Bayesian Network based Cost Benefit Factor Inference in E-Services

Guangquan Zhang, Chenggang Bai, Jie Lu and Chengqi Zhang

Abstract—This paper applies Bayesian network technique to model and inference the uncertain relationships among cost factors and benefit factors in E-services. A cost-benefit factor-relation model proposed in our previous study is considered as domain knowledge and the data collected through a survey is as evidence to conduct inference. Through calculating conditional probability distribution among factors and conducting inference, this paper identifies that certain cost factors are significantly more important than others to certain benefit factors. In particular, this study found that 'increased investment in maintaining E-services' would significantly contribute to 'enhancing perceived company image' and 'gaining competitive advantages', and 'increased investment in staff training' would significant contribute to 'realizing business strategies'. These results have the potential to improve the strategic planning of companies by determining more effective investment areas and adopting more suitable development activities where E-services are concerned.

Index Terms—About four key words or phrases in alphabetical order, separated by commas (see "Subject Categories" in http://www.ieee.org/web/developers/webhies/index.htm).

I. INTRODUCTION

The term 'E-services' is typically used to describe a variety of electronic interactions, ranging from basic services, such as the delivery of news and stock quotes, to smart services, such as the delivery of context-aware emergency services [3], and provide of recommendations. The ability of E-services to fulfill customer demands is assisting businesses in reducing service costs and obtaining more benefits. Companies in the earlier stages of employing E-services have had little data, knowledge and experience of the potential of Bservices for organizational impacts and benefits. After several years experience of E-services, companies can obtain related knowledge and provide related data. They urgently need to weight the costs involved in moving services online against the benefits received by adopting E-services. They must identify what kinds of investment effectively contribute to particular benefit aspects of an E-service application.

Recent reports concerning the success, quality, usability and benefit of E-services have led researchers to express increasing interest in conducting an evaluation of the use of E-service applications [20]. In general, various research methods and techniques used in the research of E-service evaluation, such as surveys, cases and modelling, are instigated under three major categories. The first is website feature, function or usability evaluation. Typical approaches in the categories are testing, inspection and inquiry [6], and they are often used together in conducting a web search or a desk survey [18]. The second category is investment analysis which has been conducted for justifying investment in an E-service application, and for exploring the changes that take place in organizational operations. Significant results have been reported in, such as, Giaglis et al. [5], Drinjak et al. [4], and Amir et al. [1]. The third is the establishment of evaluation frameworks or models. For example, Lee et al. [13] created a framework for evaluating the business value of B2B E-service through five propositions. Zhang and von Dran [21] developed a two-factor model for website evaluation. More generally, Hahm et al. [6] presented a value-driven framework for E-commerce website evaluation.

Our research reported in Lu [14], Lu & Zhang [15, 16] identified the inter-relationships and interactive impacts among E-service functions, E-service development attributes, the benefits received via adopting an E-service, and the costs to move service operations online. In particular, Lu & Zhang [16] examined E-services' investment in which aspects have a more significant contribution to particular benefit items. As a further study of previous work, this paper aims to apply Bayesian network technique to inference the uncertain relationships among cost factors and benefit factors in E-services.

After the introduction, this paper reviews our previous work including the research framework for E-service evaluation, data collection and the cost-benefit factor-relation model in Section 2. Section 3 introduces Bayesian network techniques. Section 4 reports the inference results among cost factors and benefit factors conducted by applying Bayesian network. Conclusions are discussed in Section 5.

II. PREVIOUS RESEARCH REVIEW

A. FCBD research framework

Lu & Zhang [16] established a conceptual research framework for assessing E-service applications which has four categories: E-service function (F), E-service cost (C), E-service benefit (B), and E-service development attribute (D), called FCBD research.
framework. E-service function here is concerned with the capability and quality of the E-services. Cost is the expenses incurred in adopting E-services. E-service benefit is concerned with the benefits gained through employing E-services, which development attribute takes into account the strategies, policies and types of companies involved when developing E-service applications. Each category consists of a set of evaluation factors (Fig. 1). \( H_a \) \( \ldots \) \( H_b \) imply a set of respective hypotheses.

![Fig. 1: FCBD research framework with factors](image)

**B. Data collection**

This study collected data concerning E-service development attributes, functions, costs and benefits from a sample of Australia companies (E-service providers). These sample companies were selected from two industry categories: Tourism (including Travel, Accommodation, Entertainment and Health care) and IT/Communication services (IT services and Information services) in Australia.

This study conducted a web search first to determine a sample of companies which have adopted E-services on an appropriate level and volunteers were obtained from these companies. A total of 100 websites were randomly selected from company websites registered in the Yellow Pages Online (NSW, Australia) [http://www.yellowpages.com.au](http://www.yellowpages.com.au) under Tourism/Travel and IT/Communication categories. A questionnaire survey was then conducted with the sample companies from February to March 2002. As a pre-test survey, an initial questionnaire was sent to three subjects as a way of setting initial feedback. Based on the pre-test results, the questionnaire was refined. The final questionnaire was then posted, emailed, or faxed, to the 100 selected companies. Out of 34 questions in the questionnaire, some items were related to E-service functions and development attributes, some were related to the costs of developing E-service applications, and some were related to the benefits obtained from developing E-service applications. The survey assumes that respondents represent their colleagues and they should not be asked directly about hypotheses. A total of 50 responses were obtained, and the results shown in this paper are based on 48 completed responses. All cost factor questions listed in the questionnaire use a five-point Likert scales: 1—very important. For example, if a company thinks the cost of maintaining an E-service is very important it records the degree of importance as 4 or 5. A 5-point scale is also used for present benefit assessment: 1—low benefit, 5—very high benefit. For example, if a company considers that, currently, their E-service only builds very basic customer relations, and the company would ideally prefer to build closed relations, then the company would score perhaps 3 on the present benefit assessment for B. The survey result was used to identify why companies adopt E-service applications, how they evaluate an E-service application, what the main benefit factors are, and what kinds of benefits have been obtained. It also identified the major costs and barriers of E-service applications and, most importantly,
which cost items significantly contributed to particular benefit items.

C. Cost-benefit factor-relation model

By completing a group of ANOVA tests, a set of 'effect' relationships between cost and benefit factors are obtained. These relationships reflect that certain cost factors have a significant effect on certain benefit factors. These effects are presented in a cost-benefit factor-relation model (Fig. 2). The lines in the model express the 'effect' relationships between related cost factors and benefit factors. Although every cost factor makes direct or indirect contributions to all benefit factors to a certain degree, some cost factors are more important for the improvement of particular benefit factors than others.

![Fig. 2: Cost-benefit factor-relation model](image)

III. BAYESIAN NETWORK

Bayesian network is a powerful knowledge representation and reasoning tool under conditions of uncertainty [19]. A Bayesian network $B = (N, A, \Theta)$ is a directed acyclic graph (DAG) $<N, A>$ with a conditional probability distribution (CPD) for each of its nodes, collectively represented by $\Theta$, for each node $n \in N$ represents a variable, and each arc $a \in A$ between nodes represents a probabilistic dependency [19]. In a practical problem, the nodes of a Bayesian network represent uncertain variables, and the arcs are the causal or influential links between the variables. The association with each node is a set of CPDs that model the uncertain relationships between the node and its parent nodes.

The benefits of using Bayesian network to model uncertain relationships have been well discussed [8, 11]. Many applications have proven that Bayesian network is an extremely powerful technique for reasoning the relationships among a number of variables under uncertainty. For example, Heckerman [7] applied Bayesian network successfully into lymph-node pathology diagnosis. Breese & Blake [2] applied Bayesian network successfully into computer default diagnosis.

Comparing with other inferencing analysis approaches, Bayesian network has four good features in its applications. Firstly, unlike neural network approach, which usually appears to user as a "black box", all the parameters in a Bayesian network have an understandable semantic interpretation [17]. This makes users to construct a Bayesian network directly by using domain expert knowledge. Secondly, Bayesian network has ability to learn the relationships among its variables. This not only lets users observe the relationships among variables easily, but also can handle data missing issue [10]. Thirdly, Bayesian network can conduct inference inversely. Feed-forward neural networks and fuzzy logic approaches are strictly one-way, that is, when a model is given a set of inputs it can predict the output, but not vice versa. Fourthly, Bayesian network can combine prior information with current knowledge to conduct inference as it has both causal and probabilistic semantics. This is an ideal representation for users, for example, experts, to give prior knowledge which often comes in a causal form [10].

IV. BAYESIAN NETWORK BASED COST-BENEFIT FACTOR ANALYSIS

In general, there are three main steps when applying Bayesian network approach in a practical problem: creating a graphical structure, calculating the conditional probabilities, and finally using the model to do inference.

A. Creating a graphical structure

As $B_1, B_2, B_3, B_4, B_5$, and $B_6$ in Fig. 2 don't have any connections to any cost factor nodes, an initial Bayesian network structure of cost and benefit factors relationships can be created by deleting these notes, shown in Fig. 3. The lines in the structure express the 'effect' relationships between these factors.

![Fig. 3: Initial cost-benefit factor-relation Bayesian network](image)
study uses a local search algorithm, greedy Hill-climbing [9].
This algorithm starts at a specific point in a space, checks all
neighboring scores, and then moves to the neighbor that has the
highest score. If all neighbors' scores are less than the current
point, that is, a local maximum is reached, the algorithm will stop
and/or restart in another point of the space. By running the
Hill-climbing algorithm for structure learning from data, Fig. 4 is
obtained where the link between C2 and B1 is deleted.

Fig. 4: Cost-benefit factor-relation Bayesian network after
structure learning from data collected

B. Calculating the conditional probability distributions
(CPD)
Let \( X = (X_1, \ldots, X_n) \) be a note set, \( X_i \) \( (i=0,1,\ldots,m) \) is a
discrete node, in a Bayesian network \( E, m=19 \) for Fig. 4. The CPD
of node \( X_i \) is defined as \( \theta^a_{i,p} = P(X_i = x_i | P_a = p_a) \) [9],
where \( P_a \) is the parent set of node \( X_i \), \( P_a \) is a configuration
(set of values) for the parent set \( P_a \) of \( X_i \), and \( x_i \) is a value
that \( X_i \) \( (i=0,1,\ldots,m) \) takes. Based on the data collected from
the survey, CPDs of all nodes are calculated and shown in Fig. 4.
This paper only shows and discusses the inference results related to nodes C2 and G which are assumed
respectively to be 'high (4)'. Similar results can be obtained
when these notes get other values. Table 3 shows the
probabilities of all nodes under the evidence \( C_2=4 \) (high).

Table 3: Probabilities of all nodes when \( C_2=4 \) (high)

\[
\begin{array}{cccccc}
\text{Pr}(\text{state}) & \text{node} & 1 & 2 & 3 & 4 & 5 \\
\hline
C_2 & 0.1469 & 0.1265 & 0.2082 & 0.3510 & 0.1673 \\
C_3 & 0.3102 & 0.2082 & 0.3102 & 0.1469 & 0.0245 \\
C_4 & 0.1265 & 0.1469 & 0.3102 & 0.2694 & 0.1469 \\
C_5 & 0.1469 & 0.1729 & 0.2694 & 0.3494 & 0.1469 \\
B_1 & 0.1478 & 0.1469 & 0.2694 & 0.3494 & 0.1469 \\
B_2 & 0.1265 & 0.1469 & 0.2694 & 0.3494 & 0.1469 \\
B_3 & 0.1478 & 0.1472 & 0.2694 & 0.3494 & 0.1469 \\
B_4 & 0.1478 & 0.1469 & 0.2694 & 0.3494 & 0.1469 \\
B_5 & 0.1478 & 0.1472 & 0.2694 & 0.3494 & 0.1469 \\
\end{array}
\]

It can be found from the two tables that the relationships
among notes C2 and B3, C3 and B3 are hardly linear. Therefore,
using conditional probabilities to express these relationships
will be more suitable than using traditional linear regression
methods.

C. Inference
The cost-benefit factor-relation Bayesian network has been
created with both structure and conditional probabilities are
defined. It can be thus used for inference the effects and
relationships between cost and benefit factors. Basically,
inference is done by fixing the states of observed variables, and
then propagating the beliefs around the network until all the
beliefs (in the form of conditional probabilities) are consistent.
The desired probability distributions can be read directly from
the network.

There are a number of algorithms for doing the inference in
Bayesian networks, which make different tradeoffs between
speed, complexity, generality, and accuracy. Junction-tree
algorithm, developed by Lauritzen & Spiegelhalter [12], is one of
the most popular algorithms and used in this study. It is based
on a deep analysis of the connections between graph theory and
probability theory. It uses an auxiliary data structure called a
junction tree, and suitable for medium and small size of samples.

Through running Junction-tree algorithm for inference, a
group of valuable results between C and B (i=1 to m) are
obtained. This paper only shows and discusses the inference
results related to notes C2 and G which are assumed
deeply connected to B3 (Table 1), and C3 to B3 (Table 2).

Table 1: The conditional probabilities for node B1 based on C2

\[
\begin{array}{cccccc}
\text{Pr}(\text{state}) & \text{node} & 1 & 2 & 3 & 4 & 5 \\
\hline
C_2 & 0.0065 & 0.2390 & 0.0965 & 0.4903 & 0.1677 \\
C_3 & 0.1020 & 0.0039 & 0.2980 & 0.4941 & 0.1020 \\
C_4 & 0.1672 & 0.2492 & 0.4313 & 0.0033 & 0.1672 \\
C_5 & 0.1767 & 0.0605 & 0.4093 & 0.2930 & 0.0605 \\
C_6 & 0.0125 & 0.3250 & 0.3250 & 0.0125 & 0.3250 \\
\end{array}
\]

Table 2: The conditional probabilities for node B3 based on C1

\[
\begin{array}{cccccc}
\text{Pr}(\text{state}) & \text{node} & 1 & 2 & 3 & 4 & 5 \\
\hline
C_1 & 0.0043 & 0.1130 & 0.3304 & 0.4391 & 0.1130 \\
C_2 & 0.1130 & 0.0043 & 0.5478 & 0.1130 & 0.2217 \\
C_3 & 0.0542 & 0.2625 & 0.3667 & 0.2104 & 0.1063 \\
C_4 & 0.1130 & 0.1130 & 0.4391 & 0.3304 & 0.0043 \\
C_5 & 0.0182 & 0.0182 & 0.4727 & 0.4727 & \\
\end{array}
\]

<table>
<thead>
<tr>
<th>BI1</th>
<th>BI2</th>
<th>BI3</th>
<th>BI4</th>
<th>BI5</th>
<th>BI6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0553</td>
<td>0.1469</td>
<td>0.3918</td>
<td>0.2694</td>
<td>0.1265</td>
<td></td>
</tr>
<tr>
<td>0.0023</td>
<td>0.1186</td>
<td>0.3512</td>
<td>0.2930</td>
<td>0.2349</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 5 shows the effect of observing when the value of C (maintaining E-service) is 'high' (=4). The probability of a high B1 is increased from 0.2427 to 0.3494, suggesting that C and B1 are correlated to some extent, that is, a high C tends to "cause" a high B1. It is also found that the probability of a high B2 (enhancing perceived company image) is increased its value from 0.2670 to 0.3667, B3 (gaining and sustaining competitive advantages) is increased from 0.2490 to 0.2930. Same as above these results mean that C is correlated with B4 and B5, when a high investment in E-service maintenance (C1) will tend to "cause" a high enhancement of company image (B4) and gain competitive advantages (B5).

Table 4: Probabilities of the nodes when C=4 (high)

<table>
<thead>
<tr>
<th>CI</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>C8</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
<th>B4</th>
<th>B5</th>
<th>B6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0653</td>
<td>0.1469</td>
<td>0.3918</td>
<td>0.2694</td>
<td>0.1265</td>
<td>0.0023</td>
<td>0.1186</td>
<td>0.3512</td>
<td>0.2930</td>
<td>0.2349</td>
<td>0.0449</td>
<td>0.1469</td>
<td>0.3306</td>
<td>0.2490</td>
</tr>
</tbody>
</table>

Table 5: Prior and posterior probability when C=4 (high)

Fig. 5: Prior and posterior probability when C=4 (high)

Table 4 shows the probabilities of all the other nodes under the evidence C=4 (high). Fig. 6 shows the effect of observing when the value of C=4 (training cost) is 'high'. The probability of a high B1 (realizing business strategies) has increased from 0.2694 to 0.3306, B2 (enhancing perceived company image) is increased from 0.2490 to 0.2930.

Table 4 shows the probabilities of all the other nodes under the evidence C=4 (high). Fig. 6 shows the effect of observing when the value of C (training cost) is 'high'. The probability of a high B1 (realizing business strategies) has increased from 0.2694 to 0.3306, B2 (enhancing perceived company image) is increased from 0.2490 to 0.2930. Same as above these results mean that C is correlated with B4 and B5, when a high investment in E-service maintenance (C1) will tend to "cause" a high enhancement of company image (B4) and gain competitive advantages (B5).

Table 5: Prior and posterior probability when C=4 (high)

By applying Bayesian network technique this paper identifies that certain cost factors were significantly more important than others to certain benefit factors. For example, increased investment in maintaining E-services would significantly contribute to 'enhancing perceived company image' and 'gaining competitive advantages'. This indicates that in order to improve the perceived company image it would be appropriate to invest in 'maintaining E-services'. Another significant finding is that increased investment in staff training would significant contribute to 'realizing business strategies' in E-service application. These results provide an insight into whether investment on certain E-service aspects are perceived as more important that others for specific business objectives.

The findings shown in the study will provide practical recommendations to the following: (1) E-service providers, when forming strategies to reduce E-service costs, increase benefits, enhance E-service functionality and attract customers; (2) E-service application developers, when designing new applications; and (3) E-service managers, for maintaining current E-service applications which provide better services and more effective operations.

REFERENCES


Second International Conference on Information Technology & Applications

Conference Organising Committee

Conference Chair Prof Shi Guanglan Heilongjiang University, China
Conference Vice-Chair Dr. Fu Yuzuo Shanghai Jiao-tong University, China
Technical Chair Dr. Sean He University of Tech Sydney, Australia
International Advisor Dr Dapeng Tien Charles Sturt University, Australia

ICITA 2004 is organised by:

Heilongjiang University, Harbin, China
http://www.hlju.edu.cn

And supported by:

Shanghai Jiao Tong University, Shanghai, China
http://www.sjtu.edu.cn
IEEE, NSW Section, Australia
http://esf.iss.org/118/nsw/
IEEE, CS Chapter, Beijing, China
http://www.cie-china.org/ieee-beijing/
Saura Inc., Japan
http://www.saura.com/