlying granular distributions are combinable, that is, have at least one parent relation which is conflict-free.

An obvious corollary of this conjecture is the following:

If there exists a granule A_s in Δ_1 that is disjoint from all granules A_t in Δ_2, or vice-versa, then Δ_1 and Δ_2 are not combinable.

An immediate consequence of this corollary is that distinct probability distributions are not combinable and, hence, that the Dempster-Shafer rule is not applicable to such distributions. This explains why the example given in Zadeh (1979b, 1984) leads to counterintuitive results.

Concluding Remarks

The relational view of the Dempster-Shafer theory that is outlined here exposes the basic ideas and assumptions underlying the theory and makes it much easier to understand. Furthermore, it points to extensions of the theory for use in various AI-oriented applications and, especially, in expert systems. Among such extensions, which are discussed in Zadeh (1979a), is the extension to second-order relations in which (1) the data entries are not restricted to crisp sets and (2) the distributions of data entries are specified imprecisely. This extension provides a three-way link between the Dempster-Shafer theory, the theory of information granularity (Zadeh, 1979a, 1981) and the theory of fuzzy relational databases (Zemankova-Leech and Kandel, 1984) Another important extension relates to the combination of sources of information with unequal credibility indexes. Extension to such sources necessitates the use of graded possibility distributions in which possibility, like probability, is a matter of degree rather than a binary choice between perfect possibility and complete impossibility.

As far as the validity of the Dempster-Shafer rule is concerned, the relational point of view leads to the conjecture that it cannot be applied until it is ascertained that the bodies of evidence are not in conflict; that is, there exists at least one parent relation which is conflict-free. In particular, under this criterion, it is not permissible to combine distinct probability distributions—which is allowed in the current versions of the Dempster-Shafer theory.

References

Barnett, J A (1981) *Computational methods for a mathematical theory of evidence* IJCAI 7: 868-875

Biskup, J (1980) A formal approach to null values in database relations In H Gallaire & J M Nicolas (Eds) *Formal bases for data bases* New York: Plenum Press

Dempster, A P (1967) Upper and lower probabilities induced by a multivalued mapping *Annals Mathematics Statistics* (38)325-339

Dillard, R , (1983) Computing confidences in tactical rule-based systems by using Dempster-Shafer theory Tech Doc 649, Naval Ocean Systems Center.

Dubois, D , & Prade, H (1982) On several representations of an uncertain body of evidence In M M Grupta & E. Sanehez, (Eds.) *Fuzzy information and processes* Amsterdam: North Holland, 167-181

Garvey, T , & Lowrance, J (1983) Evidential reasoning: implementation for multisensor integration Tech Rep 305, SRI International Artificial Intelligence Center

Gordon, J & Shortliffe, E (1983) The Dempster-Shafer theory of evidence In B Buchanan, B & E. Shortliffe, (Eds) *Rule based systems.* Menlo Park, Calif : Addison-Wesley

Joshi, A K , & Webber, B L , (1982) *Taking the initiative in natural language database interactions* European Artificial Intelligence Conference, Paris

Joshi, A. K (1982) Varieties of cooperative responses in a question-answer system In F Kiefer, (Ed.) *Questions & answers* Dordrecht, Netherlands: Reidel, 229-240

Kaplan, S J (1982) Cooperative responses for a portable natural language query system *Artificial Intelligence* 19: 165-187

Kempson, R M. (1975) *Presupposition and the delimitation of semantics.* Cambridge: Cambridge University Press

Nguyen, H T (1978) On random sets and belief functions *Journal of Mathematical Analysis and Applications* (65): 531-542

Prade, H , (1984) Lipski's approach to incomplete information data bases restated and generalized in the setting of Zadeh's possibility theory *Information Systems* 9(1): 27-42

Shafer, G (1976) *A mathematical theory of evidence.* Princeton, N J : Princeton University Press

Wesley, L., Lowrance, J , & Garvey T (1984) Reasoning about control: an evidential approach Tech Rep. 324, SRI International Artificial Intelligence Center

Zadeh, L A (1979a) Fuzzy sets and information granularity. In M. M Gupta, R Ragade, and R Yager, (Eds) *Advances in fuzzy set theory and applications* Amsterdam: North Holland, 3-18

Zadeh, L A (1979b) On the validity of Dempster's rule of combination of evidence ERL Mem M79/24, Department of EECS, University of California at Berkeley

Zadeh, L A (1981) Possibility theory and soft data analysis In L Cobb & R M Thrall, (Eds) *Mathematical frontiers of the social and policy sciences* Boulder, Colo: Westview Press, 69-129

Zadeh, L A. (1984) Review of Shafer's a mathematical theory of evidence *AI Magazine* 5(3): 81-83

Zemankova-Leech, M & Kandel, A (1984) *Fuzzy relational databases—a key to expert systems* Cologne, West Germany: Verlag TUV, Interdisciplinary Systems Research