Diagnosing E-commerce Website Quality Based on DEA

Zijuan Liu
College of Information Engineering
Capital Normal University
Beijing, China
jzi-2008@163.com

***Abstract*—The website quality affects the success of E-commerce directly, and therefore reasonable quality evaluation and diagnosis of E-commerce website is of great practical significance. In this paper, DEA was adopted to evaluate the quality of E-commerce website and a quality diagnosis method was proposed based on DEA. Moreover, the system of quality evaluation target was setup. Finally, this method was proved to be effective and scientific by example analysis.**

Keywords-DEA; e-commerce website; system of evaluation target; quality diagnosis

I. INTRODUCTION

In the network economic times, the website quality affects the success of e-commerce directly [1-2] and the quality of the sites has become a common topic concerned by e-commerce operators and users. Quality is the life, and only to improve the quality of the website can the firms gain a competitive advantage and benefits. With the rapid development and wide application of e-commerce, academia and the business are badly in need of scientific and effective method for website evaluation so as to compare the relative merits of websites and guide website construction. The scientific quality diagnosis for e-commerce will contribute to the improvement of overall level, and promote the healthy development of e-commerce.

At present, the Chinese scholars have less research on the quality of e-commerce website [3]. In this paper, a quality diagnosis method was proposed based on DEA. Moreover, the system of quality evaluation target was setup and the detailed steps of quality diagnosis were described. Finally, this method was proved to be effective and scientific by example.

II. DEA PRINCIPLE AND EVALUATION MODEL

A. DEA Principle

The thought of Data envelopment analysis can be traced back to 1957 in which Farrell proposed the concept of technical efficiency and the thought of envelope when he carried out empirical research on agricultural productivity in the United States [4]. In 1978 A. Charnes, a famous operations researcher proposed a non-parametric DEA method based on the concept of relative efficiency and mathematical programming tools [5]. This method not only can use linear programming to determine whether the point of decision making units is in the production frontier surface, but also can get a lot of useful management information. DEA is a purely technical production frontier non-parametric analysis method, whose evaluation results entirely rely on the objective data and won't be impacted by the difference of input and output indicators dimension. As an objective multi-criteria decision making approach, DEA doesn't require pre-given weighting coefficient of each indicator or to make some subjective judgments to determine the weight coefficients in advance. This method highlights the superiority in the relative effectiveness evaluation of Decision-making units (DMU) with multiple inputs and outputs indicators [6]. Its results reflect the actual state of DMU and also contain profound economic implications which can be used to improve and enhance the DMU's efficiency.

B. Evaluation Model

Before presenting the evaluation model, we first need to outline the conceptual model which provides the basis for our analytical method. We conceptualize consumer-website interaction during online shopping as a production process in which the customer conducts a purchase transaction by utilizing some major functionalities provided by the e-commerce website. In economics, the production process defines the technical means by which inputs are converted into outputs. This technical relationship is represented by the production function, which articulates the maximum level of outputs produced for each given level of inputs. Deviations from the production frontier reflect inefficiencies in production [7].

For our current analyses, we employ the input-oriented BCC model as the evaluation model. There are a variety of DEA models, which evaluate DMU from different angles. Since the decision-making is easier to control in the analysis of site quality and goal-directed purchasing is more prevalent in consumer behavior, we select the input-oriented DEA model which is appropriate for modeling online purchase situations where goal-directed purchasing is prevalent . Besides, the CCR model[8] allows for constant returns to scale, whereas the BCC model allows for variable returns to scale in the production process[9].Our analyses show that the online shopping production process does in fact exhibit variable returns to scale, hence the BCC model formulation is appropriate here. The evaluation model is as follows:

$$\begin{cases} \min[\theta - \varepsilon(E_1^T S^- + E^T S^+)] = V_D \\ s.t. \\ \sum_{j=1}^{n} X_j \lambda_j + S^- = \theta X_{j_0} \\ \sum_{j=1}^{n} Y_j \lambda_j - S^+ = Y_{j_0} \\ \sum_{j=1}^{n} \lambda_j = 1 \\ \lambda_j \geq 0, \quad j=1,2,...n \\ S^+, \quad S^- \geq 0 \end{cases}$$

Where

θ = the efficiency index,

E1, E = column vector of m-dimensional and s-dimensional for the elements with full amount,

ε = the non-Archimedean infinitesimal smaller than any positive number of non-negative, which has often been applied to take a small enough positive number, such as 10^{-6},

λ_j = the right weight of DMU, $j = 1,2,...n$

Y_j = vector of s outputs (y_{sj}) resulting from the production, s>0, $y_{sj} \geq 0$

X_j = vector of m inputs (x_{mj}) used in the production process, m>0, $x_{mj} \geq 0$

S^+ = vector of s deviations from production frontier (s_s^+), s>0, $s_s^+ \geq 0$

S^- = vector of m deviations from production frontier (s_m^-), m>0, $s_m^- \geq 0$.

III. THE SYSTEM OF QUALITY EVALUATION TARGET

Building the system of quality evaluation target, not only need to have the users' behavior as a basic starting point, but also need to consider the accessibility of the data and the actual size of the workload. The potential customer will become a site-value customer with the purchase behavior in e-commerce website. Generally, the customers' transaction information will be stored in the database, which we can use to analyze the customers' access behavior so as to evaluate and diagnose the quality of e-commerce sites.

A. DMU

In DEA, the unit of analysis is called the decision-making unit (DMU) which represents a production unit. A DMU may be defined as an individual or as broadly as an industry, a firm or even as an economy. DEA estimates the relative efficiencies of DMUs from observed measures of inputs and outputs. The productivity of a DMU is evaluated by comparing it against a hypothetical DMU that is constructed as a convex combination of other DMUs in the dataset. In this paper, we define a DMU as a whole that registered users visit e-commerce sites to complete the transactions during a week.

B. Input and Output Variables

DEA is only as good as the initial selection of input and output variables. The inputs must represent the resources consumed by the DMUs and the outputs must characterize the end results of the production by the DMUs. Another conceptualization of the outputs is unit performance. This works so long as DEA's axioms of production are satisfied. In online shopping, inputs consist of customers' use of various website functionalities and the output consists of a checkout of a basket of products. The input and output variables are summarized in Table I.

TABLE I. EVALUATION MODEL'S INPUT AND OUTPUT VARIABLES

CATEGORY	VARIABLE	MEASURE	DESCRIPTION
Inputs	x_1	Product list page	Number of customers visiting product list page
	x_2	Promotion page	Number of customers visiting promotion page
	x_3	Product page	Number of customers visiting product page
	x_4	Shopping cart page	Number of customers visiting Shopping cart page
	x_5	Order process pages	Number of customers submitting order
	x_6	Checkout pages	Number of customers who have been pay success
Output	y_1	Basket Size	Number of different products at checkout

Taken together, the six input variables (x_1 through x_6) represent all major website functionalities a customer has used in conducting her purchase transaction. The output measure, the number of different products at checkout (y_1 = Basket Size), represents the level of performance of the online shopping production process.

C. Evaluation Metrics

We need to delve deeper into the potential causes of the major website functionalities' inefficiencies. One straightforward way to do this is to examine the severity (or scale) of the observed inefficiencies for each major website functionality[10]. Besides, we can analyze the conversion rate for customers between two important function pages. Toward this goal, we propose four metrics:

- Definition1 (Unit Inefficiency):The Unit Inefficiency of website functionality on output represents how much the inefficiencies due to the particular website functionality are with respect to a unit of output. It can be calculated by the following formula (1).

- Definition2 (Con.Rate1): the conversion rate from product page to Shopping cart page for customers. It can be calculated by the following formula (2).
- Definition3 (Con.Rate2): the conversion rate from Shopping cart page to Order process pages for customers. It can be calculated by the following formula (3).
- Definition4 (Con.Rate3): the conversion rate from Order process pages to checkout pages for customers. It can be calculated by the following formula (4).

The four metrics above can be easily computed from the DEA results. The calculation formula for each metric is as follows:

$$\text{Unit Inefficiency} = s_m^- / y_{sj} \quad (1)$$

$$\text{Con.Rate1} = x_4 / x_3 \quad (2)$$

$$\text{Con.Rate2} = x_5 / x_4 \quad (3)$$

$$\text{Con.Rate3} = x_6 / x_5 \quad (4)$$

IV. The Steps of Quality Diagnosis

After building the system of quality evaluation target, we can follow the steps below to diagnose the e-commerce websites.

Step1: Collecting web log data so as to get the value of input and output variables for each DMU after data preparation and data analysis.

Step2: Using MatLab programming to solve the DEA model and getting the slack variable values (s_m^-)of all input variables for each DMU from the DEA results.

Step3: Calculating the Unit Inefficiency of all input variables for each DMU, and then obtaining the sum and average value of Unit Inefficiency for every input variable.

Step4: Analyzing the major website functionalities' inefficiencies according to the results in step 3.

Step5: Calculating Con.Rate1, Con.Rate2 and Con.Rate3 of each DMU, and then obtaining the average value of each conversion rate.

Step 6: Analyzing the important factor affecting the quality of website based on the above results, and diagnosing the major website functionalities' inefficiencies in the end.

V. Empirical Application

A. Research site and data collection

Data for this study were collected from a B2C website. Clickstream data were collected directly from the B2C site's web servers. The website uses HTTP session cookies downloaded onto the visitor's computer to track the customer's shopping behavior at the website. Typical data pre-processing procedures for using web server logs were used to extract navigation path sequences for visitors from the clickstream data [11]. The navigation sessions were combined to identify purchase transactions. Then website evaluation metrics were extracted to measure the extent to which major functionalities of the website were used in each of the purchasing processes. Our analysis focuses on the registered customers' transaction data for consecutive 12 weeks. The data of input and output variables are shown in Table II.

TABLE II. The Data of Input and Output Variables

DMU	Inputs						Output
	x_1	x_2	x_3	x_4	x_5	x_6	y_1
DMU1	538620	653516	362525	13548	9922	9635	4
DMU2	526768	635865	370669	13256	11543	10652	6
DMU3	652408	715462	431452	16682	14355	12893	8
DMU4	642509	746208	432428	16214	13981	11567	5
DMU5	672486	779879	486062	19687	17168	16952	9
DMU6	516758	622850	386314	14985	11998	10096	5
DMU7	476056	571896	366119	15231	11543	10820	7
DMU8	762209	880106	551641	19658	17519	15783	13
DMU9	652486	748653	488913	17056	14179	14025	9
DMU10	712208	812312	534246	18342	15664	14936	9
DMU11	752305	848375	578817	19023	16562	15894	12
DMU12	723642	819876	545555	16524	13340	13125	7

B. Results

To gain insights into the potential causes of the major website functionalities' inefficiencies, we analyzed the inefficiencies by website functionality with the metrics proposed earlier. The Unit Inefficiency can be computed from the DEA results. (See Table III).

TABLE III. Unit Inefficiency of Website Functionality

VARIABLE	WEBSITE FUNCTIONALITY	UNIT INEFFICIENCY	
		TOTAL	*AVERAGE*
x_1	Product list	0.5736	0.0478
x_2	Promotion	0.7222	0.0602
x_3	Product information	0.4094	0.0341
x_4	Shopping cart	0.0190	0.0016
x_5	Order process	0.005	0.0004
x_6	Checkout	0.0091	0.0008

The results of other three metrics is as follows:

Con.Rate1(average)=3.65% ;

Con.Rate2(average)=84.43% ;

Con.Rate3(average)=93.12% .

As we can see from the Table III, Promotion, Product List and Product information are most problematic in incurring inefficiencies at the B2C website. Inefficiencies due to Shopping cart, Order process and Checkout are quite insignificant. According to the conclusions, we can consider whether the goods and services in promotion satisfy users' needs and the Information Architecture of home page is sufficient to allow users to obtain the desired information in the shortest time. Moreover, the average of Con.Rate1 is the lowest (Con.Rate1(average)=3.65%)which means that Product information is the most important problem influencing the customers from product page to checkout pages. Therefore, we should demonstrate the relevant details of commodity services in the product page as logical as possible and highlight the buy button so as to convince customers to buy the products.

VI. CONCLUSION

In this paper, a quality diagnosis method was proposed based on DEA. Moreover, the system of quality evaluation target was setup from a microscopic point and the detailed steps of quality diagnosis were described. Finally, this method was proved to be effective by example. Studies have shown that the proposed method can help enterprises to find fundamental and local factors impacting site quality. Besides, it provides important scientific basis for improving the site and contributes to the improvement of the overall level and quality for e-commerce websites.

REFERENCE

[1] Torkzadeh G, Dhillon G (2002) "Measuring factors that influence the success of internet commerce," Information System Research (2)13, pp. 187- 204.

[2] McKinney V, Yoon K, Zahedi F (2002) "The measurement of web - customer satisfaction: An expectation and disconfirm ation approach," Information System Research (3)13, pp.296- 315.

[3] CHANG Jinling, XIA Guoping (2006) "Modeling E-commerce website quality management based on Bayesian network," Journal of Tsinghua University (S1) 46, pp.1002-1006.

[4] FARRELL, MJ (1957) "The measurement of productive efficiency," Journal of Royal statistical Society (3)120, pp.253-290.

[5] CHARNES A, COOPER, W W. RHODES E (1978) "Measuring the efficiency of decision making units," European Journal of Operational Research (6) 2, pp.429-444.

[6] WEI Quanling (2004) "Data envelopment analysis," Beijing Science Press (45)17, pp.1793-1807.

[7] Aigner, D. J. and S. F. Chu (1968) "On Estimating the Industry Production Function," American Economic Review (58)4, pp. 826-839.

[8] Charnes, A., W. W. Cooper and E. Rhodes (1981) "Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through," Management Science (27)6, pp. 668-697.

[9] Banker, R. D., A. Charnes and W. W. Cooper (1984) "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis," Management Science (30)9, pp. 1078-1092.

[10] Jungpil Hahn, Robert J. Kauffman (2004) "Identifying E-Commerce Website Design Inefficiencies: A Business Value-Driven Approach Using DEA," Journal of Pre-ICIS Annual Workshop on HCI Research (1) 3, pp.1-30.

[11] Cooley, R., B. Mobasher and J. Srivastava (1999) "Data Preparation for Mining World Wide Web Browsing Patterns," Journal of Knowledge and Information Systems (1)1, pp. 5-32.