Conditional Belief Functions

Liping Liu, *Susquehanna University, Selinsgrove, PA 17870*
Prakash P. Shenoy, *University of Kansas, Lawrence, KS 66045*

ABSTRACT

We define the notion of conditional belief functions using the semantics of evidential support. We use three properties to demonstrate that the notion of conditional belief functions generalizes both Dempster-Shafer theory of belief functions and Bayesian theory of probability functions.

INTRODUCTION

A belief function was thought to be the generalization of both logical and probabilistic models. It corresponds to a single logical statement if it has only one focal proposition and reduces to a probability distribution if all the foci are singletons. The theory of belief functions represents logical knowledge using joint rather conditional belief functions. For example, let H and T be two binary zero-one variables. The logical statement "if T = 1, then H = 1" is represented by the belief function with m({(0,0), (0,1), (1,1)}) = 1. In general, the logical statement "if A, then B" is expressed by $m(A^c \cup B) = 1$. However, such a joint representation is difficult or at least cumbersome to represent inexact if-then rules. The difficulty is due to the lack of a mechanism equivalent to conditioning with which context dependencies can be encoded. As Lewis [2] has shown, the information conveyed in a conditional probability statement P(A | B) = p can not be represented by assigning probabilities to some Boolean function of A and B or to any set of Boolean formulas.

In this paper we generalize the concept of belief functions by defining conditional belief functions. We treat conditionals as a primitive concept instead of a derivative from a conditioning rule [8]. We directly assign basic probabilities to logical statements by following the semantics of evidential support. We show that the notion of conditional belief functions generalizes both Dempster-Shafer theory of belief functions and Bayesian theory of probabilities.

Recently there has been extensive research on asymmetric decision problems using influence diagrams and valuation networks [3,4,6,7]. Asymmetry can be viewed as an incomplete numerical specification of probabilities and utilities. From this perspective, an asymmetric problem can be perfectly represented by a set of conditional belief functions. As Liu and Shenoy [4] show, a valuation network using conditional belief functions does not include any artificial acts and states, degenerate probabilities and utilities. It exploits the structural asymmetry of problems as well as the coalescence of numerical specifications. It can save up to 80% of input data and is a most natural and compact way of representing asymmetric Bayesian decision problems.

An outline of the paper is as follows. To motivate the reader, Section 2 provides a brief review of Dempster-Shafer theory of belief functions. Section 3 defines the concept of conditional belief functions and explores its basic properties.

DEMPSTER-SHAFER THEORY

The notion of belief functions can be traced to the work of Jakob Bernoulli on pooling pure evidence. In modern language, an item of pure evidence proves a claim with a certain probability but has no opinions on its negation. Probabilities in accordance with pure evidence are not additive. For example, suppose I find a scrap of newspaper predicting a blizzard tomorrow, which I regard as infallible. Also, suppose I am 75% certain that the newspaper is today's. Then, I am 75% sure of a blizzard tomorrow. However, if the newspaper is not today's, both blizzard and no-blizzard could happen since the newspaper carries no information on tomorrow's weather. The degree of support for {no blizzard} is 0 and for {blizzard, no-blizzard} is 25%.

Bernoulli's idea of non-additive probabilities has now been well developed by Dempster [1], Shafer [5], and many others, under the name of the Dempster-Shafer theory of belief functions. In that theory, a piece of evidence is encoded as a probability measure. The degree of belief for a claim is interpreted as a degree of the evidential support. Degrees of belief from independent items of evidence are combined by Dempster's rule of combination. Let X be a set of discrete variables and ω_X its finite sample space. Let x denote a possible value (or configuration) of X in ω_X, which represents the assertion that the true value of X is x. Let $\boldsymbol{x}$ denote a subset of ω_X, which is interpreted as the assertion that the true value of X is in $\boldsymbol{x}$. Then, the degree of evidential support for $\boldsymbol{x}$ is represented by $m(\boldsymbol{x})$. The assignment of $m(\boldsymbol{x})$ is in accordance with a certain item of evidence and satisfies the following axioms:

$$0 \le m(\boldsymbol{x}) \le 1, \ m(\phi) = 0, \ \Sigma\{m(\boldsymbol{x}) \mid \boldsymbol{x} \subset \omega_X\} = 1. \quad (1)$$

A subset $\boldsymbol{x}$ is called a focal element iff $m(\boldsymbol{x}) > 0$. Due to lack of evidence justifying a more specific allocation, a portion of our total belief allocated to a focal element $\boldsymbol{x}$ does not necessitate the allocation of any partial belief to its subset. For the above newspaper example, we can encode the evidence by a probability measure with m({blizzard}) = 0.75 and m({blizzard, no-blizzard}) = 0.25. Thus, {blizzard} and {blizzard, no-blizzard} are the two focal elements. The 25% of belief for {blizzard, no-blizzard} does not imply any reallocation of the belief to its subsets {blizzard} and {no-blizzard}.

If all the focal elements are singletons, we call the belief function *Bayesian*. On the other hand, if the sample space is the only focal element, we call the belief function *vacuous*. One advantage of the belief function modeling is its ability to represent ignorance and partial ignorance. In Bayesian inference, complete ignorance is often represented by a uniform prior or a prior with large-scale parameters such as a Gaussian distribution with large variance. Such priors often lack theoretical or empirical bases and sometimes imply vanishingly small prior probability for regions of practical interest. The belief function formalism represents ignorance by vacuous belief functions. It clearly distinguishes lack of belief from disbelief. For example, a vacuous belief function with m({rainy, not-rainy}) = 1 will be regarded as totally different from the one with m({rainy}) = 1/2 and m({not-rainy}) = 1/2.

Another advantage of the belief function formalism is its ability to pool independent pieces of evidence by Dempster's rule. Each piece of evidence is encoded as a probability measure. The pooling of two independent pieces of evidence can be encoded as the product of two probability measures. From this perspective, Dempster [7] developed a rule for combining belief functions that represent independent pieces of evidence. Suppose there are two belief functions Bel_1 and Bel_2 respectively for sets X and Y. Their basic probability assignments are respectively $m_1(\boldsymbol{x})$ and $m_2(\boldsymbol{y})$. Then, by Dempster's rule, the combined belief function, denoted by $Bel_1 \otimes Bel_1$, is for set $Z = X \cup Y$ and has basic probability assignment:

$$m(\boldsymbol{z}) = \alpha^{-1}\Sigma\{ m_1(\boldsymbol{x})m_2(\boldsymbol{y}) \mid \boldsymbol{z}^{\downarrow X} = \boldsymbol{x} \text{ and } \boldsymbol{z}^{\downarrow Y} = \boldsymbol{y}), \quad (2)$$

where $\boldsymbol{z}^{\downarrow X}$ is the projections of $\boldsymbol{z}$ to X, defined as $\boldsymbol{z}^{\downarrow X} = \{z^{\downarrow X} \mid z \in \boldsymbol{z}\}$, where $z^{\downarrow X}$ is a value in ω_X projected from z by dropping the coordinates of z that are not in X. $\boldsymbol{z}^{\downarrow Y}$ can be interpreted similarly.

The parameter α in (2) is a normalization constant given by

$$\alpha = \Sigma\{ m_1(\boldsymbol{x})m_2(\boldsymbol{y}) \mid \boldsymbol{x}^{\downarrow X\cap Y} \cap \boldsymbol{y}^{\downarrow X\cap Y} \neq \phi\}. \quad (3)$$

Note that $\boldsymbol{x}^{\downarrow X\cap Y} \cap \boldsymbol{y}^{\downarrow X\cap Y} = \phi$ indicates the conflict between the two assertions $\boldsymbol{x}$ and $\boldsymbol{y}$. One of them must be false and a joint assertion is qualitatively impossible. Therefore, α measures the total belief committed to all the joint assertions that are qualitatively possible. If $\alpha = 0$, the two belief functions are *incompatible* because they have no joint assertions qualitatively possible.

Combination corresponds to the integration of knowledge. Sometimes we are interested in drawing partial knowledge from a full body of knowledge. That corresponds to the coarsening of knowledge, obtained by the marginalization of a belief function. Suppose Bel is a belief function for X with basic probability assignment m($\boldsymbol{x}$) and Y is a subset of X. Then we define the marginal of Bel to Y, denoted by $Bel^{\downarrow Y}$, as a belief function for Y with basic probability assignment $m^{\downarrow Y}$ satisfying

$$m^{\downarrow Y}(\boldsymbol{y}) = \Sigma\{m(\boldsymbol{x}) \mid \boldsymbol{x}^{\downarrow Y} = \boldsymbol{y}\} \quad (4)$$

CONDITIONAL BELIEF FUNCTIONS

In this section, we define conditional belief functions by following the above semantics of evidential support. For any two disjoint sets of variables H and T, let $\boldsymbol{h} \mid \boldsymbol{t}$ denote the logical assertion that "the true H is in $\boldsymbol{h}$ if the true T is in $\boldsymbol{t}$." Given two logical assertions, $\boldsymbol{h}^1 \mid \boldsymbol{t}^1$ and $\boldsymbol{h}^2 \mid \boldsymbol{t}^2$, we say $\boldsymbol{h}^1 \mid \boldsymbol{t}^1$ is stronger than $\boldsymbol{h}^2 \mid \boldsymbol{t}^2$ if $\boldsymbol{h}^1 \subset \boldsymbol{h}^2$ and $\boldsymbol{t}^1 \cap \boldsymbol{t}^2 \neq \phi$. Similar to Dempster-Shafer theory of belief functions, we can allocate a non-zero belief m($\boldsymbol{h} \mid \boldsymbol{t}$) to logical assertion $\boldsymbol{h} \mid \boldsymbol{t}$ if there is a piece of evidence that partially supports $\boldsymbol{h} \mid \boldsymbol{t}$ but does not support any stronger assertion than $\boldsymbol{h} \mid \boldsymbol{t}$. We call a logical assertion $\boldsymbol{h} \mid \boldsymbol{t}$ a *focal element* iff m($\boldsymbol{h} \mid \boldsymbol{t}$) > 0. Correspondingly, we call $\boldsymbol{h}$ the focal head and $\boldsymbol{t}$ the focal tail.

The notion of conditional belief functions has three non-trivial special cases. First, if every focal head is ω_H, then it is a vacuous belief function of H given T. Second, if every focal tail is ω_T, then our knowledge about H is irrelevant to that about T. It is a marginal belief function of H. Finally, if every focal head is a singleton, then it is a Bayesian probability function.

Two conditional belief functions of H given T, m_1 and m_2, are called *equivalent* and denoted by $m_1 \sim m_2$, if for every focal element of m_1, $\boldsymbol{h}^1 \mid \boldsymbol{t}^1$, there exists a focal element of m_2, $\boldsymbol{h}^2 \mid \boldsymbol{t}^2$, such that $\boldsymbol{t}^1 \cap \boldsymbol{t}^2 \neq \phi$, $\boldsymbol{h}^1 = \boldsymbol{h}^2$, and $m_1(\boldsymbol{h}^1 \mid \boldsymbol{t}^1) = m_2(\boldsymbol{h}^2 \mid \boldsymbol{t}^2)$. The notion of equivalence addresses the non-uniqueness of encoding the same piece of evidence. Suppose H and T are two binary zero-one variables and $m_1(\{1\} \mid \{0, 1\}) = 0.9$, $m_1(\{0, 1\} \mid \{0\}) = 0.1$, and $m_1(\{0\} \mid \{1\}) = 0.1$. This belief function can be equivalently represented as $m_2(\{1\} \mid \{0\}) = 0.9$, $m_2(\{1\} \mid \{1\}) = 0.9$, $m_2(\{0, 1\} \mid \{0\}) = 0.1$, and $m_2(\{0\} \mid \{1\}) = 0.1$. Note that $m_1(\{1\} \mid \{0, 1\}) = 0.9$ is decomposed into $m_2(\{1\} \mid \{0\}) = 0.9$ and $m_2(\{1\} \mid \{1\}) = 0.9$. They represent exactly the same belief. However, the former takes advantage of numerical coalescence, which is very important in representing an asymmetric Bayesian decision and reasoning problem [4].

It is easy to see that the equivalence relation ~ is reflexive, symmetric, and transitive. Therefore, all equivalent belief functions form a class. When modeling or computing uncertainties using conditional belief functions, we cannot ensure the results are unique. However, we ensure that the different results, if any, are equivalent and belong to the same class.

In every class of conditional belief functions, there is one member function whose focal tails are all singletons. We call such a member *the atomic belief function.* For every belief function with focal elements $\boldsymbol{h} \mid \boldsymbol{t}$ and the mass function m, then its equivalent atomic belief function has focal elements $\boldsymbol{h} \mid t$, where $t \in \boldsymbol{t}$ and the mass function $m_0(\boldsymbol{h} \mid t) = m(\boldsymbol{h} \mid \boldsymbol{t})$ if $t \in \boldsymbol{t}$.

With the concepts of equivalence and atomic belief functions, now we can formally stipulate the axioms of conditional belief functions. A basic probability assignment function m for H given T is a non-negative, real-valued function m on logical assertions $\boldsymbol{h} \mid \boldsymbol{t}$ such that its atomic equivalent m_0 satisfies the following: For any $t \in \omega_T$,

$0 \le m_0(h \mid t) \le 1$, $m_0(\phi \mid t) = 0$,

$\Sigma\{m_0(h \mid t) \mid h \subset \omega_H\} = 1$

Suppose m is a mass function of H given T. We define the corresponding belief and plausibility functions as follows: For any logical assertion $h \mid t$, where t is nonempty,

$$Bel(h \mid t) = \sum\{m(h' \mid t') \mid h' \subseteq h, t' \supseteq t\} \quad (4)$$

$$Pl(h \mid t) = \sum\{m(h' \mid t') \mid h' \cap h \neq \phi, t' \supseteq t\} \quad (5)$$

Property 1. For any $h \subseteq \omega_H$ and nonempty $t \subseteq \omega_T$,

$$0 \le Bel(h \mid t) \le Pl(h \mid t) \le 1 \quad (6)$$

$$Pl(h \mid t) \le 1 - Bel(h^c \mid t) \quad (7)$$

The equality in (7) holds if the belief function is atomic.

Property 2. If a belief function is Bayesian, i.e., all focal heads are singletons. Then for any $t \in \omega_T$ and $h \subseteq \omega_H$,

$$Bel(h \mid t) = Pl(h \mid t).$$

That is, both belief and plausibility functions reduce to a probability function.

Property 3. Suppose all focal tails are ω_T. Then for any $h \subseteq \omega_H$ and $t \subseteq \omega_T$,

$$Bel(h \mid t) = Bel(h \mid \omega_T),\ Pl(h \mid t) = Pl(h \mid \omega_T) \quad (8)$$

$$Pl(h \mid t) = 1 - Bel(h^c \mid t) \quad (9)$$

That is, conditional belief functions reduce to marginal ones if T is irrelevant for H.

The above properties demonstrate that the concept of conditional belief functions is a natural extension of both Bayesian theory of probabilities and Dempster-Shafer theory of belief functions.

Example: A game uses two fair but illegible dice a and b and a biased coin. The die a has face numbers 1, 3, and 6 blurred and illegible. The die b has numbers 2, 3, and 5 blurred and illegible. The coin is biased toward heads and its probability of landing heads is 2/3. The coin is tossed and the gambler must decide whether he wants to see the outcome or not. If he wants the outcome to be disclosed, he will be awarded \$100 if the coin lands a head and otherwise he will be punished by \$10. Before he decides to reveal, he has a choice of buying a hint with \$5. The hint is another gamble as follows. The gambler is told that he will be asked to throw die a if the outcome is a head or die b if it is a tail. However, whether the die is a or b is not identifiable. The only information the gambler can acquire is the outcome of throwing the die.

In this problem, let T denote the outcome of tossing the coin and H the outcome of throwing a die. Then the knowledge about the hint game can be expressed as a conditional belief function with the basic probability assignment as follows:

$m(\{1,3,6\} \mid \{head\}) = 1/2$,

$m(\{2,3,5\} \mid \{tail\}) = 1/2$, $m(\{4\} \mid \omega_T) = 1/6$,

$m(\{5\} \mid \{head\}) = m(\{2\} \mid \{head\}) = 1/6$,

$m(\{1\} \mid \{tail\}) = m(\{6\} \mid \{tail\}) = 1/6$.

Then we can compute the belief and plausibility for all logical assertions. For example, we have

$Bel(\{1,2,4\} \mid \{head\}) = m(\{2\} \mid \{head\}) + m(\{4\} \mid \omega_T) = 1/3$,

$Pl(\{1,2,4\} \mid \{head\}) = m(\{2\} \mid \{head\}) + m(\{4\} \mid \omega_T)$

$+ m(\{1,3,6\} \mid \{head\}) = 5/6$,

$Bel(\{3,5,6\} \mid \{head\}) = m(\{5\} \mid \{head\}) = 1/6$,

$1 - Pl(\{1,2,4\} \mid \{head\}) = 1/6 = Bel(\{3,5,6\} \mid \{head\})$,

$Bel(\{3,5,6\} \mid \omega_T) = 0 < 1 - Pl(\{1,2,4\} \mid \omega_T) = 5/6$.

References

[1] Dempster, A. P. 1967. Upper and lower probabilities induced by a multivalued mapping, *Annals of Mathematical Statistics* **38**, 325–339.

[2] Lewis, D. 1976. Probabilities of conditionals and conditional probabilities. *Phil. Rev.* **85**, 297–315.

[3] Liu, Liping. and P. P. Shenoy 1993. Configuration networks for asymmetric decision problems. Working Paper 254, School of Business, University of Kansas.

[4] Liu, Liping. and P. P. Shenoy 1998. Representing asymmetric decision problems using conditional belief functions. Working Paper 256, School of Business, University of Kansas.

[5] Shafer, G. 1976. *A Mathematical Theory of Evidence*. Princeton: Princeton University Press.

[6] Shenoy, P. P. 1993. Representing and solving asymmetric decision problems using valuation networks. Working Paper, School of Business, University of Kansas, Lawrence, Kansas.

[7] Smith, J. E., S. Holtzman and J. E. Matheson 1993. Structuring conditional relationships in influence diagrams. *Operations Research*. **41**(2), 280–297.

[8] Xu, H. and P. Smets 1994, Reasoning in evidential networks with conditional belief functions, *International Journal of Approximate Reasoning* **11**, 1-31.