Decision Support Systems 49 (2010) 486–497

Contents lists available at ScienceDirect

Decision Support Systems

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

Using data envelopment analysis and decision trees for efficiency analysis and recommendation of B2C controls

Sangjae Lee

College of Business Administration, Sejong University, 98 Kunja-dong, Kwangjin-gu, Seoul 143-747, Republic of Korea

ARTICLE INFO

Article history:
Received 11 September 2009
Received in revised form 30 March 2010
Accepted 12 June 2010
Available online 17 June 2010

Keywords:
B2C applications
B2C controls
Data envelopment analysis (DEA)
Decision trees
Implementation of B2C applications

ABSTRACT

Appropriate guidelines for controls in B2C (business-to-consumer) applications (hereafter B2C controls) should be provided such that these guidelines accomplish efficiency of controls in the context of specific system environments, given that many resources and skills are required for the implementation of such controls.

This study uses a two-step process for the assessment of B2C controls, i.e., efficiency analysis and recommendation of controls. First, using a data envelopment analysis (DEA) model, the study analyzes the efficiency of B2C controls installed by three groups of organizations: financial firms, retail firms, and information service providers. The B2C controls are composed of controls for system continuity, access controls, and communication controls. DEA model uses B2C controls as input and three variables of implementation of B2C applications, i.e., volume, sophistication, and information contents as output. Second, decision trees are used to determine efficient firms and generate rules for recommending levels of controls.

The results of the investigation of the DEA model indicate that retail firms and information service providers implement B2C controls more efficiently than financial firms do. Controls for system continuity are implemented more efficiently than access controls. In financial firms, controls for system continuity, communication controls, and access controls, in a descending order, are efficiently adopted in B2C applications.

Every company can determine its relative level of reduction in each component of controls in order to make the control system efficient. The firms that efficiently implement B2C controls are determined using a decision tree model. The decision tree model is further used to recommend the level of controls and suggest rules for controls recommendation. This suggests the possibility of using decision trees for controls assessment in B2C applications.

1. Introduction

As organizations rely increasingly on IS (information Systems) for strategic advantage and operations, management needs to pay more attention to IS security issues due to a corresponding increase in the impact of IS security abuses. Controls of e-business applications should ensure the security, integrity, auditability, and controllability of the configured software, data, and support organization [16].

As controls are not only expensive to put in place and operate, but also increase audit efforts and slow down the execution of business processes, it is necessary to automate controls in an optimum manner from a cost and regulatory perspective [28]. How internal auditors or security administrators make decisions of controls assessment or adjustment is in large part a matter of judgment and experience. Thus, it is necessary to devise a systematic approach for controls assessment for security management that relies on a series of subjective judgments of internal auditors or security administrators.

Gordon and Loeb [13] have shown an economic model to determine the optimal level of investment in information security. Cavusoglu et al. [7] suggested a model based on game theory for strategic investment decisions in security controls; in the IT security problem, the firm and hacker are players and the firm's payoff from security investment depends on the extent of hacking it is subjected to. Industry characteristics such as system vulnerability to security risks and availability of value added networks can affect the efficiency of EDI (Electronic Data Interchange) controls [21]. Pareek [27] used an optimization algorithm based on a linear programming model to identify controls that need to be tested to address the risks.

The law of diminishing returns posits that, beyond a certain point, the effectiveness of protection provided by additional controls will diminish and no longer improve the quality of the information systems [9]. Guidance can be provided by analyzing the data collected from questionnaires used to measure controls. In view of the state of implementation, and given the high cost and resources needed to develop controls embedded in the system, it is necessary to analyze the efficiency of controls in B2C (business-to-consumer) applications (hereafter B2C controls). Depending on who performs the analysis, however, a wide range of security measures may be implemented, resulting in either too few or too many B2C controls. This study intends to investigate the assessment of B2C controls, i.e., efficiency analysis and recommendation of controls, using a data envelopment analysis (DEA)

E-mail address: sangjae@sejong.ac.kr.

0167-9236/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.dss.2010.06.002

model and decision trees. Previous studies have combined the use of DEA and decision trees in analyzing organizational units [30,31,34]. Using decision trees may support IS managers and convince them of the kinds of control measures that are necessary under given system circumstances. The firms that efficiently implement B2C controls are determined using a decision tree model. The decision tree model is further used to recommend the level of controls and to suggest rules for controls recommendations. This suggests the possibility of using decision trees for controls assessment in B2C applications.

2. Theoretical background

2.1. Process of controls assessment

Audit management staff members face a constant need to cut time in completing controls assessment and testing. Automated data mining tools are capable of investigating a large amount of information and searching for patterns that may not be identified easily by manual means [26]. The US Sarbanes-Oxley Act has already had a significant influence on the assessment and evaluation of internal controls [4]. Section 404 demands that management annually conducts an assessment of the design and operation of the company's internal controls and procedures for financial reporting. The objective of controls assessment is to assure customers, stakeholders, and government agencies that controls are in place and effective.

Too few security measures make the firm's IT environment vulnerable to a wide range of potentially damaging risks, and too many security measures lead to an increase in costs, to slowdowns and to delays in processing applications [9]. The assessment of controls done is to ensure adequacy of controls for a balance between costs incurred for implementing controls and the resulting benefits derived.

Controls assessment provides a method for management to determine the current status of their information security programs; this assessment involves controls improvement, if necessary through the timely detection and correction of weak controls [4]. Assessment offers a means of identifying problems of controls and recommendations for improvement. Further, assessment is done to reduce or eliminate costly and inefficient controls while creating valuable alternatives.

The examples of controls assessment models includes the NIST (National Institute of Standards and Technology) model from the US National Institute of Standards and Technology, the COBIT (Control Objectives for Information and Related Technology) model from the Information Systems Audit and Controls Association, and the Business Process Model. The NIST model suggested a questionnaire to be applied for any organization. The COBIT model is an IT governance tool that helps organizations understand and manage the risks associated with IT. The Business Process Model is based on the identification of risks associated with each business process. Sound knowledge of business processes, information control objectives, and company environments are crucial factors in the success of introduction of the controls assessment model.

2.2. DEA analysis

Assessing an IT security investment has been a sticking point, and a rational methodology is required to analyze security investments [7]. This study used DEA to analyze the efficiency of B2C controls. DEA is a mathematical programming formulation based technique that provides an efficient frontier to suggest an estimate of the relative efficiency of each decision making unit (DMU) in a problem set [10]. DEA is developed around the concept of evaluating the efficiency of a decision alternative based on its performance of creating outputs in means of input consumption. The efficiency of each DMU, relative to its peers, is defined as the ratio of that member's weighted sum of outputs to its weighted sum of inputs. No functional form relating the input to output variables is necessary. The parametric approach, such as regression equation and discriminant analysis, however, requires specific assumptions about the functional form and the distribution of error terms (e.g., independently or normally distributed). Those DMUs not on the frontier are scaled against a convex combination of the DMUs on the frontier facet closest to them.

DEA is used in a wide range of contexts, such as software projects [23], information technology investments [33], technology commercialization projects [34], EDI controls [21], Internet companies [32], branches of banks [37], service delivery processes [31], data warehouse operations [24] and supplier evaluation and selection ([8].

2.3. Decision trees

In order to ensure successful audits, organizations should maintain a minimized list of controls that is manageable and easy to understand [3]. For instance, network security (especially, updated firewalls and secure wireless connections), virus and spyware protection, and backup procedures are the three most important controls for a small business [6]. Auditors can use data mining techniques such as statistical modeling to uncover patterns that can help organizations identify process improvements, detect fraud, and improve risk management [26]. As auditing based on data mining may require additional resources to ensure management data analysis for a continuous assurance process, auditors should plan adequately and have a reasonable perspective before embarking on a data mining exercise.

Decision trees are a rapid and effective method of classifying data set entries, and can offer good decision support capabilities. A decision tree is a tree in which each non-leaf node denotes a test on an attribute of cases, each branch corresponds to an outcome of the test, and each leaf node denotes a class prediction. The quality of a decision tree depends on both its classification accuracy and its size.

Classification using decision trees categorizes a set of cases in a database into different classes according to a classification model. Two kinds of data sample are used for the classification task. A training sample (i.e., a set of cases whose class labels are known) is first analyzed and a classification model is constructed based on the features available in the data of the training sample. Such a classification model is then used to categorize a test sample (i.e., a set of cases whose class labels are unknown). For example, we can use the classification model learned from the existing customers' data to predict what services a new customer would like.

A case in the training sample set consists of multiple attributes (independent and dependent factors) and a known class label associated with them. The independent factors are represented as an attribute-value vector, $x = (\chi_1\chi_2,\ldots,\chi_j)$. Assume that the cases can fall into *j* classes, that is, $C = (c_1c_2,\ldots,c_j)$. Then, a training sample can be denoted by $M = \{(x_m,y_m)\}$ where $x_m \in X$ (all possible attribute space) and $y_m \in C$ (all possible cases), $m = 1,\ldots M$ (the size of the model set). On the other hand, since all the cases in a test sample have no known class levels, a test sample is denoted by $S = \{(x_\varepsilon,y_\varepsilon)\}$ where $x_s \in X$ and $y_s \in \emptyset$, $s = 1,\ldots,S$ (the size of the test sample). A decision tree can be induced that will make it possible to assign a class to the dependent factor of a new case in the test sample based on the values of independent factors.

Applications of a decision tree based classification include target marketing, churn prediction, medical diagnosis, and so on. For instance, Bernstein and Provost [5] used decision trees in the development of a knowledge discovery assistant, in order to categorize different methods used to solve a specific problem. Endou and Zhao [12] investigated a decision tree implementation method that relied on an evolution of the training data set used. The training data set was developed to give the best coverage of the domain knowledge. Markey et al. [25] adapted a simple decision tree to the classification for lung cancer patients of clinical specimens as diseased/non-diseased. Zmazek et al. [39] used decision trees to predict radon gas concentration from other environmental factors, leading to a possible future earthquake prediction