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 Eservices

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

I. INTRODUCTION

The term 'Eservices' 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 fulfil 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 E-services 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 Eservices. They must identify what kinds of

This work was supported in part by the UTS IRG Grant.

Dr G. Zhang, Dr C. Bai, Dr Lu and Prof. C. Zhang are with the Faculty of IT, University of Technology, Sydney, Broadway, NSW 2007 Australia (e-mail: {zhangg, jielu, chengqi}@it.uts.edu.au).

Dr C. Bai is a visiting scholar from Department of Automatic Control, Beijing University of Aeronautics and Astronautics (email: bcg@buaa.edu.cn)

ICITA2004 ISBN 0-646-42313-4

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 Eservice 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 Eservice through five propositions. Zhang and von Dran [21] developed a two-factor model for website evaluation. More generally, Hahn 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), Eservice benefit (B), and E-service development attribute (D), called FCBD research framework. Eservice function here is concerned with the capability and quality of the Eservices. 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} , ..., H_{G} imply a set of respective hypotheses.

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

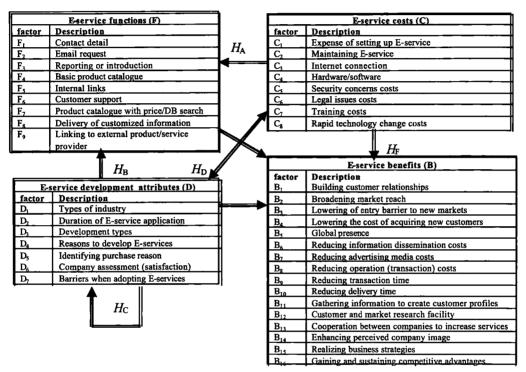


Fig. 1: FCBD research framework with factors

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 Eservices 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 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,

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--not important at all, 5--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 Eservice 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 Eservice applications and, most importantly,

which cost items significantly contributed to particular benefit items.

C. Cost-benefit factor-relation mo del

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.

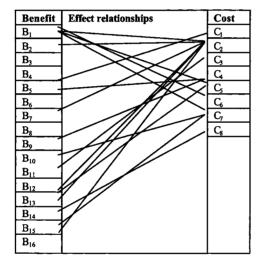


Fig. 2: Cost-benefit factor-relation model

III. BAYESIAN NETWORK

Bayesian network is a powerful knowledge representation and reasoning tool under conditions of uncertainty [19]. A Bayesian network $B = \langle N, A, \Theta \rangle$ is a directed acyclic graph (DAG) $\langle N, A \rangle$ with a conditional probability distribution (CPD) for each of its nodes, collectively represented by Θ , 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 Bayes ian 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_3 , B_4 , B_6 , B_8 , and B_{12} 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.

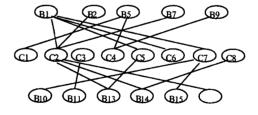


Fig. 3: Initial cost-benefit factor-relation Bayesian network

The notes and relationships shown in Fig. 3 are considered as a result from domain knowledge. Data collected through the survey, discussed in Section 2, is used to complete structure learning of the Bayesian network. A number of structure learning algorithms for Bayesian network have been proposed. This

study uses a local search algorithm, greedy Hill-climbing [9]. This algorithm starts at a specific point in a space, checks all nearest neighbors, 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 C_2 and B_1 is deleted.

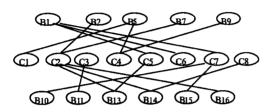


Fig. 4: Cost-benefit factor-relation Bayesian network after structure learning from data colleted

B. Calculating the conditional probability distributions (CPD)

Let $X = (X_0, \dots, X_m)$ be a note set, X_i $(i=0,1,\dots m)$ is a discrete node, in a Bayesian network B, m=19 for Fig. 4. The CPD of node X_i is defined as $\theta_{x_i|Pa_i}^B = P(X_i = x_i \mid Pa_i = pa_i)$ [9], where Pa_i is the parent set of node X_i , pa_i is a configuration (set of values) for the parent set Pa_i of X_i , and x_i is a value that X_i $(i=0,1,\dots m)$ takes. Based on the data collected from the survey, CPDs of all nodes are calculated and shown in Fig. 4. This paper only shows the results related the relationships among C_2 to C_3 (Table 1), and C_3 to C_3 (Table 2).

Table 1: The conditional probabilities for node B₇ based on C₂

Pr(B ₇ /	B ₇ =1	B ₇ =2	B ₇ =3	B ₇ =4	B ₇ =5
C ₂)					
C ₂ =1	0.0065	0.3290	0.0065	0.4903	0.1677
C ₂ =2	0.1020	0.0039	0.2980	0.4941	0.1020
C ₂ =3	0.1672	0.2492	0.4131	0.0033	0.1672
C ₂ =4	0.1767	0.0605	0.4093	0.2930	0.0605
C ₂ =5	0.0125	0.3250	0.3250	0.0125	0.3250

Table 2: The conditional probabilities for node B_{15} based on C_7

Pr(B ₁₅	B ₁₅ =1	B ₁₅ =2	B ₁₅ =3	B ₁₅ =4	B ₁₅ =5
/C ₇)					
C ₇ =1	0.0043	0.1130	0.3304	0.4391	0.1130
C ₇ =2	0.1130	0.0043	0.5478	0.1130	0.2217
C ₇ =3	0.0542	0.2625	0.3667	0.2104	0.1063
C ₇ =4	0.1130	0.1130	0.4391	0.3304	0.0043
C ₇ =5	0.0182	0.0182	0.0182	0.4727	0.4727

It can be found from the two tables that the relationships among notes C_2 and B_7 , C_7 and B_{15} 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 middle and small size of samples.

Through running Junction-tree algorithm for inference, a group of valuable results between G and B_i (i=1 to m) are obtained. This paper only shows and discusses the inference results related to notes 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 C₂=4 (high).

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

		г -	1	ı	
Pr() state	1	2	3	4	5
node					
Cı	0.1469	0.1265	0.2082	0.3510	0.1673
C ₂				1	
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	0.1878	0.2694	0.2490	0.1469
C ₆	0.1878	0.2490	0.3102	0.1673	0.0857
C ₇	0.1878	0.1878	0.3918	0.1878	0.0449
C ₈	0.1673	0.1469	0.2286	0.3918	0.0653
В1	0.1413	0.1684	0.2603	0.2273	0.2027
B ₂	0.0023	0.1186	0.5256	0.1767	0.1767
B ₅	0.1729	0.2388	0.1930	0.2201	0.1753
B ₇	0.1767	0.0605	0.4093	0.2930	0.0605
B ₉	0.1469	0.2082	0.2898	0.1878	0.1673
B ₁₀	0.1469	0.1878	0.3510	0.1469	0.1673
B ₁₁	0.2082	0.1469	0.2898	0.2694	0.0857
B ₁₃	0.1050	0.1666	0.2741	0.3494	0.1050
B ₁₄	0.0030	0.0517	0.2290	0.3667	0.3496

B ₁₅	0.0653	0.1469	0.3918	0.2694	0.1265
B ₁₆	0.0023	0.1186	0.3512	0.2930	0.2349

Fig. 5 shows the effect of observing when the value of Q (maintaining Eservice) is 'high' (=4). The probability of a high B_{13} (cooperation between companies to increase services) is increased from 0.2427 to 0.3494, suggesting that Q and Q are correlated to some extent, that is, a high Q tends to "cause" a high Q to 0.3667, Q tends to "cause" and sustaining competitive advantages) is increased from 0.2490 to 0.2930. Same as above these results mean that Q is correlated with Q and Q will tend to "cause" a high enhancement of company image Q will tend to "cause" a high enhancement of company image Q and gain competitive advantages Q.

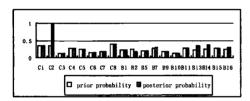


Fig. 5: Prior and posterior probability when C₂=4 (high)

Table 4 shows the probabilities of all the other nodes under the evidence C_7 =4 (high). Fig. 6 shows the effect of observing when the value of C_7 (training cost) is 'high'. The probability of a high B_5 (realizing business strategies) has increased from 0.2694 to 0.3304, suggesting that C_7 and B_{15} are correlated to some extent, that is, a high C_7 tends to "cause" a high B_{15} .

Table 4: Probabilities of the nodes when $C_7=4$ (high)

state Pr() node	1	2	3	4	5
C ₁	0.1469	0.1265	0.2082	0.3510	0.1673
C ₂	0.1265	0.2082	0.2490	0.3510	0.0653
C ₃	0.3102	0.2082	0.3102	0.1469	0.0245
C ₄	0.1265	0.1469	0.3102	0.2694	0.1469
C₅	0.1469	0.1878	0.2694	0.2490	0.1469
C ₆	0.1878	0.2490	0.3102	0.1673	0.0857
C ₇	0	0	0	1	0
C ₈	0.1673	0.1469	0.2286	0.3918	0.0653
Bı	0.1547	0.1408	0.3343	0.2230	0.1471
B ₂	0.0449	0.1061	0.4327	0.2490	0.1673
B ₅	0.1729	0.2388	0.1930	0.2201	0.1753
B ₇	0.1265	0.1469	0.3306	0.2694	0.1265
B ₉	0.1469	0.2082	0.2898	0.1878	0.1673
B ₁₀	0.2217	0.0043	0.5478	0.1130	0.1130
B ₁₁	0.2082	0.1469	0.2898	0.2694	0.0857

B ₁₃	0.1289	0.1785	0.3468	0.2427	0.1031
B ₁₄	0.0412	0.0948	0.2403	0.2670	0.3566
B ₁₅	0.1130	0.1130	0.4391	0.3304	0.0043
B ₁₆	0.0449	0.1469	0.3306	0.2490	0.2286

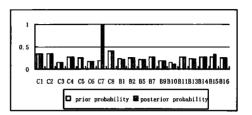


Fig. 6: Prior and posterior probability when C₇=4 (high) V. CONCLUSIONS

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

- Amir Y., Awerbuch B. and Borgstrom R. S. (2000), A Cost-Benefit framework for online management of a metacomputing system, *Decision Support Systems*, Vol 28, No 1-2, pp. 155-164.
- [2] Breese J. & Blake R. (1995), Automating computer bottleneck detection with belief nets. Proceedings of the Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, San Francisco, CA, pp36-45.
- [3] Chidambaram L. (2001), The editor's column: Why e-Service Journal, e-Service Journal, Vol 1, No 1, pp.1-3.
- [4] Drinjak J., Altmann G. and Joyce P. (2001), Justifying investments in electronic commerce, Proceedings of The Twelfth Australia conference on Information Systems, 4-7, December 2001, Coffs Habour, Australia, pp. 187-198.
- [5] Giaglis G. M., Paul R. J. and Doukidis G. I. (1999), Dynamic modelling to assess the business Value of electronic commerce, *International Journal of Electronic Commerce*, Vol 3, No3, pp. 35-51.
- [6] Hahn J. and Kauffman R.J. (2002), Evaluating selling web site performance from a business value perspective, Proceedings of

- international conference on e-Business, May 23-26, 2002, Beijing, China, pp. 435-443.
- [7] Heckerman D. (1990), An empirical comparison of three inference methods. Uncertainty in Artificial Intelligence, edited by Shachter, R., Levitt, T., Kanal, L., & Lemmer, J., North-Holland, New York, pp283-302.
- [8] Heckerman D., Mamdani A. &Wellman M. (1995), Real-world applications of Bayesian networks, Comm ACM, 38(3), pp25-26.
- [9] Heckerman D. (1996), A tutorial on learning Bayesian networks. Technical Report MSRTR95-06, Microsoft Research.
- [10] Heckerman D. (1997), Bayesian Networks for Data Mining, Data Mining and Knowledge Discovery, Vol.1, No.1, pp79-119.
- [11] Jensen F.V. (1996), An Introduction to Bayesian Networks, UCL Press..
- [12] Lauritzen S. & Spiegelhalter D. (1988), Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems (with discussion), J. R. Statis. Soc. B, Vol.50, No. 2, pp 157-224.
- [13] Lee C., Seddon P. and Corbitt B. (1999), "Evaluating business value of internet-based business-to-business electronic commerce", Proceedings of 10th Australia Conference on Information Systems, pp. 508-519.
- [14] Lu J. (2001), Measuring cost/benefits of E-business applications and customer satisfaction, Proceedings of 2nd International We-b Conference, 29-30 November 2001, Perth, Australia, pp. 139-147.
- [15] Lu J. and Zhang G. (2002), Which factors are affecting Esservice benefit -- A research framework, Proceedings of International Conference on E-business (ICEB2002), 23-26 May 2002, Beijing, China, 128-135.
- [16] Lu J. and Zhang G. (2003) Cost benefit factor analysis in E-services, accepted by International Journal of Service Industry Management (IJSIM).
- [17] Myllymaki P. (2002), Advantages of Bayesian Networks in Data Mining and Knowledge Discovery, http://www.bayesit.com/docs/advantages.html.
- [18] Ng H., Pan Y. J. and Wilson T. D. (1998), Business use of the world wide web: A report on further investigations, *International Journal* of Management, Vol 18, No 5, pp. 291-314.
- [19] Pearl J. (1988), Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kauffman, Palo Alto, CA.
- [20] Smith A. G. (2001), Applying evaluation criteria to New Zealand government websites, *International Journal of Information Management*, Vol 21, pp. 137-149.
- [21] Zhang P. and von Dran G., (2000), Satisfiers and dissatisfiers: a two-factor model for website design and evaluation, *Journal of American Association for Information Science (JASIS)*, Vol 51, No 14, pp.1253-1268.

Second International Conference on Information Technology & Applications

Conference Organising Committee

Conference Chair Prof Shi Guangian Heliongliang 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:

Hellongijang University, Harbin, China

http://www.hlju.edu.cn

And supported by:

Shanghat Hao Tong University, Shanghat, China

http://www.situ.edu.cn

IEEE NSW Section Australia

http://ewh.ieeo.org/r10/nsw/

IEEE, CS Chapter, Beljing, China

http://www.cie-china.org/leee-beijing/

Saora Inc., Japan

http://www.sacra.com/

