Information Fusion – Basics on ET

Combination Techniques for Uncertain Information in Measurement and Signal Processing

3.2 Dempster-Shafer Theory

Information Fusion Prof. Dr.-Ing. Volker Lohweg

Lectures – Contents

Information Fusion Only valid with verbal explanations Prof. Dr.-Ing. Volker Lohweg 2

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – Introduction

- Dempster-Shafer Theory (DST) is a mathematical theory of evidence. The seminal work on the subject is [Shafer, 1976], which is an expansion of [Dempster, 1967].
- In a finite discrete space, Dempster-Shafer theory can be interpreted as a generalization of probability theory where probabilities are assigned to *sets* as opposed to mutually exclusive singletons.
- In traditional probability theory, evidence is associated with only one possible event. In DST, evidence can be associated with multiple possible events, e.g., sets of events.
- One of the most important features of Dempster-Shafer theory is that the model is designed to cope with varying levels of precision regarding the information and no further assumptions are needed to represent the information. It also allows for the direct representation of uncertainty of system responses where an imprecise input can be characterized by a set or an interval and the resulting output is a set or an interval.

Karl Sentz and Scott Ferson**: SANDIA REPORT** SAND2002-0835, Release April 2002, **Combination of Evidence in Dempster-Shafer Theory,** Sandia National Laboratories Albuquerque, New Mexico 87185 and Livermore, California 94550

Information Fusion Only valid with verbal explanations Prof. Dr.-Ing. Volker Lohweg 3

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – Introduction

- DST models the uncertainty interval in such a way that it can not be allocated for a certain hypothesis or against a certain hypothesis.
- Ignorance can be modeled explicitly.
- Therefore, we will not define probabilities for events, but so called **evidences***.
- Evidence can be interpreted as a generalization of probability where, instead of an one-dimensional measure a two-dimensional measure is used.
- Evidence as a measure of uncertainty is based upon
  - the ***degree of belief*** of a sensor source information and
  - the ***plausibility*** of the sensor source information.

*Beleg, Nachweis

Information Fusion Only valid with verbal explanations Prof. Dr.-Ing. Volker Lohweg 4

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – Introduction

Dempster and Shafer's ***support interval*** **:**

- The key idea of support intervals is that one should logically separate the arguments ***for*** and the arguments ***against*** a given hypothesis, rather than assuming one to be the complement of the other.
- Two concepts: *belief* and *plausibility*
- The belief *Bel(H)* represents the weight of the facts supporting the working hypothesis *H.*
- The plausibility *Pl(H)* is one minus the weight of the facts speaking against *H.*

$$Pl(H) = 1 - Bel(H^C)$$

Information Fusion Only valid with verbal explanations Prof. Dr.-Ing. Volker Lohweg 5

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – Introduction

- Illustrative Example 3.2-1 "Weather report"
  - Let us assume that the actual weather report forecasts good weather for tomorrow. From expert's experience we know: The certainty of this prognosis is 80%.
  - In classical probability theory we would define P[good] = 0.8 and P[bad] = 0.2.
  - However, P[bad] = 0.2 is not logical, because we have no information whether the weather is good or bad. We are not able to give a stable reason in either direction.
  - The a/m topic distinguishes Probability theory and DST:
    - For P[bad] we are not able to give a stable reason in either direction.
  - In DST the P[bad] = 0.2 is not allocated to the complement P[good^c], but to all possible hypotheses.

After: Andreas Garzotto, Vollautomatische Erkennung von Schriftzeichen in gedrucktem Schriftgut, Diss, Universität Zürich, 1994

Information Fusion Only valid with verbal explanations Prof. Dr.-Ing. Volker Lohweg 6

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – Introduction

- Illustrative Example 3.2-1 "Weather report", cont'd

- Comparison
  - Probability Theory:
    - P[good] = 0.8; P[bad] = 0.2.

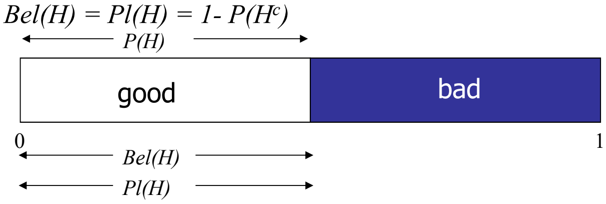
  - DST:
    - P[good] = [0.8 … 1.0] (Plausibility interval, all information which is supported).
    - P[good] = 0.8 (Belief, this information is supported), 0.2 is uncertain.
    - P[bad] = P[good^c] = 0 (Belief of bad weather, because there is no supporting information).

Institut für industrielle Informationstechnik
Institute Industrial IT

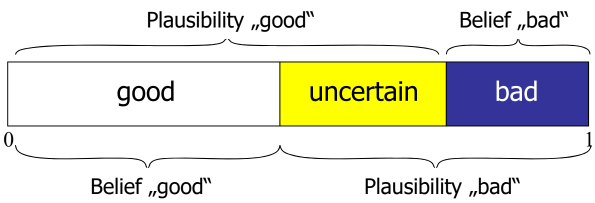
3.2 Evidence Theory – Introduction

Bayesian probability theory

$P(H) = Bel(H) = Pl(H) = 1 - P(H^c)$



Dempster and Shafer: $Bel(H) \leq Pl(H)$



Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – Introduction

- Illustrative Example 3.2-1 "Weather report", cont'd

  - Now let us assume we have two information sources:
    - The **weather forecast {A}** with 80% certainty, and a **weather house {B}** with 60% certainty.
    - Both information sources predict good weather with 80% resp. 60% certainty.
  - The following table shows the aggregation (fusion) of the source information (resulting evidences):

Hypothesis: Good weather	Weather forcast certain (80%)	Weather forcast uncertain (20%)
Weather house certain (60%)	Weather good: $0.8 \cdot 0.6 = 0.48$ (48%)	Weather good: $0.2 \cdot 0.6 = 0.12$ (12%)
Weather house uncertain (40%)	Weather good: $0.8 \cdot 0.4 = 0.32$ (32%)	Weather uncertain: $0.2 \cdot 0.4 = 0.08$ (8%)

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – Introduction

- Illustrative Example 3.2-1 "Weather report", cont'd
  - Three of four combinations support the hypothesis "good weather". This leads to the fact that the belief in "good weather" is calculated by the sum of all supporting evidences (basic probability assignments, masses), that is:
    - *Bel* [good weather] = 0.48 + 0.12 + 0.32 = 0.92 (92%).
  - Therefore, the prognosis for "good weather" is more certain by fusing two information sources (80% → 92%).
  - No prognosis is possible for 8%. It is plausible that with 8% we have good ***or*** bad weather.
  - The belief in "bad weather" is 0%, because there is no supporting hypothesis for "bad weather".

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – Introduction

- Illustrative Example 3.2-1 "Weather report", cont'd
  - Now let us assume we have two **conflicting** information sources:
    - Source {A} predicts good weather with 80% certainty.
    - Source {B} predicts bad weather with 60% certainty.
  - The following table shows the aggregation (fusion) of the source information (resulting evidence):

good / bad	Weather forcast certain (80%)	Weather forcast uncertain (20%)
Weather house certain (60%)	Impossible: 0.8 · 0.6 = 0.48 (48%)	Weather bad: 0.2 · 0.6 = 0.12 (12%)
Weather house uncertain (40%)	Weather good: 0.8 · 0.4 = 0.32 (32%)	Weather uncertain: 0.2 · 0.4 = 0.08 (8%)

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – Introduction

- Illustrative Example 3.2-1 "Weather report", cont'd
  - The fact that two conflicting information sources are both right is of course impossible!
  - DST rescales all evidences in a way that the sum of all possible resulting evidences equals 1, that is:
    - 0.32 + 0.12 + 0.08 = 0.52.
    - Weather bad: 0.12/0.52 ≈ 0.23
    - Weather good: 0.32/0.52 ≈ 0.62
    - Weather uncertain: 0.08/0.52 ≈0.15

good / bad	Weather forcast certain (80%)	Weather forcast uncertain (20%)
Weather house certain (60%)		Weather bad: approx. 0.23 (23%)
Weather house uncertain (40%)	Weather good: approx. 0.62 (62%)	Weather uncertain: approx. 0.15 (15%)

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – Introduction

- Illustrative Example 3.2-1 "Weather report", cont'd

good / bad	Weather forcast certain (80%)	Weather forcast uncertain (20%)
Weather house certain (60%)		Weather bad: approx. 0.23 (23%)
Weather house uncertain (40%)	Weather good: approx. 0.62 (62%)	Weather uncertain: approx. 0.15 (15%)

- *Bel* (good) = 0.62; *Pl* (good) = 0.62 + 0.15 = 0.77
- *Bel* (bad) = 0.23; *Pl* (bad) = 0.23 + 0.15 = 0.38
- *Igr* (good) = *Pl* (good) - *Bel* (good) = 0.15
- *Igr* (bad) = *Pl* (bad) - *Bel* (bad) = 0.15

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – Introduction

- If *H* denotes the hypothesis then the belief and the plausibility are interconnected as follows (cf. one slide before):

$$Pl(H) = 1 - Bel(H^C) \rightarrow Pl(H) + Bel(H^C) = 1$$

  - Numerical examples:
    - *Pl* (good) = 1 - *Bel* (*bad*) = 1 – 0.23 = 0.77
    - *Pl* (bad) = 1 - *Bel* (*good*) = 1 – 0.62 = 0.38

- The belief for the counterhypothesis *Bel* (H^c) is sometimes referred to as *doubt* $D(H)$ with respect to the working hypothesis *H*. Hence we have:

$$Pl(H) = 1 - D(H) \rightarrow Pl(H) + D(H) = 1$$

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – DST Definitions

- Frame of discernment
  - The so called **Frame of Discernment*** ***Ω*** is the set of all *mutually exclusive and complete* hypotheses or classes for which an expert or a system can decide on the basis of measurement results.
    - Example $\Omega = \{A, B, C\}$
  - The set must be complete regarding the sensing units, that is, all classes which can appear must be covered in a set.
  - Eg., referring to our weather report, {A} could be defined as sunshine, {B} is cloudy, and {C} is rainy.

mutually exclusive and complete = sich gegenseitig ausschließend und vollständig

*ger.: Wahrnehmungsrahmen

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – DST Definitions

- Frame of discernment
  - The so called **Power Set*** (potency set) defines the hierarchical ordered class evidences.
  - A gross class evidence combines *disjunctively* (or) a set of fine evidences.
    - Example $\Omega = \{A, B, C\}$
    - Power Set

$$\Theta = 2^{\Omega} = \left\{ \varnothing, \{A\}, \{B\}, \{C\}, \{A,B\}, \{A,C\}, \{B,C\}, \underbrace{\{A,B,C\}}_{\Omega} \right\}$$

C	B	A
0	0	0
0	0	H
0	H	0
0	H	H
H	0	0
H	0	H
H	H	0
H	H	H

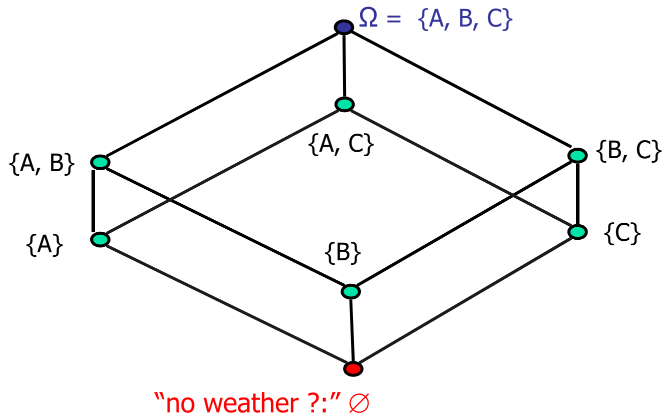
  - The gross class {A, B} combines the classes {A} and {B} in a disjunctive way. It means, that sunshine as well as a cloudy sky may be the case.

*ger.: Potenzmenge

3.2 Evidence Theory – DST Definitions

- Frame of discernment
  - Graphical representation

A = sunshine
B = cloudy
C = rainy



Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – DST Definitions

- Basic Probability Assignment (BPA), Basic Belief Assignment (BBA), Mass m
  - The theory of evidence assigns a belief mass to each subset of the power set. Formally, a function m: $2^{\Omega} \rightarrow [0,1]$, is called a

basic belief assignment* (BB(P)A) or **mass *m***,

when it verifies two axioms. First, the mass of the empty set is zero:

$$m(\varnothing) = 0$$

  - Second, the masses of the remaining members of the power set add up to a total of 1: $\sum_{X \in 2^{\Omega}} m(X) = 1$
  - The mass function m(X) defines a measure for the support of a certain class or a hypothesis based on sensing sources.

*basic probability assignment

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – DST Definitions

- Basic Probability Assignment (BPA), Basic Belief Assignment (BBA), Mass m
  - The mass *m*(*X*) of a given member of the power set, *X*, expresses the proportion of all relevant and available evidence that supports the claim that the actual state belongs to *X* but to no particular subset of *X*.
  - The value of *m*(*X*) pertains *only* to the set *X* and makes no additional claims about any subsets of *X*, each of which has, by definition, its own mass.
  - If all supporting masses are distributed to the appropriate classes, Ω gets the remaining (rest) masses → rest evidences, that is the difference between all distributed masses and 1.
  - This rest evidence represents ignorance. The set Ω represents the fact that all hypotheses are possible → Ignorance.

Information Fusion Only valid with verbal explanations Prof. Dr.-Ing. Volker Lohweg 19

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – DST Definitions

- Basic Probability Assignment (BPA), Basic Belief Assignment (BBA), Mass m
  - If a sensing unit is not able to deliver plausible measurement results, it does mean the data fusion process can be modeled in such a way that

$$m(\Omega) = 1 \,. \quad \text{(Complete Uncertainty)}$$

  - A counterhypothesis is, opposed to the traditional probability theory, not modeled as a traditional complement $P^c = 1 - P$, but as the set complement of the hypothesis.
  - The complement of {A} is {B,C} if Ω ={A, B, C}. E.g. the complement of {sunshine} is {cloudy or rainy}.

Information Fusion Only valid with verbal explanations Prof. Dr.-Ing. Volker Lohweg 20

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – DST Definitions

- Basic Probability Assignment (BPA), Basic Belief Assignment (BBA), Mass m
  - Example 3.2-2 "weather report, revisited"
    - We assume two weather reports (WR) are existing.
      - Hypothesis 1: WR1 forecasts for the next day: 60% rainy (0.6); 30% alternating cloudy and sunny (0.3) .
      - Hypothesis 2: WR2 forecasts for the next day: 80% sunny (0.8).

$$m_{WR1}(\{C\}) = 0.6, \quad m_{WR1}(\{A,B\}) = 0.3, \quad m_{WR1}(\{A,B,C\}) = 0.1$$
$$m_{WR1}(\{A,B,C\}) = 1 - (m_{WR1}(\{C\}) + m_{WR1}(\{A,B\}))$$
$$m_{WR2}(\{A\}) = 0.8, \quad m_{WR2}(\{A,B,C\}) = 0.2$$

Nothing can be said about other combinations, Therefore: → Ω

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – DST Definitions

- Basic Probability Assignment (BPA), Basic Belief Assignment (BBA), Mass m
  - Focal Element
    - The focal elements of the uncertain quantities are defined as

$$\{X : X \subset 2^{\Omega}, m(X) > 0\}$$

    - A subset of the Frame of discernment Ω is called the Focal Element of a belief function *Bel* over Ω if m(X)>0.

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – DST Definitions

- Belief Function
  - A function *Bel* : $2^{\Omega} \rightarrow [0,1]$ is called **belief function***, if the following applies:

$$Bel(X) = \sum_{A \subset X} m(A)$$

  - The belief function collects all BPAs which supports a specific X by adding all m-values of the subsets of X.
  - *Bel* is also called lower probability function.

*ger.: Vertrauensfunktion

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – DST Definitions

- Belief Function
  - It is possible to obtain the basic probability assignment from the *Belief* measure with the following inverse function:

$$m(X) = \sum_{A \subset X} (-1)^{|X-A|} Bel(A)$$

where $|X\text{-}A|$ is the difference of the cardinality of the two sets.

*ger.: Vertrauensfunktion

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – DST Definitions

- Plausibility Function
  - A function *Pl* : 2^{Ω} → [0,1] is called **plausibility function***, if the following applies:

$$Pl(X) = \sum_{A \cap X \neq \varnothing} m(A)$$

  - The plausibility function defines a measure for the maximum possible support for X by the addition of all the m-values for which the intersection A ∩ X is not empty.
  - *Pl* is also called upper probability function.

*ger.: Plausibilitätsfunktion

Information Fusion Only valid with verbal explanations Prof. Dr.-Ing. Volker Lohweg 25

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – DST Definitions

- Plausibility Function - dubiety
  - A function *D* : 2^{Ω} → [0,1] is called **doubt (dubiety) function***, if the following applies:

$$D(A) = Bel\left(A^C\right)$$

  - The dubiety function is seldom used on its own. More common are the following relations:

$$Pl(A) = 1 - D(A) = 1 - Bel(A^C)$$

$$Bel(A) = 1 - Pl(A^C)$$

  - This definition of Plausibility in terms of Belief comes from the fact that all basic assignments must sum to 1.

*ger.: Zweifelhaftigkeitsfunktion

Information Fusion Only valid with verbal explanations Prof. Dr.-Ing. Volker Lohweg 26

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – DST Definitions

- Plausibility Function
  - Some proofs: $Pl(X) = 1 - Bel(X^c) = 1 - \sum_{A \subset X^c} m(A) = \cdots$

$$\sum_{A \subset \Theta(X)} m(A) - \sum_{A \subset X^c} m(A) = \sum_{A \cap X \neq \varnothing} m(A) \geq \sum_{A \subset X} m(A) = Bel(X)$$

$$Pl(X) + Pl(X^c) = Pl(X) + 1 - Bel(X) = 1 + Pl(X) - Bel(X) = 1 + Igr(X) \geq 1$$

$$Bel(X) + Bel(X^c) = Bel(X) + 1 - Pl(X) = 1 + Bel(X) - Pl(X) = \cdots$$

$$\cdots = 1 - (Pl(X) - Bel(X)) = 1 - Igr(X) \leq 1$$

$\boxed{Pl(X) \geq Bel(X)}$ $\boxed{Pl(X) + Pl(X^c) \geq 1}$ $\boxed{Bel(X) + Bel(X^c) \leq 1}$

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Evidence Theory – DST Definitions

- Comments
  - As a consequence, given any one of these measures (*m*(*A*), *Bel*(*A*), *Pl*(*A*)) it is possible to derive the values of the other two measures.
  - The precise probability of an event (in the classical sense) lies within the lower and upper bounds of *Belief* and *Plausibility,* respectively.
  - The probability is uniquely determined if *Bel (A) = Pl(A)*. In this case, which corresponds to classical probability, all the probabilities, *P(A)* are uniquely determined for all subsets *A* of the universal set *X* [Yager, 1987, p.97].

$$Bel(X) = Pl(X) = P(X) \quad \text{(Probability Theory)}$$

  - Otherwise, *Bel (A)* and *Pl(A)* may be viewed as **lower and upper bounds on probabilities**, respectively, where the actual probability is contained in the interval described by the bounds.

Institut für industrielle Informationstechnik
Institute Industrial IT

3.2 Theory of Evidence – Summary

- One of the computational advantages of the Dempster-Shafer framework is that priors and conditionals need not be specified.
- Dempster-Shafer theory allows one to specify a degree of ignorance in this situation instead of being forced to supply prior probabilities which add to unity.
- The DST rule of combination is under some criticism.
- The rules of combination are an active research topic.