International Journal of Approximate Reasoning 55 (2014) 1113–1128

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Forecasting using belief functions: An application to marketing econometrics

Orakanya Kanjanatarakul [a,*], Songsak Sriboonchitta [b], Thierry Denœux [c]

[a] Faculty of Management Sciences, Chiang Mai Rajabhat University, Thailand
[b] Faculty of Economics, Chiang Mai University, Thailand
[c] UMR 7253 Heudiasyc, Université de Technologie de Compiègne, CNRS, France

ARTICLE INFO

Article history:
Received 28 December 2013
Received in revised form 27 January 2014
Accepted 27 January 2014
Available online 3 February 2014

Keywords:
Prediction
Dempster–Shafer theory
Evidence theory
Bass model
Innovation diffusion
Sales forecasting

ABSTRACT

A method is proposed to quantify uncertainty on statistical forecasts using the formalism of belief functions. The approach is based on two steps. In the estimation step, a belief function on the parameter space is constructed from the normalized likelihood given the observed data. In the prediction step, the variable Y to be forecasted is written as a function of the parameter θ and an auxiliary random variable Z with known distribution not depending on the parameter, a model initially proposed by Dempster for statistical inference. Propagating beliefs about θ and Z through this model yields a predictive belief function on Y. The method is demonstrated on the problem of forecasting innovation diffusion using the Bass model, yielding a belief function on the number of adopters of an innovation in some future time period, based on past adoption data.

1. Introduction

Forecasting may be defined as the task of making statements about events that have not yet been observed. As such statements can usually not be guaranteed to be true, handling uncertainty is a critical issue in any forecasting task. Forecasting methods can be classified as statistical, causal or judgmental, depending on the kind of information used (respectively, data, causal relations or expert opinions). Whatever the approach used, a forecast cannot be trusted unless it is accompanied by some measure of uncertainty. Most of the time, forecast uncertainty is described by subjective probabilities or prediction intervals.

Recently, new formal frameworks for handling uncertainty have emerged and have become increasingly used in various application areas. One such framework is the Dempster–Shafer theory of belief functions [6,33,35]. In this approach, a piece of evidence about some question of interest is represented by a belief function, which is mathematically equivalent to a random set [29]. Independent pieces of evidence are then combined using an operation called Dempster's rule to obtain a unique belief function quantifying our state of knowledge about the question of interest. Since probability measures are special belief functions, and Bayesian conditioning can be seen as a special case of Dempster's rule, Dempster–Shafer theory is formally an extension of Bayesian probability theory. In particular, both approaches yield the same conclusions from the same initial information. However, the theory of belief function has greater expressive power and it can be argued to yield more sensible results in the presence of deep uncertainty. The reader is referred to, e.g., [34,37] for detailed discussions on the comparison between belief function and probabilistic reasoning. Recent examples of applications of Dempster–Shafer theory can be found, e.g., in Refs. [27,25,24,3,11], among others.

* Corresponding author.

0888-613X/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ijar.2014.01.005

Although the theory of belief functions has gained increasing popularity in the last few years, applications to forecasting have been, until now, very limited. The purpose of this paper is to demonstrate the application of belief function theory to statistical forecasting problems, with emphasis on situations where data are scarce and, consequently, uncertainty is high and needs to be quantified.

As an important application area, we will consider marketing econometrics and, more specifically, the forecasting of innovation diffusion. This has been a topic of considerable practical and academic interest in the last fifty years [28]. Typically, when a new product is launched, sale forecasts can only be based on little data and uncertainty has to be quantified to avoid making wrong business decisions based on unreliable forecasts [21,20,22]. The approach described in this paper uses the Bass model for innovation diffusion [2] together with past sales data to quantify the uncertainty on future sales using the formalism of belief functions. The forecasting method exemplified here can be applied to any forecasting problem when a statistical model can be postulated and some historical data is available.

The rest of this paper is organized as follows. The notion of likelihood-based belief function will first be recalled in Section 2 and the forecasting problem will be addressed in Section 3. These notions will then be applied to innovation diffusion in Section 4. Finally, the relationship with existing approaches will be discussed in Section 5 and Section 6 will conclude the paper.

2. Likelihood-based belief function

Basic knowledge of the theory of belief functions will be assumed throughout this paper. A complete exposition in the finite case can be found in Shafer's book [33] and the relation with random sets is explained in [30]. The reader is referred to [3] for a quick introduction on those aspects of this theory needed for statistical inference. In this section, the definition of a belief function from the likelihood function and its justification as proposed in [10] will first recalled. The forecasting problem will then be addressed in the next section.

Let $X \in \mathbb{X}$ denote the observable data, $\theta \in \Theta$ the parameter of interest and $f_\theta(x)$ the probability mass or density function describing the data-generating mechanism. Statistical inference consists in making meaningful statements about θ after observing the outcome x of the random experiment. This problem has been addressed in the belief function framework by many authors, starting to Dempster's seminal work [5–7,9]. In contrast with Dempster's approach relying on an auxiliary variable (see Section 3 below), Shafer proposed, on intuitive grounds, a more direct approach in which a belief function Bel_x^Θ on Θ is built from the likelihood function. This approach was further elaborated by Wasserman [39] and discussed by Aickin [1], among others. It was recently justified by Denœux in [10], from the following three basic principles:

Likelihood principle. This principle states that all the relevant information from the random experiment is contained in the likelihood function, defined by $L_x(\theta) = \alpha f_\theta(x)$ for all $\theta \in \Theta$, where α is any positive multiplicative constant [15,14]. This principle was shown by Birnbaum [4] to result from the principles of sufficiency and conditionality, which are of immediate intuitive appeal. This principle entails that Bel_x^Θ should be defined only from the likelihood function.

Compatibility with Bayesian inference. This principle states that, if a Bayesian prior $\pi(\theta)$ is available, combining it with Bel_x^Θ using Dempster's rule [33] should yield the Bayesian posterior. It follows from this principle that the contour function $pl_x(\theta)$ associated to Bel_x^Θ should be proportional to the likelihood function:

$$pl_x(\theta) \propto L_x(\theta). \tag{1}$$

Least commitment principle. According to this principle, when several belief functions are compatible with some constraints, we should select the least committed one according to some informational ordering [13,36].

In [10], it was shown that the least committed belief function verifying (1) according to the commonality ordering [12] is the consonant belief function Bel_x^Θ whose contour function is the relative likelihood function:

$$pl_x(\theta) = \frac{L_x(\theta)}{\sup_{\theta' \in \Theta} L_x(\theta')}. \tag{2}$$

This belief function is called the likelihood-based belief function on Θ induced by x. The corresponding plausibility function can be computed from pl_x as:

$$Pl_x^\Theta(A) = \sup_{\theta \in A} pl_x(\theta), \tag{3}$$

for all $A \subseteq \Theta$. The focal sets of Bel_x^Θ are the levels sets of $pl_x(\theta)$ defined as follows:

$$\Gamma_x(\omega) = \left\{\theta \in \Theta \;\middle|\; pl_x(\theta) \geqslant \omega\right\}, \tag{4}$$

for $\omega \in [0,1]$. These sets may be called *plausibility regions* and can be interpreted as sets of parameter values whose plausibility is greater than some threshold ω. The belief function Bel_x^Θ is equivalent to the random set induced by the Lebesgue measure λ on $[0,1]$ and the multi-valued mapping Γ_x from $[0,1]$ to 2^Θ [30]. In particular, the following equalities hold: