Introduction to Dempster-Shafer Theory of Belief Functions

David Mercier & Eric Lefevre

Université d'Artois

Tutorial PFIA 2022 - Lundi 27 juin 2022

UNIVERSITÉ D'ARTOIS

- Dempster-Shafer (DS) theory of belief functions is a flexible mathematical framework for dealing with imperfect information.

- Dempster-Shafer (DS) theory of belief functions is a flexible mathematical framework for dealing with imperfect information.
- Also known as Dempster-Shafer (DS) theory or Evidence theory or Belief Function Theory.

- Dempster-Shafer (DS) theory of belief functions is a flexible mathematical framework for dealing with imperfect information.
- Also known as Dempster-Shafer (DS) theory or Evidence theory or Belief Function Theory.
- It was introduced by A. P. Dempster in the 1960's for statistical inference, and developed by G. Shafer in the late 1970's into a general theory for reasoning under uncertainty.

UNIVERSITÉ D'ARTOIS

- Dempster-Shafer (DS) theory of belief functions is a flexible mathematical framework for dealing with imperfect information.
- Also known as Dempster-Shafer (DS) theory or Evidence theory or Belief Function Theory.
- It was introduced by A. P. Dempster in the 1960's for statistical inference, and developed by G. Shafer in the late 1970's into a general theory for reasoning under uncertainty.
- DS encompasses probability theory and set-membership approaches as special cases.

UNIVERSITÉ D'ARTOIS

- Dempster-Shafer (DS) theory of belief functions is a flexible mathematical framework for dealing with imperfect information.
- Also known as Dempster-Shafer (DS) theory or Evidence theory or Belief Function Theory.
- It was introduced by A. P. Dempster in the 1960's for statistical inference, and developed by G. Shafer in the late 1970's into a general theory for reasoning under uncertainty.
- DS encompasses probability theory and set-membership approaches as special cases.
- It is very general: many applications in AI (expert systems, machine learning), engineering (information fusion, uncertainty quantification, risk analysis), statistical inferences, etc.
- Evidential reasoning can be applied to very large problems.

UNIVERSITÉ D'ARTOIS

- **Uncertainty**

  Example: "I think John is 1.8m tall"
  In this case, the piece of information John is 1.8m tall is precise but uncertain

- **Imprecision**

  Example: "John is between 1.7m and 1.9m tall"
  In this case, the piece of information "John is between 1.7m and 1.9m tall" is certain but imprecise

UNIVERSITÉ D'ARTOIS

- **Uncertainty**

  Example: "I think John is 1.8m tall"
  In this case, the piece of information John is 1.8m tall is precise but uncertain
- **Imprecision**

  Example: "John is between 1.7m and 1.9m tall"
  In this case, the piece of information "John is between 1.7m and 1.9m tall" is certain but imprecise
- **Imprecision and uncertainty**

  Example: "I think John is between 1.7m and 1.9m tall"
  In this case, the piece of information "John is between 1.7m and 1.9m tall" is both uncertain and imprecise

UNIVERSITÉ D'ARTOIS

- **Uncertainty** $\Rightarrow$ Classicaly tackled with probabilities

  Example: "I think John is 1.8m tall"

  In this case, the piece of information John is 1.8m tall is precise but uncertain
- **Imprecision**

  Example: "John is between 1.7m and 1.9m tall"

  In this case, the piece of information "John is between 1.7m and 1.9m tall" is certain but imprecise
- **Imprecision and uncertainty**

  Example: "I think John is between 1.7m and 1.9m tall"

  In this case, the piece of information "John is between 1.7m and 1.9m tall" is both uncertain and imprecise

- **Uncertainty** ⇒ Classicaly tackled with probabilities

  Example: "I think John is 1.8m tall"

  In this case, the piece of information John is 1.8m tall is precise but uncertain

- **Imprecision** ⇒ Classicaly tackled with sets

  Example: "John is between 1.7m and 1.9m tall"

  In this case, the piece of information "John is between 1.7m and 1.9m tall" is certain but imprecise

- **Imprecision and uncertainty**

  Example: "I think John is between 1.7m and 1.9m tall"

  In this case, the piece of information "John is between 1.7m and 1.9m tall" is both uncertain and imprecise

UNIVERSITÉ D'ARTOIS

Aleatory uncertainty (Randomness)

vs

Epistemic uncertainty (Lack of knowledge)

Introduction

Difficulties to represent ignorance with probabilities

"Le tiercé c'est mon dada" (O. Sharif)

- Consider a **horse race** with three horses h_1, h_2 and h_3

Expert 1: "All three horses have an equal chance of winning (same level)"

Model: $p(\{h_1\}) = p(\{h_2\}) = p(\{h_3\}) = \frac{1}{3}$

Expert 2: "I have no idea (complete ignorance)"

Model: $p(\{h_1\}) = p(\{h_2\}) = p(\{h_3\}) = \frac{1}{3}$

Le Derby d'Epsom (Géricault)

UNIVERSITÉ D'ARTOIS

- Consider a **horse race** with three horses h_1, h_2 and h_3

Expert 1: "All three horses have an equal chance of winning (same level)"

Model: $p(\{h_1\}) = p(\{h_2\}) = p(\{h_3\}) = \frac{1}{3}$

Expert 2: "I have no idea (complete ignorance)"

Model: $p(\{h_1\}) = p(\{h_2\}) = p(\{h_3\}) = \frac{1}{3}$

Le Derby d'Epsom (Géricault)

- Problem: Two distinct pieces of information are modeled identically.
- There is a need for a richer model.

Outline

Outline

UNIVERSITÉ D'ARTOIS

- Let us consider a variable of interest X taking its values into a finite set of hypotheses $\Omega = \{\omega_1, ..., \omega_K\}$ called the universe or the frame of discernment.
  - Example: the horse that will win the race. $\Omega = \{h_1, h_2, h_3\}$
- A piece of information regarding the value ω_0 taken by this variable can be represented using a mass function (MF) m defined as a mapping $m : 2^\Omega \to [0, 1]$ verifying

$$\sum_{A \subseteq \Omega} m^\Omega(A) = 1 \ .$$

- The real $m(A)$ represents the part of belief allocated to the hypothesis that the searched true value ω_0 belongs to A and nothing more.
- A set A s.t. $m(A) > 0$ is called a focal set of m.

UNIVERSITÉ D'ARTOIS

"Si vous avez perdu au tiercé, vengez-vous. Mangez du cheval." (P. Dac)

- Let us consider again the **horse race** example with $\Omega = \{h_1, h_2, h_3\}$

Expert 1: "All three horses have an equal chance of winning (same level)"

Model: $m(\{h_1\}) = m(\{h_2\}) = m(\{h_3\}) = \frac{1}{3}$

Expert 2: "I have no idea (complete ignorance)"

Model: $m(\{h_1, h_2, h_3\}) = 1$

Le Derby d'Epsom (Géricault)

- If the evidence tells us that the truth is in $A \subseteq \Omega$ for sure, then we have a logical or categorical mass function m_A s.t. $m_A(A) = 1$.
- m_Ω represents the total ignorance, it is called the vacuous mass function
- If all focal sets of m are singletons, m is said to be Bayesian. It is equivalent to a probability distribution.
- A mass function can thus be seen as:
  - a generalized set
  - a generalized probability distribution

UNIVERSITÉ D'ARTOIS

- A MF m is in one-to-one correspondence (each function represents the same information) with :
  - a belief function Bel defined for all $A \subseteq \Omega$ by:

$$Bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B),$$

    $Bel(A)$ represents the total degree of belief supporting the fact that $\omega_0 \in A$ (Total support in A)

  - a plausibility function Pl defined for all $A \subseteq \Omega$ by:

$$Pl(A) = \sum_{A \cap B \neq \emptyset} m(B) = Bel(\Omega) - Bel(\overline{A})$$

    with $\overline{A} = \Omega \setminus A$.
    $Pl(A)$ represents the total sum of beliefs that are not in contradiction with A (Consistency with A)

Representation of information

Example

With $\Omega = \{a, b, c\}$, $m(\{a\}) = 0.3$, $m(\{b\}) = 0.4$ and $m(\Omega) = 0.3$

Let us compute $Bel(\{a, b\})$ as an example, we have

$$Bel(\{a, b\}) = \sum_{B:\emptyset \neq B \subseteq \{a,b\}} m(B)$$
$$= m(\{a\}) + m(\{b\}) + m(\{a, b\}) = .7$$

For Pl, let us compute $Pl(\{a, b\})$ as an example as well

$$Pl(\{a, b\}) = \sum_{B:B \cap \{a,b\} \neq \emptyset}$$
$$= m(\{a\}) + m(\{b\}) + m(\{a, b\}) + m(\{a, c\}) + m(\{b, c\}) + m(\Omega)$$
$$= 1$$

UNIVERSITÉ D'ARTOIS

With $\Omega = \{a, b, c\}$, $m(\{a\}) = 0.3$, $m(\{b\}) = 0.4$ and $m(\Omega) = 0.3$

bin.order		*m*	*Bel*	*Pl*
000	$\emptyset$			
001	a	.3		
010	b	.4		
011	a, b			
100	c			
101	a, c			
110	b, c			
111	a, b, c	.3		

```
> library(ibelief)

> m=c(0,.3,.4,0,0,0,0,.3)

> pl=mtopl(m)

> bel=mtobel(m)
```


UNIVERSITÉ D'ARTOIS

Example: With $\Omega = \{a, b, c\}$, $m(\{a\}) = 0.3$, $m(\{b\}) = 0.4$ and $m(\Omega) = 0.3$

bin.order		*m*	*Bel*	*Pl*
000	$\emptyset$			
001	a	.3	.3	.6
010	b	.4	.4	.7
011	a, b		.7	1
100	c			.3
101	a, c		.3	.6
110	b, c		.4	.7
111	a, b, c	.3	1	1

UNIVERSITÉ D'ARTOIS

- $Bel(\emptyset) = Pl(\emptyset) = 0$
- $Bel(\Omega) \leq 1$ and $Pl(\Omega) \leq 1$ (as $m(\emptyset)$ could be positive)
- $Bel(A) \leq Pl(A)$
- $Pl(A) = 1 - Bel(\overline{A})$
- If m is Bayesian (i.e. all focal elements are singletons) then $Bel = Pl$ is a probability measure

- The uncertainty about a proposition A is represented by two numbers: $Bel(A)$ and $Pl(A)$, with $Bel(A) \leq Pl(A)$.
- The intervals $[Bel(A), Pl(A)]$ have maximum length when m is vacuous ($m = m_\Omega$).
  - In this case: $Bel(A) = 0$ for all $A \neq \Omega$ and $Pl(A) = 1$ for all $A \neq \emptyset$
- The intervals $[Bel(A), Pl(A)]$ have minimum length when m is Bayesian.
  - In this case, for all A:

$$Bel(A) = Pl(A) = \sum_{\omega \in A} m(\omega)$$

and Bel and Pl are probability measures.

- If m has its focal elements nested ($A_1 \subset A_2 \subset \ldots \subset A_n$, with A_i, $i \in \{1, 2, \ldots, n\}$ the focal elements of m), m is said to be consonant.
- In this case, for all $A \subseteq \Omega$, $B \subseteq \Omega$:

$$Bel(A \cap B) = min(Bel(A), Bel(B))$$

and

$$Pl(A \cup B) = max(Pl(A), Pl(B))$$

meaning Pl is a possibility measure and Bel is its dual necessity measure.

Outline

Representation of information

Combining information

Decision making

Conclusion

Cunjunctive Rule of Combination

- Two mass functions m_1 and m_2 from two reliable and distinct sources of information can be combined using the conjunctive rule of combination (CRC) defined by:

$$(m_1 \circledcap m_2)(A) = m_{1 \circledcap 2}(A) = \sum_{B \cap C = A} m_1(B) \cdot m_2(C), \quad \forall A \subseteq \Omega \,.$$

UNIVERSITÉ D'ARTOIS

Cunjunctive Rule of Combination

Example

With $\Omega = \{a, b, c\}$, let us consider a MF m_1 and another independent MF m_2 s.t.

$$\begin{cases} m_1(\{b\}) & = & .4 \\ m_1(\{a,b\}) & = & .6 \end{cases} \text{ and } \begin{cases} m_2(\{b,c\}) & = & .3 \\ m_2(\{a,b,c\}) & = & .7 \end{cases}$$

CRC	$m_2(\{b,c\}) = .3$	$m_2(\{a,b,c\}) = .7$
$m_1(\{b\}) = .4$	$\{b\} \cap \{b,c\} = \{b\}$ $.4 \times .3 = .12$	$\{b\} \cap \{a,b,c\} = \{b\}$ $.4 \times .7 = .28$
$m_1(\{a,b\}) = .6$	$\{a,b\} \circledcap \{b,c\} = \{b\}$ $.6 \times .3 = .18$	$\{a,b\} \circledcap \{a,b,c\} = \{a,b\}$ $.6 \times .7 = .42$

The CRC $m = m_1 \circledcap m_2$ is given by

- $m(\{b\}) = .4 \times .3 + .4 \times .7 + .6 \times .3 = .58$
- $m(\{a,b\}) = .6 \times .7 = .42$

UNIVERSITÉ D'ARTOIS

Dempster's rule

- **If and only if m_1 and m_2 are two reliable and distinct mass functions** (Axiomatic justifications, see e.g. Smets 2007)
- Dempster's rule := CRC normalized: $m_{1\oplus 2}(\emptyset) = 0$ and

$$(m_1 \oplus m_2)(A) = m_{1\oplus 2}(A) = \frac{1}{1-\kappa} \sum_{B \cap C = A} m_1(B) \cdot m_2(C), \quad \forall A \neq \emptyset$$

with $\kappa = \sum_{B \cap C = \emptyset} m_1(B) \cdot m_2(C)$ (called degree of conflict).

Disjunctive Rule of Combination

- If the sources are distinct but only one of the sources is reliable (and we don't know which one), the disjunctive rule of combination (DRC) defined as follows can be applied:

$$(m_1 \ocircle\!\!\!\!\cup\, m_2)(A) = m_{1 \ocircle\!\!\!\!\cup 2}(A) = \sum_{B \cup C = A} m_1(B) \cdot m_2(C), \quad \forall A \subseteq \Omega \,.$$

UNIVERSITÉ D'ARTOIS

Properties for these rules

- With these rules $\circledcap$, $\oplus$ and $\circledcup$, the order the sources are combined does not change the results
- $m_1 \circledcap m_2 = m_2 \circledcap m_1$ (Commutativity) (Likewise for $\oplus$ and $\circledcup$)
- $(m_1 \circledcap m_2) \circledcap m_3 = m_1 \circledcap (m_2 \circledcap m_3)$ (Associativity) (Likewise for $\oplus$ and $\circledcup$)

Combining information
Example

	m_1	m_2	$m_1 \circledcap m_2$	$m_1 \oplus m_2$	$m_1 \circledcup m_2$
$\emptyset$					
a	.3				
b	.4				
a, b		.5			
c					
a, c		.1			
b, c					
a, b, c	.3	.4			

```
> m1=c(0,.3,.4,0,0,0,0,.3)

> m2=c(0,0,0,.5,0,.1,0,.4)

> mcunjunctive = DST(cbind(m1,m2),1)

> mdempster = DST(cbind(m1,m2),2)

> mdisjunctive = DST(cbind(m1,m2),4)
```

Combining information

Example

	m_1	m_2	$m_1 \circledcap m_2$	$m_1 \oplus m_2$	$m_1 \circledcup m_2$
$\emptyset$			.04		
a	.3		.30	.312	
b	.4		.36	.375	
a, b		.5	.15	.156	.35
c					
a, c		.1	.03	.031	.03
b, c					
a, b, c	.3	.4	.12	.125	.62

UNIVERSITÉ D'ARTOIS

Misconception about Dempster's rule

- Following an old report from Zadeh (1979) - it is still nowadays repeated that "Dempster's rule yields counterintuitive results" (usually used as a justification to introduce new combination rules)
- Zadeh's example: $\Omega = \{a, b, c\}$, two experts reporting:
  - Expert 1: $m_1(\{a\}) = 0.99$, $m_1(\{b\}) = 0.01$ and $m_1(\{c\}) = 0$
  - Expert 2: $m_2(\{a\}) = 0$, $m_2(\{b\}) = 0.01$ and $m_2(\{c\}) = 0.99$
- Then $m_{1 \textcircled{\cup} 2}(b) = 1$, which is claimed to be "counterintuitive" by some authors because both experts considered b as very unlikely.
- But:
  - Both experts are totally reliable.
  - Expert 1 indicates that c is absolutely impossible.
  - Expert 2 indicates that a is absolutely impossible.
- Then b is the only possibility. We are in a situation, which is possible for both experts, where the true answer is b.
- Dempster's rule does produce sound results when used cordance with the axioms, from which it derived.

Discounting

A simple correction example (Shafer, 1976).

Discounting of a mass function (MF) m is defined by (Shafer,1976):

$$\begin{cases} {}^{\alpha}m(A) & = & (1-\alpha)m(A), \quad \forall A \subset \Omega, \\ {}^{\alpha}m(\Omega) & = & (1-\alpha)m(\Omega) + \alpha \ , \end{cases}$$

where $\alpha \in [0,1]$ is the discount rate.

Discounting of a mass function (MF) m is defined by (Shafer,1976):

$$\begin{cases} {}^{\alpha}m(A) &= (1-\alpha)m(A), \quad \forall A \subset \Omega, \\ {}^{\alpha}m(\Omega) &= (1-\alpha)m(\Omega) + \alpha \ , \end{cases}$$

where $\alpha \in [0,1]$ is the discount rate.

Example:

- $\Omega = \{a, b, c\}$
- $m(\{a\}) = .2$, $m(\{b\}) = .4$ and $m(\{a,b\}) = .4$
- With discount rate $\alpha = .2$:

$$\begin{cases} {}^{\alpha}m(\{a\}) &= .8 \times .2 &= .16 \\ {}^{\alpha}m(\{b\}) &= .8 \times .4 &= .32 \\ {}^{\alpha}m(\{a,b\}) &= .8 \times .4 &= .32 \\ {}^{\alpha}m(\Omega) &= .8 \times .0 + .2 &= .20 \end{cases}$$

UNIVERSITÉ D'ARTOIS

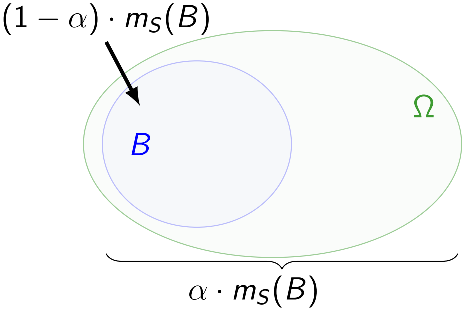
```
> m=c(0,.2,.4,.4,0,0,0,0)

> mdisc = discounting(m,.8)
```

L'argument placé dans cette fonction est $1 - \alpha = .8$ qui est le degré de fiabilité de la source (80% des masses sont gardées dans ce cas)

UNIVERSITÉ D'ARTOIS

For each focal element B of m_S:



- A part $(1-\alpha) \cdot m_S(B)$ remains on B.
- A part $\alpha \cdot m_S(B)$ is transferred to Ω.

UNIVERSITÉ D'ARTOIS

Discounting

Matrix representation (Smets, 2002)

Discounting ${}^{\alpha}m$ is a generalization of m (${}^{\alpha}m \sqsupseteq_s m$):

$$ {}^{\alpha}m(A) = \sum_{B \subseteq \Omega} {}^{\alpha}G(A,B)m(B) \ , $$

with ${}^{\alpha}\mathbf{G}$ a generalisation matrix defined by:

$$ {}^{\alpha}G(A,B) = \begin{cases} 1-\alpha & \text{if } A = B \neq \Omega, \\ \alpha & \text{if } A = \Omega \text{ and } B \subset A, \\ 1 & \text{if } A = B = \Omega \\ 0 & \text{otherwise.} \end{cases} $$

$$ {}^{\alpha}\mathbf{G} = \begin{pmatrix} 1-\alpha & 0 & \dots & 0 & 0 \\ 0 & 1-\alpha & \dots & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \dots & 1-\alpha & 0 \\ \alpha & \alpha & \dots & \alpha & 1 \end{pmatrix} $$

UNIVERSITÉ D'ARTOIS

Discounting

Matrix representation: example

With $\alpha = .2$, $\beta = 1 - \alpha = .8$ and $\Omega = \{a, b, c\}$:

$$
\begin{array}{ll}
 & {}^{\alpha}m \quad = \quad {}^{\alpha}G(A,B) \quad \cdot \quad m \\
\begin{array}{l}
000 : \emptyset \\
001 : \{a\} \\
010 : \{b\} \\
011 : \{a, b\} \\
100 : \{c\} \\
101 : \{a, c\} \\
110 : \{b, c\} \\
111 : \{a, b, c\}
\end{array}
&
\begin{pmatrix} .0 \\ .16 \\ .32 \\ .32 \\ .0 \\ .0 \\ .0 \\ .20 \end{pmatrix}
=
\begin{pmatrix}
\beta & & & & & & & \\
 & \beta & & & & & & \\
 & & \beta & & & & & \\
 & & & \beta & & & & \\
 & & & & \beta & & & \\
 & & & & & \beta & & \\
 & & & & & & \beta & \\
\alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & 1
\end{pmatrix}
\cdot
\begin{pmatrix} .0 \\ .2 \\ .4 \\ .4 \\ .0 \\ .0 \\ .0 \\ .0 \end{pmatrix}
\end{array}
$$

UNIVERSITÉ D'ARTOIS

Outline

UNIVERSITÉ D'ARTOIS

- A way to make a **hard decision** is to choose a decision $d = \omega \in \Omega$ maximizing a probability transform of m, as for example using the Pignistic transform $BetP$ defined by:

$$BetP(\{\omega\}) = \sum_{A \subseteq \Omega, \omega \in A} \frac{m(A)}{|A|\,(1 - m(\emptyset))}, \quad \forall \omega \in \Omega\ .$$

- Example: With $\Omega = \{a, b, c\}$, $m(\{a\}) = 0.3$, $m(\{b\}) = 0.4$ and $m(\Omega) = 0.3$

$$BetP = \begin{cases} \{a\} \mapsto 0.3 + \frac{0.3}{3} = 0.4 \\ \{b\} \mapsto 0.4 + \frac{0.3}{3} = 0.5 \\ \{c\} \mapsto 0.0 + \frac{0.3}{3} = 0.1 \end{cases}$$

UNIVERSITÉ D'ARTOIS

```
> m=c(0,.3,.4,0,0,0,0,.3)
> betp = mtobetp(m)
```


UNIVERSITÉ D'ARTOIS

- A way to make a **partial decision** is to choose a set-valued decision $d = A \subseteq \Omega$ composed of elements of Ω, which are **not dominated according to a preference relation**:

1. The relation of **strong dominance or interval dominance** defined by

$$\omega \succeq_{sd} \omega' \iff Bel(\{\omega\}) \geq Pl(\{\omega'\})$$

2. The relation of **weak dominance** defined by

$$\omega \succeq_{wd} \omega' \iff Bel(\{\omega\}) \geq Bel(\{\omega'\}) \; and \; Pl(\{\omega\}) \geq Pl(\{\omega'\})$$

UNIVERSITÉ D'ARTOIS

Decision making

An example with partial decisions using the strong and weak dominance criteria

Example: With $\Omega = \{a, b, c\}$, $m(\{a\}) = 0.3$, $m(\{b\}) = 0.4$ and $m(\Omega) = 0.3$

$$Bel = \begin{cases} \{a\} \mapsto 0.3 \\ \{b\} \mapsto 0.4 \\ \{c\} \mapsto 0.0 \end{cases} \qquad Pl = \begin{cases} \{a\} \mapsto 0.6 \\ \{b\} \mapsto 0.7 \\ \{c\} \mapsto 0.3 \end{cases}$$

Relation SD	Non-dominated
$Bel(\{b\}) = .4 \not\geq Pl(\{a\}) = .6$ and $Bel(\{c\}) = 0 \not\geq Pl(\{a\}) = .6$	a
$Bel(\{a\}) = .3 \not\geq Pl(\{b\}) = .7$ and $Bel(\{c\}) = 0 \not\geq Pl(\{b\}) = .7$	b
$Bel(\{a\}) = .3 \geq Pl(\{c\}) = .3$ (so $a \succeq_{sd} c$)	-

Conclusion using SD: $d = \{a, b\}$.

Relation WD	Dominated
$Bel(\{b\}) = 0.4 \geq Bel(\{a\}) = 0.3$ and $Pl(\{b\}) = 0.7 \geq Pl(\{a\}) = 0.6$ (so $b \succeq_{wd} a$)	a
$Bel(\{b\}) = 0.4 \geq Bel(\{c\}) = 0.0$ and $Pl(\{b\}) = 0.7 \geq Pl(\{c\}) = 0.3$ (so $b \succeq_{wd} c$)	c

Conclusion using WD: $d = \{b\}$.

UNIVERSITÉ D'ARTOIS

Outline

UNIVERSITÉ D'ARTOIS

Conclusion

- Dempster-Shafer (DS) theory of belief functions is a flexible mathematical framework for dealing with imperfect information.
- It encompasses probability theory and set-membership approaches as special cases.
- Belief functions can be seen as weighted opinions.

References to start learning Belief functions

[1] https://bfasociety.org/#schools

[2] T. Denoeux' talks: https://www.hds.utc.fr/~tdenoeux/dokuwiki/en/talks

[2] T. Denoeux' teaching: https://www.hds.utc.fr/~tdenoeux/dokuwiki/en/teaching

[3] G. Shafer. *A mathematical theory of evidence*. Princeton, N.J.: Princeton University Press, 1976

[4] P. Smets. "The application of the matrix calculus to belief functions". In: *International Journal of Approximate Reasoning* 31.1–2 (2002), pp. 1–30

[5] P. Smets. "Analyzing the Combination of Conflicting Belief Functions". In: *Information Fusion* 8.4 (2007), pp. 387–412

[6] T. Denœux. "Decision-Making with Belief Functions: a Review". In: *International Journal of Approximate Reasoning* 109 (2019), pp. 87–110

UNIVERSITÉ D'ARTOIS

Thank you for your attention.