Using Dempster–Shafer Theory in Data Mining

Malcolm J. Beynon
Cardiff University, UK

INTRODUCTION

The origins of Dempster-Shafer theory (DST) go back to the work by Dempster (1967) who developed a system of upper and lower probabilities. Following this, his student Shafer (1976), in their book "A Mathematical Theory of Evidence" developed Dempster's work, including a more thorough explanation of belief functions, a more general term for DST. In summary, it is a methodology for evidential reasoning, manipulating uncertainty and capable of representing partial knowledge (Haenni & Lehmann, 2002; Kulasekere, Premaratne, Dewasurendra, Shyu, & Bauer, 2004; Scotney & McClean, 2003).

The perception of DST as a generalisation of Bayesian theory (Shafer & Pearl, 1990), identifies its subjective view, simply, the probability of an event indicates the degree to which someone believes it. This is in contrast to the alternative frequentist view, understood through the "Principle of I sufficient reasoning", whereby in a situation of ignorance a Bayesian approach is forced to evenly allocate subjective (additive) probabilities over the frame of discernment. See Cobb and Shenoy (2003) for a contemporary comparison between Bayesian and belief function reasoning.

The development of DST includes analogies to rough set theory (Wu, Leung, & Zhang, 2002) and its operation within neural and fuzzy environments (Binaghi, Gallo, & Madella, 2000; Yang, Chen, & Wu, 2003). Techniques based around belief decision trees (Elouedi, Mellouli, & Smets, 2001), multi-criteria decision making (Beynon, 2002) and non-paramnteric regression (Petit-Renaud & Denœux, 2004), utilise DST to allow analysis in the presence of uncertainty and imprecision. This is demonstrated, in this article, with the 'Classification and Ranking belief Simplex' (CaRBS) technique for object classification, see Beynon (2005a).

BACKGROUND

The terminology inherent with DST starts with a finite set of hypotheses Θ (the frame of discernment). A *basic probability assignment* (*bpa*) or mass value is a function m: $2^{\Theta} \to [0, 1]$ such that $m(\varnothing) = 0$ ($\varnothing$ - the empty set) and

$$\sum_{A \in 2^{\Theta}} m(A) = 1 \text{ (}2^{\Theta}\text{ - the power set of }\Theta\text{).}$$

If the assignment $m(\varnothing) = 0$ is not imposed then the transferable belief model can be adopted (Elouedi, Mellouli, & Smets, 2001; Petit-Renaud & Denœux, 2004). Any $A \in 2^{\Theta}$, for which $m(A)$ is non-zero, is called a focal element and represents the exact belief in the proposition depicted by A. From a single piece of evidence, a set of focal elements and their mass values can be defined a *body of evidence* (BOE).

Based on a BOE, a *belief* measure is a function *Bel*: $2^{\Theta} \to [0, 1]$, defined by,

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

for all $A \subseteq \Theta$. It represents the confidence that a specific proposition lies in A or any subset of A. The *plausibility* measure is a function *Pls*: $2^{\Theta} \to [0, 1]$, defined by,

$$Pls(A) = \sum_{A \cap B \neq \varnothing} m(B),$$

for all $A \subseteq \Theta$. Clearly $Pls(A)$ represents the extent to which we fail to disbelieve A. these measures are directly related to one another, $Bel(A) = 1 - Pls(\neg A)$ and $Pls(A) = 1 - Bel(\neg A)$, where $\neg A$ refers to its compliment 'not A'.

To collate two or more sources of evidence (e.g. $m_1(\cdot)$ and $m_2(\cdot)$), DST provides a method to combine them, using Dempster's rule of combination. If $m_1(\cdot)$ and $m_2(\cdot)$ are two independent BOEs, then the function $(m_1 \oplus m_2)$: $2^{\Theta} \to [0, 1]$, defined by:

$$(m_1 \oplus m_2)(y) = \begin{cases} 0 & y = \varnothing \\ (1-\kappa)^{-1} \sum_{A \cap B = y} m_1(A) m_2(B) & y \neq \varnothing \end{cases}$$

where $\kappa = \sum_{A \cap B = \varnothing} m_1(A) m_2(B)$,

is a mass value with $y \subseteq \Theta$. The term $(1 - \kappa)$, can be interpreted as a measure of conflict between sources. It is important to take this value into account for evaluating the quality of combination: when it is high, the combination may not make sense and possible lead to questionable decisions (Murphy, 2000). One solution to mitigate conflict is to assign noticeable levels of ignorance to all evidence, pertinently the case when low level measurements are considered (Gerig, Welti, Guttman, Colchester, & Szekely, 2000).

To demonstrate the utilization of DST, the example of the murder of Mr. Jones is considered, where the murderer was one of three assassins, Peter, Paul and Mary, so the frame of discernment Θ = {Peter, Paul, Mary}. There are two witnesses. Witness 1, is 80% sure that it was a man, the concomitant BOE, defined $m_1(\cdot)$, includes m_1({Peter, Paul}) = 0.8. Since we know nothing about the remaining mass value it is considered ignorance and allocated to Θ, hence m_1({Peter, Paul, Mary}) = 0.2. Witness 2, is 60% confident that Peter was leaving on a jet plane when the murder occurred, so a BOE defined $m_2(\cdot)$ includes, m_2({Paul, Mary}) = 0.6 and m_2({Peter, Paul, Mary}) = 0.4.

The aggregation of these two sources of information (evidence), using Dempster's combination rule, is based on the intersection and multiplication of focal elements and mass values from the BOEs, $m_1(\cdot)$ and $m_2(\cdot)$. Defining this BOE $m_3(\cdot)$, it can be found; m_3({Paul}) = 0.48, m_3({Peter, Paul}) = 0.32, m_3({Paul, Mary}) = 0.12 and m_3({Peter, Paul, Mary}) = 0.08. This combined evidence has a more spread-out allocation of mass values to varying subsets of the frame of discernment Θ. Further, there is a general reduction in the level of ignorance associated with the combined evidence. In the case of the belief (*Bel*) and plausibility (*Pls*) measures, considering the subset {Peter, Paul}, then Bel_3({Peter, Paul}) = 0.8 and Pls_3({Peter, Paul}) = 1.0. Smets (1990) offers a comparison on a variation of this example with how it would be modelled using traditional probability and the transferable belief model.

A second larger example supposes that the weather in New York at noon tomorrow is to be predicted from the weather today. We assume that it is in exactly one of the three states: dry (D), raining (R) or snowing (S). Hence the frame of discernment is represented by Θ = {D, R, S}. Let us assume that two pieces of evidence have been gathered: *i*) The temperature today is below freezing, and *ii*) The barometric pressure is falling; i.e., a storm is likely. These pieces of evidence are represented by the two BOE, $m_{\text{freeze}}(\cdot)$ and $m_{\text{storm}}(\cdot)$, respectively, and are reported in Table 1.

For each BOE in Table 1, the exact belief (mass) is distributed among the focal elements (excluding $\varnothing$). For $m_{\text{freeze}}(\cdot)$, greater mass is assigned to {S} and {R, S}, for $m_{\text{storm}}(\cdot)$, greater mass is assigned to {R} and {R, S}. Assuming that $m_{\text{freeze}}(\cdot)$ and $m_{\text{storm}}(\cdot)$ represent items of evidence which are independent of one another, a new BOE $m_{\text{both}}(\cdot)$ is given by Dempster's rule of combination; with $m_{\text{both}}(\cdot) = m_{\text{freeze}}(\cdot) \oplus m_{\text{storm}}(\cdot)$, shown in Table 2.

The BOE $m_{both}(\cdot)$ represented in Table 2 has a lower level of local ignorance ($m_{both}(\Theta)$ = 0.0256), than both of the original BOEs, $m_{\text{freeze}}(\cdot)$ and $m_{\text{storm}}(\cdot)$. Amongst the other focal elements, more mass is assigned to {R} and {S}, a consequence of the greater mass assigned to the associated focal elements in the two constituent

Table 1. Mass values and focal elements for $m_{freeze}(\cdot)$ and $m_{storm}(\cdot)$

BOE	$\varnothing$	{D}	{R}	{S}	{D, R}	{D, S}	{R, S}	Θ
$m_{\text{freeze}}(\cdot)$	0.0	0.1	0.1	0.2	0.1	0.1	0.2	0.2
$m_{\text{storm}}(\cdot)$	0.0	0.1	0.2	0.1	0.1	0.1	0.3	0.1

Table 2. Mass values and focal elements for the BOE $m_{both}(\cdot)$

BOE	$\varnothing$	{D}	{R}	{S}	{D, R}	{D, S}	{R, S}	Θ
$m_{\text{freeze}}(\cdot)$	0.0	0.1282	0.2820	0.2820	0.0513	0.0513	0.1795	0.0256

Related Content

Process Mining to Analyze the Behaviour of Specific Users
Laura Maruster (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1589-1597).*
www.irma-international.org/chapter/process-mining-analyze-behaviour-specific/11031

Data Transformation for Normalization
Amitava Mitra (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 566-571).*
www.irma-international.org/chapter/data-transformation-normalization/10877

XML-Enabled Association Analysis
Ling Feng (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 2117-2122).*
www.irma-international.org/chapter/xml-enabled-association-analysis/11112

Symbiotic Data Miner
Kuriakose Athappilly (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1903-1908).*
www.irma-international.org/chapter/symbiotic-data-miner/11079

Semi-Supervised Learning
Tobias Scheffer (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1787-1793).*
www.irma-international.org/chapter/semi-supervised-learning/11060