Wybrane zagadnienia sztucznej inteligencji
Dempster-Shafer Theory

Halina Kwaśnicka

Politechnika Wrocławska

halina.kwasnicka@pwr.wroc.pl

Outline

- Imprecise knowledge – short repetition
- Dempster-Shafer theory – introduction
- How it works – examples
- Problems and possible ways of overcoming them
- Summary

Students activity

- If, and why we should or not should consider uncertain knowledge in Knowledge Based Systems?
- What methods do you know?

https://www.global-english.com/news/tips-for-tefl-discussion-classes/

Dempster-Shafer Theory – Introduction (1)

- Alternative Models of Dealing with Uncertainty Information/Evidence
- Designed to Deal with the distinction between uncertainty and ignorance
  - we compute rather **the probability that evidence support a proposition**, not the probability of the proposition
- Applicability of D-S
  - it is useful when we do not have sufficient data to accurately estimate the prior and conditional probabilities to use Bayes rule
  - a model is incomplete
  - it is not possible to estimate probabilities
  - belief intervals are used to **estimate how close the evidence is to determining the truth of a hypothesis**

Dempster-Shafer Theory – Introduction (2)

- Enables to represent ignorance about support provided by evidence
- Enables a reasoning system to be skeptical
- Example:
  - We are informed that one of three terrorist groups, A, B, or C has planned a bomb building
  - We have some evidence that group C is guilty (P(c)=0.8
  - We do not have info that groups A and B being guilty with probabilities 0.1
  - In probability theory: P(X) + P(not X) =1 -> probability of A and B is distributed equally
  - D-S allows to leave relative beliefs unspecified

Dual nature of uncertainty

- **Random (Aleatory) Uncertainty**
  - uncertainty results from the fact that a system can behave in **random** ways
  - known also as: Stochastic uncertainty, Type A uncertainty, Irreducible uncertainty, Variability, Objective uncertainty
- **Epistemic Uncertainty**
  - uncertainty results from the **lack of knowledge** about a system
  - is a property of the analysts performing the analysis
  - known also as: Subjective uncertainty, Type B uncertainty, Reducible uncertainty, State of Knowledge uncertainty, Ignorance
- Traditionally, probability theory has been used to characterize both types of uncertainty
- Traditional probability theory is not capable of capturing epistemic uncertainty

Application of traditional probability – Principle of Insufficient Reason

- Application of traditional probabilistic methods to epistemic or subjective uncertainty – Bayesian probability
- Such analysis requires the information on the probability of all events
- When not available, the uniform distribution function is often used, justified by Laplace's Principle of Insufficient Reason
- Interpretation: all simple events for which a probability distribution is not known in a given sample space are equally likely
- Example:
  - A system failure with three possible components that could cause failure
  - An expert assigns a probability of failure of Component A with 0.3
  - The expert knows nothing about Components B and C
  - Following the Principle of Insufficient Reason, the expert could assign a probability of failure of 0.35 to each of the two remaining components (B and C).
- This would be a very precise statement about the probability of failure of these two components in the face of complete ignorance regarding these components on the part of the expert

Why use the Dempster-Shafer theory

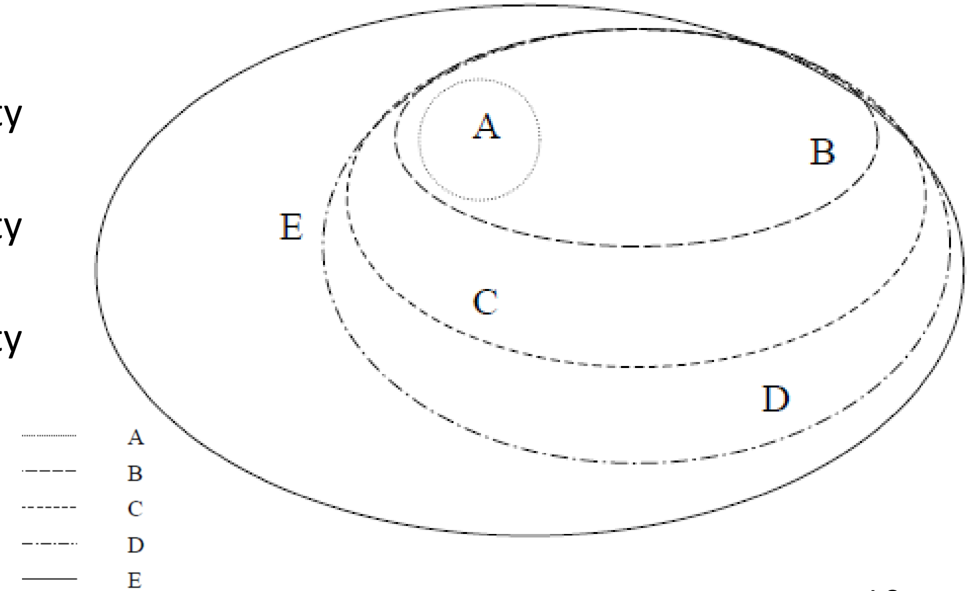
- The relatively high degree of theoretical development among the nontraditional theories for characterizing uncertainty
- The relation of Dempster-Shafer theory to traditional probability theory and set theory
- The large number of examples of applications of Dempster-Shafer theory in engineering in the past years
- The versatility of the Dempster-Shafer theory to represent and combine different types of evidence obtained from multiple sources

Types of evidence

- Two critical and related issues concerning the combination of evidence obtained from multiple sources:
  - (i) the type of evidence involved
  - (ii) how to handle conflicting evidence
- Types of evidence from multiple sources impacts the choice of how information is to be combined
- Four types of evidence from multiple sources:
  - Consonant evidence,
  - Consistent evidence,
  - Arbitrary evidence,
  - Disjoint evidence

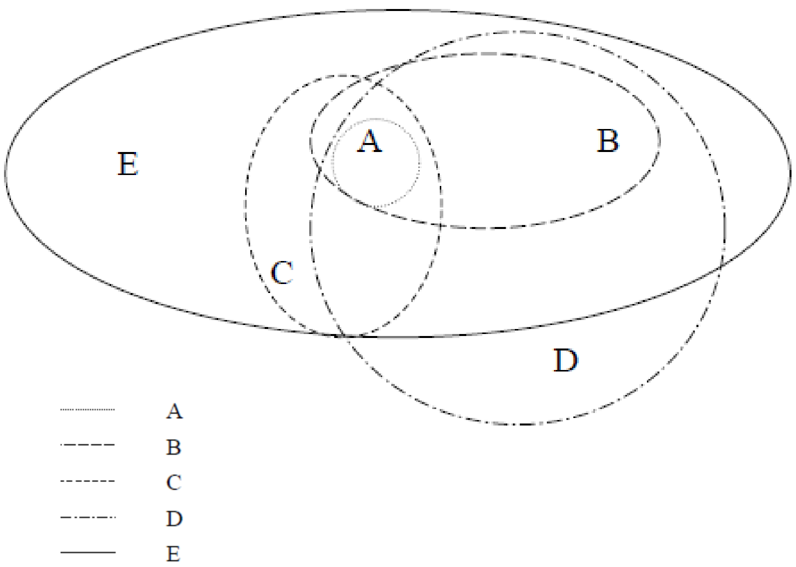
Consonant evidence

- Can be represented as a ***nested structure*** of subsets where the elements of the smallest set are included in the next larger set… all of whose elements are included in the next larger set and …
- It corresponds to the situation where information is obtained over time that increasingly narrows or refines the size of the evidentiary set
- Example from target identification: there are five sensors with varying degrees of resolution, it could look like:
  - Sensor 1 detects one target in vicinity A
  - Sensor 2 detects 2 targets: one in vicinity A and one in vicinity B
  - Sensor 3 detects 3 targets: one in vicinity A, one in B, one in C
  - Sensor 4 detects 4 targets: one in vicinity A, one in B, one in C, one in D
  - Sensor 5 detects five targets: one in A, one in B, one in C, one in, one in E



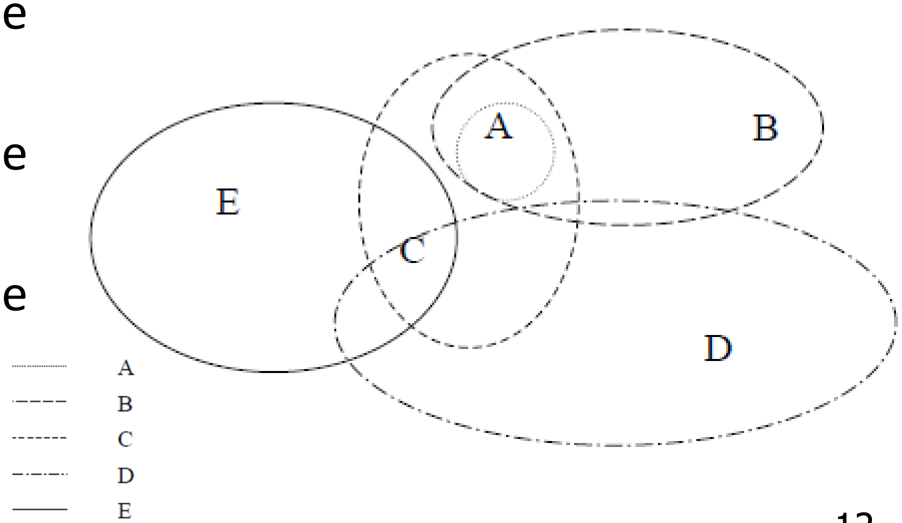
Consistent evidence

- There is ***at least*** one element that is common to ***all*** subsets
- From our target identification, this could look like:
  - Sensor 1 detects one target in vicinity A
  - Sensor 2 detects two targets: one in vicinity A and one in vicinity B
  - Sensor 3 detects two targets: one in vicinity A, one in vicinity C
  - Sensor 4 detects three targets: one in vicinity A, one in vicinity B, one in vicinity D
  - Sensor 5 detects four targets: one in vicinity A, one in vicinity B, one in vicinity C, one in vicinity E



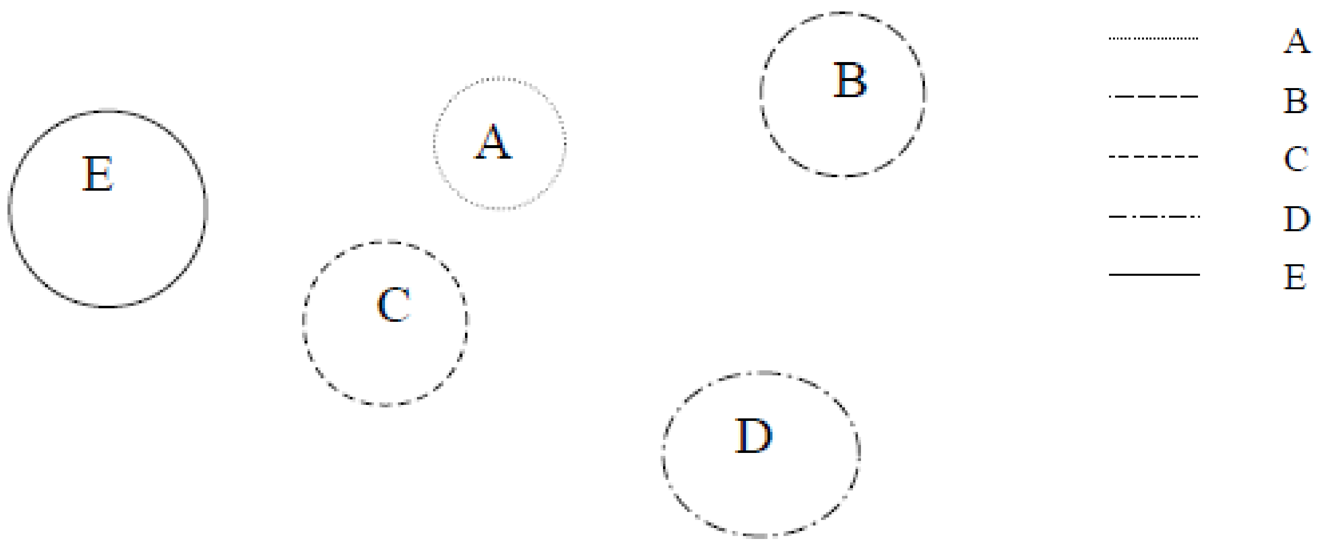
Arbitrary evidence

- It corresponds to the situation where there is ***no element*** common to ***all* subsets** (some subsets may have elements in common)
- From our target identification, this could look like:
  - Sensor 1 detects a target in vicinity A
  - Sensor 2 detects two targets: one in vicinity A and one in vicinity B
  - Sensor 3 detects two targets: one in vicinity A, one in vicinity C
  - Sensor 4 detects two targets: one in vicinity C, one in vicinity D
  - Sensor 5 detects two targets: one in vicinity C, one in vicinity E



Disjoint evidence

- It implies that any two subsets have no elements in common
  - Sensor 1 detects a target in vicinity A
  - Sensor 2 detects a target in vicinity B
  - Sensor 3 detects a target in vicinity C
  - Sensor 4 detects a target in vicinity D
  - Sensor 5 detects a target in vicinity E



Dempster-Shafer (D-ST) vs probability theory (PT) (1)

- Each of the possible configurations of evidence from multiple sources has different implications on the level of conflict associated with the situation
- Consonant evidence
  - the situation where each set is supported by the next larger set
  - implies an agreement on the smallest evidential set
  - however, there is conflict between the additional evidence that the larger set represents in relation to the smaller set
- Consistent evidence implies an agreement on at least one evidential set
- Arbitrary evidence: there is some agreement between some sources but there is no consensus among sources on any one element
- Disjoint evidence: all of the sources supply conflicting evidence

Dempster-Shafer (D-ST) vs probability theory (PT)

- Traditional PT cannot handle consonant, consistent, or arbitrary evidence without resorting to further assumptions of the probability distributions within a set
- PT cannot express the level of conflict between these evidential sets
- Dempster-Shafer theory can handle these various evidentiary types by combining a notion of probability with the traditional conception of sets
- In Dempster Shafer theory, there are many ways to incorporate conflict when combining multiple sources of information

Dempster-Shafer Theory

- Three important functions in Dempster-Shafer theory:
- the ***basic probability assignment*** function (bpa or *m*)
- the ***Belief*** function (*Bel*)
- the ***Plausibility*** function (*Pl*)
- The basic probability assignment (bpa)
  - is a primitive of evidence theory
  - term “basic probability assignment” does *not* refer to probability in the classical sense
  - bpa, represented by *m*, defines a mapping of the power set to the interval between 0 and 1, where
    - bpa of the null set is 0
    - the summation of the bpa’s of **all the subsets of the power set is 1**

In mathematics, the power set (or powerset) of any set S is the set of all subsets of S, including the empty set and S itself

The ***basic probability assignment*** – formally

- The bpa value for a given set A (denoted as $\boldsymbol{m(A)}$), expresses the proportion of *all relevant and available evidence* that supports the claim that a particular element of X (the universal set) *belongs* to the set A
- $m(A)$ value pertains only to the set A and makes no additional claims about any subsets of A
- Any further evidence on the subsets of A would be represented by another bpa, i.e. $B \subset A$, $m(B)$ – the bpa for the subset B
- Formally, m can be represented with the three equations:

(1): $m: P(X) \rightarrow [0,1]$; (2): $m(\emptyset) = 0$;

(3): $$\sum_{A \in P(X)} m(A) = 1$$

where $P(X)$ – the power set of X,
$\emptyset$ is the null set, and
A is a set in the power set ($A \in P(X)$)

Interpretation of *m*(*x*)

- Random switch model: Think of the evidence as a switch that oscillates randomly between two or more positions. The function *m* represents the fraction of the time spent by the switch in each position
- Voting model: *m* represents the proportion of votes cast for each of several results of the evidence (possibly including that the evidence is inconclusive)
- Envelope model: Think of the evidence as a sealed envelope, and m as a probability distribution on what the contents may be

http://mas.cs.umass.edu/classes/cs683/683-2004/lectures/lecture16.pdf

Measures *Belief* and *Plausibility* (1)

- From the bpa, the upper and lower bounds of an interval can be defined
- This interval contains the precise probability of a set of interest (in the classical sense) and is bounded by two nonadditive continuous measures called ***Belief*** and ***Plausibility***
- The lower bound ***Belief*** for a set A ($A \in P(X)$) is defined as the sum of all the basic probability assignments of the proper subsets (B) of the set of interest (A) ($B \subseteq A$)

(4): $$Bel(A) = \sum_{B|B \subseteq A} m(B)$$

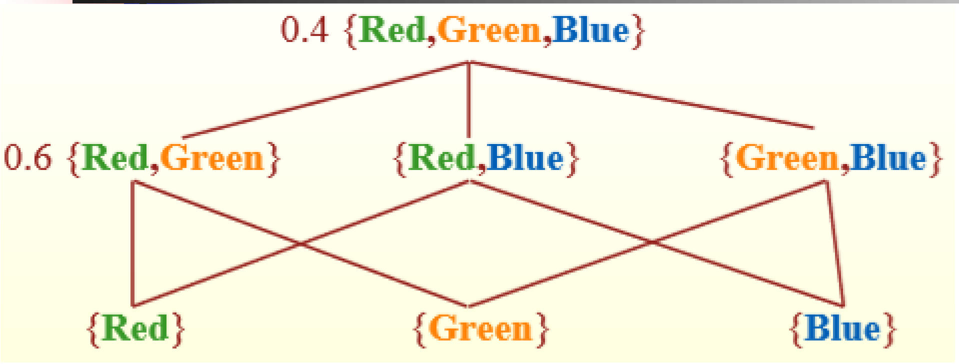
Measures *Belief* and *Plausibility* (2)

- The upper bound, ***Plausibility***, is the sum of all the basic probability assignments of the sets (B) that intersect the set of interest (A) ($B \cap A \neq \emptyset$)
- Formally, for all sets A that are elements of the power set ($A \in P(X)$):

(5): $$Pl(A) = \sum_{B \mid B \cap A \neq \emptyset} m(B)$$

- Measures ***Belief*** and ***Plausibility*** are nonadditive, so it is not required for the sum of all the *Belief* measures to be 1 and similarly for the sum of the *Plausibility* measures

Belief subsets



Assume: the evidence supports {red,green} to degree 0.6.
The remaining support will be assigned to {red, green, blue} while in Bayesian model the remaining support is assigned to the negation of the hypothesis (or its complement) {blue}

- Given a population F =(blue, red, green) of mutually exclusive elements, exactly one of which (*f*) is true, a basic probability assignment (*m*) assigns a number in [0,1] to every subset of F such that the sum of the numbers is 1
  - Mass as a representation of evidence support
- There are $2^{|F|}$ propositions, corresponding to "the true value of *f* is in subset A"
  - (blue), (red), (green), (blue, red), (blue, green), (red, green), (red, blue, green), (empty set)
- A belief in a subset entails belief in subsets containing that subset
  - Belief in (red) entails Belief in (red, green), (red, blue), (red, blue, green)

Interpretation of *Belief* and *Plausibility*

- *Belief* (or *possibility*) is the probability that *B* is provable (supported) by the evidence
  - $Bel(A) = \Sigma_{\{B \text{ in } A\}} m(B)$ (Support committed to A)
  - *Bel* ((red,blue)) = *m*((red))+*m*((blue))+*m*((red,blue))
- *Plausibility* is the probability that *B* is compatible with the available evidence (cannot be disproved)
  - Upper belief limit on the proposition
- $Pl(A) = \Sigma_{\{B \cap A \neq \{\}\}} m(B)$ Support that can move into A
  - Pl((red,blue)) = *m*((red))+*m*((blue))+*m*((red,blue))+*m*((red,green))+ *m*((red,blue,green))+*m*((blue,green))
- Pl(A) = 1-Bel(¬ A)

Confidence Interval

- Belief Interval [Bel(A),Pl(A)], confidence in A
  - Interval width is good aid in deciding when you need more evidence
- [0,1] no belief in support of proposition
  - total ignorance
- [0,0] belief the proposition is false
- [1,1] belief the proposition is true
- [.3,1] partial belief in the proposition is true
- [0,.8] partial disbelief in the proposition is true
- [.2,.7] belief from evidence both for and against proposition

DS's Rule of Combination

- Given two basic probability assignment functions m_1 and m_2 how to combine them
  - Two different sources of evidence
  - $P(s_i,s_j|d) = P(s_i|d)\ P(s_j|d)$; assume conditional independence
- Bel (C) = Sum ($m_1(A_i)\ m_2(B_j)$) where A_i intersect B_j = C
- **Normalized by** amount of non-zero mass left after intersection Sum($m_1(A_i)\ m_2(B_j)$) where A_i intersect B_j not empty

Continuation:

- Combination (called the joint m_{12}) is calculated from the aggregation of two bpa's m_1 and m_2:

$$m_{12}(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - K} \quad \text{when } A \neq \emptyset \qquad m_{12}(\emptyset) = 0$$

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

- K indicates a level of conflict

Example of Rule Combination

- Suppose that m_1(D)=.8 and m_2(D)=.9

	{D} .9	{¬D} 0	{D,¬D} .1
{D} .8	{D} .72	{} 0	{D} .08
{¬D} 0	{} 0	{¬D} 0	{¬D} 0
{D,¬D} .2	{D} .18	{¬D} 0	{D,¬D} .02

- m_{12}(D) = .72+.18+.08=.98
- m_{12}(¬D)=0
- m_{12}(D,¬D)=.02
- Using intervals: [.8,1] and [.9,1] = [.98,1]

$$m_{12}(A) = \frac{\sum_{B\cap C=A} m_1(B)m_2(C)}{1-K} \quad \text{when } A\neq\varnothing \qquad K = \sum_{B\cap C=\varnothing} m_1(B)m_2(C)$$

continuation

- Suppose that m_1(D)=.8 and m_2(¬D)=.9

	{D} 0	{¬D} .9	{D,¬D} .1
{D} .8	{D} 0	{} .72	{D} .08
{¬D} 0	{} 0	{¬D} 0	{¬D} 0
{D,¬D} .2	{D} 0	{¬D} .18	{D,¬D} .02

- Need to normalize by K=0.72:
- m_{12}(D) = (0+0.08+0)/[1-0.72] ≈ 0.29
- m_{12}(¬D)=(0+0+0.18)/[1-0.72] 0.64
- m_{12}(D,¬D)=0.02/0,28 ≈ 0,7

$$m_{12}(A) = \frac{\sum_{B\cap C=A} m_1(B)m_2(C)}{1-K} \quad \text{when } A\neq\varnothing \qquad K = \sum_{B\cap C=\varnothing} m_1(B)m_2(C)$$

Dempster-Shafer Example (1)

- Let Θ be:
  - All : allergy
  - Flu : flu
  - Cold : cold
  - Pneu : pneumonia
- When we begin, with no information *m* is:
  - {Θ} (1.0)
- suppose m_1 corresponds to our belief after observing fever:
  - {Flu, Cold, Pneu } (0.6)
  - {Θ} (0.4)
- Suppose m_2 corresponds to our belief after observing a runny nose:
  - {All, Flu, Cold} (0.8)
  - Θ (0.2)

Dempster-Shafer Example (2)

- Then we can combine m_1 and m_2:

	$\{A,F,C\}$	(0.8)	Θ	(0.2)
$\{F,C,P\}$ (0.6)	$\{F,C\}$	(0.48)	$\{F,C,P\}$	(0.12)
Θ (0.4)	$\{A,F,C\}$	(0.32)	Θ	(0.08)

- So we produce a new, combined m_3;
  - {Flu, Cold} (0.48)
  - {All, Flu, Cold} (0.32)
  - {Flu, Cold, Pneu} (0.12)
  - Θ (0.08)
- Suppose m_4 corresponds to our belief that the problem goes away on trips and thus is associated with an allergy:
  - {All} (0.9)
  - Θ (0.1)

Dempster-Shafer Example (3)

- Applying the numerator of the combination rule yields:

		{A}	(0.9)	Θ	(0.1)
{F,C}	(0.48)	{}	(.432)	{F,C}	(0.048)
{A,F,C}	(0.32)	{A}	(0.288)	{A,F,C}	(0.032)
{F,C,P}	(0.12)	{}	(.108)	{F,C,P}	(0.012)
Θ	(0.08)	{A}	(0.072)	Θ	(0.008)

- Normalizing to get rid of the belief of 0.54 associated with {} gives m_5:
  - {Flu,Cold } (0.104)=.048/.46
  - {Allergy,Flu,Cold } (0.0696)= .032/.46
  - {Flu,Cold,Pneu } (0.026)= .012/.46
  - {Allergy } (0.782)=(.288+.072)/.46
  - Θ (0.017)=.008 /.46
- What is the [belief, possibility] of Allergy?

Dempster Dempster-Shafer Pros

- Addresses questions about necessity and possibility that Bayesian approach cannot answer
- Prior probabilities not required, but uniform distribution cannot be used when priors are unknown
- Useful for reasoning in rule-based systems

Dempster-Shafer Cons

- Source of evidence not independent; can lead to misleading and counter-intuitive results
- The normalization in Dempster's Rule loses some meta-information, and this treatment of conflicting evidence is controversial
  - Difficult to develop theory of utility since Bel is not defined precisely with respect to decision making
- Bayesian approach can also do something akin to confidence interval by examining how much one's belief would change if more evidence acquired
  - Implicit uncertainty associated with various possible changes

Conflicting Evidence and Normalization Problems with Problems with Dempster-Shafer Theory

- Normalization process can produce strange conclusions when conflicting evidence is involved
- Let: cause of failure could be component A, B, or C
- Experts opinions are:
  - $\Theta = \{A,B,C\}$
  - $m_1 = \{A\}\ (0.99), \{B\}\ (0.01)$
  - $m_2 = \{C\}\ (0.99), \{B\}\ (0.01)$
  - $m_1 + m_2 = \{B\}\ (1.0)$
- Certain of B even though neither piece of evidence supported it well – see next slide
- No representation of inconsistency and resulting uncertainty/ ignorance

Dempster Combination of Expert 1 (A, B) and Expert 2 (B, C)

			Expert 1			
			A	**B**	**C**	Failure Cause
			0.99	0.01	0	m_1
Expert 2	Failure Cause	m_2				
	A	0	$m_1(A)\ m_2(A) = 0$	$m_1(B)\ m_2(A) = 0$	$m_1(C)\ m_2(A) = 0$	
	B	0.01	$m_1(A)\ m_2(B) = 0.0099$	$m_1(B)\ m_2(B) = 0.0001$	$m_1(C)\ m_2(B) = 0$	
	C	0.99	$m_1(A)\ m_2(C) = 0.9801$	$m_1(B)\ m_2(C) = 0.0099$	$m_1(C)\ m_2(C) = 0$	

If the intersection is nonempty, the masses for a particular set from each source are multiplied, e.g: $m_{12}(B)$ = (0.01)(0.01) = 0.0001;

If the intersection is empty, this represents conflicting evidence and should be calculated as well. For the empty intersection of the two sets A and C associate with Expert 1 and 2, respectively, there is: m_1(A) m_2(C)=(0.99)(0.99) =(0.9801).

Sum the masses for all sets and the conflict.

The only nonzero value is for the combination of B: m_{12}(B) = 0.0001.

For K: three cells contribute to conflict represented by empty intersections. K = (0.99)(0.01) + (0.99)(0.01) + (0.99)(0.99) = 0.9999 empty are: AB, AC, BC

Calculate the joint, m_1(B) m_2(B) = (.01)(.01) / [1-0.9999] =1

Go back to the theory

bpa, *Belief* and *Plausibility measures – inverse equations*

(6): $$m(A) = \sum_{B|B\subseteq A} (-1)^{|A-B|} Bel(B)$$

- where $|A - B|$ is the difference of the cardinality of the two sets

(7): $Pl(A) = 1 - Bel(\overline{A})$ where $\bar{A}$ is the classical complement of A

- Eq. (7) comes from the fact: all basic assignments must sum to 1

(8): $$Bel(\overline{A}) = \sum_{B|B\subseteq \overline{A}} m(B) = \sum_{B|B\cap A=\varnothing} m(B)$$

(9): $$\sum_{B|B\cap A\neq\varnothing} m(B) = 1 - \sum_{B|B\cap A=\varnothing} m(B)$$

- From the definitions of *Belief* and *Plausibility*, it follows that *Pl*(*A*) = 1- *Bel*(*A*)
- A consequence of eq. 6 and 7: given one of (*m*(*A*), *Bel*(*A*), *Pl*(*A*)) it is possible to derive the values of the other two measures

The relation to precise PT

- Probability of an event lies within the lower and upper bounds of *Belief* and *Plausibility, r*espectively

(10): $Bel(A) = P(A) = Pl(A)$

- The probability is uniquely determined if $Bel\ (A) = Pl(A)$
  - which corresponds to classical probability
  - all probabilities, for all subsets *A* of the universal set *X* are uniquely determined
- Otherwise:
  - *Bel (A)* and *Pl(A)* may be viewed as lower and upper bounds on probabilities, respectively
  - the actual probability is contained in the interval described by the bounds
- Upper and lower probabilities derived by the other frameworks in generalized information theory can *not* be directly interpreted as *Belief* and *Plausibility* functions

Rules for the combination of evidence (1)

- Familiar examples of aggregation techniques: arithmetic averages, geometric averages, harmonic averages, maximum values, and minimum values
- Combination rules are the special types of aggregation methods for data obtained from *multiple* sources
- D-S theory is based on the assumption that the sources are *independent*
- These rules can potentially occupy a continuum between conjunction (AND-based on set intersection) and disjunction (OR-based on set union)
- If *all* sources are reliable, a conjunctive operation is appropriate ($A \cap B \cap C$…)
- If there is one reliable source among many, we can use of a disjunctive combination operation (A or B or C …)
- Many combination operations lie between these above two extremes
- These three types of combinations are described as *conjunctive pooling*, *disjunctive pooling*, and *tradeoff* (many ways to achieve tradeoff between $A \cap B$ and $A \cup B$)
- There are multiple operators available in each category of pooling by which a corpus of data can be combined

Rules for the combination of evidence (2)

- Initial Dempster rule – a generalization of Bayes' rule (AND operators)
- To employ the combination rules in an application, we must understand how conflict should be treated in that particular application context
- Modified Dempster rules represents the degree of conflict in the final result
- Four modified Dempster rules: Yager's; Inagaki's unified combination; Zhang's center combination; and Dubois and Prade's disjunctive pooling rule
- Three types of averages will be considered:
  - discount and combine;
  - convolutive averaging; and
  - mixing
- All of the combination rules will be considered relative to four algebraic properties:
  - commutativity, A * B = B * A
  - idempotence, A * A = A;
  - continuity, A * B ≈ A¢ * B, where A¢ ≈ A (A¢ is very close to A);
  - associativity, A * (B * C) = (A * B) * C; where * denotes the combination operation

Dempster Rule of Combination (1)

- *Belief* and *Plausibility* are derived from the combined basic assignments
- Dempster's rule combines multiple belief functions through their basic probability assignments (m)
- Belief functions are defined on the same frame of discernment, but are based on *independent* arguments or bodies of evidence
- The Dempster rule of combination is purely a conjunctive operation AND
- The combination rule is based on conjunctive pooled evidence
- Combination (joint m_{12}) is calculated from the aggregation of two bpa's m_1 and m_2 in the following manner:

(11): $$m_{12}(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - K}$$ when $A \neq \varnothing$ (12): $m_{12}(\varnothing) = 0$, where

(13): $$K = \sum_{B \cap C = \varnothing} m_1(B) m_2(C)$$ K represents basic probability mass associated with conflict

Dempster Rule of Combination (2)

- This rule is commutative, associative, but not idempotent or continuous
- 1-*K* is a normalization factor, it *completely* ignores conflict and attributes any probability mass associated with conflict to the null set
- It yields counterintuitive results in the face of significant conflict
- The problem with conflicting evidence and Dempster's rule was originally pointed out by Lotfi Zadeh
- Example of erroneous results:
  - Physicians A says that the patient's has either meningitis with a probability 0.99, or a brain tumor, prob. 0.01
  - Physicians B agrees with A that the probability of brain tumor is 0.01, but believes that it is the probability of concussion rather than meningitis that is 0.99
  - Using the values to calculate the *m* (brain tumor) with Dempster's rule, we find that *m*(brain tumor) = *Bel* (brain tumor) = 1
- Clearly, this rule of combination yields a result that implies complete support for a diagnosis that both physicians considered to be very unlikely
- Important are: the level or degree of conflict and the relevance of conflict

Discount+Combine Method – tradeoff methods, Shafer 1976

- One can discount the sources first, and then combine the resulting functions with Dempster's rule (or an alternative rule) using a discounting function
- Discounting function must account for the absolute reliability of the sources
- Shafer applies the discounting function to each specified *Belief*
- Let $1-\alpha_i$ be the degree of reliability attributable to a particular belief function, *A* (Shafer calls this a degree of trust), where $0 \leq \alpha_i \leq 1$ and *i* is an index used to specify the particular discounting function associated with a particular belief measure
- $Bel^{\alpha i}(A)$ represents the discounted belief function defined by:

(14): $$Bel^{\alpha_i}(A) = (1-\alpha_i)Bel(A)$$

- Shafer averages all the belief functions associated with set A to obtain an average of *n* Bel, denoted by $\overline{Bel}$

(15): $$\overline{Bel}(A) = \frac{1}{n}(Bel^{\alpha}{}_1(A) + \ldots + Bel^{\alpha}{}_n(A))$$ for all subsets A of the universal set X

Yager's Modified Dempster's Rule (1)

- Yager (1987):
  - important feature is the ability to update an already combined structure when new information becomes available
  - in many cases a non-associative operator is necessary for combination (a familiar example is the arithmetic average, it is not associative cannot be updated)
- Arithmetic average can be updated by adding the new data point to the sum of the pre-existing data points and dividing by the total number of data points
- This is the concept of a ***quasi-associative*** operator introduced by Yager
- Quasi-associativity: operator can be broken down into associative suboperations
- Yager develops a general framework to look at combination rules where associative operators are a proper subset
- Yager starts with an important distinction between the basic probability mass assignment (*m*) and the ***ground probability mass assignment*** (designated by ***q***)
- The major differences between the bpa and the *ground probability assignment* are in the normalization factor and the mass attributed to the universal set
- The combined *ground probability assignment* is defined in eq. 16

Yager's Modified Dempster's Rule (2)

(16): $q(A) = \sum_{B \cap C = A} m_1(B) m_2(C)$ where A is the intersection of subsets B and C (both in the power set P (X)) (no normalization factor)

- The combined structure $q(A)$ can be used to include any number of pieces of evidence
- Assume:
- $m_1, m_2, \ldots m_n$ – the basic probability assignments for *n* belief structures
- F_i – the set of focal elements associated with the *i*th belief structure (m_i) which are subsets of the universal set *X*
- A_i represents an element of the focal set
- The combination of *n* basic probability assignment structures is:
- (17): $$q(A) = \sum_{\cap_{i=1}^{n} A_i = A} m_1(A_1) m_2(A_2) \ldots m_n(A_n)$$

Yager's Modified Dempster's Rule (3)

- Updating $q(A)$ based on new evidence (eq.16), then converting the ground probability assignments to basic probability assignments - eqs. 19-21
- Absence of the normalization (1-K), Yager circumvents normalization by allowing the ground probability mass assignment of the null set to be greater than 0, i.e.

(18): $q(\emptyset) \geq 0$, $q(\emptyset)$ is calculated exactly as Dempster's K (eq.13)

- Yager adds the the conflict value represented by $q(\emptyset)$ to the ground probability assignment of the universal set, $q(X)$, to yield the conversion of the ground probabilities to the basic probability assignment of the universal set $m^Y(X)$:

(19): $m^Y(X) = q(X) + q(\emptyset)$ Yager's rule does not change the evidence by normalizing out the conflict

Ground probability assignment functions

- Yager's rule of combination yields the same result as Dempster's rule when conflict is equal to zero, (K= 0 or $q(\emptyset)$ = 0)
- The basic algebraic properties that this rule satisfies is commutativity and quasi-associativity, but not idempotence or continuity
- The ground probability assignment functions (q) for the null set, Ø, and an arbitrary set A, are converted to the basic probability assignment function associated with this Yager's rule (m^Y):

(20): $m^Y(\emptyset) = 0$; (21): $m^Y(A) = q(A)$

- The bpa associated with Yager's rule (m^Y) are not the same as with Dempster's rule (m):
- (22): $m(\emptyset)=0$; (23): $m(X) = \dfrac{q(X)}{1 - q(\emptyset)}$ (24): $m(A) = \dfrac{q(A)}{1 - q(\emptyset)}$

Summary of Yager's rule of combination

- The introduction of the general notion of quasi-associative operators and the expansion of the theoretical basis for the combination and updating of evidence where the associative operators are a proper subset of the quasi-associative operators
- The introduction of the *ground probability assignment* functions (q) and their relation to the basic probability assignments (m^Y) associated with Yager's rule and the bpa (m) associated with Dempster's rule
- The rule does not filter or change the evidence through normalization
- The allocation of conflict to the universal set (X) instead of to the null set ($\emptyset$), thus mass associated with conflict is interpreted as the degree of ignorance

Inagaki's Unified Combination Rule (1)

- Inagaki takes advantage of the function (q) to define a continuous parametrized class of combination operations which subsumes both Dempster's rule and Yager's rule
- Specifically, Inagaki argues that every combination rule can be expressed as:

(25): $m(C) = q(C) + f(C)q(\emptyset)$, where $C \neq \emptyset$; (26): $\sum_{C \subset X, C \neq \emptyset} f(C) = 1$

(27): $f(C) \geq 0$

- From eq. 25, f can be interpreted as a scaling function for $q(\emptyset)$, where the conflict (represented by the parameter k) is defined by:

(28): $k = \dfrac{f(C)}{q(C)}$, for any $C \neq X, \emptyset$

Inagaki's Unified Combination Rule (2)

- Inagaki restricts consideration to the class of combination rules that satisfy the property:

(29): $\frac{m(C)}{m(D)} = \frac{q(C)}{q(D)}$ for any nonempty sets *C* and *D* which are distinct from *X* or Ø

- A result of this restriction: Inagaki's rule applies only to the situations where there is no information regarding the credibility or reliability of the sources
- Applying a weighting factor to the evidence based on some extra knowledge about the credibility of the sources would change the ratio and the equality would not hold

Inagaki's Unified Combination Rule (3)

- From the general eq. 25 and the restriction eq. 26 and the definition of *k* eq. 28, Inagaki derives his unified combination rule denoted by m^U
- (30): $m_k^U(C) = [1 + kq(\emptyset)]\, q(C)$ where $C \neq X, \emptyset$
- (31): $m_k^U(X) = [1 + kq(\emptyset)]\, q(X) + [1 + kq(\emptyset) - k]\, q(\emptyset)$
- (32): $0 \leq k \leq \dfrac{1}{1 - q(\emptyset) - q(X)}$, parameter *k* is used for normalization
- The determination of k is an important step in the implementation of this rule
- Lack of well-justified procedure for determining k (through experimental data, simulation, or the expectations of an expert in the context of a specific application)

Inagaki's Unified Combination Rule (4)

- Inagaki discusses the rules in the context of an application where he demonstrates the values of *Belief* and *Plausibility* as a function of k and the implications on the choice of a safety control policy.
- The value of k directly affects the value of the combined basic probability assignments and will collapse to either Dempster's rule or Yager's rule under certain circumstances
- When $k = 0$, unified combination rule coincides with Yager's rule
- When $k=1/(1 - q(\emptyset))$ the rule corresponds to Dempster's rule
- The parameter k gives rise to an entire parametrized class of possible combination rules that interpolate or extrapolate Dempster's rule

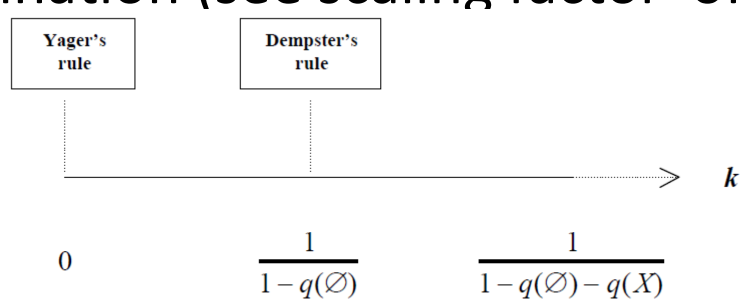
The interpretation of the „extreme rule" of Inagaki's class (1)

- The most extreme rule ("the extra rule" denoted by the parameter k_{ext}) is when k is equal to the upper bound:

(33): $$m^{U}_{k_{ext}}(C) = \frac{1-q(X)}{1-q(X)-q(\emptyset)}\, q(C), \quad \text{for } C \neq X$$

(34): $$m^{U}_{k_{ext}}(X) = q(X)$$

- Both conflict (represented by $q(\emptyset)$) and the degree of ignorance (represented by the probability mass associated with the universal set, $q(X)$) are used to scale the resulting combination (see scaling factor of $q(C)$ in eq. 33



The interpretation of the „extreme rule" of Inagaki's class (2)

- Yager's rule (k=0):
  - assigns conflict to the universal set and does not change the evidence
- Dempster's rule ($k=1/[1-q(\emptyset)]$):
  - tremendously filters the evidence by ignoring all conflict
- Inagaki's extreme rule ($k=1/[1-q(\emptyset)-q(X)]$):
  - also filters the evidence by scaling both conflict and ignorance
  - but the degree of influence of the scaling is determined by the relative values of $q(X)$ and q(Æ)
  - k has the effect of scaling the importance of conflict
- The greater the value of k, the greater the change to the evidence
- A well-justified procedure for the selection of k is as essential step toward implementing this rule in an application

Summary of Inagaki's Unified Combination Rule

- The important contributions of Inagaki's Unified rule of combination:
- The use of Yager's ground functions to develop a parametrized class of combination rules that subsumes both Dempster's rule and Yager's rule
- Inagaki compares and orders three combination rules: Dempster's rule, Yager's rule, and the Inagaki extra rule, in terms of the value of *m* in the context of an application

Zhang's Center Combination Rule (1)

- Zhang points out that Dempster's rule fails to consider how focal elements intersect
- He introduces a measure of the intersection of two sets *A* and *B* assuming finite sets
- It is defined as the ratio of the cardinality of the intersection of two sets divided by the product of the cardinality of the individual sets
- Zhang denotes this relation with *r*(*A*,*B*):

(36): $$r(A,B) = \frac{|A \cap B|}{|A||B|} = \frac{|C|}{|A||B|}$$ where $A \cap B = C$

- The resulting combination rule scales the products of the basic probability assignments of the intersecting sets ($A \cap B = C$) by using a measure of intersection, *r*(*A*,*B*)

Zhang's Center Combination Rule (2)

- The above is repeated for every intersecting pair that yields *C*
- The scaled products of the masses for all pairs whose intersection equals *C* are summed and multiplied by a factor *k*
- In this case, *k* is a renormalization factor that is independent of *C*, m_1, and m_2
- This renormalization factor provides that the sum of the basic assignments to add to 1

(37): $$m(C) = k \sum_{A \cap B = C} \left[\frac{|C|}{|A||B|} m_1(A) m_2(B)\right]$$

- The case where $|C| = |A||B|$, this rule will correspond to the Dempster rule

Zhang's Center Combination Rule (3)

- Many combination rules could be devised in the spirit of Zhang's center combination rule by defining a reasonable measure $r(A,B)$
- Example: by dividing the cardinality of intersection of *A* and *B* by the cardinality of the union of sets *A* and *B*
- This particular rule is commutative but not idempotent, continuous, or associative
- The important contributions of Zhang's work:
  - The two frame interpretation of Dempster-Shafer theory
  - The introduction of a measure of intersection of two sets (r(A,B)) based on cardinality
  - The center combination rule based on a measure of intersection of two sets that could be modified by any other reasonable measure of intersection

Dubois and Prade's Disjunctive Consensus Rule (1)

- They take a set-theoretic view of a body of evidence to form their disjunctive consensus rule
- They define the union of the basic probability assignments $m_1 \cup m_2$ (denoted by $m_\cup(C)$) by extending the set-theoretic union:

(38): $$m_\cup(C) = \sum_{A \cup B = C} m_1(A) m_2(B)$$ for all A of the power set X

- The union does not generate any conflict and does not reject any of the information asserted by the sources, so no normalization procedure is required

Dubois and Prade's Disjunctive Consensus Rule (2)

- The drawback is that it may yield a more imprecise result than desirable
- The union can be more easily performed via the belief measure:
- Let $Bel1 \cup Bel2$ be the belief measure associated with $m_1 \cup m_2$
- Then for every subset A of the universal set X:

(39): $Bel_1(A) \cup Bel_2(A) = Bel_1(A)\ Bel_2(A)$

- The disjunctive pooling operation is commutative, associative, but not idempotent

Mixing or averaging (1)

- Mixing (or *p*-averaging or averaging) is a generalization of averaging for probability distributions
- This describes the frequency of different values within an interval of possible values in the continuous case or in the discrete case, the possible simple events
- The formula for the "mixing" combination rule is:

(40): $$m_{1\ldots n}(A) = \frac{1}{n}\sum_{i=1}^{n} w_i m_i(A)$$ where m_i's are the bpa's for the belief structures being aggregated and the w_i's are weights assigned according to the reliability of the sources

Mixing or averaging (2)

- Insofar as averaging of probability distributions via mixing is regarded as a natural method of aggregating probability distributions, it might also be considered as a reasonable approach to employ with Dempster-Shafer structures, and that is why it is considered here
- Like mixing of probability distributions, mixing in Dempster-Shafer theory is idempotent and commutative, also quasi-associative

Convolutive X-Averaging (1)

- Convolutive x-averaging (c-averaging) is a generalization of the average for scalar numbers
- It is given by the formula:

(41): $$m_{12}(A) = \sum_{\frac{B+C}{2}=A} m_1(B) m_2(C)$$

- Like the mixing average, this can be formulated to include any number of bpa's, *n*:

(42): $$m_{1....n}(A) = \sum_{\frac{A_1+...A_n}{n}=A} \prod_{i=1}^{n} m_i(A_i)$$

Convolutive X-Averaging (2)

- Assume: input Dempster-Shafer structures are scalar numbers (i.e., both structures consist of a single element where all mass is at a single point)
- The result of convolutive average operation is a Dempster-Shafer structure all of whose mass is at a single point, the same point one gets by simply averaging the two scalar numbers
- None of the other Dempster-Shafer aggregation rules would give this answer
- Insofar as "averaging" is regarded as a natural method of aggregating disparate pieces of information, it might also be considered as a reasonable approach to employ with Dempster-Shafer structures
- The convolutive average is commutative, not associative, although it is quasi-associative, and not idempotent
- There are some other combination rules

Demonstration of combination rules – Example 1

- Two experts E-1 and E-2 are consulted regarding a system failure
- Failure can be caused by: Component A (C-A), Component B (C-B)or Component C (C-C)
- E-1 believes: failure is due to C-A with probability 0.99 (denoted as m_1(A)) or C-B with probability 0.01 (denoted as m_1(B))
- E-2 believes: failure is due to C-C with probability 0.99 (denoted as m_2(C)) or C_B with probability 0.01 (denoted as m_2(B))
- The distributions can be represented by the following:
- Expert 1:
  - $m_1(A) = 0.99$ (failure due to Component A)
  - $m_1(B) = 0.01$ (failure due to Component B)
- Expert 2:
  - $m_2(B) = 0.01$ (failure due to Component B)
  - $m_2(C) = 0.99$ (failure due to Component C)

Dempster's Rule – Example 1 (1)

- The combination of the masses associated with the experts is summarized in Table 1

			Expert 1			
			A	**B**	**C**	Failure Cause
			0.99	0.01	0	m_1
Expert 2	Failure Cause	m_2				
	A	0	$m_1(A)\ m_2(A) = 0$	$m_1(B)\ m_2(A) = 0$	$m_1(C)\ m_2(A) = 0$	
	B	0.01	$m_1(A)\ m_2(B) = 0.0099$	$m_1(B)\ m_2(B) = 0.0001$	$m_1(C)\ m_2(B) = 0$	
	C	0.99	$m_1(A)\ m_2(C) = 0.9801$	$m_1(B)\ m_2(C) = 0.0099$	$m_1(C)\ m_2(C) = 0$	

- Using Equations 11-13:

Dempster's Rule – Example 1 (2)

1. To calculate the combined basic probability assignment for a particular cell, simply multiply the masses from the associated column and row
2. Where the intersection is nonempty, the masses for a particular set from each source are multiplied, e.g., $m_{12}(B) = (0.01)(0.01) = 0.0001$
3. Where the intersection is empty, this represents conflicting evidence and should be calculated as well; for the empty intersection of the two sets A and C associate with E-1 and E-2, respectively, there is a mass associated with it: $m_1(A)\ m_2(C)=(0.99)(0.99) =(0.9801)$
4. Then sum the masses for all sets and the conflict
5. The only nonzero value is for the combination of B, $m12(B) = 0.0001$; in this example there is only one intersection that yields B, but in a more complicated example it is possible to find more intersections to yield B
6. For K, there are three cells that contribute to conflict represented by empty intersections; using Eq.13, K = (0.99)(0.01) + (0.99)(0.01) + (0.99)(0.99) = 0.9999
7. Using Equation 11, calculate the joint, $m_1(B)\ m_2(B) = (.01)(.01) / [1-0.9999] = 1$

Dempster's Rule – Example 1 (3)

- Though there is highly conflicting evidence, the basic probability assignment for the failure of Component B is 1, which corresponds to a $Bel(B) = 1$
- This is the result of normalizing the masses to exclude those associated with conflict
- This points to the inconsistency when Dempster's rule is used in the circumstances of significant relevant conflict that was pointed out by Zadeh

Yager's Rule – Example 1 (1)

- Yager's rule yields the almost the same matrix as with Dempster's rule
- However, there are some important exceptions in the nomenclature and eventually the allocation of conflict:
- 1. Instead of basic probability assignments (*m*), Yager calls these ground probability assignments (*q*)
- 2. Instead of using *K* to represent the conflict, Yager uses the $q(\emptyset)$ which is calculated in the exact same way as *K*. (Eq. 13)
- Using Eq. 16, the combination is calculated: $q12(B) = m_{12}(B) =$ (.01)(.01) = .0001
- Here the combination is not normalized by the factor (1-*K*)

Yager's Rule – Example 1 (2)

- Yager converts the ground probability assignments (*q*) to the basic probability assignments (*m*), the mass for a particular joint remains the same and the mass associated with conflict is attributed to the universal set *X* that represents the degree of ignorance (or lack of agreement)
- So in this case the *m*(*X*) is 0.9999
- To convert the basic probability assignment to the lower bound *Bel*, the *Bel*(*B*) is equal to the *m*(*B*) (*Bel* (*B*) = .0001), as this is the only set that satisfies the criteria for *Belief* (*B* Í*B*)
- This approach results in a significant reduction of the value for *Belief* and a large expansion of *Plausibility*
- Note that the value of *Belief* is substantially smaller than either the experts' estimates would yield individually and in such a case, this may be counterintuitive

Inagaki's Rule – Example 1 (1)

- The matrix is calculated in the same manner as in case of the Dempster rule
- Inagaki uses the ground probability functions similar to Yager
- Ultimately, the value of $m_{12}(B)$ obtained by Inagaki's rule depends on the value of k which is now a parameter
- The literature suggests that the value of k should be determined experimentally or by expert
- Figure 6 demonstrates the behavior of the Inagaki combination as a function of the value of k for this problem

m_{12} as a function of paremeter k – *Inagaki rule*

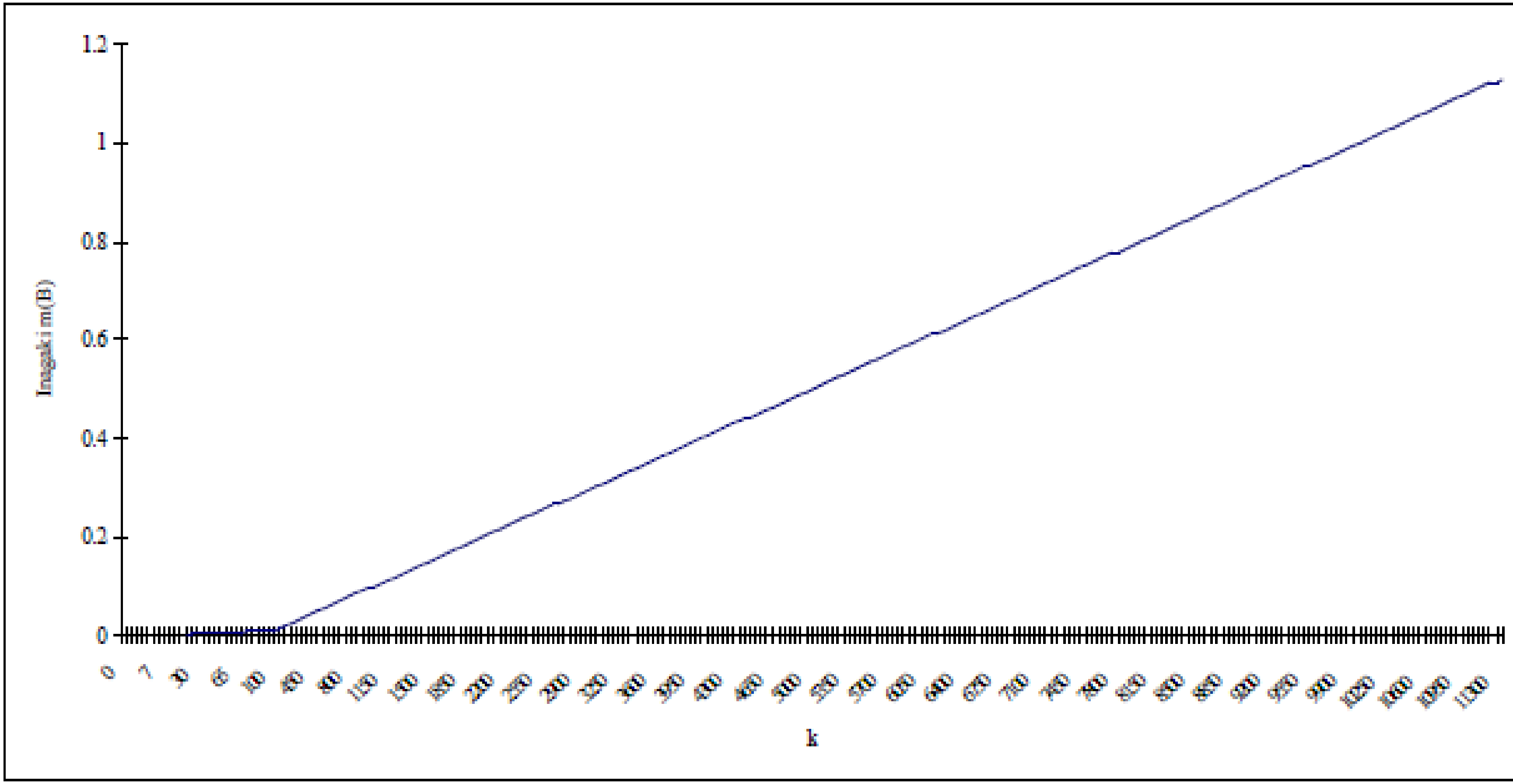


Figure 6: The value of $m_{12}(B)$ as a function of k in Inagaki's rule

Inagaki's combination – discussion

- When $k = 0$, Inagaki's combination will obtain the same result as Yager's ($m_{12}(B) = 0.0001$)
- When $k = 1/(1- q(\emptyset)) = 1/(1 – 0.9999) = 10000$, Inagaki's rule corresponds to Dempster's rule ($m_{12}(B) = 1$)
- Because there is no mass associated with the universal set $q(X)$, in this case, Inagaki's extra rule is the same as Dempster's rule
- Although, the calculation can be extended beyond Dempster's rule, any value for the combination greater than 1 does not make sense because sums of all masses must be equal to 1
- Corresponding to the increasing value of k, is the increase in the filtering of the evidence

Zhang’s Rule – Example 1

- In addition to calculating the product of the masses (Table 1), we must calculate the measure of intersection based on the cardinality of the sets
- The cardinality of each of the sets A, B, and C is 1
- In this case we find that the only nonzero intersection of the sets is set B obtained from the evaluation of B by both Experts 1 and 2
- Since $|B|=|B||B|$, we find that the Zhang combination corresponds to the Dempster combination
- This points to two problems with Zhang’s measure of intersection:
- 1. The equivalence with Dempster’s rule when the cardinality is 1 for all relevant sets or when the $|C|=|A||B|$ in the circumstance of conflicting evidence (this should not pose a problem if there is no significant conflict)
- 2. If the cardinality of B was greater than 1, even completely overlapping sets will be scaled

Mixing – Example 1

- The formulation for mixing corresponds to the sum of $m_1(B)(1/2)$ and $m_2(B)(1/2)$
- From Equation 40:
- $m_{12}(A) = (1/2)(0.99) = 0.445$
- $m_{12}(B) = (1/2)(0.01) + (1/2)(0.01) = 0.01$
- $m_{12}(C) = (1/2)(0.99) = 0.445$

Dubois and Prade's Disjunctive Consensus Pooling – Example 1

- The unions of multiple sets based on the calculations from Table 1 that can be summarized in Table 2

Union	$m_{\cup}$	Linguistic Interpretation
$A \cup A$	0	Failure of Component A
$A \cup B$	0.0099	Failure of Component A or B
$A \cup C$	0.9801	Failure of Component A or C
$B \cup B$	0.0001	Failure of Component B
$B \cup C$	0.0099	Failure of Component B or C
$C \cup C$	0	Failure of Component C
$A \cup B \cup C$	1	Failure of Component A or B or C

Table 2: Unions obtained by Disjunctive Consensus Pooling

Instead of summary

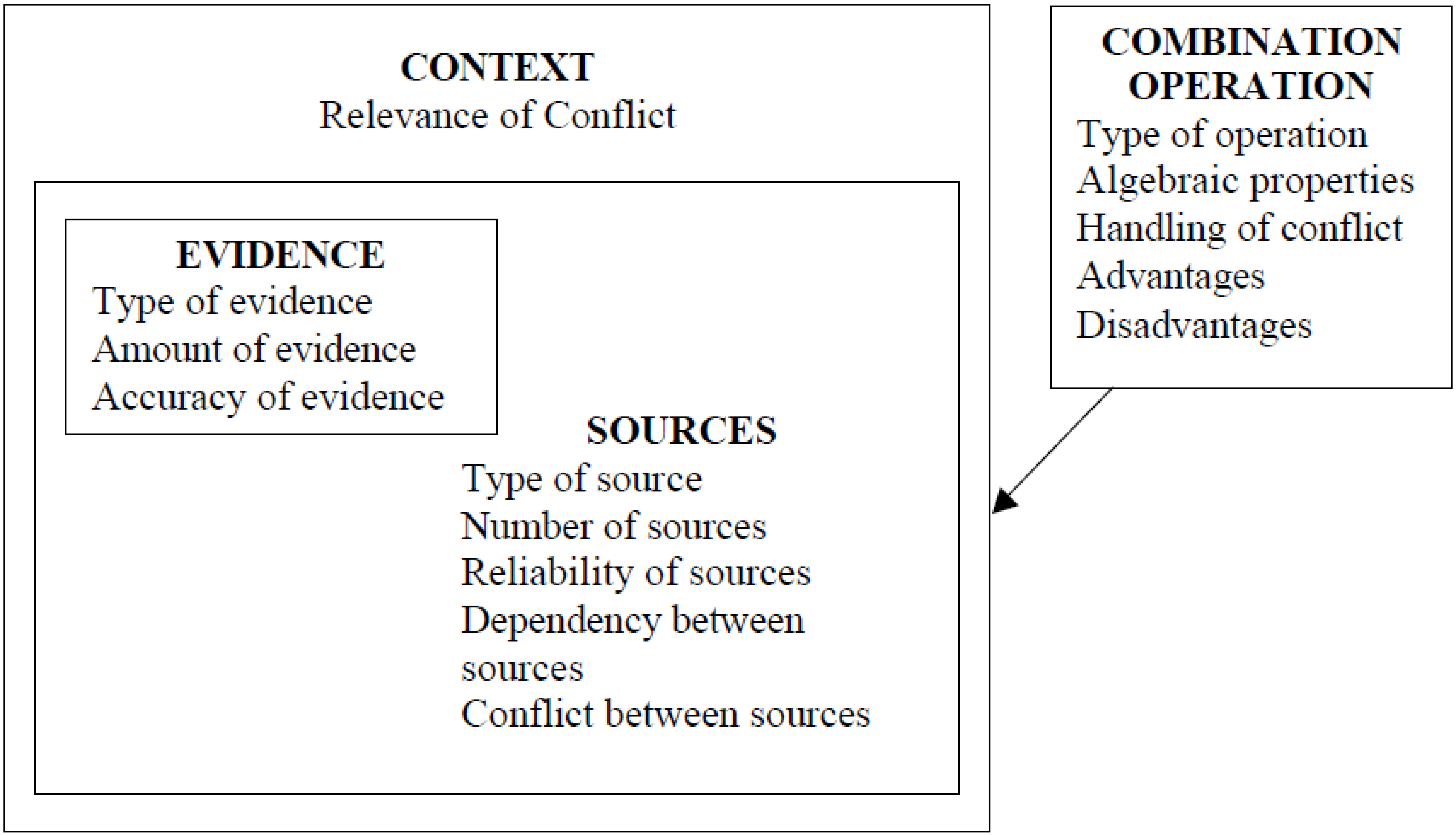


Figure 25: Important Issues in the Combination of Evidence

Summary

- When minimal conflict or irrelevant conflict and all of the sources can be considered reliable, a Dempster combination might be justified
- When there is a situation of no conflict, two of the Dempster-type rules (Yager, Inagaki (k=0, k=1), provide the same answer as Dempster's rule
- As the level of relevant conflict increases, Yager's rule might more appropriate as the conflict is not ignored
- Important issues discussed in literature with respect to the combination of evidence in Dempster-Shafer theory are linked to the characterization of conflict
- This as the most critical concern for the specific selection of a combination operation
- Specifically, what is the degree and contextual relevance of conflict and how is this handled by a particular combination rule

Thank you for your attention, especially for your activity

Each student must have remote access to the WRUST library