Expert Systems with Applications 38 (2011) 4326–4338

Contents lists available at ScienceDirect

Expert Systems with Applications

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

The hybrid model of neural networks and genetic algorithms for the design of controls for internet-based systems for business-to-consumer electronic commerce

Sangjae Lee [a,*], Hyunchul Ahn [b,*]

[a] Department of E-business, College of Business Administration, Sejong University, 98 Kunja-dong, Kwangjin-gu, Seoul 143-747, Republic of Korea
[b] School of Management Information Systems, Kookmin University, 861-1, Jeongneung-dong, Seongbuk-gu, Seoul 136-702, Republic of Korea

ARTICLE INFO

Keywords:
ISB2C
ISB2C controls
Neural networks
Genetic algorithms
ISB2C implementation

ABSTRACT

As organizations become increasingly dependent on Internet-based systems for business-to-consumer electronic commerce (ISB2C), the issue of IS security becomes increasingly important. As the usage of security controls is related to the implementation of ISB2C, the extent of ISB2C controls can be adjusted in order to enable the greatest extent of implementation of ISB2C. This study intends to propose ISB2C-NNGA (ISB2C-controls design using neural networks and genetic algorithms), a hybrid optimization model using neural networks and genetic algorithms for the design of ISB2C controls, which uses back-propagation neural networks (BPN) model as a prediction of controls using system environments, and GA as a pattern directed search mechanism to estimate the exponent of independent variables (i.e., ISB2C controls) in multivariate regression analysis of *power* model. The effect of system environments on controls can be estimated using BPN model which outperformed linear regression analysis in terms of square root of mean squared error. The effect of each mode of controls on implementation (volume) can be identified using exponents and standardized coefficients in the GA-based nonlinear regression analysis in ISB2C-NNGA. ISB2C-NNGA outperformed conventional linear regression analysis in prediction accuracy in terms of the average *R* square and sum of squared error. ISB2C can suggest the best set of values for controls to be recommended from several candidate sets of values for controls by identifying the set of values for controls which produce greatest extent of ISB2C implementation. The results of study will support the design of ISB2C controls effectively.

1. Introduction

The Internet has changed the way people think, do business, and communicate with each other. Internet-based systems for business-to-consumer electronic commerce (ISB2C) is an application of electronic commerce that allows firms conduct business transactions with consumers over Internet-based information systems such as shopping malls, portals, and web-based systems.

The downside to this is that while online, all Internet-based electronic commerce is vulnerable to misuse either by unauthorized users penetrating the system or by authorized users abusing their privileges. The 2008 Computer Security Institute/Federal Bureau of Investigation (CSI/FBI) found that respondents' estimate of the losses caused by various types of computer security incident was $288,618 for the 522 respondents (Computer Security Institute, 2008). The CSI study indicated that almost half of companies had experienced one to five security incidents in the previous year. Although IS security is not the only one slowing down the proliferation of e-commerce, lack of security is still one of the most likely reason for the low utilization of online selling and electronic payment systems (Lee, Han, & Kym, 1998; Suh & Han, 2003).

In the context of ISB2C, this paper provides a novel approach to selecting and recommending the appropriate controls for successful implementation. The approach provides a back-propagation neural networks and genetic algorithm (GA) based approach, i.e., ISB2C-NNGA (ISB2C controls design using neural networks and genetic algorithms), a hybrid optimization model using neural networks and genetic algorithms for the design of ISB2C controls, which uses back-propagation neural networks (BPN) model as a prediction of controls using system environments, and GA as a pattern directed search mechanism to estimate the exponent of independent variables (i.e., ISB2C controls) in nonlinear regression analysis of power model. The effect of system environments on controls can be estimated using BPN model. The effect of each mode of controls on implementation (volume) can be identified using exponents and standardized coefficients in the GA-based nonlinear regression analysis in ISB2C-NNGA. ISB2C can suggest the best set of values for controls to be recommended from several candidate sets of values for controls by identifying the set of values for controls which produce greatest extent of ISB2C implementation. This approach is a hybrid approach that combines BPN and

* Corresponding authors.
E-mail addresses: sangjae@sejong.ac.kr (S. Lee), hcahn@kookmin.ac.kr (H. Ahn).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.09.102

GA to identify the extent of controls that maximizes the ISB2C implementation. GA based nonlinear regression analysis is used to identify the set of values for controls which lead to greatest ISB2C implementation.

In order to evaluate the effectiveness of ISB2C-NNGA, the prediction accuracy of ISB2C-NNGA is compared with that of multivariate linear regression analysis.

2. Theoretical background

2.1. Neural networks

Neural network technology has been used for supporting business decisions. Neural networks adapt well to changing circumstances and environmental factors, and handle the fuzziness and bias aspects of human decision making (Hill & Remus, 1994). The problems where neural networks have been applied include image recognition (letter and word recognition), data compression, signal classification, financial prediction (stock market prediction, prediction of credit card fraud, prediction of bankruptcy and financial distress), diagnostic networks for medical diseases, and function approximation (Fanning & Cogger, 1994; Trippi & Turban, 1993). It has been proposed that the hybrid model that uses statistical techniques (e.g., logit, stepwised regression, and discriminant analysis) to select input variables in neural network model has a potential to achieve a high predictive performance (Lee, Han, & Kwon, 1996). Neural networks have the potential for decision support tool in nonconservative/unstructured domains that either have a poorly defined domain model or lack a known model (Dutta, Shekhar, & Wong, 1994).

A neural network contains no domain knowledge in the beginning, but it can be trained to make decisions by mapping exemplar pairs of input data into exemplar output vectors, and adjusting its weights so that it maps each input exemplar vector into the corresponding output exemplar vector, approximately (Hecht-Nielsen, 1990). A knowledge base pertaining to the internal representations (i.e., the weight values) is automatically constructed from the facts presented to train the network.

The major obstacle for rule-based expert systems lies in the difficulty of generating a knowledge base. Neural networks circumvent some of the limitations associated with rule-based expert systems development approach. Neural networks automatically extract the functional relationships between input and output using a learning algorithms. The knowledge about the mapping from input to output is encoded in the size of the weights for the connections through the learning process. The major limitation of neural networks, however, is the lack of "readability" of the process the system used to make the knowledge base. Since knowledge is distributed over the entire network, it is difficult to explain why a particular conclusion is reached (Yoon, Brobst, Bergstresser, & Peterson, 1989).

2.2. Genetic algorithms

GA is most commonly applied to global optimization, especially in complex nonlinear problems and in situations when mathematics determinate solution cannot be obtained. It adopts random Heuristic algorithm to search complex solution space for the similar optimal solution. GA, originally developed by Holland (1975), is a method for optimizing machine learning algorithms inspired by the processes of natural selection and genetic evolution (Goldberg, 1989). GA is a simple yet powerful tool for finding the global solution to an optimization problem. It is suitable for large-scale optimization problems, has tendency to find the global optimal solution. GA as a tool for search and optimizing methodology has now reached a mature stage. The particular steps are described as follows:

(1) Initialize a usually random population in which individuals characterized by chromosomes represent a set of possible solutions.
(2) Produce the fitness function which reflects the degree of "goodness" of the individuals for the problems, and evaluate the fitness of all individuals of the population.
(3) Search randomly the fittest individuals to be parents for reproducing offspring. The searching mechanism ensures that the fitter individuals have a tendency to yield good-quality offspring.

Because GA simultaneously evaluates many points in a parameter space, it is more likely to ultimately converge on the global solution. In particular, there is no requirement that the search space is differentiable or continuous, and the algorithm can iterate several times on each data point. Accordingly, it is a very suitable approach for time-varying nonlinear functions.

GAs have been used in a number of applications in engineering and social science. Recently, they applied to, for example, the design of EDI controls (Lee, 2000), the hybrid with association rule algorithms for tree induction in predicting the student course performance (Hsu, Lai, & Chiu, 2003), support of rule induction dealing with both continuous and categorical data types (Chiu & Chiu, 2004), parallel machine-scheduling problem (Chang, Chen, & Lin, 2005), matching information security vulnerabilities to organizational security profiles (Gupta, Rees, Chaturvedi, & Chi, 2006), optimization of the parameters of support vector machine for predicting bankruptcy (Wu, Tzeng, Goo, & Fang, 2007), parallel searching for optimal feature subset (Li, Zhang, & Zeng, 2009), efficient selection and assignment of Material Handling Equipment (Mirhosseyni & Webb, 2009).

2.3. ISB2C Controls for ISB2C implementation

The security effectiveness depends on various organizational factors such as size, top management support, industry type, managerial attitudes toward security risks, IT resource posture, and executive management support (Kankanhalli, Teo, Tan, & Wei, 2003; Kotulic & Clark, 2004). Using the theories of innovation adoption, some researchers have suggested various factors on security adoption such as firm size, industry type, top management support, moral compatibility, peer influence, and computing capacity (Lee & Kozar, 2005). Industry type and organizational use of IT were regarded as the two factors that influence security adoptions (Yeh & Chang, 2007). This study suggests five factors representing organizational and IS related factors. The factors include top management support, system compatibility, IS infrastructure, IS expertise, and information contents.

While automating organizational functions, cost of errors can become quite high. In order to give a belief that system is safe and accurate to users or increase the capability of implementation, management should establish security of information using policies, organizational structures, practices, and procedures for the controls of ISB2C before expanding the implementation of ISB2C. If powerful control mechanisms into computers and networks, both hardware-based as well as software-based, are not ensured, some internal applications planned for computerization might be performed manually (Parker, 1994). Before further implementation of the system is decided, "control assurance" in ISB2C must be provided to stakeholders such as internal users, customers, and industrial associations, in terms of contractual obligations or agreements. Thus, an "adequate" level of controls is a factor affecting successful integration and utilization of systems (Lee et al., 1998). The relationships among system environments, controls and implementation for ISB2C are suggested in Fig. 1.