Dempster Shafer Theory

Vishal yadav, Atul Kaushik, Rajat Verma

Dronacharya College of Engineering, India

Abstract: **The theory of belief functions, also referred to as evidence theory or Dempster–Shafer theory (DST), is a general framework for reasoning with uncertainty, with understood connections to other frameworks such as probability, possibility and imprecise probability theories. In this paper we give a detailed view of the theory.**
Keyword: **Dempster shafer theory ,belief , plausibility , etc.**

I. INTRODUCTION

The Dempster-Shafer theory or the theory of belief functions is a mathematical theory of evidence which can be interpreted as a generalization of probability theory in which the elements of the sample space to which nonzero probability mass is attributed are not single points but sets . The sets that get nonzero mass are called focal elements. The sum of these probability masses is one, however, the basic difference between Dempster-Shafer theory and traditional probability theory is that the focal elements of a Dempster-Shafer structure may overlap one another. The Dempster-Shafer theory also provides methods to represent and combine weights of evidence.

Dempster–Shafer theory is a generalization of the Bayesian theory of subjective probability. Belief functions base degrees of belief (or confidence, or trust) for one question on the probabilities for a related question. The degrees of belief itself may or may not have the mathematical properties of probabilities; how much they differ depends on how closely the two questions are related. Put another way, it is a way of representing epistemic plausibilities but it can yield answers that contradict those arrived at using probability theory. Often used as a method of sensor fusion, Dempster–Shafer theory is based on two ideas: obtaining degrees of belief for one question from subjective probabilities for a related question, and Dempster's rule for combining such degrees of belief when they are based on independent items of evidence. In essence, the degree of belief in a proposition depends primarily upon the number of answers (to the related questions) containing the proposition, and the subjective probability of each answer. Also contributing are the rules of combination that reflect general assumptions about the data.

In this formalism a degree of belief (also referred to as a mass) is represented as a belief function rather than a Bayesian probability distribution. Probability values are assigned to sets of possibilities rather than single events: their appeal rests on the fact they naturally encode evidence in favor of propositions. Dempster–Shafer theory assigns its masses to all of the non-empty subsets of the propositions that compose a system. (In set-theoretic terms, the Power set of the propositions.) For instance, assume a situation where there are two related questions, or propositions, in a system. In this system, any belief function assigns mass to the first proposition, the second, both or neither.

II. BELIEF AND PLAUSIBILITY

Shafer's formalism starts from a set of possibilities under consideration, for instance numerical values of a variable, or pairs of linguistic variables like "date and place of origin of a relic" (asking whether it is antique or a recent fake). Shafer's framework allows for belief about such propositions to be represented as intervals, bounded by two values, belief (or support) and plausibility:

$$\text{belief} \leq \text{plausibility}$$

In a first step, subjective probabilities (masses) are assigned to all subsets of the frame; usually, only a restricted number of sets will have non-zero mass (focal elements). Belief in a hypothesis is constituted by the sum of the masses of all sets enclosed by it. It is the amount of belief that directly supports a given hypothesis or a more specific one, forming a lower bound. Belief (usually denoted Bel) measures the strength of the evidence in favor of a proposition p. It ranges from 0 (indicating no evidence) to 1 (denoting certainty). Plausibility is 1 minus the sum of the masses of all sets whose intersection with the hypothesis is empty. Or, it can be obtained as the sum of the masses of all sets whose intersection with the hypothesis is not empty. It is an upper bound on the possibility that the hypothesis could be true, i.e. it "could possibly be the true state of the system" up to that value, because there is only so much evidence that contradicts that hypothesis. Plausibility (denoted by Pl) is defined to be Pl(p)=1-Bel(~p). It also ranges from 0 to 1 and measures the extent to which evidence in favor of ~p leaves room for belief in p.

For example, suppose we have a belief of 0.5 and a plausibility of 0.8 for a proposition, say "the cat in the box is dead." This means that we have evidence that allows us to state strongly that the proposition is true with a confidence of 0.5. However, the evidence contrary to that hypothesis (i.e. "the cat is alive") only has a confidence of 0.2. The remaining mass of 0.3 (the gap between the 0.5 supporting evidence on the one hand, and the 0.2 contrary evidence on the other) is "indeterminate," meaning that the cat could either be dead or alive. This interval represents the level of uncertainty based on the evidence in your system.

Hypothesis	Mass	Belief	Plausibility
Null (neither alive nor dead)	0	0	0
Alive	0.2	0.2	0.5
Dead	0.5	0.5	0.8
Either (alive or dead)	0.3	1.0	1.0

The null hypothesis is set to zero by definition (it corresponds to "no solution"). The orthogonal hypotheses "Alive" and "Dead" have probabilities of 0.2 and 0.5, respectively. This could correspond to "Live/Dead Cat Detector" signals, which have respective reliabilities of 0.2 and 0.5. Finally, the all-encompassing "Either" hypothesis (which simply acknowledges there is a cat in the box) picks up the slack so that the sum of the masses is 1. The belief for the "Alive" and "Dead" hypotheses matches their corresponding masses because they have no subsets; belief for "Either" consists of the sum of all three masses (Either, Alive, and Dead) because "Alive" and "Dead" are each subsets of "Either". The "Alive" plausibility is 1 − m (Dead) and the "Dead" plausibility is 1 − m (Alive). In other way, the "Alive" plausibility is m(Alive) + m (Either) and the "Dead" plausibility is m(Dead) + m(Either). Finally, the "Either" plausibility sums m(Alive) + m(Dead) + m(Either). The universal hypothesis ("Either") will always have 100% belief and plausibility —it acts as a checksum of sorts.

Here is a somewhat more elaborate example where the behavior of belief and plausibility begins to emerge. We're looking through a variety of detector systems at a single faraway signal light, which can only be coloured in one of three colours (red, yellow, or green):

Hypothesis	Mass	Belief	Plausibility
Null	0	0	0
Red	0.35	0.35	0.56
Yellow	0.25	0.25	0.45
Green	0.15	0.15	0.34
Red or Yellow	0.06	0.66	0.85
Red or Green	0.05	0.55	0.75
Yellow or Green	0.04	0.44	0.65
Any	0.1	1.0	1.0

III. COMBINING BELIEFS

Beliefs from different sources can be combined with various fusion operators to model specific situations of belief fusion, e.g. with Dempster's rule of combination, which combines belief constraints that are dictated by independent belief sources, such as in the case of combining hints or combining preferences. Note that the probability masses from propositions that contradict each other can be used to obtain a measure of conflict between the independent belief sources. Other situations can be modeled with different fusion operators, such as cumulative fusion of beliefs from independent sources which can be modeled with the cumulative fusion operator.

Dempster's rule of combination is sometimes interpreted as an approximate generalisation of Bayes' rule. In this interpretation the priors and conditionals need not be specified, unlike traditional Bayesian methods, which often use a symmetry (minimax error) argument to assign prior probabilities to random variables (e.g. assigning 0.5 to binary values for which no information is available about which is more likely). However, any information contained in the missing priors and conditionals is not used in Dempster's rule of combination unless it can be obtained indirectly—and arguably is then available for calculation using Bayes equations. Dempster–Shafer theory allows one to specify a degree of ignorance in this situation instead of being forced to supply prior probabilities that add to unity. This sort of situation, and whether there is a real distinction between risk and ignorance, has been extensively discussed by statisticians and economists.

IV. FORMAL DEFINITION

Let X be the universe: the set representing all possible states of a system under consideration. The power set

$$2^X$$

is the set of all subsets of X, including the empty set $\emptyset$. For example, if:

$$X = \{a, b\}$$

then

$$2^X = \{\emptyset, \{a\}, \{b\}, X\}.$$

The elements of the power set can be taken to represent propositions concerning the actual state of the system, by containing all and only the states in which the proposition is true.

The theory of evidence assigns a belief mass to each element of the power set. Formally, a function

$$m : 2^X \to [0, 1]$$

is called a basic belief assignment (BBA), when it has two properties. First, the mass of the empty set is zero:

$$m(\emptyset) = 0.$$

Second, the masses of the remaining members of the power set add up to a total of 1:

$$\sum_{A \in 2^X} m(A) = 1$$

The mass m(A) of A, a given member of the power set, expresses the proportion of all relevant and available evidence that supports the claim that the actual state belongs to Abut to no particular subset of A. The value of m(A) pertains only to the set A and makes no additional claims about any subsets of A, each of which have, by definition, their own mass.

From the mass assignments, the upper and lower bounds of a probability interval can be defined. This interval contains the precise probability of a set of interest (in the classical sense), and is bounded by two non-additive continuous measures called belief (or support) and plausibility:

$$\operatorname{bel}(A) \le P(A) \le \operatorname{pl}(A).$$

The belief bel(A) for a set A is defined as the sum of all the masses of subsets of the set of interest:

$$\operatorname{bel}(A) = \sum_{B \mid B \subseteq A} m(B).$$

The plausibility pl(A) is the sum of all the masses of the sets B that intersect the set of interest A:

$$\operatorname{pl}(A) = \sum_{B \mid B \cap A \ne \emptyset} m(B).$$

The two measures are related to each other as follows:

$$\operatorname{pl}(A) = 1 - \operatorname{bel}(\overline{A}).$$

And conversely, for finite A, given the belief measure bel(B) for all subsets B of A, we can find the masses m(A) with the following inverse function:

$$m(A) = \sum_{B \mid B \subseteq A} (-1)^{|A-B|} \operatorname{bel}(B)$$

where |A − B| is the difference of the cardinalities of the two sets.

It follows from the last two equations that, for a finite set X, you need know only one of the three (mass, belief, or plausibility) to deduce the other two; though you may need to know the values for many sets in order to calculate one of the other values for a particular set. In the case of an infinite X, there can be well-defined belief and plausibility functions but no well-defined mass function.

V. DEMPSTER'S RULE OF COMBINATION

The problem we now face is how to combine two independent sets of probability mass assignments in specific situations. In case different sources express their beliefs over the frame in terms of belief constraints such as in case of giving hints or in case of expressing preferences, then Dempster's rule of combination is the appropriate fusion operator. This rule derives common shared belief between multiple sources and ignores all the conflicting (non-shared) belief through a normalization factor. Use of that rule in other situations than that of combining belief constraints has come under serious criticism, such as in case of fusing separate beliefs estimates from multiple sources that are to be integrated in a cumulative manner, and not as constraints. Cumulative fusion means that all probability masses from the different sources are reflected in the derived belief, so no probability mass is ignored.

Specifically, the combination (called the joint mass) is calculated from the two sets of masses m_1 and m_2 in the following manner:

$$m_{1,2}(\emptyset) = 0$$
$$m_{1,2}(A) = (m_1 \oplus m_2)(A) = \frac{1}{1-K} \sum_{B \cap C = A \neq \emptyset} m_1(B) m_2(C)$$

where

$$K = \sum_{B \cap C = \emptyset} m_1(B) m_2(C).$$

K is a measure of the amount of conflict between the two mass sets.

VI. CONCLUSION

In a narrow sense, the term Dempster–Shafer theory refers to the original conception of the theory by Dempster and Shafer. However, it is more common to use the term in the wider sense of the same general approach, as adapted to specific kinds of situations. In particular, many authors have proposed different rules for combining evidence, often with a view to handling conflicts in evidence better. The early contributions have also been the starting points of many important developments, including the Transferable Belief Model and the Theory of Hints

VII. REFERENCES

1. https://en.wikipedia.org/wiki/Dempster%E2%80%93Shafer_theory
2. http://en.wikibooks.org/wiki/Expert_Systems/Dempster-Shafer_Theory
3. http://projecteuclid.org/euclid.ss/1280841734
4. http://www.sciencedirect.com/science/article/pii/S030504839900033X
5. http://what-when-how.com/artificial-intelligence/the-dempster-shafer-theory-artificial-intelligence/