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 his book "A Mathematical Theory of Evidence" added to Dempster's work, including a more thorough explanation of belief functions. 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 generalization 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 Insufficient Reason", whereby in a situation of ignorance a Bayesian approach is forced to evenly allocate subjective (additive) probabilities over the frame of discernment.

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-parametric regression (Petit-Renaud & Denœux, 2004), utilize DST to allow analysis in the presence of uncertainty and imprecision. This is demonstrated with the CaRBS (Classification and Ranking Belief Simplex) system for object classification, see Beynon (2005).

BACKGROUND

The terminology inherent within DST starts with a finite set of hypotheses Θ (frame of discernment). A *basic probability assignment* (*bpa*) or mass value is a function m: $2^{\Theta} \rightarrow [0,1]$ such that: $m(\varnothing) = 0$ ($\varnothing$ - empty set) and $\sum_{A \in 2^{\Theta}} m(A) = 1$ (2^{Θ} the power set of Θ). If the constraint $m(\varnothing) = 0$ 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 one source 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*: $2^{\Theta} \rightarrow [0,1]$, defined by $Bel(A) = \sum_{B \subseteq A} m(B)$, for all $A \subseteq \Theta$. It represents the confidence that a specific proposition lies in A or any subset of A. A *plausibility* measure is a function *Pls*: $2^{\Theta} \rightarrow [0,1]$, 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 clearly 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 independent BOEs, then the function $m_1 \oplus m_2$: $2^{\Theta} \rightarrow [0,1]$, defined by:

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

where $k = \sum_{A \cap B = \varnothing} m_1(A) m_2(B)$, is a mass value with $y \subseteq \Theta$. The term $(1-k)$, can be interpreted as a measure of conflict between the 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 possibly lead to questionable decisions (Murphy, 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, frame of discernment $\Theta = \{$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 allocated to Θ, $m_1(\{$Peter, Paul, Mary$\}) = 0.2$. Witness 2, is 60% confident that Peter was leaving on a jet plane when the murder occurred, 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, using Dempster's combination rule, is based on the inter-

section 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; $m_3(\{\text{Paul}\}) = 0.48$, $m_3(\{\text{Peter, Paul}\}) = 0.32$, $m_3(\{\text{Paul, Mary}\}) = 0.12$ and $m_3(\{\text{Peter, Paul, Mary}\}) = 0.08$. This combined evidence has a more spread-out allocation of mass values to varying subsets of 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 $Bel_3(\{\text{Peter, Paul}\}) = 0.8$ and $Pls_3(\{\text{Peter, Paul}\}) = 1.0$.

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 three states: dry (D), raining (R) or snowing (S), hence the frame of discernment $\Theta = \{D, R, S\}$. Let us presume 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 BOEs $m_{\text{freeze}}(\cdot)$ 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 evidence which are independent of each other, the BOE from the combination of this evidence, defined $m_{\text{both}}(\cdot)$, is made up of the mass values reported in Table 2.

The BOE $m_{\text{both}}(\cdot)$ represented in Table 2 has a lower level of ignorance ($m_{\text{both}}(\Theta) = 0.0256$), 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 other two BOEs. The other focal elements all exhibit net losses in their mass values. As with the assassin example, measures of belief (*Bel*) and plausibility (*Pls*) could be found, to offer evidence (confidence) on combinations of states representing tomorrow's predicted weather.

This section is closed with some cautionary words still true to this day (Pearl, 1990), 'Some people qualify any model that uses belief functions as Dempster-Shafer. This might be acceptable provided they did not blindly accept the applicability of Dempster's rule of combination (and others). Such critical - and in fact often inappropriate - use of these rules explain most of the errors encountered in the so-called Dempster-Shafer literature'.

MAIN THRUST

This section outlines one of many different methods which utilizes DST, within a data mining environment. The CaRBS system is a data mining technique for the classification (and subsequent prediction) of objects to a given hypothesis (*x*) and its compliment ($\neg x$), using a series of characteristic values (Beynon, 2005). The rudiments of CaRBS are based on DST, and with two exhaustive outcomes works on a binary frame of discernment (BFOD). Subsequently, the aim of CaRBS is to construct a BOE for each characteristic value in the evidential support for the classification of an object to {x}, {¬x} and concomitant ignorance {x, ¬x} (considered as uncertainty in subjective judgements), see Figure 1.

In Figure 1, stage *a*) shows the transformation of a characteristic value $v_{j,i}$ (*j*th object, *i*th characteristic) into a confidence value $cf_i(v_{j,i})$, using a sigmoid function, with control variables k_i and θ_i. Stage *b*) transforms a $cf_i(v_{j,i})$ into a characteristic BOE $m_{j,i}(\cdot)$, made up of the three mass values $m_{j,i}(\{x\})$, $m_{j,i}(\{\neg x\})$ and $m_{j,i}(\{x, \neg x\})$, and from Gerig, Welti, Guttman, Colchester, & Szekely (2000) are defined by:

$$m_{j,i}(\{x\}) = \frac{B_i}{1 - A_i} cf_i(v_{j,i}) - \frac{A_i B_i}{1 - A_i},$$

$$m_{j,i}(\{\neg x\}) = \frac{-B_i}{1 - A_i} cf_i(v_{j,i}) + B_i,$$

and $m_{j,i}(\{x, \neg x\}) = 1 - m_{j,i}(\{x\}) - m_{j,i}(\{\neg x\})$,

where A_i and B_i are two further control variables. When either $m_{j,i}(\{x\})$ or $m_{j,i}(\{\neg x\})$ are negative they are set to zero

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

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 $m_{\text{both}}(\cdot)$

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

Related Content

Negative Association Rules in Data Mining
Olena Dalyand David Taniar (2005). *Encyclopedia of Data Warehousing and Mining (pp. 859-864).*
www.irma-international.org/chapter/negative-association-rules-data-mining/10717

Homeland Security Data Mining and Link Analysis
Bhavani Thuraisingham (2005). *Encyclopedia of Data Warehousing and Mining (pp. 566-569).*
www.irma-international.org/chapter/homeland-security-data-mining-link/10661

Semi-Structured Document Classification
Ludovic Denoyerand Patrick Gallinari (2005). *Encyclopedia of Data Warehousing and Mining (pp. 1015-1021).*
www.irma-international.org/chapter/semi-structured-document-classification/10745

Ontology-Based Interpretation and Validation of Mined Knowledge: Normative and Cognitive Factors in Data Mining
Ana Isabel Canhoto (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 2316-2337).*
www.irma-international.org/chapter/ontology-based-interpretation-validation-mined/7765

X-WACoDa: An XML-Based Approach for Warehousing and Analyzing Complex Data
Hadj Mahboubi, Jean-Christian Ralaivao, Sabine Loudcher, Omar Boussaïdand Fadila Bentayeb (2010). *Data Warehousing Design and Advanced Engineering Applications: Methods for Complex Construction (pp. 38-54).*
www.irma-international.org/chapter/wacoda-xml-based-approach-warehousing/36607