

Transferrable Plausibility Model - A Probabilistic Interpretation of Mathematical Theory of Evidence

Mieczysław Alojzy Kłopotek

Institute of Computer Science, Polish Academy of Sciences, Warszawa, Poland,
kłopotek@ipipan.waw.pl

Abstract. This paper suggests a new interpretation of the Dempster-Shafer theory in terms of probabilistic interpretation of plausibility. A new rule of combination of independent evidence is shown and its preservation of interpretation is demonstrated.¹

1 Introduction

Dempster Rule of Independent Evidence Combination has been criticized for its failure to conform to probabilistic interpretation ascribed to belief and plausibility function. Among those verifying DST (Dempster-Shafer-Theory, [2,12]) critically were Kyburg [7], Fagin [3], Halpern [6], Pearl [9], Provan [10], Cano [1], just to mention a few.

As a way out of those difficulties, we proposed in a recent book co-authored by S.T.Wierzchoń [18] three proposals for an empirical model of DST:

- "the marginally correct approximation".
- "the qualitative model"
- "the quantitative model"

The marginally correct approximation assumes that the belief function shall constitute lower bounds for frequencies, though only for the marginals, and not for the joint distribution. Then, the reasoning process is expressed in terms of the so-called Cano et al. conditionals - a special class of conditional belief functions that are positive. This approach implies modification of the reasoning mechanism, because the correctness is maintained only by reasoning forward. Depending on the reasoning direction we need different "Markov trees" for the reasoning engine.

Note that lower/upper bound interpretations have a long tradition for DST [2,7] and have been heavily criticized [6]. The one that we presented in our

¹ This is a preliminary version of the paper:

M.A. Kłopotek: Transferable Plausibility Model - A Probabilistic Interpretation of Mathematical Theory of Evidence O.Hryniewicz, J. Kacprzyk, J.Koronacki, S.Wierzchoń: Issues in Intelligent Systems Paradigms Akademicka Oficyna Wydawnicza EXIT, Warszawa 2005 ISBN 83-87674-90-7, pp.107–118

book differs from the known ones significantly as we insist on different reasoning schemes (hypertrees) depending on which are our target variables, whose values are to be inferred. This assures overcoming of the basic difficulties with lower/upper bound interpretations.

Our qualitative approach is based on the earlier rough set interpretations of DST, but makes a small and still significant distinction. All computations are carried out in a strictly "relational" way, that is, indistinguishable objects in a database are merged (no object identities). The behavior under reasoning fits strictly the DST reasoning model. Factors of well established hypergraph representation (due to Shafer and Shenoy [14]) can be expressed by relational tables. Conditional independence is well defined. However, there is no interpretation for conditional belief functions in this model.

Rough set interpretations [15] were primarily developed for interpreting the belief function in terms of decision tables. However, the Dempster-rule of evidence combination was valid there only for the "extended decision tables", not easily derived from the original ones. In our interpretation, both the original tables and the resultant tables dealt with when simulating Dempster-rule are conventional decision tables and the process of combining of decision tables is a natural one (relational join operator).

Our rough set based interpretation may be directly applied in the domain of multiple decision tables: independence of decision variables or Shenoy's conditional independence in the sense of DST may serve as an indication of possibility of decomposition of the decision table into smaller but equivalent tables.

Furthermore, it may be applied in the area of Cooperative Query Answering [11]. The problem there is that a query posed to a local relational database system may contain an unknown attribute. But, possibly, other co-operating database systems know it and may explain it to the queried system in terms of known attributes, shared by the various systems. The uncertainties studied in the decision tables arise here in a natural way and our interpretation may be used to measure these uncertainties in terms of DST (as a diversity of support). Furthermore, if several co-operating systems respond, then the queried system may calculate the overall uncertainty measure using DST combination of measures of individual responses.

The quantitative model assumed that the objects possess multivalued properties which are then lost in some physical properties and these physical processes are described by DST belief functions (see e.g. [8]).

The quantitative model assumes that during the reasoning process one attaches labels to objects hiding some of their properties. There is a full agreement with the reasoning mechanism of DST. Conditional independence and conditional belief functions are well defined. We have also elaborated processes that can give rise to well-controlled graphoidally structured belief functions. Thus, sample generation for DST is possible. We elaborated also learning procedures for discovery of graphoidal structures from data.

The quantitative model seems to be the best fitting model for belief functions created so far.

This frequency model differs from what was previously considered [16,17] in that it assumes that reasoning in DST is connected with updating of variables for individual cases. This is different from e.g. reasoning in probability where reasoning means only selection of cases. In this way, failures of previous approaches could be overcome.

Many authors [13,16] question the need for an empirical model for DST and point rather to theoretical properties of DST considered within an axiomatic framework seeking parallels with the probability theory. Though it is true that the probability theory may be applied within the framework of Kolmogorov axioms and quite useful results are derived in this way, one shall still point out that the applicability of probability theory is significantly connected with frequencies. Both frequencies considered as "naive probabilities", or ones being probabilities "in the limit". Statistics is clearly an important part of the probabilistic world.

All three interpretations share a common drawback they are not sensu stricto probabilistic. In the current paper we make an attempt of a purely probabilistic vision of plausibility function.

2 Basics of the Dempster-Shafer Theory

We understand DST measures in a standard way (see [12]). Let Ξ be a finite set of elements called elementary events. Any subset of Ξ is a composite event, or hypothesis. Ξ be called also the frame of discernment.

Definition 1. [12] Let Ω be a finite set of elements called elementary events. The set Ω is called frame of discernment. Any subset of Ω be a composite event. A basic probability assignment (bpa) function is any function $m:2^\Omega \rightarrow [0, 1]$ such that

$$\sum_{A \in 2^\Omega} m(A) = ONE \quad m(\emptyset) = 0, \quad \forall A \in 2^\Omega \quad 0 \leq \sum_{A \subseteq B} m(B)$$

We say that a bpa is vacuous iff $m(\Omega) = ONE$ and $m(A) = 0$ for every $A \neq \Omega$.

If ONE is equal 1, then we say that the belief function is normalized, otherwise not (but ONE must be positive).

Definition 2. [12] Let a belief function be defined as $Bel:2^\Omega \rightarrow [0, 1]$ so that $Bel(A) = \sum_{B \subseteq A} m(B)$. Let a plausibility function be $Pl:2^\Omega \rightarrow [0, 1]$ with $\forall A \in 2^\Omega \quad Pl(A) = ONE - Bel(\Omega - A)$, a commonality function be $Q:2^\Omega - \{\emptyset\} \rightarrow [0, 1]$ with $\forall A \in 2^\Omega - \{\emptyset\} \quad Q(A) = \sum_{A \subseteq B} m(B)$.

Definition 3. [12] The Rule of Combination of two Independent Belief Functions Bel_{E_1}, Bel_{E_2} Over the Same Frame of Discernment (the so-called Dempster-Rule), denoted

$$Bel_{E_1, E_2} = Bel_{E_1} \oplus Bel_{E_2}$$

is defined as follows: :

$$m_{E_1, E_2}(A) = c \cdot \sum_{B, C; A=B \cap C} m_{E_1}(B) \cdot m_{E_2}(C)$$

(c - constant normalizing the sum of m to 1).

Under multivariate settings Ξ is a set of vectors in n -dimensional space spanned by the set of variables $\mathbf{X} = \{X_1, X_2, \dots, X_n\}$. If $A \subseteq \Xi$, then by projection $A^{\downarrow \mathbf{Y}}$ of the set A onto a subspace spanned by the set of variables $\mathbf{Y} \subseteq \mathbf{X}$ we understand the set B of vectors from A projected onto \mathbf{Y} . Then marginalization operator of DST is defined as follows:
 $m^{\downarrow \mathbf{Y}}(B) = \sum_{A: B=A^{\downarrow \mathbf{X}}} m(A)$.

Definition 4. (See [?]) Let B be a subset of Ξ , called evidence, m_B be a basic probability assignment such that $m_B(B) = 1$ and $m_B(A) = 0$ for any A different from B . Then the conditional belief function $Bel(.||B)$ representing the belief function Bel conditioned on evidence B is defined as: $Bel(.||B) = Bel \oplus Bel_B$.

3 New Rule of Evidence Combination

Let us suggest now a totally new approach to understanding belief functions.

We assume the following interpretation of the plausibility function: $Pl_{\xi}(A)$ is the maximum probability that an element from the set of events A occurs, given the evidence ξ , where we assume the apriorical probability of all elementary events is equal. Let ξ_1 and ξ_2 be two independent bodies of evidence, which are represented numerically by plausibility functions Pl_{ξ_1} and Pl_{ξ_2} over some frame of discourse Ω . We would like to obtain such an evidence updating rule \oplus_{Pl} that $Pl_3 = Pl_{\xi_1} \oplus_{Pl} Pl_{\xi_2}$ would have the semantics that under that interpretation $Pl_3(A)$ is the maximum probability that an element from the set of events A occurs, given the evidence Pl_1, Pl_2 under the least conflicting evidence.

Let us study in detail this assumption. First of all we have to tell what we mean by independent evidence. Let ω be an elementary event from the frame of discernment Ω . The body of evidence ξ_1 is independent of the body ξ_2 if, for each $\omega \in \Omega$, the probability of occurrence of evidence ξ_1 is independent of the occurrence of evidence ξ_2 . So we say that $Pr(\xi_1 \wedge \xi_2 | \omega) = Pr(\xi_1 | \omega) \cdot Pr(\xi_2 | \omega)$.

How shall we understand the evidence, however. For any $A \subseteq \Omega$ should hold $Pl_{\xi}(A) \geq Pr(A | \xi)$. Consequently, by the way, $Pl_{\xi}(A) + Pl_{\xi}(\Omega/A) \geq 1$.

Now observe that $Pr(\omega_1 \vee \omega_2 | \xi) = Pr(\omega_1 | \xi) + Pr(\omega_2 | \xi)$. As a consequence, we have always that $Pl_{\xi}(\{\omega_1\}) + Pl_{\xi}(\{\omega_2\}) \geq Pl_{\xi}(\{\omega_1, \omega_2\})$.

Let us now turn to combining independent evidence.

$$Pr(\omega | \xi_1 \wedge \xi_2) = Pr(\xi_1 \wedge \xi_2 | \omega) \cdot \frac{Pr(\omega)}{Pr(\xi_1 \wedge \xi_2)} Pr(\xi_1 | \omega) \cdot Pr(\xi_2 | \omega) \cdot \frac{Pr(\omega)}{Pr(\xi_1 \wedge \xi_2)}$$

$$Pr(\omega | \xi_1) \cdot Pr(\omega | \xi_2) \cdot \frac{Pr(\xi_1) \cdot Pr(\xi_2)}{Pr(\xi_1 \wedge \xi_2) \cdot Pr(\omega)}$$

So we can conclude that $Pl_{\xi_1 \wedge \xi_2}(\omega) = Pl_{\xi_1}(\omega) \cdot Pl_{\xi_2}(\omega) \cdot c$ where c is a normalizing factor (which needs to be chosen carefully).

But what about $Pr(\omega_1 \vee \omega_2 | \xi_1 \wedge \xi_2)$? We know that $Pr(\omega_1 \vee \omega_2 | \xi_1 \wedge \xi_2) = Pr(\omega_1 | \xi_1 \wedge \xi_2) + Pr(\omega_2 | \xi_1 \wedge \xi_2)$ hence

$$Pr(\omega_1 \vee \omega_2 | \xi_1 \wedge \xi_2) \\ Pr(\omega_1 | \xi_1) \cdot Pr(\omega_1 | \xi_2) \cdot \frac{Pr(\xi_1) \cdot Pr(\xi_2)}{Pr(\xi_1 \wedge \xi_2) \cdot Pr(\omega_1)} + Pr(\omega_2 | \xi_1) \cdot Pr(\omega_2 | \xi_2) \cdot \frac{Pr(\xi_1) \cdot Pr(\xi_2)}{Pr(\xi_1 \wedge \xi_2) \cdot Pr(\omega_2)}$$

As $Pr(\omega)$ is the same for all the ω s, we get

$$Pr(\omega_1 \vee \omega_2 | \xi_1 \wedge \xi_2) \\ (Pr(\omega_1 | \xi_1) \cdot Pr(\omega_1 | \xi_2) + Pr(\omega_2 | \xi_1) \cdot Pr(\omega_2 | \xi_2)) \cdot \frac{Pr(\xi_1) \cdot Pr(\xi_2)}{Pr(\xi_1 \wedge \xi_2) \cdot Pr(\omega)}$$

We can easily check that this translates to:

$$Pl_{\xi_1 \wedge \xi_2}(\{\omega_1, \omega_2\}) = \\ \max(Pl_{\xi_1}(\omega_1) \cdot Pl_{\xi_2}(\omega_1) + (Pl_{\xi_1}(\{\omega_1, \omega_2\}) - Pl_{\xi_1}(\omega_1)) \cdot (Pl_{\xi_2}(\{\omega_1, \omega_2\}) - Pl_{\xi_2}(\omega_1)) \\ , Pl_{\xi_1}(\omega_1) \cdot (Pl_{\xi_2}(\{\omega_1, \omega_2\}) - Pl_{\xi_2}(\omega_2)) + (Pl_{\xi_1}(\{\omega_1, \omega_2\}) - Pl_{\xi_1}(\omega_1)) \cdot Pl_{\xi_2}(\omega_2) \\ , Pl_{\xi_1}(\omega_2) \cdot (Pl_{\xi_2}(\{\omega_1, \omega_2\}) - Pl_{\xi_2}(\omega_1)) + (Pl_{\xi_1}(\{\omega_1, \omega_2\}) - Pl_{\xi_1}(\omega_2)) \cdot Pl_{\xi_2}(\omega_1) \\ , Pl_{\xi_1}(\omega_2) \cdot Pl_{\xi_2}(\omega_2) + (Pl_{\xi_1}(\{\omega_1, \omega_2\}) - Pl_{\xi_1}(\omega_2)) \cdot (Pl_{\xi_2}(\{\omega_1, \omega_2\}) - Pl_{\xi_2}(\omega_2))) \cdot c$$

where c is the normalizing factor mentioned earlier.

These formulas easily generalize for subsets of Ω with higher cardinality. The normalizing factor should be chosen in such a way that $Pl_{\xi_1 \wedge \xi_2}(\Omega) = 1$.

The generalization of \oplus_{Pl} for frames of discourse with cardinality higher than 3 runs along the following lines. To combine Pl_1 with Pl_2 we calculate:

- for each subset X of Ω
 $Pl_{result}(X) = PL_1^{\downarrow * X} \otimes_V Pl_2^{\downarrow * X};$

The operator $\downarrow * X$ does only a change of the domain of the Pl function keeping the values of Pl for each subset of X and presuming that the discourse frame consists only of X . In this way we get unnormalized Pl s here, which are not normalized during this operation.

The operator \otimes_V , returning a numerical value, attempts identify such combinations of mass assignments m_a and m_b to singleton sets that will not violate the constraints imposed by plausibility functions Pl_1 and Pl_2 resp. and such that the sum $\sum_{X: X \text{ singleton}} m_a(X) \cdot m_b(X)$ is maximal.

This is done by the operation of so-called pushing down the plausibilities to singleton sets. Independently for Pl_1 and Pl_2 candidate m_a and m_b are obtaining via "pushing-down" recursively a singleton ω of Ω . A candidate m_a is obtained if all singletons are pushed down. Different candidates are obtained by different sequences of pushing down. It is easy to imagine that the process is time-consuming and its complexity grows exponentially with the number of elements of a set. Nonetheless for small domains the operation is feasible.

The idea of the push-down operator $\downarrow +$ is as follows: Let Pl be a plausibility function. If A does not contain ω , $Pl^{\downarrow+\omega}(A) = \min(Pl(A), Pl(A \cup \{\omega\}) - Pl(\{\omega\}))$, and otherwise $Pl^{\downarrow+\omega}(A) = Pl(A)$.

Under these conditions it is obvious that we do not seek actually the maximum product over the whole domain, but rather in some "corner points". We will give a formal proof elsewhere that this check is in fact sufficient to establish the maximum. Here we only want to draw attention to the analogy with linear programming, where we seek the maximum subject to linear constraints. Whenever we fix "pushdown" of one of the plausibility distributions, we in fact have a linear optimization case with the other. If found, we can do the same with the other.

The \oplus_{Pl} operator is characterized by commutativity and associativity. The commutativity is easily seen because all the operations are in fact symmetrical with respect to left and right hand of the operators. The associativity is more difficult to grasp, and a formal proof will be subject of another publication. Nonetheless we can give here brief common-sense guidelines how it can be established. We can essentially concentrate on the associative properties of the maximum operator. Starting with the expression of combination of all the three plausibility functions, we can show that we can equivalently denote the same optimization task when drawing behind braces the first or the third operand.

In the next section we show some properties of the new operator compared with Dempster rule of combination for some illustrative examples.

4 Examples

Let us consider the bodies of evidence in the tables 1, 2, 3.

Table 1. mass function for the body of evidence ξa

m value	for the set of elements
0.0	{ }
0.25	{ red }
0.25	{ blue }
0.25	{ red , blue }
0.25	{ green }
0.0	{ red , green }
0.0	{ blue , green }
0.0	{ red , blue , green }

We can check the commutativity and obtain the results as in the table 4.

The associativity has been verified in table 5.

It is worth noting, that the new operator is different from Dempster rule, compare tables 4 and 6

Table 2. mass function for the body of evidence ξb

m value	for the set of elements
0.0	{ }
0.2	{ red }
0.4	{ blue }
0.1	{ red , blue }
0.0	{ green }
0.0	{ red , green }
0.3	{ blue , green }
0.0	{ red , blue , green }

Table 3. mass function for the body of evidence ξc

m value	for the set of elements
0.0	{ }
0.0	{ red }
0.15	{ blue }
0.25	{ red , blue }
0.35	{ green }
0.25	{ red , green }
0.0	{ blue , green }
0.0	{ red , blue , green }

Table 4. mass function for the body of evidence $(\xi a \oplus_{PL} \xi b) = (\xi b \oplus_{PL} \xi a)$

m value	for the set of elements
0.0	{ }
0.1111111111111105	{ red }
0.4999999999999999	{ blue }
0.2222222222222232	{ red , blue }
0.0	{ green }
0.0	{ red , green }
0.1666666666666674	{ blue , green }
0.0	{ red , blue , green }

Table 5. mass function for the body of evidence $((\xi a \oplus_{PL} \xi b) \oplus_{PL} \xi c) = (\xi a \oplus_{PL} (\xi b \oplus_{PL} \xi c))$

m value	for the set of elements
0.0	{ }
0.0	{ red }
0.4214285714285715	{ blue }
0.3214285714285714	{ red , blue }
0.014285714285714124	{ green }
0.07142857142857151	{ red , green }
0.1357142857142858	{ blue , green }
0.0357142857142857	{ red , blue , green }

Table 6. mass function for the body of evidence $(\xi a \oplus \xi b)$

m value	for the set of elements
0.0	{ }
0.2083333333333337	{ red }
0.6249999999999999	{ blue }
0.04166666666666663	{ red , blue }
0.12500000000000003	{ green }
0	{ red , green }
0	{ blue , green }
0	{ red , blue , green }

With this and other experiments we see clearly the tendency of Dempster rule to move mass downwards to singleton sets, whereas the new rule is much more cautious here and in fact does not introduce the feeling of certainty where it is not justified.

5 Conclusions

We have introduced in this paper a new DST operator for combining independent evidence providing a clear probabilistic definition of the plausibility function, which is preserved under this rule of combination.

We have also provided several toy examples to give an impression what results are returned by the new operator.

Though the strict theoretical proof of properties like cummutativeness, associativeness is still to be provided, the computations for test examples show that the properties really hold. It is also obvious from the examples that the new rule differs from the Dempster rule of evidence combination. An interested reader is invited to visit the Web page <http://www.ipipan.waw.pl/~klopotek/DSTnew/DSTdemo.html> to try out himself.

References

1. Cano J., Delgado M., Moral S.: An axiomatic framework for propagating uncertainty in directed acyclic networks, *Int. J. of Approximate Reasoning*. 1993:8, 253-280.
2. A.P.Dempster: Upper and lower probabilities induced by a multi-valued mapping. *Ann. Math. Stat.* **38** (1967), 325-339
3. R.Fagin, J.Y.Halpern: Uncertainty, belief, and probability, *Proc. Int. Joint Conf. AI, IJCAI89*, Detroit, 1989, 1161-1167.
4. R.Fagin, J.Y.Halpern: A new approach to updating beliefs, in: J.F. Lemmer, L.N. Kanal eds: *Uncertainty in Artificial Intelligence 6* (North-Holland Amsterdam, 1991), 347-374.
5. R. Fagin, J.Y. Halpern: Uncertainty, belief, and probability. *Comput. Intell.* 7:1991, 160-173.
6. J.Y. Halpern, R. Fagin: Two views of belief: belief as generalized probability and belief as evidence. *Artificial Intelligence* 54:1992, 275-317
7. H.E. Kyburg Jr: Bayesian and non-Bayesian evidential updating.
8. M.A.Kopotek, S.T.Wierzcho: Quest on New Applications for Dempster-Shafer Theory: Risk Analysis in Project Profitability Calculus. In: P. Grzegorzewski, O.Hryiewicz, M.A.Gil Eds.: *Soft Methods in Probability, Statistics and Data Analysis*, Advances in Soft Computing Series, Physica-Verlag/Springer Verlag, 2002, ISBN 3-7908-1526-8, pp. 302-309
9. Pearl, 1990 J. Pearl: Bayesian and Belief-Function formalisms for evidential reasoning: A conceptual analysis. in: G. Shafer, J. Pearl (eds): *Readings in Uncertain Reasoning*, (ISBN 1-55860-125-2, Morgan Kaufmann Publishers Inc., San Mateo, California, 1990), 540-569.

10. Provan, 1990 G.M.Provan: A logic-based analysis of Dempster-Shafer Theory, *International Journal of Approximate Reasoning* 41990, 451-495.
11. Z.W. Ras: Query processing in distributed information systems, *Fundamenta Informaticae Journal*, Special Issue on Logics for Artificial Intelligence, IOS Press, Vol. XV, No. 3/4, 1991, 381-397
12. G.Shafer: *A Mathematical Theory of Evidence* , Princeton University Press, Princeton, 1976
13. G.Shafer: Perspectives on the theory and practice of belief functions, *International Journal of Approximate Reasoning*, 1990:4, 323-362.
14. P.P.Shenoy: Valuation networks and conditional independence. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol.2, No.2, June 1994.
15. A.Skowron, J.W.Grzymała-Busse: From rough set theory to evidence theory. in:R.R.Yager, J.Kasprzyk and M.Fedrizzi, eds, *Advances in the Dempster-Shafer Theory of Evidence* J. Wiley, New York (1994), 193-236.
16. Smets, 1992 Ph.Smets: Resolving misunderstandings about belief functions, *International Journal of Approximate Reasoning* 1992:6:321-344.
17. Smets, Kenne, 1994 Ph.Smets, R.Kennes: The transferable belief model, *Artificial Intelligence* 66 (1994), 191-234
18. S.T.Wierzcho, M.A.Kopotek: *Evidential Reasoning. An Interpretative Investigation*. Wydawnictwo Akademii Podlaskiej, Siedlce, 2002 PL ISSN 0860-2719, 304 pages,