7

MODELING WEB USABILITY DIAGNOSTICS ON THE BASIS OF USAGE STATISTICS

Avi Harel
Ergolight Ltd., Haifa, Israel

Ron S. Kenett*
KPA Ltd., Raanana, Israel, and Department of Applied Mathematics and Statistics, University of Torino, Torino, Italy

Fabrizio Ruggeri
CNR IMATI, Milano, Italy

7.1 BACKGROUND ON E-COMMERCE USABILITY

This chapter presents a method for usability diagnosis of webpages based on time analysis of clickstream data. The resulting diagnostic reports enable website managers to learn about possible usability barriers. Different website design deficiencies are associated with different patterns of exceptional navigation. This chapter presents a method based on the integration of stochastic Bayesian and Markov models with models for estimating and analyzing visitors' mental activities during their interaction with a website. Based on this approach, a seven-layer model for data analysis is proposed and an example of a log analyzer that implements this model is presented. The chapter describes state-of-the-art techniques and tools implementing these methods and maps areas for future research. We begin with some definitions and

*Corresponding author; email: ron@kpa.co.il.

Statistical Methods in e-Commerce Research. Edited by W. Jank and G. Shmueli

an introduction to key concepts in e-commerce usability. The web analytics models are presented later, followed by a case study.

7.1.1 Usability

Usability is a term used to denote the ease with which people can employ a particular tool, or other human-made objects, in order to achieve a particular goal. The International Organization for Standardization document ISO 9241-11 (1998), Guidance on Usability defines usability as

> the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.

In commercial websites, usability is about the choices of site visitors. Bucklin et al. (2002) provide a two-by-two categorization of applications delineating search versus purchase, on the one hand, and within-site versus across-site choices, on the other. This chapter can be associated with the category of research on within-site choices, namely, site navigation. In e-commerce, we are concerned about visitors purchasing the website's deliverables. A 1999 study of Web users asked respondents to list the five most important reasons to shop on the Web. Even though low prices definitely attract customers, pricing was only the third most important issue for respondents. Most of the answers were related to making it easy, pleasant, and efficient to buy. The top reason was "Easy to place an order" for 83% of the respondents (Nielsen et al. 1999). The increasing importance of e-commerce is apparent in a study conducted at the Georgia Institute of Technology (GVU 1997). Over half of the 10,000 respondents to questions in eight separate surveys reported having purchased items online. The most frequently cited reason for using the Web for personal shopping was convenience (65%), followed by availability of vendor information (60%), no pressure from salespeople (55%), and saving time (53%). The purchase process follows a stage of users' market research (Moe 2006a). Desired actions during market research browsing may be viewing a key page on the site or downloading a whitepaper. Desired actions at the purchase stage include submitting a sales lead and making a purchase. Usability in e-commerce includes both stages: market research browsing and final order submission.

7.1.2 Usability and Marketing: Beyond Conversion Rates

Usability contributes to both short- and long-range profitability of websites. Short-term profitability is about site visits that end with the purchase of a product or service. The percentage of short-term successes is known in the e-commerce literature as the *conversion rate*. Long-term profitability is about site visits in which the site visitor is satisfied with the navigation results. The distinction between short-term and long-term profitability is very important. Short-term profits can be achieved through banners and popup promotions, intended to attract (distract) the visitors during their original surfing intentions, according to specific marketing goals. Most research is about short-term attributes. Setting a banner to achieve to a particular e-commerce goal typically hampers usability (Rhodes 2001). Many task-oriented visitors might find banners distracting and might abandon the site too early. Subsequently, users who deviated from their original intention might be unwilling to visit the same site again. Moe (2006a) summarized the state of the art about the effect of interruption. Speier and Valacich (1999) found negative effects on visitors and proposed an explanation based on the theory of limited mental capacity in situations of high workload. On the other hand, Zijlstra et al. (1999) found positive effects, and explained the results by a model of overcompensation and by extending mental capacity. However, we note that in spite of the maturity of e-commerce, there is no substantial research proving that the benefits of marketing campaigns involving banners and popup promotions exceed the long-term loss of income due to the negative effects of these methods.

7.1.3 Barriers to Website Usability

Users may abandon a website after being dissatisfied with its content or behavior. Major design problems found in an IBM study of the usability of e-commerce sites (Tilson et al. 1998) included the following:

- The sites did not indicate effectively where and how to add an item to the shopping list (the list of items the user plans to buy).
- The sites did not provide effective feedback when items were and were not saved in the shopping list.
- The sites did not indicate effectively if and when users needed to register or log in to order.
- The sites did not facilitate easy navigation from the shopping list to other parts of the site.

7.1.4 Usability Assurance

The term *usability assurance* refers to methods for improving ease of use during the development process. As an example of the benefits of usability assurance, consider the two designs presented in Kohavi (2006). Figure 7.1a shows the website before usability assurance and Figure 7.1b shows the website afterward.

Design B involves nine usability changes in design A that produced a higher conversation rates. The changes are as follows:

1. The space between the top “Proceed to Checkout” button line and the next line was closed.
2. The top “Continue Shopping” button was removed.
3. The “Update” button underneath the quantity box was removed.
4. The “Total” box was moved down a line. Text and amount appear in different boxes.
5. Above the “Total” box is a “Discount” box, with the amount in a box next to it.
6. Above the “Shipping Method” line is “Enter Coupon Code” with a box to enter it.

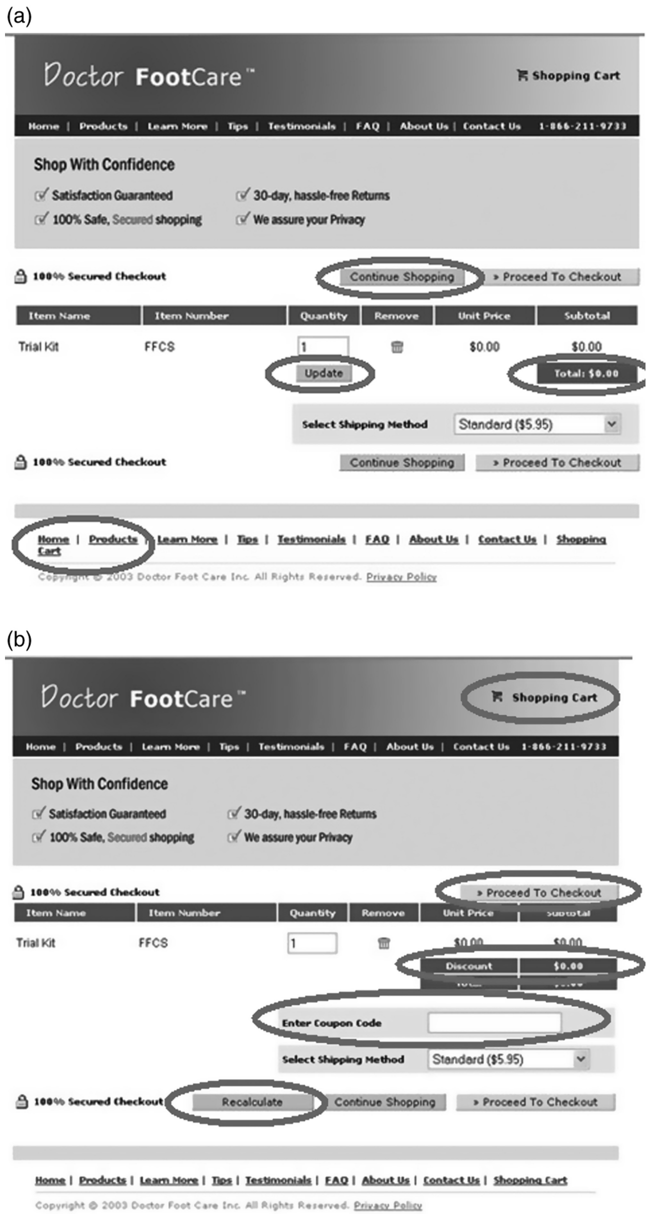

Figure 7.1 (a) Design A—before usability assurance; (b) Design B—after usability assurance. Based on Kohavi (2006).

7. There is a new “Recalculate” button to the left of “Continue Shopping.”
8. The bottom tool bar appears on two lines.
9. The shopping cart icon is one space farther away from the words “Shopping Cart.”

Usability assurance is a key dimension in product marketing success. Marcus (2002) reviewed many studies on usability return on investment (ROI) in user interface (UI) design. These studies show that investment in usability may profit by:

- Saving development time and reducing development, redesign, training, documentation, support, and maintenance costs.
- Increasing success rates: effectiveness, market share, traffic, user productivity, and efficiency, resulting in increased user satisfaction and job satisfaction.

For example, before 1999, IBM’s Web presence traditionally consisted of a difficult-to-navigate labyrinth of disparate subsites. A major website redesign made it more cohesive and user-friendly. According to IBM, the massive redesign effort quickly paid dividends. The company announced, one month after the February 1999 relaunch, that traffic to the Shop IBM online store had increased 120%, and sales went up 400% (Battey 2001).

7.1.5 Methodologies for Usability Assurance

Usability assurance is conducted throughout the development cycle. At the design stage, it is based on methodologies and practices for:

- Anticipating how users may behave when using the product or the service (as opposed to assessment of how users should behave, according to the product and service developer).
- Designing the user interface to ensure seamless interaction with the product.

An example of a methodology for usability assurance by design is presented in Dustin et al. (2001). At the validation stage, it consists of methodologies and practices for verifying that the users behave as intended and that the user interface responds gracefully in cases of deviations from the designers’ intention. An example of a book presenting common testing practices is the one by Duma and Redish (1999).

7.1.6 Usability Research

Good navigation and website design make it easier for users to find what they’re looking for and allow them to buy it once they’ve found it (Donahue 2001). The primary goal of usability research is to define design rules to be applied to requirement specifications in order to ensure that product and service websites are usable. An example of research-based Web design and usability guidelines is the official online booking of the U.S. Department of Health and Human Services. In website design, the goal is to ensure that the sites are easy to navigate. In e-commerce, we also want to ensure that the users will reach particular target pages, fill in correct purchase details, and submit their orders. Examples of Web usability research may be found in reports by User Interface Engineering (2007). For a recent guidelines book, see Nielsen and Loranger (2006).

7.1.7 Predicted Page Usability

Predicted page usability is the designer's impact on page usability. It is the prediction of page usability when specific design rules are applied based on prior research. Predicted usability is an indication of usability applicable to the specification and design stages of website development. The predicted usability attributes of main website pages are as follows:

- *Responsiveness*: The proper time for feedback to the visitor's link or control activation is between 0.1 and 1 second. Response time of less than 0.1 second is too fast, which means that visitors might not notice the feedback (Dabrowski and Munson 2001). Response time greater than 1 second is too slow, which means that visitors might change their focus to a different task.
- *Performance*: The proper download time should normally be 3–5 seconds. Beyond that visitors might change their focus to other tasks, if they are not determined to explore the page, and look for alternative sites.
- *Predicted readability*: Readability indices evaluate the readability of text on the basis of predictive models (Kenett 1996; Kenett and Baker 1999). Such metrics, based on linguistic attributes, can be used instead of running statistical surveys with actual human readers, also known as *readability surveys*, in order to get an indication of webpage readability. Readability scores are based on characteristics such as average word length and sentence length (as a proxy for syntactic complexity). Well-known indices (some included in Microsoft Word) are the Flesch Kincaid metric, the passive sentence index, and the SMOG Index.
- *Predicted relevance*: It is common practice for usability professionals to design websites according to the expected visitor's workflow, which is obtained through task analysis. Predicted page relevance is high if the pages are designed according to the expected visitor's activity, designed according to these workflows. It is low if the pages are designed according to functions or features, regardless of the visitor's workflow.

7.1.8 Limitations of Predicted Usability Attributes

Usability practitioners typically claim that, because they focus on users, they can represent the visitor's needs better than other designers. However, they also admit that users' behavior is often unpredictable. Eventually, after becoming involved in the project, it is difficult even for the best designers to predict how visitors behave on their first visits. Therefore, developers should not rely on user-centered design alone; user testing is also required. The predicted usability attributes are important for the design, but they are insufficient for launching sites with seamless navigation. Typically, the user's experience is not the same as predicted at design time. For example:

- *Responsiveness*: Certain visitors who are eager to see the page content may be willing to wait longer than other page visitors who do not expect much of this page.
- *Performance*: Certain visitors who expect to find the information they need on a particular page may be willing to wait longer than other visitors who do not expect to find any valuable information there.
- *Predicted readability*: Readability scores used by computer scientists use a limited set of rules about usage of language, mainly syntactic, involving dictionaries and grammar testing. Obviously, this is insufficient to obtain a reasonable approximation to the ways a human reader comprehends text. Actual readability depends on knowledge that people gain after years of daily experience. It is impractical to try to find all the rules that could measure the difference in readability of the two sentences "Monkeys like bananas" and "Bananas like monkeys." Yet, in usability testing, it may be obvious that the first sentence is highly readable, while the second sentence might impose a high mental load on the reader.
- *Predicted relevance*: At design time, site traffic is still unknown. It is difficult to anticipate the knowledge and needs of the actual visitors.

7.1.9 Usability Validation

Not surprisingly, theoretical research does not answer all practical design questions. Usability practitioners know quite well that they often fail to predict the user's behavior, and they almost always recommend validating the design before launching the site by testing it with real users. Usability assurance practices that are applicable to the validation stage consist of methodologies and practices for verifying that the users behave as intended and that the user interface responds gracefully in cases of deviations from the designers' intention. Common usability validation methods are based on user reports and testing. User reports capture major usability deficiencies of which the site visitors are aware. However, usability is the outcome of the aggregation of many tiny details, each of them contributing to the goal of seamless navigation. The limitation of user reports is that most users are busy overcoming the many barriers that they encounter; in terms of cognitive science, the amount of detail they need to mentally process exceeds the capacity of working memory. Typically, users are unaware of most of the design deficiencies. Once they are made aware of a minor usability problem, they prefer not to report it in order not to bother management and not to look stupid. The other method for usability validation involves user testing. A typical product user interface (UI) development cycle has three testing phases: integration testing, alpha testing, and beta testing. Validation in e-commerce is similar to UI validation, with special extra needs.

7.1.10 Integration Testing

The first testing phase is during integration of UI with all the product components. The primary goal is to make sure that the product works under normal conditions. Usability validation typically relies on the skills of programmers involved in the product's development and on consulting with usability specialists. This common practice is also typical to website usability validation. Marketing-related issues of e-commerce websites are typically tested by reference to the sites requirement specifications.

7.1.11 Alpha Testing

The second testing phase is alpha testing, primarily intended to verify that the product resists exceptional conditions typical to normal operation, by technical members of the testing group, in the developer's labs, following the predefined operational procedures. Usability testing is typically conducted by usability professionals, who observe users while they try to do predefined tasks. The testing is sometimes conducted in special usability labs at the developer's site, using special equipment (Duma and Redish 1999). Video cameras are used to capture users' behavior, scan converters are used to capture screens, and special setup and software is used to log the testers' observations and to synchronize all the input for failure analysis. Recently, eye tracking equipment has been used in special laboratory setups to learn about the ways users scan the screens. This common UI testing practice is also applicable to website usability validation. Marketing-related issues of e-commerce websites are typically tested by the marketing people involved in the site's requirement analysis and specification.

7.1.12 Beta Testing

The third testing phase is beta testing, intended to verify that the product resists exceptional conditions typical to normal operation by users, in their real operational conditions, doing their real tasks. Usability testing is conducted by usability professionals, who observe the users' behavior either by visiting them at the users' site or remotely, by special video and digital communication means. This common practice is also typical of website usability validation. A few testing companies also invite user representatives to volunteer as panelists, to try the site and comment on usability problems that they encounter while using it. Marketing-related issues of e-commerce websites are validated using special website analyzers.

7.1.13 Ongoing Testing

After a product has been released, the users test it regularly during the interaction, but only severe problems are typically reported back to the developers. With websites the situation is much better, thanks to the server log files that are regularly generated in most websites. Special software programs, called *web analytic tools*, enable site administrators to analyze users' behavior. Special marketing tools enable marketing administrators to view the marketing attributes of the users' behavior. So far, no method has been found for extracting users' expectations from server log files.

7.1.14 Subjective Page Usability

What are the attributes of actual usability, that is, the attributes that describe actual users' behavior? Subjective page usability focuses on the attributes of the visitors' navigation experience. The attributes of subjective page usability answer the following questions:

- *Perceived responsiveness*: To what degree does the page responsiveness suit the visitors' expectations?
- *Perceived performance*: To what degree does the page performance suit the visitors' expectations?
- *Subjective readability*: To what degree do the page visitors succeed in reading and comprehending the page content?
- *Subjective relevance*: To what degree do the page visitors perceive the page as relevant to their needs?

To distinguish between positive and negative attributes of the user experience, we need statistics. This chapter deals with the statistics required to measure, interpret, and classify these and other attributes of the user experience.

7.2 WEB ANALYTICS

Web analytics is the study of the behavior of website visitors. In general terms, web analytics is the process of collecting data about the activities of website visitors and mining those data for information that can be used to improve the website (Peterson 2005). In a commercial context, web analytics refers especially to the use of data collected from a website to determine which aspects of the website promotes of the business—objectives for example, which landing pages encourage people to make a purchase. Before expanding on web analytics in the context of usability studies, we present a brief introduction to measurement systems by focusing on the popular GQM approach.

GQM (goal/question/metric) is a goal-oriented approach that can be used to manage the whole measurement process. It is widely applied in the field of software process and product measurement (see Basili and Weiss 1984; Kenett and Baker 1999; and http://ivs.cs.uni-magdeburg.de/sw-eng/us/java/GQM). GQM aims at evaluating the achievement of goals by making them measurable. Therefore, metrics are needed, which become evident by asking the questions necessary to verify the goal. One starts by making up a list of goals that should be evaluated and asks the relevant questions. Then one collects metrics defined to answer questions used for evaluating the achievement of the goal. The GQM approach can be used as a systematic way to tailor and integrate process measurement's objectives into measurement goals and refine them into measurable values (Figure 7.2). Carried out through observation, interviews with users, and/or workshops, this process is iterative, systematic, and provides rapid identification of the structure for a software improvement program. It creates a foundation of repeatable procedures

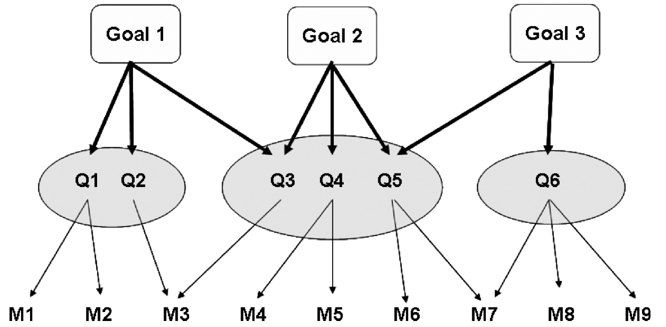

Figure 7.2 The goal questions metrics approach.

for single projects or even for an entire organization. The stages are goal identification, measurement planning, measurement performance, and validation, concluding with analysis and interpretation of the results. It derives the metrics to be used from a thorough analysis of goals and associated questions to be answered quantitatively. By following GQM guidelines, it is possible to:

- Establish the goals of the measurement process.
- Find a proper set of questions to allow one to reach the goals.
- Choose the right set of metrics in order to answer the questions.
- Plan and execute the data collection phase.
- Evaluate the results.

The meaning of the GQM main components—the Gs, the Qs, and the Ms—can be described as follows:

- The Goal describes the measurement's purpose; stating explicit goals gives the measurement program a precise and clear context.
- The set of Questions refines the goal and highlights the quality focus: "What should I know to be able to reach a certain goal?"
- The set of Metrics is used to answer each question. Metric data may result from objective or subjective measurement.

GQM is giving us a context for website evaluation using web analytics. A standard tool used in website management is a recorder, which stores records of the site visits in server log files. Several log analyzers are extensively used to provide usage statistics to website managers, enabling them to learn which pages are visited more frequently than others; how such visits change during the day, the week, or the year; and so on. Certain usage statistics, primarily about rates of site exits and backward navigation, are especially informative. High rates indicate potential usability deficiencies in the page design, which may be very important when they are in the critical path of purchasing procedures. Yet, these measures are weak in diagnostics and do not tell the site managers the reasons for the exceptional rates. High exit rates might indicate dissatisfaction with site navigation but also task completion. Also, the reasons for a high backward navigation rate may be that a page is useless, that the links to this page are wrongly labeled, or that this page is used for additional information to move to another page.

7.2.1 Web Analytics Technologies

There are two main technological approaches to collecting web analytics data. The first method, *log file analysis*, reads the log files in which the web server records all its transactions. The second method, *page tagging*, installs a JavaScript applet on each page view to notify the server about designers' events, such as local activity on the browser.

7.2.2 Server Logs

A straightforward means to collect activity data from website visitors is by server logs. These logs are used routinely by most websites as a main source for backtracking site activity. This chapter offers a method for extracting useful usability information from server logs. Server log files are usually not accessible to general Internet users, only to the webmaster or an other administrator. The two most popular platforms for web hosting, Linux and Windows, support user-defined formats by selection from a set of predefined hit attributes. The most frequently selected attributes are the user IP address, the timestamp (date and time), and the URL. These are also the attributes required for the methodology described here. Additional common attributes, optional in this chapter, include an HTTP code (useful for reducing noise due to technical problems), user agent (browser identification, useful for reducing noise due to robot hits), download size (useful for calculating download speed and measures of textual content), and referrer (useful for better discrimination between internal and external links).

7.2.3 Web Log Analysis Software

Web log analysis software (also called a *web log analyzer*) is software that parses a log file from a web server (like Apache), and, based on the values contained in the log file, derives indicators about who, when, and how a web server is visited. Indicators reported by most web log analyzers include:

- Number of visits and number of unique visitors
- Visit duration and last visit
- Authenticated users and last authenticated visits
- Days of week and rush hours
- Domains/countries of host's visitors
- Host's list
- Most viewed, entry, and exit pages
- Files type

- Operating system used
- Browsers used
- Robots
- Search engines, key phrases, and keywords used to find the analyzed website
- HTTP errors

7.2.4 Web Statistics

Many statistical software packages have been developed to take advantage of the popularity of server logs, providing valuable statistics about different aspects of site navigation, such as traffic analysis, including changes in page popularity during the day or the year, by referrers, or by user agents.

7.2.5 Sales-Oriented Analytics

In a commercial context, web analytics refers especially to the use of data collected from a website to determine which aspects of the website work toward the business objectives; for example, they are often used to calculate trends in page popularity during marketing campaigns and to find out which landing pages encourage people to make a purchase. Two common measures of sales-oriented design are the click-through rate and the conversion rate. The marketing department of any organization that owns a website should be trained to understand these tools.

Section 7.3 describes a theoretical framework for website usability validation, presenting an implementation example. Section 7.4 presents case studies based on real usability data extracted from server logs. Section 7.5 presents conclusions and suggestions for further research.

7.3 MODELING WEBSITE USABILITY

This section presents a theoretical framework for extracting usability diagnostic information from server log files. The framework consists of goal definition and of models used to achieve this goal.

7.3.1 Goal of Website Usability Validation

The theoretical framework is intended to provide two diagnostics about pages and links: page diagnostics and link diagnostics.

Page Diagnostics

- Which pages are likely to be abandoned due to long download time
- Which pages are difficult to read or comprehend
- Which pages are either not interesting or irrelevant to the page visitors
- Which pages are sensitive to performance fluctuation

- Which pages are distracting potential customers from their goals
- Which forms are difficult to fill in or submit

Link Diagnostics

- Which links are likely to be misinterpreted
- Which links are likely to be disregarded
- Which links do not behave according to the visitors' expectations
- Which links are connected to the wrong page

7.3.2 Models Used in Web Usability Diagnostics

This section presents four models used for usability diagnostics. We start with the Markov chain model to represent the overall site view and then go to the page views invoking models of mental activities in page processing. Next, we consider the need for evidence of cause-effect relationships by applying Bayesian networks. Finally, we conclude with statistical analysis of the browsing experience. The models are as follows:

Model 1—Markov Processes: Montgomery et al. (2004) used a dynamic multinomial probit model of Web browsing to show how path information can help to predict visitors' behavior. In analyzing web data, we need to introduce a time dimension so that the dynamics of the navigation sessions are captured. Markov processes model the time dimension in server requests, in which the states represent the webpages and the transitions between states represent the hypertext activity.

Model 2—Mental Activity: Moe (2006b) noted that "though clickstream data do contain a lot of information pertaining to the consumer buying process, it is often difficult to draw cause-and-effect conclusions." To draw such conclusions, we need a model of how certain situations evolve to certain effects. To draw conclusions about the barriers to seamless navigation, we need a model of normal mental activities involved in page handling and possible deviations from normal activities due to design deficiencies. A user-centered model of intrapage mental activities is used to describe the barriers to successful navigation.

Model 3—Bayesian Networks: Bayesian networks map cause and defect relationships between key variables. Here they are used to validate our hypotheses about the relationships between usability design defects and visitors' behavior in response to occurrences of usability barriers.

Model 4—Statistical Analysis: The diagnosis of usability design deficiency involves certain measurements of site navigation, statistical calculations, and rules for statistical decision making. A seven-layer model of statistical manipulations of server log data is described, which enables statistical decisions about possible usability design deficiencies to be made.

7.3.3 Model 1—Markov Processes

In analyzing web data, we need to introduce a time dimension so that the dynamics of the navigation sessions are captured. Markov processes model the time dimension in server requests, in which the states represent the webpages and the transitions between states represent the hypertext activity. Different stochastic models can be introduced to describe and analyze movement within a specific website. We will discuss choices of various Markov processes and illustrate the rationale behind them; then, for illustrative purposes, we will focus on homogeneous Markov chains and present simple estimation techniques, stressing the Bayesian perspective on Markov processes. Each webpage can be identified as a state, and the movement pattern of a visitor within the website can be considered as a sequence of passages from one state to another, with an extra state added to denote the case in which the visitor is outside the website. Markov chains can be used to describe the very simple case of moving from one webpage to another. In this case, from a given webpage, only transitions to another webpage or outside the website are possible. The probabilities of such transitions are the object of the Markov process statistical analysis. A Markov process is a stochastic process *X(t)*, on a probability space with t in the parameter space T, realizing specific values in the state space S, where the Markov property

$$\begin{aligned} P[X(t) \leq x | X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, \ldots, X(t_0) = x_0] \\ = P[X(t) \leq x | X(t_n) = x_n] \end{aligned}$$

holds for all $(t_0, \ldots, t_n)$, $t_0 < \cdots < t_n$, $t_j \in T$, $j = 0, \ldots, n$.

When $T = N$, the Markov process becomes a Markov chain and we denote the process by $\{X_n, \ n \in T)$; X_n denotes which page (state) the visitor is in at the n-th visited page within the web (see Karlin and Taylor 1975, 1981). Suppose that the website has m pages, and we consider the exit from the website as the state $m + 1$. We consider the home page of the company as state 1 and the webpage with the completion of the action of interest (e.g., order) as state m. The transition probabilities of the Markov chain are represented by a $m + 1$-dimensional square matrix A, whose elements, $p_{ij}(n)$, $i, j = 1, m + 1$, are such that $p_{ij}(n) = P(X_{\mathrm{n}+1} = j | X_n = i)$ and it holds that $\sum_j p_{ij} = 1$ for all i. The transition probabilities depend on the stage n of the visitor's visit to the website. The assumption can be realistic, since a visitor might spend his or her first stages in the website browsing for all possible offers and details, whereas later he or she will be more likely to make a decision about either leaving the website without taking any action (e.g., ordering or paying) or going quickly through all the webpages needed to complete the action. A Markov process also can be used to account for the time spent in a state. As discussed elsewhere in this chapter, the length of stay at a webpage could be an important index of poor usability. Another possible Markov process model considers the transition probabilities as a function of some covariates often available in clickstream files of e-commerce sites, such as the sex and age of the visitor, number of years doing business with the company, etc. Since this chapter aims to show how usage statistics can be helpful in analyzing usability in practice, we only consider the simple case of a homogeneous Markov chain where the transition probabilities do not depend on stage n. In this case the transition probability becomes p_{ij}, which is the object of our statistical analysis. From the log file of the accesses to the company's website, it is possible to observe N, the total number of transitions, the number n_i of transitions from state i, and the number n_{ij} of transitions from state i to state j, $i, j = 1, m+1$. The likelihood function of what we observe is given by $\prod_{i,j} p_{ij}^{n_{ij}}$ and the maximum likelihood estimates are given by $\hat{p}_{ij} = n_{ij}/n_i$, using a multinomial model for the number of transitions from state i, for all i. Based on the estimated probabilities, it is possible to identify the critical transitions and investigate if usability is the cause of unexpected behaviors. In particular, it is important for the study of the probabilities p_{im} and p_{im+1}, i.e., the transition probabilities to completion of the action (e.g., ordering) and exit from the website, respectively. Experts are therefore asked to interpret a posteriori the results of the statistical analysis. It is also possible to use experts' opinions in making inferences about the probability, following a Bayesian approach. We concentrate on the probabilities $p_i = (p_{i1}, \ldots, p_{im+1})$ for each row i of the transition matrix A. We consider a Dirichlet prior $D(\alpha_{i1}, \ldots, \alpha_{im+1})$ on p_i which is conjugate with respect to the multinomial model, so that the posterior distribution will be a Dirichlet $D(\alpha_{i1} + n_{i1}, \ldots, \alpha_{im+1} + n_{im+1})$. Considering the posterior mean, i.e., the Bayesian (optimal) estimator under the squared loss function, then, we obtain $\hat{p}_{ij} = (\alpha_{ij} + n_{ij})/(\sum_k \alpha_{ik} + n_i)$. A more complex Bayesian model is considered by Di Scala, et al. (2004), who assume a multinomial model for the number of transitions from each state i and a multivariate logit transform of the (nonnull) transition probabilities P_{ij}, defined as $\gamma_{ij} = \alpha_i + \beta_j$. They interpret α_i as a measure of effectiveness of the page i, i.e., of its ability to suggest interesting links, whereas β_j measures the attractiveness of page j. A Bayesian hierarchical model is proposed so that the γ_{ij} are not treated as independent, but their estimation *gets strength* (using the Bayesian jargon) from the observations on other γ_{kl} with $k = i$ or $l = j$. For more on robust Bayesian statistics, see Rios Insua and Ruggeri (2000).

7.3.4 Model 2—User's Mental Activities in Website Navigation

To analyze the barriers to seamless navigation, we need to have a model of normal mental activities involved in page handling and possible deviations from normal activities due to website design deficiencies. A model of website navigation, describing common barriers to successful navigation used in usability diagnostics, is now described. The model assumes that the visitor enters a first page, which is often the home page, or a page referred to by an external hyperlink. Then the visitor repeats the following sequence:

- Evaluates the page download and content.
- Reacts according to the need for additional information.

7.3.4.1 Page Evaluation During Navigation. Page evaluation typically involves the following sequence of activities:

1. Wait for the page to start downloading.
2. Read the page while downloading, looking for information that meets the goal.

3. Realize that the pages have finished downloading.
4. Look for more information related to the goal.
5. Evaluate the information, looking for the best way to proceed.
6. Evaluate the relevance of the page and the overall relevance of the site to the goal.

7.3.4.2 Visitor's Reaction to Evaluation Results. The visitor's reaction to the evaluation typically results in any of the following:

- Return to the previous page if the current page is perceived as less relevant to the goal, than the previous page.
- Link to the next website page, which is perceived as a potential bridge to a goal page.
- Try another hyperlink from a (top or sidebar) main menu if the overall navigation experience is still positive.
- Exit the site if the goal has been reached or if the overall site navigation experience is negative.

7.3.4.3 Barriers to Seamless Navigation. Users may abandon a website after being dissatisfied with its content or behavior. The following is a list of common barriers to page usability:

- The page is difficult to find.
- The time from page request to initial feedback is too long or the initial feedback is unnoticeable, and therefore visitors might suspect that there are technical problems with the requested page.
- The page download time is too long. This is the most widely recognized barrier to a positive navigation experience (Nielsen 1994).
- The page download time is too short. This may happen for extremely fast pages that impose an extremely high mental load (Dabrowski and Munson 2001).
- The page does not indicate clearly when the download is completed. This is a problem in webpages with large amounts of content that download gradually.
- The page is irrelevant to the visitors' goal; the information the visitors need may not be found there.
- The page contains distracting animation, preventing users' focus on their search goals.
- The page is difficult to read or comprehend.
- The data formats in form filling impose unnecessary constrains, or they are not suited to the visitor's profile, and many visitors are unable to provide the required data.
- The rules for form filling are too restrictive, forcing extra work, such as selecting dates from long lists of numbers.

- The links on the page are not bold enough, and visitors often miss them.
- The label describing the hyperlink to this page is misleading; therefore, it is misinterpreted.
- The links may lead visitors to pages that are irrelevant to their needs.

The goal of usability assurance is to make sure that such barriers are removed.

7.3.4.4 Analysis of the Barriers to Seamless Navigation. Common design deficiencies are now listed according to the first phase—page evaluation—of the model above.

- The visitor might wait too long for the page to start downloading. A design weakness may occur when there is no indication that the new page is about to start downloading. The visitor often suspects that there might be a technical problem with this page.
- While downloading, the visitor might first see irrelevant information, banners, or just an indication that the page is downloading. Visitors often suspect that the page does not have the information they are looking for.
- The visitor might be confused if the page finishes downloading very rapidly, behaving like a local application. The visitor may be even more confused if, later, additional information is appended gradually, making the visitor wonder when to decide that the page does not have the needed information.
- The visitor might not see the desired information, even though it is displayed on screen. Or the visitor may see the information only after spending too much time reading all the text on screen.
- The visitor might misinterpret information which is irrelevant his or her needs, consequently proceeding to a wrong page.
- The visitor might conclude that the page is irrelevant to his of her needs, and the desired information might not be found at this site.

7.3.4.5 Attributes of Page Usability. In terms of the barriers to navigation described above, and invoking the GQM approach mentioned in Section 7.1, the attributes or goals of page usability can be derived by answers the following questions:

- *Accessibility*: How easy is it to find the page? Are the links to the page noticeable?
- *Responsiveness*: Does the page provide immediate feedback? Is this feedback noticeable?
- *Performance*: Does it meet the user's expectations? Is it long enough, so that the visitors notice it? Is it longer than the visitors may tolerate?
- *Download completion indication*: Can the visitors be sure that the download is completed? Can they tell when the download is not finished?
- *Relevance*: Is the page irrelevant to the visitors' goal?

- *Focus support*: Can the visitors focus on reading the page content and finding the links they need? Or does the page contain visual or sound elements that distract the visitors from their task?
- *Subjective readability*: Is the page easy to read and comprehend?
- *Compatibility of data formats*: Do the data formats in form filling suit the visitor's profile?
- *Ease of data entry*: Does the page provide support for easy data entry?
- *Link visibility*: Are the links on the page visible and noticeable?
- *Link labeling*: How faithfully do the labels describing the hyperlinks to this page represent the page content?
- *Link information*: Do the hyperlinks refer to other pages according to the descriptions and explanations?

7.3.5 Model 3—Bayesian Networks

Predictions and diagnostics rely on basic structures of cause an defect. Model 3 provides such a structure by invoking Bayesian networks (Ben Gal 2007; Kenett 2007). Bayesian networks (BNs), also known as *belief networks*, belong to the family of probabilistic *graphical models* (GMs). These graphical structures are used to represent knowledge about an uncertain domain. In particular, each node in the graph represents a random variable, while the edges between the nodes represent probabilistic dependencies among the corresponding random variables. These conditional dependencies in the graph are often estimated by using known statistical and computational methods. Hence, BNs combine principles from graph theory, probability theory, computer science, and statistics. BNs correspond to another GM structure known as a *directed acyclic graph* (DAG) that is popular in the statistics, machine learning, and artificial intelligence societies. BNs are both mathematically rigorous and intuitively understandable. They enable effective representation and computation of the joint probability distribution over a set of random variables (Pearl 2000).

The structure of a DAG is defined by two sets: the set of nodes (vertices) and the set of directed edges. The nodes represent random variables and are drawn as circles labeled by variables names. The edges represent direct dependence among the variables and are drawn by arrows between the nodes. In particular, an edge from node X_i to node X_j represents a statistical dependence between the corresponding variables. Thus, the arrow indicates that a value taken by variable X_j depends on the value taken by variable X_i or roughly speaking, that variable X_i "influences" X_j. Node X_i is then referred to as a *parent* of X_j and X_j is referred to as the *child* of X_i. An extension of these genealogical terms is often used to define the sets of *descendants*—the set of nodes that can be reached on a direct path from the node or *ancestors* nodes—the set of nodes from which the node can be reached on a direct path. The structure of the DAG guarantees that no node can be its own ancestor or its own descendant. Such a condition is of vital importance to the factorization of the joint probability of a collection of nodes, as seen below. Note that although the arrows represent a direct causal connection between the variables, the *reasoning process* can operate on a BN by propagating information in any direction.

A BN reflects a simple conditional independence statement, namely, that each variable is independent of its nondescendants in the graph given the state of its parents. This property is used to reduce, sometimes significantly, the number of parameters required to characterize the joint probability distribution (JPD) of the variables. This reduction provides an efficient way to compute the posterior probabilities given the evidence (Lauritzen et al. 1988; Pearl 2000; Jensen 2001).

In addition to the DAG structure, which is often considered the qualitative part of the model, one needs to specify the quantitative parameters. The parameters are described in a manner consistent with a Markovian property, where the conditional probability distribution (CPD) at each node depends only on its parents. For discrete random variables, this conditional probability is often represented by a table listing the local probability that a child node takes on each of the feasible values—for each combination of values of its parents. The joint distribution of a collection of variables can be determined uniquely by these local conditional probability tables (CPTs). Formally, a BN B is an annotated DAG that represents a JPD over a set of random variables $\mathbf{V}$. The network is defined by a pair $B = \langle G, \Theta \rangle$, where G is the DAG whose nodes $X_1, X_2, \ldots, X_n$ represent random variables and whose edges represent the direct dependencies between these variables. The graph G encodes independence assumptions, by which each variable X_i is independent of its nondescendants given its parents in G. The second component Θ denotes the set of parameters of the network. This set contains the parameter $\theta_{x_i|\pi_i} = P_B(x_i|\pi_i)$ for each realization x_i of X_i conditioned on π_i, the set of parents of X_i in G. Accordingly, B defines a unique JPD over V, namely:

$$P_B(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P_B(X_i|\pi_i) = \prod_{i=1}^{n} \theta_{X_i|\pi_i}.$$

For simplicity of representation we will omit the subscript B.

If X_i has no parents, its local probability distribution is said to be *unconditional*; otherwise, it is *conditional*. If the variable represented by a node is *observed*, then the node is said to be an *evidence node*; otherwise, the node is said to *be hidden* or *latent*. The complexity of a domain may be reduced by models and algorithms that describe an approximated reality. When variable interactions are too intricate for application of an analytic model, we may represent current knowledge about the problem, such as a cause generating at least one effect (Pearl 2000), where the final effect is the target of the analysis; for example, in Figure 7.3, the network topology (Lauritzen and Spiegelhalter 1988) of cause and effect is built by choosing a set of variables (e.g., "Visit Africa," "Smoking") that describe the domain (a patient presents some problems, and the physician wants to identify his or her disease and the correct therapy). The domain knowledge allows experts to draw an arc to a variable from each of its direct causes (i.e., visiting Africa may cause tuberculosis). Given a BN that specified the JPD in a factored form, one can evaluate all possible inference queries by marginalization, i.e., summing out over irrelevant variables.

Two types of inference support are often considered: *predictive support* for node X_i, based on evidence nodes connected to X_i through its parent nodes (called also

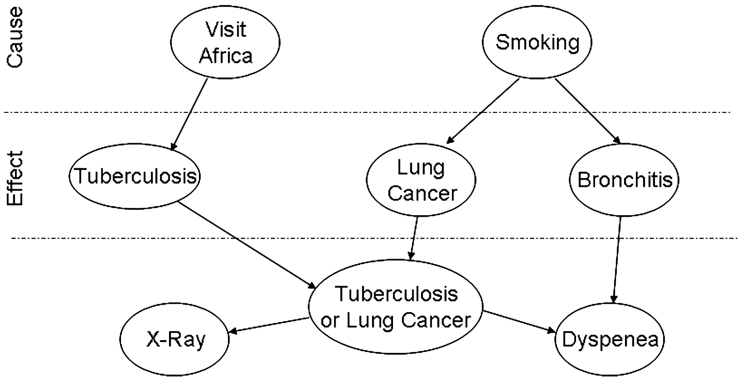

Figure 7.3 An example of a causal network (Lauritzen and Spiegelhalter 1988).

top-down reasoning), and *diagnostic support* for node X_i, based on evidence nodes connected to X_i through its children nodes (called also *bottom-up reasoning*). In general, the full summation (or integration) over discrete (continuous) variables is called *exact inference* and is known to be an NP-hard problem. Some efficient algorithms exist to solve the exact inference problem in restricted classes of networks. In many practical settings, the BN is unknown and one needs to learn it from the data. This problem is known as the *BN learning problem*, which can be stated informally as follows: Given training data and prior information (e.g., expert knowledge, causal relationships), estimate the graph topology (network structure) and the parameters of the JPD in the BN. Learning the BN structure is considered a harder problem than learning the BN parameters.

Moreover, another obstacle arises in situations of *partial observability* when nodes are hidden or when data is missing. In the simplest case of known BN structure and full observability, the goal of learning is to find the values of the BN parameters (in each CPD) that maximize the (log) likelihood of the training dataset. This dataset contains m cases that are often assumed to be independent. Given training dataset $\Sigma = \{\mathbf{x}_1, \ldots, \mathbf{x_m}\}$, where $\boldsymbol{x_l} = (x_{l1}, \ldots, x_{ln})^T$, and the parameter set $\theta = (\theta_1, \ldots, \theta_n)$, where θ_i, is the vector of parameters for the conditional distribution of variable X_i (represented by one node in the graph), the log likelihood of the training dataset is a sum of terms, one for each node:

$$\log L(\Theta|\Sigma) = \sum\nolimits_m \sum\nolimits_n \log P(x_{li}|\pi_i, \theta_i).$$

The log-likelihood scoring function *decomposes* according to the graph structure; hence, one can maximize the contribution to the log likelihood of each node independently. Another alternative is to assign a prior probability density function to each parameter vector and use the training data to compute the posterior parameter distribution and the Bayes estimates. To compensate for zero occurrences of some sequences in the training dataset, one can use appropriate (mixtures of) conjugate prior distributions, e.g., the Dirichlet prior for the multinomial case or the Wishart prior for the Gaussian case. Such an approach results in a maximum a posteriori estimate and is also known as the *equivalent sample size* (ESS) method.

BNs are gaining popularity in a wide range of application areas such as risk management (Cornalba et al. 2007) and management science in general (Godfrey et al. 2007; Kenett et al. 2008). Availability of software for analyzing BNs is further expanding their role in decision analysis and decision support systems (Jensen 2001; Bnlearn 2008; Genle 2006; Hugin 2007; SPSS 2008). BNs can be used to validate hypotheses about the relationships between usability design defects and the visitor's behavior in response to occurrences of usability barriers. This will be demonstrated in Section 7.5.

7.3.6 Model 4—Data Analysis in Usability Diagnostics

In e-commerce, we assume that site visitors are task driven, but we do not know if the visitors' goals are related to our website. Also, we have no way to tell if they know anything about the site, if they believe that the site is relevant to their goals, or if they have visited it before; it may be that the visitors are exploring the site or following a procedure to accomplish a task. Yet, their behaviors reflect their perceptions of the site's contents and their estimates of their effort in subsequent site investigation. The diagnosis of usability design deficiency involves certain measurements of site navigation, statistical calculations, and statistical decision. A model of statistical manipulations of server log data is now described which enables statistical decisions about possible usability design deficiencies.

7.3.6.1 Data Sources. Data analysis is possible only if the logs of the visitors' activities include at least three attributes:

- The IP address, used to distinguish between users.
- The page URL, used to specify the particular page.
- The timestamp, used to compute the time interval between hits.

7.3.6.2 Quantifying Website Usability Attributes. How can we tell whether the visitor encounters any difficulty in any of the evaluation stages? Server logs provide timestamps for all hits, including those of page html text files but also those of image files and scripts used for the page display. The timestamps of the additional files enable us to estimate three important time intervals:

- The time the visitors wait until the beginning of file download, used as a measure of page responsiveness.
- The download time, used as a measure of page performance.
- The time from download completion to the visitor's request for the next page, in which the visitor reads the page content but also does other things, some of them unrelated to the page content.

The challenge now is to decide whether the visitors feel comfortable with these time intervals. When do they feel that they wait too much, when is the download time too long or too short, and how do the visitors feel about what they see on screen?

7.3.6.3 Time Analysis of Mental Activities. How can we conclude that visitors consider a time interval acceptable, too short, or too long? For example, consider an average page download time of five seconds. Site visitors may regard it as too long if they expect the page to load rapidly, for example, in response to a search request. However, five seconds may be quite acceptable if the users' goal is to learn or explore specific information that they expect is related to their goal. Diagnostic-oriented time analysis observes the correlation between page download time and page exits. If the visitors care about the download time, then the page exit rate should depend on the page download time. When the page download time is acceptable, visitors may stay at the site, looking for additional information. When the download time is too long, more visitors might abandon the site and go to competitors. Longer download times imply higher exit rates; otherwise, if the visitors are indifferent about the download time, then the page exit rate should be invariant with respect to the page download time. We can now change our perspective on these variables and consider how the download time depends on the exit behavior. We compare the download time of successful visits with that of visits that ended in site exit. If the site visitors care about the download time, we should expect the average download time of those visitors who abandoned the site to be longer than the average of those who continued with site navigation. Otherwise, if the site visitors are indifferent about the download time, we should expect that the download time of the two groups will not be significantly different.

To determine the degree of the usability barrier, we need statistics. A reasonable statistic may be the correlation between download time and exit rate. To determine the significance of the usability barrier, we compare the download time of two samples: one of page views that ended in site exit and the other of all other page views. The null hypothesis is that the two samples are of the same population. If the null hypothesis is rejected, we may conclude that the page visitors' behavior depends on the download time: If the download time of the first sample exceeds that of the second sample and the difference is statistically significant, then we may conclude that the download time of the particular page is significantly too long. A simple two-tailed T-test may be sufficient to decide whether the visitors' behavior is sensitive to the page download time.

7.3.6.4 Analysis of the Visitor's Response Time. In the time interval between the page display on screen and the visitor's action to get the next page, the visitor is busy doing different things. Some of them are mental activities related to the visitor's task, and others are unrelated to the visitor's task (e.g., having a coffee break). It would useful to estimate the task-related and idle (those unrelated to the task) parts of the response time. The problem is that there is no way to distinguish between the two types of activities by manipulations of server logs. However, it makes sense to assume that the variance of the task-related activities is much smaller than that of the idle activities. We can diminish the effect of the idle activities by statistics, such as harmonic average, that weigh more on short time intervals and less on long time intervals.

7.3.6.5 Methodology for Usability Diagnostics. The goals of usability diagnostics are to identify, for each site page, all the design deficiencies that hamper the positive navigation experience at each evaluation stage. To understand the user's experience, we need to know the user's activity compared to the user's expectation. Neither one is available from the server log file, but they can be estimated by appropriate processing. The methodology here involves integration of two types of information:

- Design deficiencies, which are common barriers to seamless navigation, based on the first part of the model described above—visitor's page evaluation.
- Detectors of these design deficiencies, common indicators of possible barriers to seamless navigation, based on the second part of the model—visitor's reaction.

7.3.6.6 Usability Problem Indicators. The way to conclude that the download time of a particular page is too long is by a measure of potentially negative user experience, namely, the site exit rate. Exit rate is a preferred measure for deciding on the effect of download time, but it is irrelevant to the analysis of the visitors' response time. The reason for this is that server logs do not record the event of the visitor's leaving the site, so we cannot measure the time intervals of terminal page views. Therefore, we need to use other indicators of a potentially negative navigation experience. Model 2, the user's mental activities in website navigation described above, listed the most likely visitors' reactions to exceptional situations. Based on this model, we can list the following indicators of the visitor's tolerance of webpage design deficiencies:

- The visitor returning to the previous page may indicate that the current page was perceived as less relevant to the goal than the previous page.
- The visitor linking to the next website page may indicate that the link was perceived as a potential bridge to a goal page.
- The visitor activating the main menu may indicate that he or she is still looking for the information after failing to find it in the current page.
- The visitor exiting the site may indicate that either the goal has been reached or the overall site navigation experience became negative.

So, besides site exit, other indicators of potential usability problem are the rates of navigation back to a previous page and escape from the page to the main menu. These events, as well as the event of site exit, are called here *usability problem indicators* (UPIs).

7.3.6.7 Time Analysis of the Task-Related Mental Activities. Once we have an estimate of the task-related mental activities and a UPI suited to nonterminal activities, we can adapt the method for analyzing problematic performance to analyzing problematic task-related mental activities. How can we conclude that the response time interval is proper for the page, too short, or too long? For example, consider an average page reading time of 20 seconds. Is it too long or too short? Obviously, the reading time of long pages should be longer than that of short pages, so we should not expect absolute measures to be valuable in scoring the value of the reading time. Visitors who feel that the information is irrelevant to their needs are more likely to respond quickly, go backward, or select a new page from the main menu. Therefore, the average time of a page for visitors who navigated backward or retried the main menu should be shorter than that of the average of all visitors. On the other hand, visitors who believe that the information is relevant to their needs, but do not easily understand the page text, are likely to spend more time than average reading the page content, and the average time on the page should be longer. The following assumptions are used for the diagnosis:

- Visitors satisfied with the page display are less likely to exit the site, to return to a previous page, or to escape to the main menu than visitors not satisfied with the page display.
- The exception to the first assumption are terminal pages, at which the users' goals are achieved.

The time that visitors spend reading a page depends on various perceptual attributes, including the relevance of the page to their goals, the ease of reading and comprehending the information on the page, the ease of identifying desired hyperlinks, etc. The method described above for download time analysis may be applicable to these perceptual attributes, provided that we know how usability barriers should affect the visitor's response time. The analysis is based on the following assumptions:

- Task-driven visitors are sensitive to readability deficiencies: Both the response time and the rate of UPIs for pages that are easy to read should be lower than those for pages that are difficult to read. Casual visitors, on the other hand, are less sensitive to readability deficiencies.
- Task-driven visitors are sensitive to the relevance of the information on the page to their needs, but in a different way: The UPI rate of relevant pages should be lower than that of irrelevant pages, but the response time of relevant pages should be higher than that of irrelevant pages Casual visitors, on the other hand, are less sensitive to relevance deficiencies.

7.3.6.8 Interpreting the Time Analysis of Page Readability. The time that visitors spend reading a page depends on various perceptual attributes, including the relevance of the page to their goals, the ease of reading and comprehending the information on the page, the ease of identifying desired hyperlinks, etc. Assume that for a particular page, the average time on screen before backward navigation is significantly longer than the average time over all page hits. Such a case may indicate a readability problem due to design flaws, but it can also be due to good page content, which encouraged users who spent a long time reading the page to go back and reexamine the previous page. It would be nice if we could distinguish between reading the page and other task-related mental activities, such as evaluating a product on sale, comparing prices, etc. Unfortunately, there is no direct way to measure the time it takes for the visitor to accomplish each of the mental activities. Therefore, the convention here is that a problem identified in page readability could be attributed to the other task-related mental activities. This is yet another example of the limits of artificial intelligence and of our need to rely on the human intelligence.

7.3.6.9 Interpreting the Time Analysis of Page Relevance to the Visitors' Needs. Task-driven visitors are likely to respond quickly by a invoking a UPI if they perceive irrelevant information that threatens to lose their focus. Possible sources of such perception are inappropriate content, such as banners distracting visitors from their original tasks; inappropriate layout design; and misleading links to the page due to unethical marketing such as by portals, to poor explanations about the links or to labeling mistakes. The psychological explanation for such perceptions is the limited capacity of human working memory. Too many mental resources required for processing the information on a page might compete with the need to mentally maintain the visitor's goal. Consequently, task-driven users will always prefer navigating pages that do not require too much mental processing. Another possible reason for quick activation of a UPI is that the page is well designed and the page designers wanted visitors who are finished reading the page to navigate backward or try an item from the main menu. Apparently, the latter reason is hypothetical; the main reason for quick UPI activation is poor design, which forces the visitor to take measures in order to stay in focus.

7.3.6.10 Analysis of Misleading Links. One cause of visitors' perception of the page as irrelevant is when a link from a portal or from another page is misleading, either intentionally or by a design mistake. Once a page is identified as being perceived as irrelevant to visitors, the site administrator may want to know why and what links are responsible for the visitors' negative attitude. The problematic links can be found by time analysis similar to that of page response time. This is achieved by sampling the transitions from all pages to the tested page and by comparing the response time of the transitions that were followed by a UPI with that of the other transitions.

7.4 IMPLEMENTATION FRAMEWORK

The implementation framework in this section is based on the structure of reports derived by WebTester, a patented software tool developed by ErgoLight since 1999 to analyze server log files and generate reports about potential usability deficiencies in websites (see www.ergolight-sw.com). This tool demonstrates the feasibility of the method presented in Section 7.2. The usability data extracted using this tool are used in the next section to demonstrate the kinds of diagnostics the method can provide.

7.4.1 WebTester Reports

The reports that WebTester generates include site-level scoring, usage statistics, and diagnostic information as follows:

Page Diagnostics

- Pages that are abandoned due to long download time
- Pages that are difficult to comprehend
- Pages that are not interesting or irrelevant
- Pages that distract potential customers from their goals

Link Diagnostics

- Links that are likely to be misinterpreted
- Links that are likely to be disregarded
- Links that do not behave according to visitors' expectations
- Links that connect to the wrong page

Page Statistics

- Number of page views (view count)
- Number and percentage of site entries from the pages
- Number and percentage of site exits from the pages
- Average download time
- Average reading time
- Average search time for links to these pages

7.4.2 Data Processing

7.4.2.1 Preliminary Activity—Analysis of the Structure of Server Log Files. Server log files consist of records of optional fields. WebTester analyzes the record format and identifies most of the fields automatically according to the two main format conventions (UNIX and MS). Few fields, such as the Size field, are of common integer format and cannot be identified automatically. The operator's control is required to select the proper field and to confirm the automatic field identification.

7.4.2.2 The Lowest Layer—User Activity. WebTester processes the server logs to obtain compressed logs of user activity, in which the records correspond to significant user actions (involving screen changes or server-side processing). The data processing includes algorithms for:

- Visitor identification, including support for variable IP addresses.
- Filtering out traces of robot visits. Robots are identified by the Referrer field, by the User Agent field, or by certain characteristics of the robot sessions.

- Associating embedded objects (images, JavaScripts, etc.) with the page containing them.
- Associating pages with frames by time proximity.

7.4.2.3 The Second Layer—Page Hit Attributes. WebTester computes estimates of record-level usage attributes:

- Page size.
- Page download time.
- Page introduction time, namely, the download time of the first visit in a navigation session.
- Page processing time, calculated by subtracting the statistics of download time of elementary objects (images, JavaScripts, etc.) from the download time.
- User response time (from end-of-page download to beginning of the download of the next page).

The page response time, which may enable diagnostics of lack of or extreme system response time, is not available in server logs, and it was not included in standard usability reports.

7.4.2.4 The Third Layer—Transition Analysis. Using the statistics for page transitions WebTester identifies:

- The site main pages (those accessible through main menus).
- Repeated form submission (indicative of visitors' difficulties in form filling).

7.4.2.5 The Fourth Layer—UPI Identification. Marking certain visits as indicators of possible navigational difficulty:

- Estimates for site exit by the time elapsed until the next user action (as no exit indication is recorded on the server log file).
- Backward navigation.
- Transitions to main pages, interpreted as escaping the current subtask.

7.4.2.6 The Fifth Layer—Usage Statistics. Obtaining statistics of hit attributes over pages, such as:

- Average page size.
- Average introduction time.
- Average download time.
- Average time between repeated form submission.
- Average time on the screen (indicating content-related behavior).
- Average time on a previous screen (indicating ease of link finding).

WebTester computes the statistics over all page views and also over those views that indicate possible difficulties. The average time computations are by the harmonic average, with the intention to weigh out fluctuations due to technical problems and visitors' idle time. A typical e-commerce URI consists of a common page (html, asp, etc.) and parameters defining the deliverables. Typically, the parameters are arranged according to the hierarchy of the deliverables, namely, categories of products or services. In practice, the samples for full URIs are too small to enable significant results. Also, it is desired that the diagnostics are provided for all levels along the URI parameters to support all levels of the categorization. The statistics in WebTester are calculated for all categories, of all levels of detail.

7.4.2.7 The Sixth Layer—Statistical Decision. For each of the page attributes, WebTester compares the statistics over the exceptional page views to those over all the page views. The null hypothesis is that (for each attribute) the statistics of both samples are the same. A simple two-tailed T-test was used to reject it and therefore to conclude that certain page attributes are potentially problematic. The error level was set to 5%.

7.4.2.8 The Top Layer—Interpretation. For each of the page attributes, WebTester provides a list of possible reasons for the difference between the statistics over the exceptional navigation patterns and those over all the page hits. However, the usability analyst must decide which of the potential source of visitors' difficulties is applicable to the particular deficiency.

7.5 CASE STUDIES OF WEBSITE USABILITY DATA ANALYSIS

In this section, we present case studies of Web-based usability data analysis using models 1–4 described above. The data used in this section was obtained by applying ErgoLight WebTester to server logs of the website of www.israeliz.com. The reports generated by applying the tool are available at the ErgoLight website, in the section of report examples: http://www.ergolight-sw.com/pub/Sites/Israeliz/All/First/Reports/wtFrameSet.html. Not all visitors' actions are recorded on the server. Yet, the most significant user actions are the page views, that result in screen changes and in form submission, which have sufficient traces in the server log file. We first present a case study illustrating the application of Markov chain to web statistics and then focus on an implementation of BNs.

7.5.1 Markov Chain Applied to Web Usability Statistics

The apparent UPIs include

- Backward navigation.
- Site exit.
- Escape to the main menu.

A scheme for page event sequences is: The user links to Page -> Page starts download -> Page download complete -> User recognizes page -> User reads page -> User activates a link, about to leave the page. Table 7.1 presents the measured variables and their definitions in the dataset we have analyzed. In that table, key variables are highlighted. The time variables used for the statistics are variables that apparently have an impact on the visit experience. They are:

- The page download time, measured from the first page view until the end of subsequent embedded files (images, JavaScripts, etc.).
- The elapsed time until the next user event, interpreted as reading time.
- The elapsed time since the previous user event, interpreted as the time until the user found the link to this page.

TABLE 7.1 Variables in Case Study (Key Variables Highlighted)

ID	Variable	Definition
1	ActionCount	"Count of page visits as recorded in the log file, including actual visits and faked visits (refreshing and form submission)"
2	NuOfRealVisits	"The actual visits, after removing the faked visits."
3	**TextSize**	Size of html page
4	**PageSize**	TextSize + size of subsequent images recorded individually
5	NuOfEntries	Count of site entries through this page
6	NuOfPreExits	Count of site exit through next page
7	NuOfExits	Count of site exit through this page
8	NuOfPreBacks	Count of backward navigations from this page
9	NuOfBacks	Count of backward navigations to this page
10	NuOfDownloads	Count of page visits with positive download time
11	NuOfGetToPost	"Count of transitions from Get (typically, viewing) to Post (typically, form submission)"
12	TotalDownloadFrequency	Sum of (1/DownloadTime) over all real page visits
13	TotalReadingFrequency	Sum of (1/ReadingTime) over all real page visits
14	TotalSeekingFrequency	Sum of (1/SeekingTime) over all real page visits
15	TotalProcessingFrequency	Sum of (1/ProcessingTime) over all real page visits
16	TotalGetToPostFrequency	Sum of (1/FormFillingTime) over all real page visits
17	AvgDownloadFrequency	Average = Total/Count
18	AvgReadingFrequency	Average = Total/Count
19	AvgSeekingFrequency	Average = Total/Count
20	AvgGetToPostFrequency	Average = Total/Count
21	AvgPostToPostFrequency	Average = Total/Count
22	**AvgDownloadTime**	Harmonic average of page download time over all real page visits
23	**AvgReadingTime**	Harmonic average of page reading time over all real page visits

(*Continued*)

TABLE 7.1 *Continued*

ID	Variable	Definition
24	**AvgSeekingTime**	Harmonic average of page seeking time over all real page visits
25	AvgProcessingTime	Harmonic average of page processing time over all real page visits
26	AvgResponseTime	Harmonic average of page response time over all real page visits
27	PercentExitsSlowDownload	"Number of page visits characterized as slow downloads, followed by site exits/NuOfRealVisits * 100"
28	PercentPbackSlowDownload	"Number of page visits characterized as slow downloads, followed by backward navigation/NuOfRealVisits * 100"
29	PercentExitsSlowSeeking	"Number of page visits characterized as slow seeking, followed by site exits/NuOfRealVisits * 100"
30	PercentExitsFastSeeking	"Number of page visits characterized as fast seeking, followed by site exits/NuOfRealVisits * 100"
31	PercentPbackSlowSeeking	"Number of page visits characterized as slow seeking, followed by backward navigation/NuOfRealVisits * 100"
32	PercentPbackFastSeeking	"Number of page visits characterized as fast seeking, followed by backward navigation/NuOfRealVisits * 100"
33	PercentPexitSlowReading	"Number of page visits characterized as slow reading, followed by site exit from next page/NuOfRealVisits * 100"
34	PercentPexitFastReading	"Number of page visits characterized as fast reading, followed by site exit from next page/NuOfRealVisits * 100"
35	PercentPbackSlowReading	"Number of page visits characterized as slow reading, followed by backward navigation/NuOfRealVisits * 100"
36	PercentPbackFastReading	"Number of page visits characterized as fast reading, followed by backward navigation/NuOfRealVisits * 100"
37	**UsabilityScore**	"Normalized harmonic average of: download time, seeking time, reading time"
38	**UsabilityAlert**	Sum of all UPIs (6–11 above)
39	**AccessibilityScore**	A measure of the speed of finding the links to the current page
40	**PerformanceScore**	A measure of the download speed
41	**ReadabilityScore**	A measure of the reading speed (characters per second)

The page visitor experience is regarded as negative or positive according to whether or not a UPI was indicated afterward. Table 7.2 is a summary table relating content and reading time to visit experience. The expected relationships between the key variables are as follows:

- Page size should directly affect the page download time.
- The amount of text on the page should directly affect the page reading time.
- The ratio TextSize/PageSize may be equal to 1 if the page is pure html (has no graphics, video, etc.) or is very small (since images consume much more space than text).
- Page download time should affect the visit experience. Too long a download time should result in a negative visit experience. Research indicates that too short a download time might also result in a negative visit experience; however, this might be a problem only for extremely good service (server performance).
- Page-seeking time may affect the visit experience. Too much seeking time might have a "last straw" effect if the page is not what visitors expected to find.
- Page reading time may affect the visit experience. Users may stay on the page because it includes a lot of data (in this case study, the data are textual) or because the text they read is not easy to comprehend.

The descriptive statistics of key variables from 122 links are presented in Table 7.3.

Figure 7.4 presents a cluster analysis of the key variables using MINITAB™ version 14 (www.minitab.com). Processing time and response time are naturally closely related. The three score functions form a separate cluster indicating internal consistency.

Figure 7.5 presents histograms of TextSize, PageSize, UsabilityScore, and UsabilityAlert.

We apply the methods about Markov processes illustrated in Section 7.1 to the data at hand. For privacy protection reasons, we modified the names of the webpages. We illustrate the methods when concentrating on one page and analyzing the moves out of it. We consider the case of the webpage *content/man8.htm* and the other ones accessible from it. We identify them as *content/man1.htm*, *content/man6.htm*, *content/man7.htm*, *tool/cont200.htm*, and *tool/cont202.htm*, and we add the state corresponding to the exit from the website. The transition matrix of the Markov chain has a very high dimension since the website has many pages. We concentrate on the estimation of pages accessible

TABLE 7.2 The Effect of Content and Reading Time on Usability

Property of Page Content	Average Reading Time	Visit Experience
Reasonable text size	Short	Positive
Reasonable text size	High	Negative
Large text size	High	Positive
Animation	Short	Negative

TABLE 7.3 Descriptive Statistics of Key Variables in the Case Study Data Set

Variable	Mean	StDev	CoefVar	Minimum	Median	Maximum
TextSize	20109	15817	78.66	0.000000000	17941	72359
PageSize	79577	259545	326.16	0.000000000	40273	2352613
PercentExitLoss	14.25	15.95	111.99	0.000000000	10.00	100.00
AccessibilitySco	15.50	17.67	113.97	0.0178	8.74	76.70
PerformanceScore	46.46	25.13	54.10	3.53	41.70	100.00
ReadabilityScore	2780	4746	170.71	0.000000000	1409	43135
UsabilityScore	73.04	27.15	37.17	14.22	69.61	139.40
UsabilityAlert	58.83	39.54	67.21	0.000000000	52.50	222.00
AvgDownloadTime	2.164	4.243	196.09	0.000000000	1.000	27.000
AvgReadingTime	153	1141	746.28	0.000000000	10.0	12288
AvgSeekingTime	62.2	508.6	818.12	0.000000000	10.0	5624.0

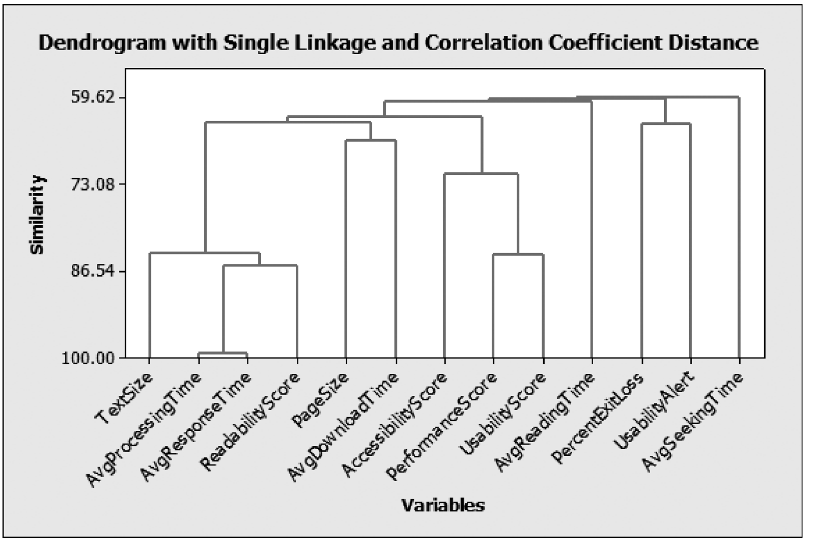

Figure 7.4 Dendrogram from cluster analysis of variables in Table 7.3.

only from *content/man8.htm* and consider a Dirichlet prior with coefficients equal to 1 for the probabilities of transition to accessible states and 0 for the others. This assumption allows only for transitions to the pages linked by *content/man8.htm* and does not take into account a possible move to another page of the website just by typing its address. In this context, we observed 16 transitions to six states. Table 7.4 gives the data and both frequentist and Bayesian estimates.

Both types' estimates have positive and negative properties. The frequentist estimate does not exploit available information, if any, but gives a unique result, not

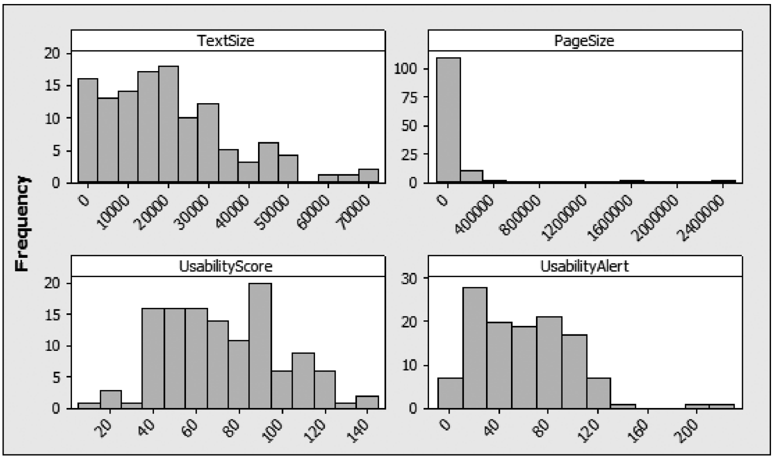

Figure 7.5 Histograms of TextSize, PageSize, UsabilityScore, and UsabilityAlert.

TABLE 7.4 Estimation of Transition Probabilities

Page	Number of Transitions	Frequentist Estimate	Bayesian Estimate
content/man1.htm	2	0.125	0.136
content/man6.htm	2	0.125	0.136
content/man7.htm	4	0.250	0.228
tool/cont200.htm	2	0.125	0.136
tool/cont202.htm	2	0.125	0.136
outside	4	0.250	0.228

questionable once this approach is taken. The Bayesian estimate incorporates the expert's opinion at the price of a strong dependence on this prior and possible wrong elicitation. A sensitivity study, in the spirit of Rios Insua and Ruggeri (2000), can attenuate the impact of a wrong statement about the prior distribution. Furthermore, it is worth observing that the prior Bayesian estimates were all equal to $1/6$ and the posterior Bayesian estimates are, as expected, between the prior Bayesian estimates and the maximum likelihood estimates. Both estimates indicate that transition probabilities to the outside and to *content/man7.htm* are highest.

7.5.2 BNs Applied to the Web Usability Statistics

The data were then reanalyzed using the basic BN presented in Figure 7.6. The network combines background information with a learned network generated using the GeNIe version 2.0 software (http://genie.sis.pitt.edu).

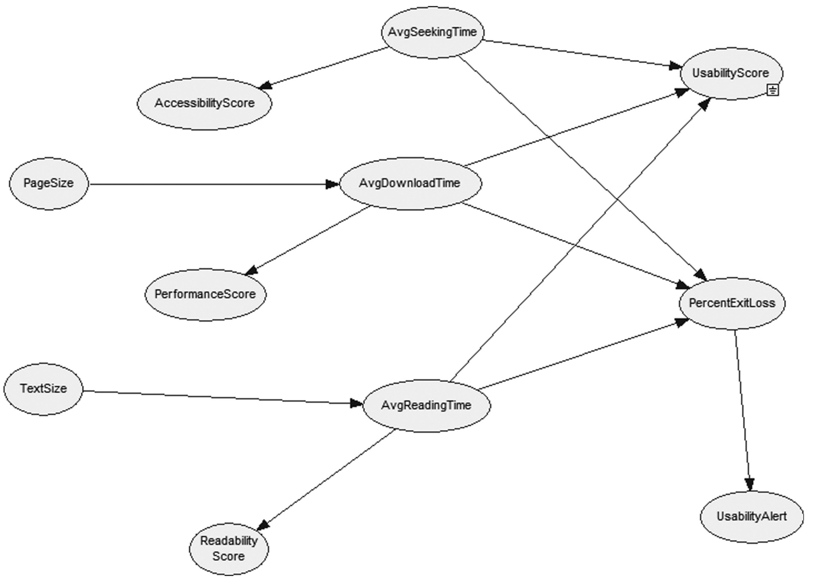

Figure 7.6 BN of key variables.

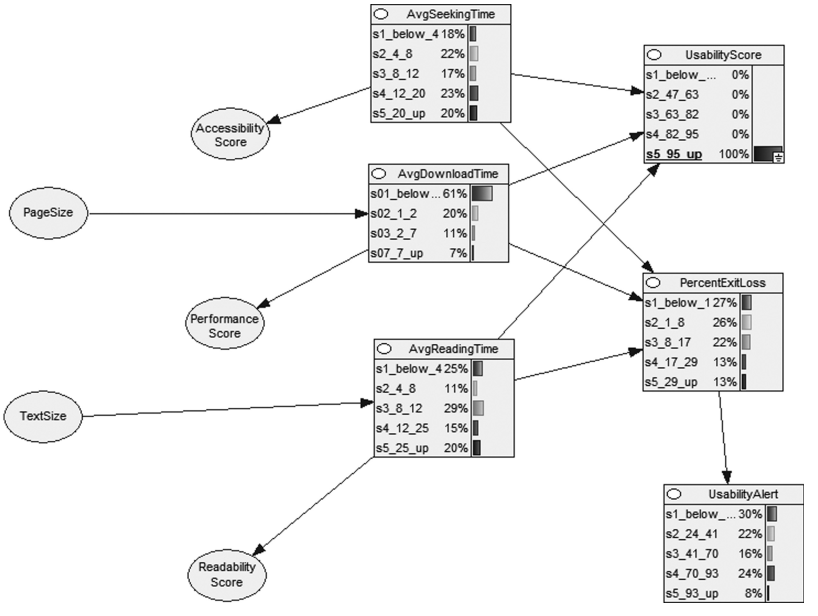

Figure 7.7 Diagnostic distributions conditioned on the UsabilityScore being at its highest level.

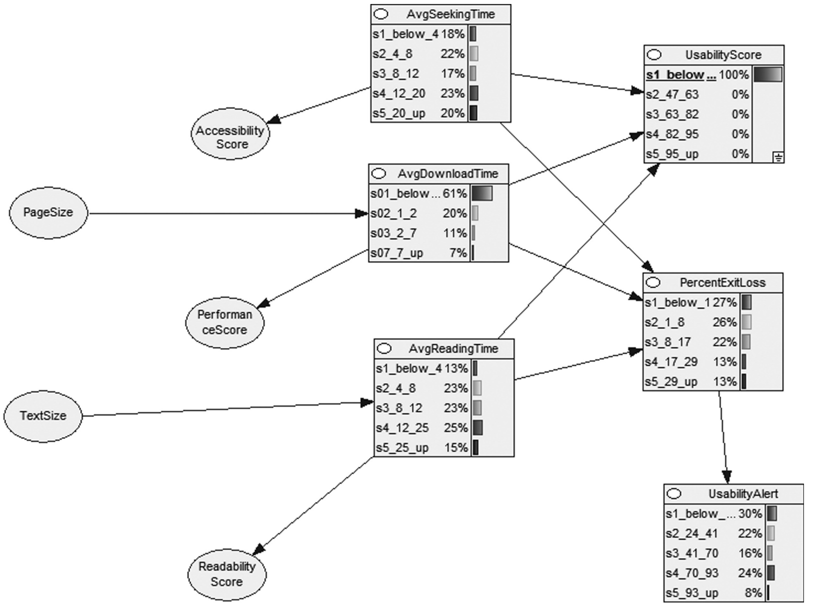

Figure 7.8 Diagnostic distributions conditioned on the UsabilityScore being at its lowest level.

One the basis of the network, we can perform various diagnostic checks. Figure 7.7 presents distributions of various variables conditioned on the UserScore being at its its highest level. Note that the percent of AvgSeekingTime, AvgDownloadTime, and AvgReadingTime at the lowest levels is 18%, 61%, **25%**, and respectively. We can now perform similar diagnostics by conditioning the UsabilityScore to be at its lowest level. Figure 7.8 provides such an analysis.

Note that the percent of AvgSeekingTime, AvgDownloadTime, and AvgReadingTime at the lowest levels is 18%, 61%, and **13%**, respectively. AvgReadingTime has dropped from 25% to 13%. The other variables have not changed. AvgReadingTime is therefore the differentiating factor between high and low values of UsabiliScore. Moreover, conditioning on UsabilityScore has no impact on UsabilityAlert.

7.6 CONCLUSIONS AND AREAS FOR FUTURE RESEARCH

This chapter has presented a method for identifying design deficiencies that result in degraded usability based on web usage data. The method is based on the integration of models from statistics with models of visitors' behavior. It was implemented in a report-generating tool, WebTester. We also introduced several modeling approaches and discussed case studies using Markov chains and BNs. Further analysis linked UPIs to conversion rates. A general approach for such an analysis involves proactive interventions and expert panels.

7.6.1 Limitations of the Example Diagnostic Reports

7.6.1.1 Standard Log Files Are Insufficient. WebTester had produced reports for about 20 websites; five of them concerned small businesses. All of these reports were based on data in server log files. The sample data presented here show the potential of applying this method. Yet, the results are not very impressive due to shortcomings of the source data. Server logs are used extensively in the industry because of their many advantages:

- The web server normally produces log files routinely, so the raw data are available for the analysis with no extra efforts.
- The web server reliably records every transaction it makes.
- The data are on the company's own servers, enabling the company to control the file format.
- The data are in a standard format, ensuring compatibility with commercial log analyzers.
- Log files contain information on visits from search engine spiders, which enables search engine optimization.
- Log files contain information on client requests that the server failed to execute.

However, server logs are ineffective for usability diagnostics. The main limitation is that server logs do not have indications about all necessary client activity. The most important problems are:

- Lack of indication of page reload from the cache, which is essential for employing backward navigation required for subjective readability and relevance analysis. If a person revisits a page, the second request will often be retrieved from the browser's cache, so no request will be received by the web server. This means that the person's path through the site is lost. Caching can be defeated by configuring the web server, but this can result in degraded performance for the visitor to the website.
- Missing records about the visitors' interaction with order forms. The server log does not include records of the visitor's clicks, only of the GET and PUT requests from the client to the server. The problem is that most form-filling activities are processed locally, without notifying the server about them, which means that they are not recorded on the server logs. This limitation is quite significant, because major usability barriers involve the visitors' difficulties in following the flow of data entry and editing and recalling the proper data formats.
- Missing records of pages loaded by third parties, such as gateways (web proxies).
- A convenient way to work around this problem is to install Java script tags in the page views so that the client will deliver a request to the server about any significant visitor event, such as data entry in form filling. This technique is used extensively in e-commerce sites. The cache problem is solved with no extra effort, because the request is sent to the server on each page view. Another limitation of server logs is the high level of noisy data. Main sources of the noisy data are:
- An IP address shared by different people, such as members of the same household. This problem is negligible in most practical cases, as the likelihood that two or more people who share the same IP address will visit the same website simultaneously is very low. Also, visitor ambiguity may be reduced by identifying the visitors not only by the IP address, but also by the User Agent (the browser used for the navigation). This method was implemented in WebTester.
- Distributed IP addressing in a multiserver operational environment, in which the client request arrives through different routing from the client browser. In e-commerce this may be a problem in b2b applications, in which the client is an enterprise operated through a complex system. A distributed IP address system has a common part and a variable part. Records from distributed IP addresses can be aggregated to the proper session by the common part of the IP address. This method was implemented in WebTester.
- Spider visits (by search engine crawlers) are recorded in the server log files. These visits can be detected and filtered out by special characteristics of the

spiders, including a special User Agent (browser identification), known referrer, and certain navigation patterns (e.g., extremely fast access to the server). These methods were implemented in WebTester.

- Obviously, the effect of noise can be reduced by enlarging the sample size. In WebTester, the size of log files used was 10–20 megabytes, and apparently the sample size for main pages was sufficient. To enable support of this capability, a new log analyzer should be developed that will use a proprietary clickstream logger, such as by Java script tagging.

7.6.1.2 Learning Effect. Johnson et al. (2000) confirmed that the power law of practice used in cognitive science also applies to website navigation, implying that users spend less time per session the more they visit the site. However, another study shows that this is not a real limitation. Bucklin and Sismeiro (2001) found that repeat visits by a user lead to fewer pages viewed per session but to no change in average page view duration. This means that the power law of practice applies to learning of the navigation path, but not to the usability attributes of subjective readability and relevance to the visitors' needs.

7.6.2 Further Analysis

7.6.2.1 Proactive Interventions. Proactive interventions such as design changes are, de facto, empirical experiments. The statistical methodology for the design of experiments involves linking variables and parameters using eight steps explained in Table 7.5 that identify factors, specify their levels, and, using appropriate combinations (experimental runs), determine an experimental array. Responses are the measured target variables of an experiment (Kenett and Zacks 1998).

7.6.2.2 Expert Panel. In order to determine the percentage of usability design flaws that can be identified based on log files, one can use an expert panel. In

TABLE 7.5 Designed Experiment Checklist

Design Concern	Design Activity
1. Problem Definition	Plain language description of the problem and how it was identified
2. Response Variables	What will be measured and how the data will be collected
3. Factors	What factors can be controlled by design decisions
4. Factor Levels	What are the current factor levels and what are reasonable alternatives
5. Experimental Array	What are the factor level combinations to be tried and in what order
6. Number of Replications	How long will the experiment run
7. Data Analysis	How the data will be analyzed, with what software
8. Budget and Project Control	Who is in charge of the project, what is the timetable, and what resources are needed

running such a panel, one first lets usability experts analyze specific site usability heuristically (few experts are required for a site, as demonstrated in prior research). Their evaluation can be, for example, recorded using a specially designed questionnaire. Second, one compares the effect of usability problems identified though web surfing usage data with the expert opinions. As an example, consider the dilemma of banners and online advertising. Clearly, advertisement on websites hampers usability. If we want to determine the effect on conversion rates, we can run an experiment with and without ads. This is a simple experiment with one factor at two levels. Measurement of conversion rates and UPIs will help us determine if we have reached the right trade-off between advertisement design and usability. Finally, in considering usability, the temporal dimension should also be analyzed. Usability can be affected by changes in performances over time. In tracking time, many approaches based on sequential statistical analysis can be implemented, such as statistical process control or more advanced procedures such as Shiryayev-Roberts optimal detection (Kenett and Pollak 1986; Kenett and Zacks 2003).

REFERENCES

Basili, V. and Weiss, D. (1984) A methodology for collecting valid software engineering data. *IEEE Transactions on Software Engineering*, 728–738.

Battey, J. (1999). IBM's redesign results in a kinder, simpler web site. Retrieved October 10, 2001, from http://www.infoworld.com/cgi-bin/displayStat.pl?/pageone/opinions/hotsites/hotextr990419.htm.

Ben Gal, I. (2007). Bayesian networks. In *Encyclopaedia of Statistics in Quality and Reliability* (F. Ruggeri, R.S. Kenett, and F. Faltin, editors in chief), Wiley, Chichester, UK.

Bnlearn Package (2008) Available at http://cran.dsmirror.nl/web/packages/bnlearn/index.html.

Bucklin, R.E., Lattin, J.M., Ansari, A., Bell, D., Coupey, E., Gupta, S., Little, J.D.C, Mela, C., Montgomery, A., and Steckel, J. (2002). Choice and the Internet: From clickstream to research Stream. *Marketing Letters*, 13(3): 245–258.

Bucklin, R.E. and Sismeiro, C. (2001). A model of web site browsing behavior estimated on clickstream data. Working Paper, Anderson School at UCLA.

Cornalba, C., Kenett, R., and Giudici, P. (2004). Sensitivity analysis of Bayesian networks with stochastic emulators. *Proceedings of the ENBIS-DEINDE Conference*, University of Torino, Turin, Italy.

Dabrowski, J.R. and Munson, E.V. (2001). Is 100 milliseconds too fast? *Conference on Human Factors in Computing Systems*, Seattle, Washington.

Di Scala, L., La Rocca, L., and Consonni, G. (2004). A Bayesian hierarchical model for the evaluation of web sites. *Journal of Applied Statistics*, 31(1): 15–27.

Donahue, G.M. (2001). Usability and the bottom line. *Software IEEE*, 18(1): 31–37.

Duma, J.S. and Redish, J.C. (1999). *A Practical Guide to Usability Testing*. Intellect Ltd, Bristol, UK.

Dustin, E., Rashka, J., and McDiarmid, D., (2001). *Quality Web Systems: Performance, Security, and Usability*. Addison-Wesley, Reading, MA.

GeNle (2006) Decision Systems Laboratory, University of Pittsburgh, USA. Available at http://genie.sis.pitt.edu.

Godfrey, A.B. and Kenett, R.S., (2007). Joseph M. Juran, a perspective on past contributions and future impact. *Quality and Reliability International*, 23: 653–663.

GVU (1997) GVU's 8th WWW User Survey. Available at http://www.gvu.gatech.edu/user_surveys/survey-1997-10.

Hugin Decision Engine (2007) Hugin Expert, Denmark. Available at http://www.hugin.com.

International Organization for Standardization (1998) ISO 9241-11, *Guidance on Usability*, Geneva, Switzerland.

Jensen, F.V. (2001). *Bayesian Networks and Decision Graphs*. New York: Springer.

Johnson, E.J., Bellman, S., and Lohse, G.L. (2000). what makes a web site "Sticky"? Cognitive lock in and the power law of practice. Working Paper, Graduate School of Business, Columbia University.

Karlin, S. and Taylor, H.M. (1975). *A First Course in Stochastic Processes* (2nd ed.). New York: Academic Press.

Karlin, S. and Taylor, H.M. (1981). *A Second Course in Stochastic Processes*. New York: Academic Press.

Kenett, R. (1996). Software specification metrics: A quantitative approach to assess the quality of documents. *Proceedings of the IEEE Convention of Electrical and Electronics Engineers in Israel*, Tel Aviv, Israel.

Kenett, R.S. (2007). Cause and effect diagrams. In *Encyclopaedia of Statistics in Quality and Reliability* (F. Ruggeri, R.S. Kenett, and F. Faltin, editors in chief), Wiley, Chichester, UK.

Kenett, R. and Baker, E. (1999). *Software Process Quality: Management and Control*. New York: Marcel Dekker.

Kenett, R., De Frenne, A., Tort-Martorell, X., and McCollin, C. (2008). The statistical efficiency conjecture. In *Applying Statistical Methods in Business and Industry—the State of the Art* (S. Coleman T. Greenfield, and D. Montgomery, eds.). Wiley, pp. 61–96.

Kenett, R. and Pollak, M. (1986). A semi-parametric approach to testing for reliability growth with an application to software systems. *IEEE Transactions on Reliability*, R-35(3): 304–311.

Kenett, R. and Zacks, S. (2003). *Modern Industrial Statistics: Design and Control of Quality and Reliability* (2nd ed.). San Francisco: Duxbury Press.

Kohavi, R. (2006). Focus the mining beacon: Lessons and challenges from the world of e-commerce. *San Francisco Bay ACM Data Mining SIG*, June 13.

Lauritzen, S.L. and Spiegelhalter, D.J. (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.

Marcus, A. (2002). Return on Investment for usable UI design. *User Experience Magazine*, pp. 25–31.

Moe, W.W. (2006a). An empirical two-stage choice model with varying decision rules applied to Internet clickstream data. *Journal of Marketing Research*, pp. 680–692.

Moe, W.W. (2006b). A field experiment assessing the interruption effect of pop-up promotions. *Journal of Interactive Marketing*, 20(1): 34–44.

Montgomery, A.L., Kannan Srinivasan, S.L., and Liechty, J.C. (2004). Modeling online browsing and path analysis using clickstream data. *Marketing Science*, GSIA Working Paper #2003-E26.

Nielsen, J. (1994). Response times: The three important limits. In *Usability Engineering*. San Francisco: Morgan Kaufmann, San Francisco, CA.

Nielsen, J. (1999). Why people shop on the Web Nielsen alert box. Available at http://www.useit.com/alertbox/990207.html.

Nielsen, J. and Loranger, H. (2006). *Prioritizing Web Usability*. New Riders Press, Indianapolis.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York.

Peterson, E. (2005). *Website Measurement Hacks, Tips and Tools to Help Optimize Your Online Business*. O'Reilly, Media Inc.

Rhodes, J.S. (2001). From banners to scumware to usable marketing. Available at http://webword.com/moving/scumware.html.

Rios Insua, D. and Ruggeri, F. (2000). *Robust Bayesian Analysis*. New York: Springer-Verlag.

Speier, C. and Valacich, J.S. (1999). The influence of task interruption on individual decision making: An information overload perspective. *Decision Sciences*, 30(2): 337–360.

Tilson, R., Dong, J., Martin, S., and Kieche, E. (1998). Factors and principles affecting the usability of four e-commerce sites. *Proceedings of the Conference on the Human Factors and the Web*, Barking Ridge, NJ.

User Interface Engineering (UIE) (2007) Research articles. Available at http://www.uie.com.

SPSS (2008) Clementine. Available at http://www.spss.com/clementine/capabilities.htm.

Zijlstra, F.R.H., Roe, R.A., Leonora, A.B., and Krediet, I. (1999). Temporal factors in mental work: Effects of interrupted activities. *Journal of Occupational and Organizational Psychology*, 72: 163–185.

ADDITIONAL BIBLIOGRAPHY

Becher, J. and Kochavi, R. (2001). Tutorial on e-commerce and clickstream mining. *First SIAM International Conference on Data Mining*, April 5.

Bevan, N. (1999). Common industry format usability tests. *Proceedings of UPA'99*, Usability Professionals Association, Scottsdale, Arizona, 29 June–2 July.

Bevan, N. and Macleod, M. (1994). Usability measurement in context. *Behavior and Information Technology*, 13: 132–145.

International Organization for Standardization (1999) ISO 13407 *Human-Centred Design Processes for Interactive Systems*, Geneva, Switzerland.

International Organization for Standardization (2006) ISO 25062 *Common Industry Format (CIF) for Usability Test Reports*, Geneva, Switzerland.

Kwan, T.T., McGrrath, R.E., and Reed, D.A. (1995). NCSA World Wide Web server: Design and performance. University of Illinois, Urbana, IL.

Lin, T.Y. and Siewiorek, D.P. (1990). Error log analysis: Statistical modeling and heuristic trend analysis. *IEEE Transactions on Reliability*, 419–432.

Mason, L. (2002). E-commerce data mining: Lessons learned and insights gained. *Proceedings of the American Statistical Association Annual Meeting*, Chicago.

Park, Y.-H. and Fader P.S. (2004). Modeling browsing behavior at multiple web sites. *Marketing Science*, 23(Summer): 280–303.

Ragavan, N.R. (2005). Data mining in e-commerce: A survey. *Sadhana*, 30 (Parts 2 and 3): 275–289.

US Department of Health and Human Services (2007) Research-based web design and usability guidelines. Available at http://www.usability.gov/pdfs/guidelines.html.