Classification and Ranking Belief Simplex

Malcolm J. Beynon
Cardiff University, UK

INTRODUCTION

The classification and ranking belief simplex (CaRBS), introduced in Beynon (2005a), is a nascent technique for the decision problems of object classification and ranking. With its rudiments based on the Dempster-Shafer theory of evidence—DST (Dempster, 1967; Shafer, 1976), the operation of CaRBS is closely associated with the notion of uncertain reasoning. This relates to the analysis of imperfect data, whether that is data quality or uncertainty of the relationship of the data to the study in question (Chen, 2001).

Previous applications which have employed the CaRBS technique include: the temporal identification of e-learning efficacy (Jones & Beynon, 2007) expositing osteoarthritic knee function (Jones, Beynon, Holt, & Roy, 2006), credit rating classification (Beynon, 2005b), and ranking regional long-term care systems (Beynon & Kitchener, 2005). These applications respectively demonstrate its use as a decision support system for academics, medical experts, credit companies, and governmental institutions.

Through its reliance on DST, the CaRBS technique allows the presence of ignorance in its analysis (Smets, 1991), in the case of the classification of objects this means that ambiguity is minimised but ignorance tolerated. Continuing the case of object classification, in the elucidation of CaRBS, two objective functions are considered here to quantify the level of classification achieved, which take into account differently the issues of classification ambiguity and classification ignorance. The use of the CaRBS technique allows the fullest visualisation of the decision support results, able through the depiction of the evidence from characteristics describing each object as a simplex coordinate in a simplex plot.

An associated issue is the ability of CaRBS to offer decision support based on incomplete data, without the need for any inhibiting external management of the missing values present (Beynon, 2005b). Within CaRBS, a missing value is considered an ignorant value and retained in the considered data, allowing the fullest real interpretation of results from the original data (Schafer & Graham, 2002). This article presents the rudiments of the CaRBS technique and an expository analysis using it on an example data set.

BACKGROUND

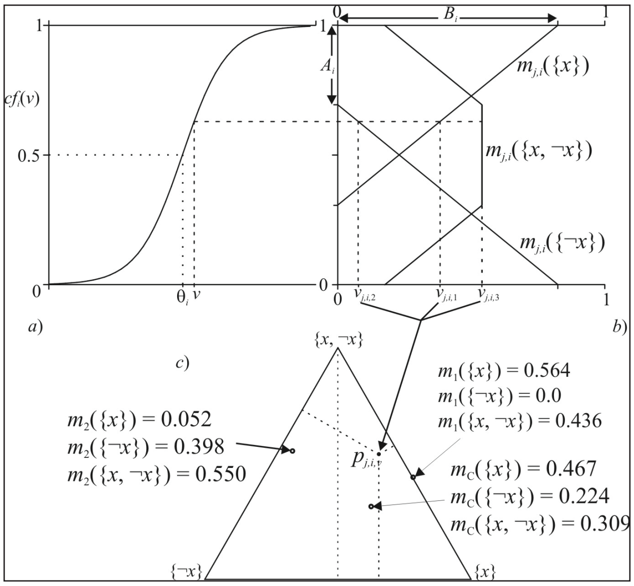
The classification and ranking belief simplex, introduced in Beynon (2005a, 2005b), is a novel object classification and ranking technique, where objects such as o_j are described by a series of characteristics $c_1, \ldots, c_n$. Considering the classification problem only here, these characteristics contribute evidence to whether the classification of an object is to a given hypothesis ($\{x\}$), its complement ($\{\neg x\}$), and a level of ignorance ($\{x, \neg x\}$). DST forms the basis for the operation of the CaRBS technique, as such it is termed around the formation of bodies of evidence (BOEs), made up of mass values representing the levels of exact belief in the focal elements of the hypothesis, not the hypothesis and concomitant ignorance.

More formally, the evidence from a characteristic c_i, for the object o_j, is defined a characteristic BOE, termed $m_{j,i}(\cdot)$, made up of a triplet of mass values; $m_{j,i}(\{x\})$, $m_{j,i}(\{x, \neg x\})$ and $m_{j,i}(\{\neg x\})$, where $\{x\}$, $\{\neg x\}$ and $\{x, \neg x\}$, are the focal elements discussed. The rudiments of the CaRBS technique can then be described by reference to Figure 1, where the construction of a characteristic BOE is shown.

In Figure 1, stage *a*) shows the transformation of a characteristic value v (j^{th} object, i^{th} characteristic) into a confidence value $cf_i(v)$, using a sigmoid function, with control variables k_i and θ_i. Stage *b*) transforms a $cf_i(v)$ value 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 $m_{j,i}(\{x, \neg x\})$, defined by (Safranek et al., 1990):

$$m_{j,i}(\{x\}) = \max\left(0, \frac{B_i}{1 - A_i} cf_i(v) - \frac{A_i B_i}{1 - A_i}\right),$$

Figure 1. Stages within the CaRBS technique to construct a characteristic BOE from a characteristic value v



$$m_{j,i}(\{\neg x\}) = \max\left(0, \frac{-B_i}{1-A_i} cf_i(v) + B_i\right),$$

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. Stage *c*) shows a BOE $m_{j,i}(\cdot)$; $m_{j,i}(\{x\}) = v_{j,i,1}$, $m_{j,i}(\{\neg x\}) = v_{j,i,2}$ and $m_{j,i}(\{x, \neg x\}) = v_{j,i,3}$, can be represented as a simplex coordinate ($p_{j,i,v}$) in a simplex plot (equilateral triangle). The point $p_{j,i,v}$ exists such that the least distance it is to each of the sides of the equilateral triangle are in the same proportion (ratio) to the values $v_{j,i,1}$, $v_{j,i,2}$ and $v_{j,i,3}$.

Within DST, Dempster's rule of combination is used to combine a series of characteristic BOEs, to produce an *object* BOE, associated with an object and their level of classification to $\{x\}$, $\{\neg x\}$, and $\{x, \neg x\}$. The combination of two BOEs, $m_i(\cdot)$ and $m_k(\cdot)$, defined $(m_i \oplus m_k)(\cdot)$, results in a combined BOE whose mass values are given by the equations in Box 1.

This process to combine two BOEs is demonstrated on $m_1(\cdot)$ and $m_2(\cdot)$, to produce $m_C(\cdot)$, see Figure 1*c*. The two BOEs, $m_1(\cdot)$ and $m_1(\cdot)$, have mass values in the vector form $[m_i(\{x\}), m_i(\{\neg x\}), m_i(\{x, \neg x\})]$, as [0.564, 0.000, 0.436] and [0.052, 0.398, 0.550], respectively. The combination of $m_1(\cdot)$ and $m_2(\cdot)$, using the expressions, is evaluated to be [0.467, 0.224, 0.309] (= $m_C(\cdot)$). In Figure 1*c*, the simplex coordinates of the BOEs, $m_1(\cdot)$ and $m_2(\cdot)$, are shown along with that of the combined BOE $m_C(\cdot)$.

The described combination process can then be used iteratively to combine the characteristic BOEs describing each object into an *object* BOE. It is noted, the CaRBS system is appropriate for a problem where each related characteristic has a noticeable level of concomitant ignorance associated with it and its contribution to the problem (Gerig, Welti, Guttman, Colchester, & Szekely, 2000).

The effectiveness of the CaRBS system is governed by the values assigned to the incumbent control variables, k_i, θ_i, A_i and B_i ($i = 1, ..., n$). This necessary configuration is defined as a constrained optimisation problem, solved here using trigonometric differential evolution—TDE (Storn & Price, 1997; Fan & Lampinen, 2003). When the classification of a number of objects is known, the effectiveness of a configured CaRBS system can be measured by a defined objective function (OB). Two objective functions are considered

Box 1.

$$(m_i \oplus m_k)(\{x\}) = \frac{m_i(\{x\})m_k(\{x\}) + m_k(\{x\})m_i(\{x,\neg x\}) + m_i(\{x\})m_k(\{x,\neg x\})}{1-(m_i(\{\neg x\})m_k(\{x\}) + m_i(\{x\})m_k(\{\neg x\}))}$$

$$(m_i \oplus m_k)(\{\neg x\}) = \frac{m_i(\{\neg x\})m_k(\{\neg x\}) + m_k(\{x,\neg x\})m_i(\{\neg x\}) + m_k(\{\neg x\})m_i(\{x,\neg x\})}{1-(m_i(\{\neg x\})m_k(\{x\}) + m_i(\{x\})m_k(\{\neg x\}))}$$

$$(m_i \oplus m_k)(\{x,\neg x\}) = 1-(m_i \oplus m_k)(\{x\}) - (m_i \oplus m_k)(\{\neg x\})$$