Object Classification Using CaRBS

Malcolm J. Beynon
Cardiff Business School, UK

INTRODUCTION

The notion of uncertain reasoning has grown relative to the power and intelligence of computers. From sources which are uncertain information and/or imprecise data, it is importantly the ability to represent uncertainty and reason about it (Shafer & Pearl, 1990). A very general problem of uncertain reasoning is how to combine information from independent and partially reliable sources (Haenni & Hartmann,forthcoming). With data mining, understanding the confirming and/or conflicting information from characteristics describing objects classified to given hypotheses is affected by their reliability. Further, the presence of missing values compounds the problem, since the reasons for their presence may be external to the incumbent reliability issues (Olinsky, Chen, & Harlow, 2003; West, 2001).

These issues are demonstrated here using the classification technique: Classification and Ranking Belief Simplex (CaRBS), introduced in Beynon and Buchanan (2004) and Beynon (2005). CaRBS operates within the domain of uncertain reasoning, namely in its accommodation of ignorance, due to its mathematical structure based on the Dempster-Shafer theory of evidence (DST) (Srivastava & Mock, 2002). The ignorance here encapsulates incompleteness of the data set (presence of missing values), as well as uncertainty in the evidential support of characteristics to the final classification of the objects.

This chapter demonstrates that a technique such as CaRBS, through uncertain reasoning, is able to uniquely manage the presence of missing values by considering them as a manifestation of ignorance, as well as allowing the possible unreliability of characteristics to be inherent. Importantly, the described process removes the need to falsely transform the data set in any way, such as through imputation (Huisman, 2000).

The example issue of credit ratings considered here has become increasingly influential since its introduction in around 1900 with the Manual of Industrial and Miscellaneous Securities (Levich, Majnoni, & Reinhart, 2002). The rating agencies shroud their operations in particular secrecy, stating that statistical models cannot be used to replicate their ratings (Singleton & Surkan, 1991), hence advocating the need for alternative analyses, including those utilising uncertain reasoning.

BACKGROUND

DST is a methodology for evidential reasoning, manipulating uncertainty, and capable of representing partial knowledge (Kulasekere, Premaratne, Dewasurendra, Shyu, & Bauer, 2004; Scotney & McClean, 2003). Early after its introduction it was considered as a generalisation of Bayesian theory.

The traditional terminology within DST begins with a finite set of hypotheses Θ (frame of discernment). A *mass value* (basic probability assignment) is a function $m: 2^{\Theta} \rightarrow [0, 1]$ such that: $m(\varnothing) = 0$ and $\sum_{A \in 2^{\Theta}} m(A) = 1$ (2^{Θ} the power set of Θ). Any $A \in 2^{\Theta}$, for which $m(A) > 0$ 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 mass values) is defined as a body of evidence (BOE).

To collate two or more sources of evidence, 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:

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

where $k = \sum_{A \cap B = \varnothing} m_1(A) m_2(B)$, is a mass value associated with $y \subseteq \Theta$. The term $(1 - k)$ can be interpreted as a measure of conflict between the sources (Murphy, 2000) and is made up of one minus the sum of the products of mass values from the two pieces of evidence with empty intersection (often the k is also called the level of conflict and not $1 - k$). The associated problem with conflict is the larger the value of k the more conflict in the sources of evidence, and subsequently the less sense there is in their combination (Murphy, 2000).

To demonstrate DST, the example of the murder of Mr. Jones is considered, where the murderer was one of three assassins, Peter, Paul, and Mary, so 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}) = 0.800. Since we know nothing about the remaining mass value it is allocated to Θ, m_1({Peter, Paul, Mary}) = 0.200. Witness 2, is 60% confident that Peter was leaving on a jet plane when the murder occurred, a BOE defined

Table 1. Intermediate combination of BOEs $m_1(\cdot)$ and $m_2(\cdot)$

$m_2(\cdot) \setminus m_1(\cdot)$	$m_1(\{\text{Peter, Paul}\}) = 0.800$	$m_1(\{\text{Peter, Paul, Mary}\}) = 0.200$
$m_2(\{\text{Paul, Mary}\}) = 0.600$	{Paul}, 0.480	{Paul, Mary}, 0.120
$m_2(\{\text{Peter, Paul, Mary}\}) = 0.400$	{Peter, Paul}, 0.320	{Peter, Paul, Mary}, 0.080

$m_2(\cdot)$, includes $m_2(\{\text{Paul, Mary}\}) = 0.600$ and $m_2(\{\text{Peter, Paul, Mary}\}) = 0.400$.

Defining the combination of these sources of evidence the BOE $m_3(\cdot)$, using Dempster's combination rule, the intermediate set intersections of focal elements of the two BOEs and multiplication of the respective mass values are given in Table 1.

In Table 1, the noticeable result is that no intersections of focal elements produce the empty set, so $k=\sum_{A\cap B=\varnothing} m_1(A)m_2(B) = 0$. It follows, with the measure of conflict $(1-k) = (1-0) = 1$, then the values in Table 1 identify the combined BOE $m_3(\cdot)$ has the form, $m_3(\{\text{Paul}\}) = 0.480$, $m_3(\{\text{Peter, Paul}\}) = 0.320$, $m_3(\{\text{Paul, Mary}\}) = 0.120$ and $m_3(\{\text{Peter, Paul, Mary}\}) = 0.080$. This combined evidence has a more spread-out allocation of mass values to varying subsets of {Peter, Paul, Mary}. Further, there is a general reduction in the level of ignorance associated with the combined evidence. Smets (2002) offers a comparison of a variation of this example with how it would be modelled using traditional probability and Transferable Belief Model.

MAIN THRUST

The main thrust of this chapter is the description and application of the CaRBS system for object classification (Beynon, 2005), which operates in the DST environment. It operates on n_O objects (o_1, o_2, …), each described by n_C characteristics (c_1, c_2, …) and classified to a given hypothesis x or its compliment $\neg x$. For the object o_j ($1 \le j \le n_O$) and i^{th} characteristic c_i ($1 \le i \le n_C$), a *characteristic* BOE, defined $m_{j,i}(\cdot)$, has the mass values, $m_{j,i}(\{x\})$, $m_{j,i}(\{\neg x\})$ and $m_{j,i}(\{x, \neg x\})$. Following Gerig, Welti, Guttman, Colchester, and Szekely (2000), they are given by:

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

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

where $cf_i(v) = \frac{1}{1+e^{-k_i(v-\vartheta_i)}}$ is the confidence value associated with a characteristic value supporting evidence on the association of objects to the given hypothesis and its complement, and k_i, θ_i, A_i and B_i are incumbent control variables. Importantly, if either $m_{j,i}(\{x\})$ or $m_{j,i}(\{\neg x\})$ are negative they are set to zero, and the respective $m_{j,i}(\{x, \neg x\})$ then calculated. In Figure 1, a characteristic value v is shown to be first transformed into a confidence value (Figure 1a), then deconstructed into its characteristic BOE (Figure 1b) and finally represented as a single simplex coordinate $p_{i,j,i,v}$ in a simplex plot (Figure 1c).

The group of characteristic BOEs $m_{j,i}(\cdot)$ $i = 1, \ldots, n_C$ associated with an object o_i and its classification to x and $\neg x$ can be combined using Dempster's combination rule into an *object* BOE, defined $m_i(\cdot)$, from which its final classification can be found.

The CaRBS system depends on the assignment of values to the incumbent control variables (with standardised characteristic values, their example domains are; $-1 \le k_i \le 2$, $-1 \le \theta_i \le 1$, $0 \le A_i < 1$ and $B_i = 0.4$, see Beynon, 2005). The configuration process then becomes a constrained optimisation problem, solved here using Trigonometric Differential Evolution (TDE) (Fan & Lampinen, 2003), with operation parameters; amplification control $F = 0.99$, crossover constant $CR = 0.85$, trigonometric mutation probability $M_t = 0.05$ and number of parameter vectors $NP = 360$. In summary, TDE develops possible optimum solutions by perturbing previous solutions with the differences between two other previous solutions.

The employed objective function (OB) attempts to minimise the ambiguity in the classification of objects but not the inherent ignorance. For sets of objects making up the equivalence classes, $E(x)$ and $E(\neg x)$, namely those associated with the hypothesis and not the hypothesis, respectively, the optimum solution is to maximise the weighted difference values $(m_j(\{x\}) - m_j(\{\neg x\}))$ and $(m_j(\{\neg x\}) - m_j(\{x\}))$, respectively. The subsequent OB is given by:

$$\text{OB} = \frac{1}{4}\left(\frac{1}{|E(x)|} \sum_{o_j \in E(x)} (1-m_j(\{x\})+m_j(\{\neg x\})) + \frac{1}{|E(\neg x)|} \sum_{o_j \in E(\neg x)} (1+m_j(\{x\})-m_j(\{\neg x\})) \right)$$

which has domain $0 \le \text{OB} \le 1$. Each $(m_j(\{x\}) - m_j(\{\neg x\}))$ and $(m_j(\{\neg x\}) - m_j(\{x\}))$ difference value measures the ambiguity in each classification, there is no attempt to minimise the $m_{j,i}(\{x, \neg x\})$ values so no inclination to directly minimise the concomitant ignorance in each objects' classification. Correct classification is graphically defined by which side of the vertical dashed line down from the $\{x, \neg x\}$ vertex, in a simplex plot, an object BOE's simplex coordinate is positioned (classifying to x (right) and $\neg x$ (left)).