Omega 59 (2016) 97–112

Contents lists available at ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

A NCaRBS analysis of SME intended innovation: Learning about the Don't Knows

Malcolm J Beynon [a], Paul Jones [b,*], David Pickernell [c], Gary Packham [d]

[a] Cardiff Business School, Cardiff University, Aberconway Building, Colum Drive, Cardiff CF10 3EU, Wales, United Kingdom
[b] Futures Entrepreneurship Centre, Mast House, Shepherds Wharf, Plymouth University, Plymouth PL4 0HJ, Devon, England, United Kingdom
[c] Faculty of Business and Society, University of South Wales, Treforest CF37 1DL, Wales, United Kingdom
[d] Lord Ashcroft International Business School, Anglia Ruskin University, East Road, Cambridge CB1 1PT, United Kingdom

ARTICLE INFO

Article history:
Received 31 August 2014
Accepted 28 April 2015
Available online 12 June 2015

Keywords:
SME
NCaRBS
Don't Know
Innovation

ABSTRACT

This study demonstrates a novel form of business analytics, respecting the quality of the data available (allowing incompleteness in the data set), as well as engaging with the uncertainty in the considered outcome variable (inclusive of Don't Know (DK) responses). The analysis employs the NCaRBS technique, based on the Dempster–Shafer theory of evidence, to investigate the relationship between Small and Medium-sized Enterprise (SME) characteristics and whether they intended to undertake future innovation. The allowed outcome response for intended innovation was either, Yes, No and DK, all of which are considered pertinent responses in this analysis. An additional consequence of the use of the NCaRBS technique is the ability to analyse an incomplete data set, with missing values in the characteristic variables considered, without the need to manage their presence. From a soft computing perspective, this study demonstrates just how exciting the business analytics field of study can be in terms of pushing the bounds of the ability to handle real 'incomplete' business data which has real, and sometimes uncertain, outcomes. Further, the findings also inform how different notions of ignorance in evidence are accounted for in such analysis.

1. Introduction

Individual Small and Medium-sized Enterprises (SMEs) have their own strategies for their survival and contribution to the associated economy [81,79,34,75], including in respect to innovation. Innovation, put simply finding a more effective way of doing something (or the application of enhanced solutions that meet new requirements), can therefore be seen to play a critical role in enabling these firms' business growth and improving performance [35]. Whilst this highlights an important applied business research area, there is an associated research problem, specifically the uncertainty of this potential future activity for the firms themselves [63]. Within a business analytics context, this study asks the question whether it is possible, and indeed relevant, to gain knowledge of firms expressing uncertain innovation plans, such as by answering 'Don't Know' (DK) to related questions. For example, if an SME gives a DK response to an intended innovation question, is there an underlying indication that the firm is more inclined to actually mean 'No' or 'Yes' to such intended innovation.

In terms of analysing such uncertainty, Francis and Busch [29] suggest, generally, which respondents with non-substantive responses, such as DK, should not be excluded from analysis, arguing such responses are not random and so exclusion would introduce bias in any undertaken analysis. The general limited investigation of the DK response problem, considered a vexing problem to researchers [26], with the slight exception of within the area of political opinion [26,31,46,51], may be due to the lack of technical approaches able to pertinently investigate this problem. Business analytics can assist in such analysis, as an area of research, it has manifested itself to cover the more general data mining and knowledge discovery terms often used (see [55]), and has been welcoming of the development of new approaches to analyse data.

This paper, demonstrates the exciting potential of business analytics, using a nascent soft computing based methodology (see later), in a multi-direction investigation of SME intended innovation in the UK. Beyond the prior mentioned intention to be inclusive of the non-substantive DK response and how other variables may relate to them, a further direction of this study is to consider the pertinent ability to analyse incomplete data, here meaning without the need to manage in any way prevalent missing values, without needing to transform the data in any way. This approach is in contrast to the perceived

* Corresponding author. Tel.: +44 1752 585523.
E-mail addresses: BeynonMJ@cardiff.ac.uk (M. Beynon), Paul.Jones@plymouth.ac.uk (P. Jones), david.pickernell@southwales.ac.uk (D. Pickernell), gary.packham@anglia.ac.uk (G. Packham).

http://dx.doi.org/10.1016/j.omega.2015.04.018
0305-0483/

inevitable problem of how to deal with the missing values, indeed Svolba (2014) [73] states this very point, going onto highlight the business point of view on the handling of missing values.

Whether it is concerned with small, medium or big data, the issue of analysing incomplete data usually means some form of data management is required ([3,64], Svolba, 2014) [73]. For example, dummies representing missing values in predictors can be incorporated into regression analysis, an example of how more traditional techniques might accommodate such incompleteness (see [32], for recent survey of literature on missing values). The level of impact of the missing value issue is succinctly described by Koslowsky [43], p. 312, who stated;

> One of the most critical issues in model formulation and marketing analytics is how to handle missing data. If not handled correctly, even the best analysis efforts can fail, and even worse, an entire database marketing strategy can be seriously damaged.

The ability to analyse incomplete data, without having to manage the missing values in some way, therefore, introduces an important dimension of intelligence to the business analytics area of research. Specifically, this identifies an interesting point, namely that intelligence here may not just be about producing a more pertinent answer, but also about more pertinently using the data available. Indeed, what is more intelligent, using an 'intelligent' method to transform the incomplete data into complete data (see, for example, [36]), or using an intelligent method that allows the use of the original incomplete data without any transformation (as in this study)? A consequence of this study includes the elucidation of two notions of ignorance in the evidence in the classification problem (ignorance due to missing values and ignorance from variable value contribution).

Such an intelligent method, however, in addition to being able to handle these two issues, of uncertain DK responses and incomplete (missing) data, would also need to still be able to analyse the important applied problem which the data has been identified as being able to help address, here SME intended innovation, producing results that are clearly interpretable. One specific feature of the unfolding popularity of business analytics is its association to producing results that can be then used in policy decisions, and for example, the ability to offer competitive advantage amongst organisations (see, for example, [42,69]). Here, the competitive advantage in the considered applied problem may be more at the policy maker level, being able to use the presented results to develop policies that inspire higher SME performance (here innovation).

The technique employed throughout this study is the N-State Classification and Ranking Belief Simplex (NCaRBS), introduced in Beynon and Kitchener [10] and Beynon et al. [9], a development from the original CaRBS [7,8]. With its methodology based on the Dempster–Shafer theory of evidence [20,68], also called theory of belief functions, the technique has a close association to soft computing (see, for example, [39]). In this study, the use of NCaRBS will demonstrate the ability to pertinently work throughout the three research directions outlined previously. Results presented will include consideration of the level of classification fit of the analysis undertaken, contribution (predictive power) of the characteristic variables considered, the ability to interpret analysis of individual objects and validation of results through re-sampling based analysis.

The structure of the rest of the paper is as follows: In Section 2, brief descriptions of soft computing, the NCaRBS analysis technique and incomplete data handling are presented. In Section 3, the incomplete FSB-innovation data set is described and research problem presented. In Section 4, an initial analysis using NCaRBS is presented, including exposition of the level of classification fit, contribution of characteristic variables and elucidation of individual objects' classification details. In Section 5, validation of the results is given with respect to a re-sampling based analysis of the data set, using in-sample and out-sample partitioned data sets. In Section 6, inferences in respect to SME innovation and business analytics are given. In Section 7, conclusions are given as well as direction for future research.

2. Soft computing, NCaRBS technique and incomplete data handling

This section is broken down into three subsections, briefly describing the issues of, soft computing, NCaRBS technique and incomplete data handling.

2.1. Soft computing

One direction contributing to the nascence of business analytics has been technical development in the area of soft computing. The understood tolerance of imprecision, uncertainty and approximation, underpinning the inspiration of soft computing in respect to modelling a wide variety of human rational decisions [67], has brought a number of non-traditional analysis techniques into the domain of business analytics.

Pertinent to this study, Azvine et al. [5] focus on soft computing as an emerging technology suitable for incorporation into business analytics applications, highlighting the often significant degree of manual intervention in preparing, presenting and analysing business data. The analysis presented in this study, will remove some of the often awkward impact of managing missing values within incomplete data, as referred to previously, with here the ability to analyse incomplete data without such management (see later).

Underlying the technique employed in this study (NCaRBS—see next subsection), and associated with soft computing, is Dempster–Shafer theory (DST—[20,68]), otherwise known as the theory of belief functions (see, for example, [21]). Liu [48] states where DST fits with other, more common, methodologies (p. 1):

> The Dempster–Shafer theory of belief functions has become a primary tool for knowledge representation that bridges fuzzy logic and probabilistic reasoning.

Further, DST is closely associated with uncertain reasoning (understanding uncertain knowledge and how to represent it). Ref. [15], confirm the association of uncertain reasoning and soft computing:

> Soft computing technologies have provided us with a unique opportunity to establish a coherent software engineering environment in which uncertainty and partial data and knowledge are systematically handled.

The technique next described and employed in this study is based on DST, and is able to demonstrate much of the qualities of uncertain reasoning/soft computing based business analytics. Throughout the analysis part of this study, the reader should be conscious of the data able to be analysed, and how the approach can be used in other areas closely associated with business analytics.

2.2. Technical description of NCaRBS

NCaRBS (N-state Classification and Ranking Belief Simplex, [9]), models the classification of n_{O} objects (o_1, o_2,…), to n_{D} decision outcomes (d_1, d_2,…), based on their description by n_{C} characteristics (c_1, c_2,…). The characteristics' evidence is expressed through the initial construction of *constituent* BOEs (bodies of evidence—see [20,68]), from characteristic values $v_{i,j}$ (ith object, jth characteristic), to discern between an object's association to (focal elements) a decision outcome (say $\{d_h\}$), its complement ($\{\neg d_h\}$) and a level of concomitant ignorance ($\{d_h, \neg d_h\}$).