G Model
ASOC-2493; No. of Pages 9

ARTICLE IN PRESS

Applied Soft Computing xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Applied Soft Computing

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

Efficiency classification by hybrid Bayesian networks—The dynamic multidimensional models

Han-Ying Kao*, Bo-Shan Chen

Department of Computer Science and Information Engineering, National Dong Hwa University, No. 1, Sec. 2, Da-Hsueh Rd., Shoufeng, Hualien 97401, Taiwan

ARTICLE INFO

Article history:
Received 27 March 2013
Received in revised form 26 August 2014
Accepted 26 August 2014
Available online xxx

Keywords:
Data envelopment analysis
Naïve Bayesian networks
Dynamic Bayesian networks
Fuzzy parameters
Case study

ABSTRACT

This study develops the hybrid models of dynamic multidimensional efficiency classification. By integrating data envelopment analysis (DEA), naïve Bayesian networks (NBN) and dynamic Bayesian networks (DBN), this work proposes a five-step design for efficiency classification: (1) performance evaluation with DEA model, (2) efficiency discretization, (3) intra-period classification by NBN, (4) inter-period classification by DBN, (5) testing and validation. Due to the Markovian property of the dynamic models, the inter-period dependency is assumed invariant over time. In data-driven parameter learning, the fuzzy parameters for incorporating the variation in dynamic dependencies are introduced. We conduct an empirical case study of higher education in Taiwan to demonstrate the usability of this design.

1. Introduction

Every organization has its peculiar missions that usually comprise various dimensions and evolve over time. From the standpoint of evaluation, an organization needs to be observed from compound perspectives and over a period of time. Non-profit organizations, in contrast with profit business, emphasize specific missions which supersede profitability. For example, research institutions aim at research and development (R&D) activities and achievements. On the other hand, the financial function plays a key role in their survival, therefore operational as well as financial performance must be considered in research institutions. Moreover, endogenous as well as exogenous factors related to organizations are constantly changing, and thus, how to diagnose the inefficiency or predict the trends in dynamic environments becomes inevitable.

This study develops the hybrid models of dynamic multidimensional efficiency classification. By integrating data envelopment analysis (DEA), naïve Bayesian networks (NBN) and dynamic Bayesian networks (DBN), this work proposes a five-step design for efficiency classification including (1) performance evaluation with DEA model, (2) efficiency discretization, (3) intra-period classification by NBN, (4) inter-period classification by DBN, (5) testing and validation. In training the dynamic models with the Markovian property, a temporal dependency is assumed invariant over time. However, in data-driven parameter learning, it is difficult to identify the stationary temporal relationships in DBN. So, fuzzy parameters for incorporating the variations in dynamic dependencies are introduced. To demonstrate the usability of this design, we have conducted an empirical case study of higher education in Taiwan.

In the remainder of this article, we review the essential background in Section 2. In Section 3, the methods and research procedures are addressed. In Section 4, we conduct a case study of higher education in Taiwan. Finally, the concluding remarks are given in Section 5.

2. Background review

In this section we briefly review the background of data envelopment analysis, Bayesian networks, and fuzzy sets and theory.

2.1. Data envelopment analysis

Data envelopment analysis (DEA) [1,2] is a celebrated efficiency evaluation technique, among which the BCC model (denominated from the inventors, Banker, Charnes and Cooper) [1], has been widely used. The BCC model assesses the relative efficiency of decision-making units (DMUs) by extending the constant-returns-to-scale CCR model (denominated from the inventors, Charnes, Cooper and Rhodes) [2] to variable returns to scale. Consider n DMUs ($j = 1, \ldots, n$) under assessment. Each DMU consumes m inputs ($i = 1, \ldots, m$) and produces s outputs ($r = 1, \ldots, s$), denoted by $X_{1j}, X_{2j}, \ldots, X_{mj}$ and $Y_{1j}, Y_{2j}, \ldots, Y_{sj}$ respectively. The efficiency of DMU_k can be computed by the BCC model as follows:

* Corresponding author. Tel.: +886 3 8634059; fax: +886 3 8634010.
E-mail addresses: teresak@mail.ndhu.edu.tw, teresak_hk@yahoo.com.tw (H.-Y. Kao).

http://dx.doi.org/10.1016/j.asoc.2014.08.061
1568-4946/

Please cite this article in press as: H.-Y. Kao, B.-S. Chen, Efficiency classification by hybrid Bayesian networks—The dynamic multidimensional models, Appl. Soft Comput. J. (2014), http://dx.doi.org/10.1016/j.asoc.2014.08.061

BCC ratio model

$$\begin{aligned}
&\text{Max} \quad E_k = \frac{\sum_{r=1}^{s} u_r Y_{rk} - u_0}{\sum_{i=1}^{m} v_i X_{ik}} \\
&\textit{subject to}\,(s.t.) \\
&\frac{\sum_{r=1}^{s} u_r Y_{rj} - u_0}{\sum_{i=1}^{m} v_i X_{ij}} \le 1, \quad j = 1, 2, \ldots, n \\
&u_r, v_i \ge \varepsilon \quad r = 1, 2, \ldots, s; \quad i = 1, 2, \ldots, m; \quad u_0 \text{ unrestricted in sign}
\end{aligned} \tag{1}$$

In the BCC ratio model, the objective function E_k is maximized for every DMU_k individually. In the model, X_{ik} and Y_{rk} are the ith input and rth output of DMU_k; u_r, v_i are the weights of the outputs and inputs, respectively; ε is a small positive value which ensures all weights are non-negative. When the intercept of the production function $u_0 > 0$ occurs, the efficiency frontier presents decreasing returns to scale; if $u_0 < 0$, it manifests increasing returns to scale; while if $u_0 = 0$, the models turn out to be constant-returns-to-scale CCR models. For computational convenience, the ratio model is normally transformed into a linear programming (LP) model by assuming $\sum_{i=1}^{m} v_i X_{ik} = 1$ [2]. Notably, the solution space of the DEA LP model is smaller than that of the ratio model due to the constraint $\sum_{i=1}^{m} v_i X_{ik} = 1$. Thus, the LP model only finds the local optimum for the ratio model which comprises fractional terms [3].

2.2. Bayesian networks

Bayesian networks (BN) [4–7] may be described by a directed acyclic graph (DAG) in which the nodes represent the variables, the arcs represent the directed causal influences between the linked variables, and the influences may be described by conditional probabilities. They are widely used knowledge representation and reasoning models for various domains under uncertainty. Because an expert system requires both predictive and diagnostic information, two types of reasoning are common in Bayesian networks, deduction and abduction. Deduction, or prediction, is a logical process from a hypothesis to deduce evidence where probabilistic relationships are involved. Abduction, or diagnosis, is a logical process that hypothetically explains experimental observations [7]. The naïve Bayesian network (NBN) model was named by Titterington et al. [8] for its simplicity. In a NBN model, the variable of interest has to be a root without a parent node.

If a Bayesian network does not involve any temporal factors, it is a static network. Static Bayesian networks can be extended into dynamic Bayesian networks (DBN) by introducing relevant temporal dependencies between representations of the static network at different times [9–12]. Two types of dependencies can be distinguished in a dynamic network: contemporaneous dependencies and non-contemporaneous dependencies [9]. Contemporaneous dependencies refer to arcs between nodes that represent variables within the same time period. Non-contemporaneous dependencies refer to arcs between nodes that represent variables at different times.

2.3. Uncertainty in Bayesian networks

Most previous studies on Bayesian networks use probability distributions associated with the random variables (nodes) as the numerical models. Before modeling the Bayesian network, we identify the sources of uncertainty that may exist in reasoning systems. The sources of uncertainty can be classified as (a) stochastic properties of the system, which manifest with random behaviors of the variables; (b) incomplete knowledge, accompanied with personal or subjective judgments; (c) semantic vagueness in system features, such as *high* manufacturing capability, *good* stock control performance, *fair* customer satisfaction, and so on. Taking different sources of uncertainty into account, probabilistic approach may be computationally intractable in the context of uncertainties (b) and (c). For example, it is difficult to describe the causal links by pure probability theory when semantic vagueness is involved. Also in a highly uncertain system with incomplete knowledge, it will be hard to estimate the probability parameters and subjective knowledge may be unavoidable. Freeling [13] claimed that fuzzy probability can be developed as an extension of probability theory, which is more promising than possibility and probability theory as the uncertainty measure. By fuzzy probabilities, the numerical models for Bayesian networks can be extended and more adaptive to various reasoning context. This study intends to incorporate the concept of possibility [14–16] and construct fuzzy parameters in Bayesian networks.

Reviewing the graphical decision model with a possibility approach, Yamada [17] addresses uncertain reasoning with multiple causes and conditional possibilities on a causal network model, which focuses on the causal effect in two layered networks. Rodríguez-Muñiz et al. [18] explore the statistical rules for modeling fuzzy random variables and utilities in influence diagrams mainly based on the value-preserving transformations. Later, López-Díaz and Rodríguez-Muñiz [19] analyze how to evaluate influence diagrams with multiple value nodes in terms of fuzzy random variables by dynamic programming. These works emphasize utility evaluation and ignore diagnostic reasoning.

Additionally, studies on fuzzy probabilities and parameters as an alternative approach in fuzzy logic framework are developed. Zadeh [20,21] constructs the generalized theory of uncertainty (GTU) and include fuzzy probability as one building block of the fuzzy logic. Li and Kao [22] use fuzzy parameters to formulate the soft constraints in learning average causal effects from imperfect experiments. Later the multiple-objective nonlinear programming models are developed to solve constrained diagnostic reasoning and decision problems [23,24]. Shipley and Johnson [25] propose a fuzzy sets-based selection of employees in order to meet the projects' goals for the preferred cognitive style. They design an algorithm based on belief in the fuzzy probability of a cognitive style fitting a defined goal. Abdelaziz and Masri [26] introduce a multistage stochastic program with triangular fuzzy probability distributions. In their study, the α-cut method is used as the defuzzification technique for the mini-max approach in decomposing the stochastic program. Later in Section 3, we will learn the fuzzy parameters for the dynamic Bayesian Networks in the inter-period classification.

3. The methods

This section presents the approaches for efficiency estimation and classification. In the proposed framework, the factors and efficiency outcomes from DEA are integrated with the hybrid models in learning the patterns of DMUs' performance. The research procedure and the integrated models of this study are depicted as Figs. 1 and 2, respectively.

Assume that n DMUs are evaluated from two perspectives, *Efficiency* 1 (E^t_{k1}) and *Efficiency* 2 (E^t_{k2}) for DMU k ($k = 1, \ldots, n$), at time t ($t = 1, \ldots, p$). In Fig. 2, the models are composed of the intra-period classification NBNs and inter-period classification DBNs. The DBNs indicate the temporal dependency between two types of efficiencies in adjacent periods, while the NBNs express the interrelation