

# Use of Bayesian belief networks to help understand online audience

Waldek Jaronski, Josee Bloemer, Koen Vanhoof and Geert Wets

Limburgs Universitair Centrum, Universitaire Campus - gebouw D  
B-3590 Diepenbeek, Belgium  
e-mail: {waldemar.jaronski, jose.bloemer, koen.vanhoof, geert.wets}@luc.ac.be

**Abstract.** Online businesses possess of high volumes web traffic and transaction data. Often, also valuable data regarding visitor opinions and attitudes towards the service and the website itself are available by means of online surveys. Additionally, sociodemographic data can provide characteristics of the audience, help differentiate between customer segments and understand drivers of loyalty with respect to each segment. Faced with the potentially rich body of the three kinds of information, companies urgently seek thereby for methods to analyze them in an efficient and insightful manner. The contribution of the present work consists of the application of Bayesian network technology for the joint analysis of all these data of the aforementioned dimensions that results in meaningful and valuable marketing knowledge. At the same time, the outlined solution yields also interesting practical results helping to understand better what is really going on on the website.

## 1. Introduction

Today's online companies face a major concern in losing their audience. Recent empirical studies by Mainspring and Bain & Company [16] have demonstrated that the average customer must shop four times at an online store before the store profits from that customer. Confronted with low customer retention rates, marketing managers of online companies try to find answers to the following questions: "What drives loyalty at my website?, What are the shortcomings of the service of my website?, What are the most important characteristics of my target audience?" Very often online e-businesses possess a high volume of web traffic- and transaction data. Also valuable information regarding visitor opinions and attitudes towards the service and the website itself are available by surveying online. Additionally, sociodemographic data can help to differentiate between customer segments. A combination of all these data may help to understand the drivers of loyalty with respect to different segments [e.g., 15]. However, to our knowledge, a specialized tool for the optimal analysis of such data is virtually non-existent. Standard tools based on aggregate website statistics fail to deliver superior knowledge facilitating taking strategic decisions. Companies are urgently seeking for intelligent methods to analyze the data in an efficient and insightful manner. The contribution of the present work

consists in applying Bayesian Belief networks for the joint analysis of traffic data combined with attitudinal data and sociodemographic data. This joint analysis must then result in meaningful and valuable marketing knowledge.

The paper is organized as follows. In Section 2 we present a brief overview of related work and current approaches in the marketing literature pertaining to the field of e-loyalty. Section 3 presents short introduction to Bayesian networks and explains advantages of applying this methodology in the context of marketing research. In Sections 4 and 5 we describe some results of our preliminary experiments of applying Bayesian nets on the example of visitor bases of two different websites. Finally, we close in Section 6 by presenting summary and potential avenues for further research.

## **2. Related work**

The marketing literature is still rather limited in theory and practical findings pertaining to the specific context of e-business as opposed to the traditional brick-and-mortar environment. There is a plethora of anecdotal evidence in the e-business community, however, to the best of our knowledge, there is a lack of empirically validated research regarding drivers of e-loyalty. The data collected by websites and on-line marketing companies as Media Metrix or MetrixLab are potentially very rich, but efficient methods for discovering useful knowledge in these data sources are virtually non-existent.

Existing state-of-the-art tools producing reports on the navigation data collected in web access logs have proven not to be so useful as they were expected [17]. These tools are able to deliver aggregate reports on navigation patterns on the website, but they fail to enhance our knowledge about why visitors stick to some web sites and switch to another.

Moreover, there is a substantial body of research [4; 17] in Web Mining pertaining predominantly to web usage analysis, user profiling or recommendation systems based on transaction and navigation data. However, in our opinion, existing work in this discipline provides little insight for marketing managers.

In [18], the authors use the clickstream data from Media Metrix in order to build a probabilistic model of repeat visiting. Their approach however is based solely on behavior data and, though highly predictive, does not offer any conceptual explanation for customer loyalty other than assuming that there are some underlying personal traits as for instance a personal visiting rate. Another recent study [8] illustrates the use of dependency networks for visualization of predictive relationships among demographic and web usage attributes of web surfers using Media Metrix data, but at the same time fail to integrate user subjective assessments into their model. In short, there is little quantitative research in the field of e-loyalty, and as a result, there is a need to support marketing managers who suffer from the lack of readily available empirical research findings.

Some authors argue that e-business requires new approaches and techniques for data modeling [13; 14]. For instance, [15] speculate that most new modeling techniques are likely to be data-driven, and that doing research on Internet-related data can enhance the field of marketing by way of whole new theoretical extensions,

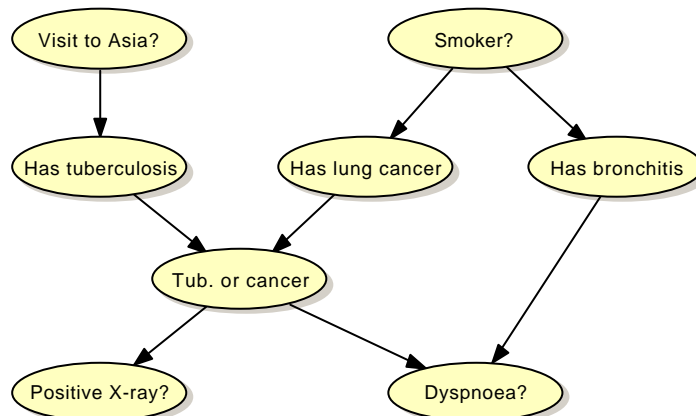
databases and methodologies. This paper is in line with these speculations as it utilizes an approach for web-related data analysis that encompasses advantages of both data- and model-driven techniques.

In contrast to the aforementioned approaches, the postulated methodology for web data analysis can help to understand the reasons which drive the repeat visiting behavior. Our approach is both data- and knowledge- driven. It is possible to encapsulate prior knowledge before building a model; the model can also be solely learned from data without specification of any prior knowledge.

### 3. Methodology

#### 3.1 Bayesian networks

Bayesian networks (BNs) have been used since their introduction in expert systems [19] as a knowledge representation formalism when knowledge was acquired from an expert. For example, Figure 1 gives a Bayesian network presentation of shortness-of-breath (dyspnoea). The network shows that this may be due to tuberculosis, lung cancer, bronchitis, a combination of these diseases or none of them. A recent visit to Asia increases the risk of tuberculosis, while smoking is known to be a risk factor for both lung cancer and bronchitis. The results of a single chest X-ray do not discriminate between lung cancer and tuberculosis, as neither does the presence or absence of dyspnoea. This example originates from [12].



**Fig. 1. Dyspnoea network.**

Within the past decade efficient algorithms for learning the models from data and for the network querying have been proposed. Bayesian discovery methods, and Bayesian networks in particular, have recently been gaining even more interest in the data mining community.

A Bayesian network consists of two components [19]: (1) a directed acyclic graph (DAG) in which nodes represent stochastic domain variables and directed arcs represent conditional dependencies between the variables, (2) a probability

distribution for each node as represented by conditional dependencies captured with the directed acyclic graph. Conditional dependencies represented with arcs link a variable called child variable with the set of its immediate predecessors called parent variables, according to the arc direction. Conditional dependency of a child variable given the configuration of the parent variables is quantified by means of the conditional probability distributions. The representation of the dependencies existing among variables in a domain with the Bayesian network yields significant savings of information required to encompass the domain knowledge. Directed arcs can, under certain assumptions, also be given causal interpretations. The variables in the network can take on values from a limited set of values (states), in which case they are regarded as discrete, or can be continuous and be described with parameters of Gaussian or other standard distribution for continuous random variables.

### **3.2 Network querying**

The first task, usually referred to as reasoning or network querying, consists in computing the posterior probability distribution of one variable under consideration, given that values of some other variables are known. In this case we assume that the network structure (the DAG component) has been already provided by the specialist or learned from data, and as a result, we only need to learn the quantitative component, viz. conditional probabilities [5]. Once all the relevant numbers have been computed, one can use the model for reasoning. We can use queries to the network for making prediction and sensitivity analysis. For example given the fact that a patient is a smoker, we will adjust our beliefs (increased risks) regarding lung cancer and bronchitis. However, our beliefs regarding tuberculosis are unchanged (i.e., tuberculosis is conditionally independent of "smoking" given the empty set of variables). Now, suppose we get a positive X-ray result for the patient. This will affect our beliefs regarding tuberculosis and lung cancer, but not our beliefs regarding bronchitis (i.e., bronchitis is conditionally independent of X-ray given smoking). However, had we also known that the patient suffers from shortness-of-breath, the X-ray result would also have affected our beliefs regarding bronchitis (i.e., "bronchitis" is not conditionally independent of "X-ray" given "smoking" and "dyspnoea"). To this end a well-grounded classical probability calculus based essentially on Bayes' theorem can be used. However, even for small sets of variables direct application of the standard probability calculus is very resource demanding. Therefore, one can use one of the proposed more efficient algorithms for propagation of the evidence entered on so called junctions tree representation of the Bayesian network [11]. More details about algorithms for probabilistic reasoning can be found in [6].

### **3.3 Network structure**

As mentioned before, the network structure can be learned from the data too. It is often the case when one is not certain about the direct dependencies among the variables in the domain. This task in general is NP-hard since the number of potential structures to review grows faster than exponentially as the function of the number of

nodes, therefore in practice additional assumptions and conditions have usually been introduced that reduce the time complexity of this task. Two different approaches for the structural learning have been proposed, each viewing a Bayesian network from a different perspective [e.g., 9]. Firstly, a Bayesian network is viewed as a structure that encodes the joint distribution of the variables. With this approach, one aims to find a structure that best fits the data in terms of some scoring metric, for instance maximum likelihood (ML), maximum a posteriori (MAP), minimum description length (MDL), Akaike information criterion (AIC) or Kullback-Leibler entropy scoring functions [5; 7]. Secondly, a Bayesian network encodes a set of conditional dependencies among the variables according to d-separation concept, which suggests learning the structure by identifying the conditional independence relationships using one of statistical tests, such as Chi-squared test or mutual information test [22; 2]. Depending on the algorithm used for learning, one can draw different conclusions and interpretations.

### 3.4 Applying Bayesian network technology

Application of the Bayesian networks for data mining in marketing is valuable for many reasons. Given the fact that the structure is known, it is easy and effective to learn from data and next investigate the quantitative relationships between the variables and make predictions. Besides the general Bayesian networks of the kind we exploit here, various other classes of Bayesian networks have been proposed that are geared especially for classification tasks, e.g. Naive Bayes, tree augmented Naive Bayes, Bayesian network augmented Naive Bayes, or Bayesian multi-nets [e.g., 3]. Clearly, classification with BNs consists in computation of probability  $P(C/e)$ , where  $C$  is the class variable,  $e$  is the configuration of the instance to be classified. The instance is assigned a state of  $C$  with the highest posterior probability.

Furthermore, inductive discovery of the network structure can provide useful knowledge of the presence or absence of direct dependencies among the particular variables in case when one is not certain about the network structure. The graphical representation of this knowledge, unlike in other predictive models such as decision trees or neural networks, is straightforward to interpret and understand by the users, who have also the possibility to modify the model for even better representation.

Other advantageous aspect of modeling with Bayesian networks is that it is easy to update the qualitative dimension of the model when new cases arrive by using one of the algorithms for adaptation [e.g., 12].

From practical point of view, it is worthwhile to notice that all reported values are given in terms of probabilities ranging from 0 to 100, having the intuitive meaning of likelihoods. No prior knowledge of any statistics is therefore required to interpret the results, although in order for a user to avoid misjudgments they should be aware of underlying assumptions, for instance of conditional independencies among variables. As regards the interpretation of the Bayesian network structure, [8] report that its semantics can be sometimes confusing for an untrained individual.

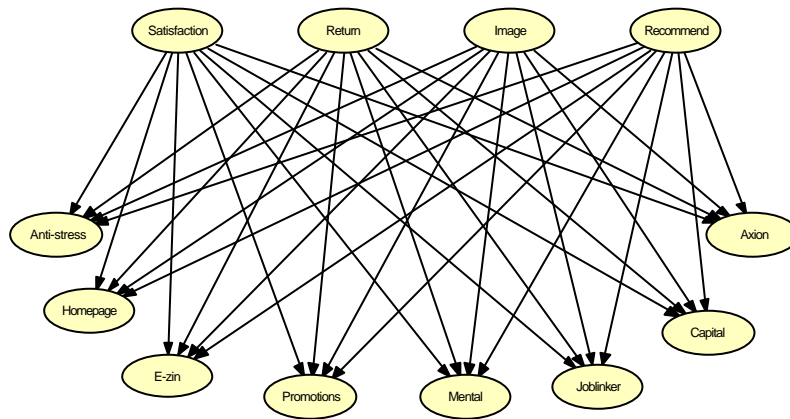
A powerful characteristic is also the ability to perform sensitivity analysis. In its most basic form sensitivity analysis shall give answers to questions like [10]: (1) which evidence is in favour of, against and/or irrelevant for a hypothesis in the focus, (2) which evidence discriminates hypothesis  $h_i$  from  $h_j$ . The former consists in

calculation of  $P(h/e')/P(h)$  for each subset  $e'$  of evidence  $e$ . In order to find out which findings discriminate between two hypothesis  $h_i$  and  $h_j$ , for each subset  $e'$  of evidence  $e$  the ratio of  $P(h_i/e')/P(h_j/e')$  is used.

In particular, application of Bayesian networks for data mining in web marketing can help provide answers to the following questions: What is my most loyal audience? What are the sociodemographics of my most loyal customers?, What drives loyalty?, i.e. what aspects of my web service do the most loyal customers value most? What causes loyal customers to be loyal?, What caused disloyal customers to be non-loyal?, and What opinions are not related to loyalty at all?

#### 4. Axion case study

Axion is a website of a major Belgian bank dedicated to young people. The data we have available describe duration of the stay in each of the eight sections of the Axion website and visitor's opinions on satisfaction, willingness to recommend, intention to return and perceived impact of the change of the website's image. Each case in this data set corresponds to a specific visitor session. The dataset consisted of 250 cases, attributes explaining duration of stay at each section were discretized into four bins of equal frequency, number of states for the judgment variables was limited from ten to four states by aggregating in order to avoid sparse probability tables.



**Fig. 2. The Bayesian network for Axionweb.**

As we have mentioned in the preceding section, domain experts often possess some domain knowledge and can provide the network structure. In the following example, we assume that the structure of the network has been correctly given by marketing specialists (see Figure 2), and we use dataset to compute relevant conditional probabilities. This Bayesian network assumes that: (1) the durations of the stay in each section during one visitor session are independent conditional on visitor's opinions towards the visit, (2) the judgment variables are marginally independent. The

justification of the first assumption resulted from the supposition that visitors were familiar with the website, so the duration of stay at each section was directly determined by visitor's experience with it, although we were unable to verify how valid this assumption was in this case. Furthermore, we assumed uniform priors and equivalent sample size of 1 for all network parameters. The marginal log likelihood of this model given data was  $-1999.08$ .

By means of querying the network one can investigate what for instance the probability distributions for the judgment variables are, given the duration of stay in some sections.

In particular, assume we were interested in the overall satisfaction of visitors who spend more time in one section than in another. Users who visit Joblinker section tend to be less satisfied, are less likely to return and unwilling to recommend the website than the total population of visitors. This information can be used to make improvements in this section but further investigations to this end are required.

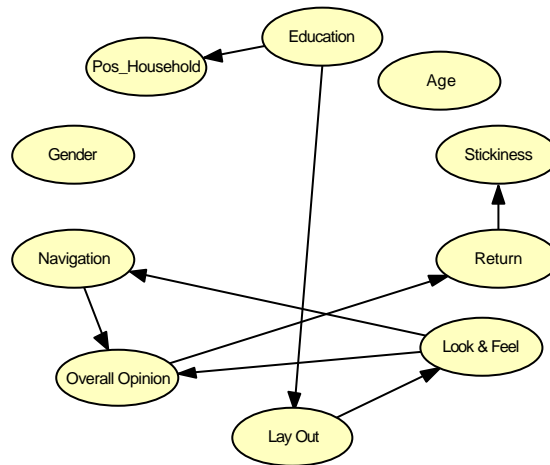
## **5. Dutch portals case-study**

In the following example we will use another data set to learn the probabilities as well as the Bayesian network structure. The data we use here have been collected by means of software installed on the users' machines, that monitors behavior of the user while surfing on the Internet. Each web page viewed by the user, along with its web address, date, time and duration is registered locally by this software in a database and transferred to the central database from time to time. Web users can download the software, called the OpinionBar, free from the Internet and install it on their computers whether at home, at work or at school. During the installation, the user is asked several questions about her/his socio-demographic profile, frequency of Internet use, etc. Occasionally, when one is in the process of visiting a website on which the bar keeps a survey, users are asked to complete a survey regarding user's opinions and judgments towards various facets of the website service and user's experience. The surfing behaviour data can be aggregated for each user as well as for the whole visitor base in order to provide e-metrics as total visit duration or stickiness. Each case in the data set corresponds to a particular user and is described with respect to sociodemographic, attitudinal and behavioural data. Any other data like that stored in web log files can be explored after appropriate processing and aggregation.

In this case study we investigated relationships between selected sociodemographic variables such as age, income, gender, visitors' opinions with regard to look and feel, lay out, ease of navigation, overall opinion, likelihood to return and basic behavioral measure of loyalty, viz. stickiness calculated as a ratio of total duration of all visits and number of visits. The data consisted of 251 cases and described visitors of a certain portal site in the Netherlands. Judgments attributes that were originally operationalized on five-item scale were next aggregated into three states. Stickiness was discretised into four bins of equal frequency, and age was discretized into four bins accordingly to common practice among the marketing community. Some judgment variables had a significant number of missing entries, of which only likelihood to return had a fraction of unobserved data slightly greater than 50%,

namely 52%. Overall opinion, lay out, look and feel had 46%, likelihood to return 52%, ease of navigation 42%, and education variable had 29% of missing values.

The first phase of the analysis consist in the examination of the results of the structural learning algorithm. In order to learn both the qualitative and quantitative dimensions of the models we used Bayesware Discoverer tool [1]. This tool implements the well known Bayesian methodology proposed in [5], which uses Maximum Likelihood scoring function, and K2 heuristic for searching through the space of possible network structures to consider. We allowed each node to have a maximum number of two parents, and provided the order, in which each node is tested as child only of the lower nodes in the order. The rationale behind the fact that we allowed only two nodes as parents was that we favoured more parsimonious models over more complex ones. The order we specified was: Stickiness, Likelihood to return, Overall opinion, Ease of navigation, Look and feel, Lay out, Position in the household, Education, Age, Gender. This ordering reflects our beliefs of the general causal nature between the variables, i.e., we expected that visitor's behaviour expressed as stickiness, and behavioral intentions accounted for by likelihood to return must be result of the visitor's judgments, which in turn were supposed to be possibly influenced by sociodemographic profile. All network structures were a priori equally likely. For dealing with missing values Bayesware Discoverer implements the Bound and Collapse method proposed by [20].



**Fig. 3. Network structure inferred from the experimental data.**

The result of the structural learning is presented in Fig. 3. The marginal log likelihood score produced by this network structure was  $-1802.54$ . Run on Pentium II with 264 MB of RAM memory, Bayesware Discoverer required 5 seconds to learn the complete model. To test the goodness of fit of the model we have carried out classification test for the overall opinion variable, and we have obtained predictive accuracy of 76.8%. We have also compared this network structure with other, more or less plausible models by adding deleting, or changing direction of some arcs, however



no improvement in the log likelihood has been achieved for these alternative models. Finally, we have tested the selected order against datasets describing visitors of three other popular portal sites in the Netherlands, in each case obtaining visually similar network structures (in all three cases there was no arc from likelihood to return to stickiness, but from ease of navigation instead, and in one case there was an arc from position in household to lay out instead of from education and arcs from gender and age pointing to position in household were added).

If we suppose that for our dataset the Causal Markov assumption holds and that there are no hidden variables, then the presented Bayesian network can be interpreted causally [9]. It thus follows for instance, that the most important judgment is the overall opinion on the website as it directly affects likelihood of return and indirectly the average duration of stay at the website (the log likelihood for a model in which there is an arc from overall opinion to stickiness instead of from likelihood to return was  $-1853.77$ ).

We can also conclude that the joint probability distribution of the variables in the domain is better represented with a model in which visitor's age and gender are not linked with any other variable. These findings suggest that age and gender are not likely to be related to loyalty, so that it does not make sense to segment visitors according to these attributes.

Perhaps one of the most interesting observations a marketing analyst would like to make are the sociodemographic characteristics of the most loyal audience. More precisely, the network structure induced from data suggests that education and position in the household do not provide any additional insight for the visitor loyalty once the visitor's opinions are known, but one might still try to predict the loyalty without knowing the visitor's website perceptions. In particular, visitors with college education, who are partners of breadwinners, tend to be somewhat less loyal than the audience as a whole. The likely reason for this may be lay out – this group of visitors is 1.42 times more likely to have negative and 0.84 times more likely to have very positive perception of lay out. Similarly, we can find out which opinions and website elements are relatively more important to the online audience than others.

Another analysis concerns the situation in which we want to assess the loyalty of the website's target group. If the target group is supposed to consist of, for instance, well-educated, breadwinning web users then we select the relevant sociodemographic states, and observe the values of the probability distribution for the loyalty variables. If loyalty turns out to be lower than average and this trend is stable it might turn out that we should redefine the target audience and advertise on some other websites. For instance, given the visitor is breadwinning college graduate, it is 1.03 times more probable that he/she will state he/she is unlikely to visit the website again.

Moreover, the reasons for which the target audience is not loyal can also be found. To this end we select the states similar as in the example above for the sociodemographic variables, and the visitor loyalty variables should be set as low. Now, the probability distributions for the relevant user opinions indicate the desired information. The opinions for which the decrease is most significant communicate the potential reasons for which the selected group of visitors is not loyal. The negative overall opinion of the chosen segment is now 2.53 times more probable, also negative opinion on look and feel is 1.90 times more likely, unfavorable opinion on ease of navigation is 1.26 times more likely, while it is 1.68 times more probable that

perception of lay out is clear. This can suggest that this group of visitors is not loyal due to unfavorable perceptions of look and feel.

Sensitivity analysis is amongst the most powerful applications of Bayesian network methodology. It is convenient to find out what elements of the website contribute most to visitor loyalty and what is the expected return on improved visitors' perceptions of various website elements as ease of navigation, look and feel, etc. Let hypothesis  $h$  of our focus be "Overall opinion = very positive". Also, let the set of evidences under consideration  $e$  be  $e = \{e_N, e_{LF}\}$ , where observation  $e_N$  is "Ease of navigation = very good", and  $e_{LF}$  is "Look&Feel = very positive". By querying the network we have  $P(h) = 0.22$  and  $P(h/e) = 0.86$ . Also  $P(h/e_N) = 0.48$  and  $P(h/e_{LF}) = 0.73$ . Therefore we can conclude that very positive perception of look & feel has a greater impact on very positive overall opinion than very good ease of navigation has. However, one can also notice that very positive perception of look & feel is not a sufficient condition for very positive overall opinion, since the ratio  $P(h/e_{LF})/P(h/e) = 0.85$  is significantly lower than 1.

It is also possible to compute probability distribution of some variables for an instance in case we are not fully sure about the states of the instance, or we have only partial knowledge about it. This task consists in entering the subjective likelihoods for certain states for variables under consideration, as opposed to selecting states with 100% certainty as in the examples above. If we were, for instance, interested in the analysis of sensitivity of customer loyalty in terms of the perceived ease of navigation, we could give slightly higher likelihoods for the states representing better perception and lower likelihood for the unfavorable states, which would correspond to overall better perception of this website element. Now, observed probability distribution of loyalty scores indicates the expected return in terms of loyalty. The ratio of the original probabilities and the ones after propagation can be interpreted also as a quantitative measure of the improvement in ease of navigation. In particular, if visitors were 2 times more likely to consider the ease of navigation as very good than as poor, the probability of likelihood to return stated as very likely would be 1.10 times higher.

## 6. Summary and further research

The Bayesian network approach for data mining in marketing we outlined here proves to be useful for researchers as well as for practitioners. Its most apparent benefits include its ability of prediction and explanation. The former aspect pertains to inferring values of unobserved variable based on the set of variables that are observed. Explanation can be considered as specification of direct and indirect relationships between the variables, specifically probabilistic conditional independence among them. The approach lends itself also to performing sensitivity analysis. By augmenting BN models with decision and utility nodes, they can also be used as tools for probabilistic decision analysis.

The presented methodology allows a flexible and meaningful analysis of web-related data. The results show the robustness and the added value of the approach. It

has been tested on several data sets, in each case similar parameters have been obtained. The same network structure can be applied for various websites.

In the future, we intend to integrate also other variables into the model. This work will include incorporation of other visitor judgments such as trust, commitment, and involvement. Another interesting contribution would be to incorporate other e-metrics, as well as information about the particular website sections or/and pages like the duration of stay at certain sections. Also, experiments with a data set containing online transaction data could yield interesting results.

### References

1. Bayesware Discoverer. Bayesware Limited. <http://www.bayesware.com>.
2. Cheng, J., D.A. Bell, and W. Liu (1997). Learning Belief Networks from Data: An Information Theory Based Approach. In: Proceedings of ACM CIKM'97.
3. Cheng, J. and R. Greiner (2001). Learning Bayesian Belief Network Classifiers: Algorithms and System. In: Proceedings of the Fourteenth Canadian Conference on Artificial Intelligence, AI'2001.
4. Cooley, R., B. Mobasher, and J. Srivastava (1997). Web Mining: Information and Pattern Discovery on the World Wide Web. In: Proceedings of IEEE Intl. Conf. Tools with AI, pp. 558-567, Newport Beach.
5. Cooper, G.F., and E. Herskovits (1992). A Bayesian Method for the Induction of Probabilistic Networks from Data. *Machine Learning*, 9, 309-347.
6. Cowell, R. G., A. P. David, S. L. Lauritzen, and D. J. Spiegelhalter (1999). *Probabilistic Networks and Expert Systems*. Springer-Verlag, New York.
7. Heckerman, D (1995). A Tutorial on Learning with Bayesian Networks. Technical Report MSR-TR-95-06, Microsoft Research.
8. Heckerman, D., D. M. Chickering, C. Meek, R. Rounthwaite, C. Kadie (2000). Dependency Networks for Inference, Collaborative Filtering, and Data Visualization. *Journal of Machine Learning Research*, 1, 49-75.
9. Heckerman, D., C. Meek, and G. Cooper (1997). A Bayesian Approach to Causal Discovery. Technical Report MSR-TR-97-05, Microsoft Research.
10. Jensen, F. V. (1996). *An Introduction to Bayesian Networks*. UCL Press.
11. Jensen, F. V., S. L. Lauritzen, and K. G. Olesen (1990). Bayesian Updating in Causal Probabilistic Networks by Local Computations. *Computational Statistics Quarterly*, 4, 269-282.
12. Lauritzen, S. L., and D. J. Spiegelhalter (1988). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, Series B (Methodological)*, 50 (2), 157-224.
13. Leeflang, P.S.H., and D.R. Wittink (2000). Building Models for Marketing Decisions: Past, Present and Future. *International Journal of Research in Marketing*, 17, 105-126.
14. Lilien, G.L., and A. Rangaswamy (2000). Modeled to Bits: Decision Models for the Digital, Networked Economy. *International Journal of Research in Marketing*, 17, 227-235.
15. Mahajan, V., and R. Venkatesh (2000). Marketing Modeling for E-Business. *International Journal of Research in Marketing*, 17, 215-225.
16. Mainspring and Bain & Company (2000).  
[http://www.nua.ie/surveys/index.cgi?f=VS&art\\_id=905355695&rel=true](http://www.nua.ie/surveys/index.cgi?f=VS&art_id=905355695&rel=true)
17. Mena, J. (1999). *Data Mining Your Website*. Butterworth-Heinemann/Digital Press.
18. Moe, W. M., and P. S. Fader (2000). Capturing Evolving Visit Behavior in Clickstream Data. Working Paper, The Wharton School, August 2000.
19. Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo (CA).

20. Ramoni, M. and P. Sebastiani (1997). Learning Bayesian Networks from Incomplete Databases. In: Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence. Morgan Kaufman, San Mateo (CA).
21. Spiegelhalter, D. J. and S. L. Lauritzen (1990). Sequential Updating of Conditional Probabilities on Directed Graphical Structures. *Networks*, 20 (5), 579-605.
22. Spirtes, P., C. Glymour, and R. Scheines (1993). *Causation, Prediction, and Search*. Springer-Verlag.