Biomedical Informatics

Building Medical Expert Systems: The Dempster-Shafer Theory of Evidence

Miguel García Remesal

Department of Artificial Intelligence

mgremesal@fi.upm.es

The Dempster-Shafer Approach

- First described by Arthur Dempster (1960) and extended by Glenn Shafer (1976)
- Useful for systems aimed to medical or industrial diagnosis
- Emulates experts’ reasoning methods:
  - They establish a set of possible hypotheses supported by evidence (symptoms, fails)

Main Features

- Emulate incremental reasoning
- Ignorance can be successfully modeled
- DS assigns subjective probabilities to sets of hypothesis
  - CF-based methods assign subjective probabilities to ***individual hypotheses***

Example

- A physician: “The patient is likely to have renal insufficiency with degree 0.6”

- Expert medical knowledge:
  - Renal insufficiency can be caused either by urine infection or nephritis

- The set [renal_insufficiency, nephritis] is assigned with degree 0.6

- Further analysis are required to be more specific

The Dempster-Shafer Approach

- When reasoning, we require a set Θ of exclusive and exhaustive hypotheses
- Θ is called the ***frame of discernment***
- Hypotheses can be organized as a lattice (partial order)

Example

- $\Theta = \{A, B, C, D\}$
  - A = “measles”
  - B = “chicken pox”
  - C = “mumps”
  - D = “influenza”

- What does $\{A\} \in 2^{\Theta}$ stand for?

- What about $\{A, B\} \in 2^{\Theta}$?

Basic Probability Assignment

- BPAs are subjective probability assignments to sets of hypotheses belonging to 2^{Θ}
  - Must be provided by experts

- Model the credibility of the different sets of hypotheses

- But… ignorance is also modelled!

Basic Probability Assignment

- A BPA *m* can be defined as a function:

$$m : 2^{\Theta} \rightarrow [0,1]$$

$$\sum_{X \in 2^{\Theta}} m(X) = 1$$

- BPA for the empty hypothesis:

$$m(\phi) = 0$$

- All subsets such that *m*(Ø) > 0 are called *focal points*

Basic Probability Assignment

- $m(\Theta)$ is the measure of total belief not assigned to any proper subset of Θ

$$m(\Theta) = 1 - \sum_{X \in 2^{\Theta} - \{\Theta\}} m(X)$$

- Example:
  - m({measles, flu}) = 0.3
    - $m(\Theta) = 1 - 0.3 = 0.7$

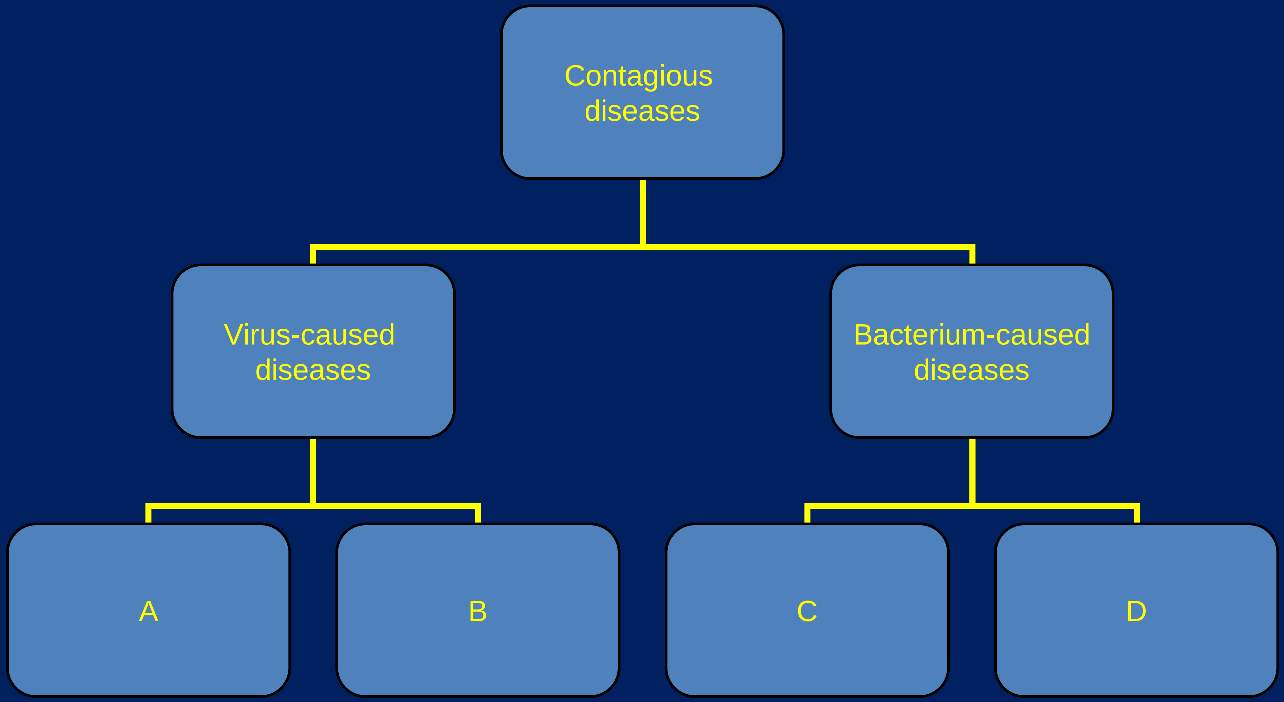
  - *m*({measles, flu}) = 0.3 is not further subdivided among the subsets {measles} and {flu} ***¿WHY?***

Example 1

- Statement:
  - Let us suppose we know that one or more diseases in $\Theta = \{A, B, C, D\}$ is the right diagnosis
  - We don’t know enough to be more specific
- Probability assignment? (i.e. focal points)

Example 2

- Suppose we have the following classification superimposed upon elements $\Theta = \{A, B, C, D\}$



Example 2

- Statement:
  - We know to degree 0.5 that the disease is caused by a virus
- Probability assignment?

Example 3

- Statement:
  – We know the disease is not A to degree 0.4
- Probability assignment?

Evidence Combination

- Diagnostic tasks are incremental and iterative. They involve:
  - Conclusions from gathered evidence
  - Decisions about what kinds of further evidence to gather

- Evidence gathered in one iteration must be combined with evidence gathered in the next one

Dempster's Rule for Evidence Combination

- The D-S theory provides a simple rule to combine evidence provided by two BPAs
- Let m_1 and m_1 be BPAs
- Dempster's rule computes a new *m* value for each A ϵ 2^Θ as follows:

$$m_1 \oplus m_2(A) = \sum_{\substack{A=X \cap Y \\ X,Y \in 2^\Theta}} m_1(X) \cdot m_2(Y)$$

Example

- $\Theta = \{A, B, C, D\}$
- $m_1(\{A, B\}) = 0.4$, $m_1(\Theta) = 0.6$
- $m_2(\{A, B\}) = 0.3$, $m_2(\Theta) = 0.7$

m_3?

BPA Renormalization

- It may turn out the following situation:
  - There are two subsets X, Y such that :
    - X and Y are disjoint
    - $m_1(X) > 0$, $m_2(Y) > 0$ (focal points)
  - This implies that $m_3(\emptyset) \neq 0$
- Problem: remember the definition of BPAs!
  - $m(\emptyset) = 0$
- Solution: renormalization

BPA Renormalization

- If $m(\emptyset) > 0$ it is necessary to carry out a renormalization

- The renormalization is performed as follows:

$$m'(X) = \frac{m(X)}{F_N} = \frac{m(X)}{1 - m(\phi)}$$

$$m(\phi) = 0$$

Example

- $m_1(\{A, B\}) = 0.3$, $m_1(\{A\}) = 0.2$, $m_1(\{\mathrm{D}\}) = 0.1$, $m_1(\Theta) = 0.4$
- $m_2(\{A, B\}) = 0.2$, $m_2(\{A\}) = 0.2$, $m_2(\{\mathrm{C}, \mathrm{D}\}) = 0.2$, $m_2(\Theta) = 0.4$

m_3?

Belief Intervals

- Given a subset, X we use an interval to quantify:
  - Uncertainty
    - Measures the available information (analysis, tests, etc.)
    - The fewer the information the higher the uncertainty
  - Ignorance
    - Measures the imprecision of the uncertainty measure
    - Example: The physician determines that P(X) is between 0.2 and 0.8
      - Thus, the level of ignorance is high (broad interval)

Credibility

- The credibility of a subset X can be defined as the sum of probabilities of all subsets that fully occur in the context of X
- It can be calculated as follows:

$$Cr(X) = \sum_{Y \subseteq X} m(Y)$$

- It can be regarded as a lower bound of the probability of X

Plausibility

- The plausibility of a subset X can be defined as the sum of probabilities of all subsets that occur either fully or partially in the context of X
- It can be calculated as follows:

$$Pl(X) = \sum_{Y \cap X \neq \Phi} m(Y)$$

- It can be regarded as an upper bound of the probability of X

Properties

- *Cr* and *Pl* satisfy (among others) the following properties:

$$Cr(\Phi) = Pl(\Phi) = 0$$

$$Cr(\Theta) = Pl(\Theta) = 1$$

$$Pl(X) \geq Cr(X)$$

$$Cr(A \cup B) \geq Cr(A) + Cr(B) - Cr(A \cap B)$$

Belief Intervals

- The interval [*Cr*(X), *Pl*(X)] reflects the uncertainty and ignorance associated to X
- Two parameters to be taken into account:
  - The actual values of Cr(X) and Pl(X)
    - Measures the uncertainty
  - The size of the interval
    - Measures the ignorance
- When new evidence is added, it is required to update the interval

Belief Intervals

CASE	CONDITION	EXAMPLE [Cr(X), Pl(X)]
IGNORANCE	Cr(X) << Pl(X)	[0, 1]
MAXIMUM INFORMATION	Cr(X) = Pl(X)	[0.6, 0.6]
CERTAINTY	Cr(X) and Pl(X) close to 1	[0.99, 1]
UNCERTAINTY	Cr(X) and Pl(X) close to 0.5	[0.49, 0.50]

- The closer to 0.5 (1), the greater (smaller) the uncertainty
- The broader (narrower) the interval, the greater (smaller) the ignorance
- Note that it is possible to have high uncertainty with zero ignorance → Cr(X) = Pl(X) = 0.5

Example

$m'_3(\Theta) = 0.186$

$m'_3(\{A, B\}) = 0.302$

$m'_3(\{C, D\}) = 0.093$

$m'_3(\{A\}) = 0.349$

$m'_3(\{D\}) = 0.070$

$m'_3(\emptyset) = 0$

Cr, Pl?

Example

	Cr	Pl
Ø	0	0
{A}	0.349	0.837
{B}	0	0.488
{C}	0	0.279
{D}	0.070	0.349
{A, B}	0.651	0.837
{A, C}	0.349	0.930
{A, D}	0.419	1
{B, C}	0	0.581
{B, D}	0.070	0.651
{C, D}	0.163	0.349
{A, B, C}	0.651	0.930
{A, B, D}	0.721	1
{B, C, D}	0.163	0.651
{A, C, D}	0.512	1
Θ	1	1

Calculating Probabilities for Individual Hypotheses

- Probabilities for individual hypotheses are calculated as follows:

$$P(h) = \sum_i \frac{m(F_i)}{|F_i|}$$

where F_i are the focal points that contain the individual hypothesis h

- Note that we are dealing with probabilities:

$$\sum_{\{h_i \subset \theta \| h_i | = 1\}} P(h_i) = 1$$