

# **The Research on the Mental Representations of Website Users' Information Seeking Behaviors of Website Interface based on Markov Chain Monte Carlo: The Case of the Location of Search Field**

*Peng Wu, School of Economics and Management, Nanjing University of Science & Technology, No 200 .Xiaolinwei road. Nanjing, Jiangsu China.210094. 86-25-84315963,dragonwu99@vip.sina.com*  
*Lijun Zhang, School of Economics and Management, Nanjing University of Science & Technology, No 200 .Xiaolinwei road. Nanjing, Jiangsu China, 210094, 86-25-84315963, paopao5028@126.com*  
*Liren Gan, School of Economics and Management, Nanjing University of Science & Technology, No 200 .Xiaolinwei road. Nanjing, Jiangsu China, 210094,86-25-84315963, gan5707@sina.com*

## **ABSTRACT**

We designed the algorithm of Markov chain Monte Carlo applied in the measurement of website users' mental representations, of which main function is to transform users' subjective mental representations to observation signals. It involved two parts, accept function and transmission kernel construction. In the algorithm, we could look the users' information seeking behavior as an accept function. From the selection between of current and proposed state, we could construct a Markov chain, with which the users' mental representations could be simulated and predicted exactly in the decision process of information seeking. We also did the empirical study of users' selection about the search field location of government websites.

**Keywords:** Markov chain Monte Carlo; Mental Representations; Mental Model; Information Seeking; Human-Machine Interaction

## **INTRODUCTION**

Research has shown the method for investigating the website users' mental representations based on Markov chain Monte Carlo (MCMC) during the process of information seeking. Many studies have pay attention to the website users' information seeking behavior, but few studies have examined the mental representations behind behavior. In the research, website users' mental representations are expressed as a non-negative function defined over a set of objects. We present a behavioral method for estimating these functions. Our approach uses website users as components of a MCMC algorithm, a sophisticated sampling method originally developed in statistical physics. An empirical study of users' selection about the search field location has been designed to verify our method, the result demonstrates that our method can simulate and predict website users' mental representations exactly, which can bridge the method of cognitive psychology and complex survey statistical method for website users' behavior study.

With the increasing of online information and more complexity of the websites designing, it becomes much more difficulty to search useful information [20].The motivation, characteristics and disciplines of information seeking behaviors of internet users are the focuses of the study now. And it is a trend that study website interaction interface based on cognitive psychology theory. The cognitive psychology theory and experimental methods are important to the user behavior studies, we can observe, analyze the user cognitive phenomena through them, and then provide the disciplines for designing the website interactive systems. A key challenge for cognitive psychology is the investigation of mental representations, such as object categories, subjective probabilities, choice utilities, and memory traces

[3][6][10]. The accuracy of the measurement of these mental representations affects the effectiveness of the user information behavior study and website information architecture optimization.

Determining how people mentally represent different concepts is one of the central targets of cognitive psychology. Identifying the content of mental representations allows us to explore the correspondence between those representations and the world, and to understand their roles in cognition. Psychologists have developed a variety of sophisticated techniques for estimating certain kinds of mental representations, such as the spatial or fractal representations underlying similarity judgments, from behavior [11][16-17]. However, the standard method used for investigating the vast majority of mental representations, such as object categories, subjective probabilities, choice utilities, and memory traces, is asking participants to make judgments on a set of stimuli that the researcher selects before the experiment begins, which increases the burden of the participants [9], so that we can't get a satisfied result through this method. In essence, mental representations can be defined as a non-negative function over a set of samples, we can use the samples to compute any statistic of interest concerning a particular function, and samples will provide a good approximation to the means, variances, covariance, and higher moments of the distribution. This makes it possible to test claims about mental representations made by different cognitive ideas.

Markov chain Monte Carlo has been particularly useful for correctly generating samples from distributions where the probability of a particular outcome is multiplied by an unknown constant. With the MCMC method, a Markov chain is constructed in such a way that it is guaranteed to converge to a particular distribution, allowing the states of the Markov chain to be used in the same way as samples from that distribution. Transitions between states are made by a series of local decisions as to whether to accept a proposal change to a given object, requiring only limited knowledge of the underlying distribution [18] [22]. During the process of website human machine interaction, website users' information seeking behaviors can act as elements of an MCMC algorithm, samples can be got from the distribution  $p(x)$  associated with the underlying representation  $f(x)$ . This task is a simple two-alternative forced choice [9]. The procedure of website users' information seeking can be looked as starting a Markov chain at some initial state. The result of the choice selection can be input as the parameters for the generating a Markov chain, and the information seeking strategy transmissions proposing a change to the state of the Markov chain and then deciding whether or not to accept this change based on the probabilities of the different states under the information seeking target distribution. After allowing enough iterations for the Markov chain to converge to its stationary distribution (known as the "burn-in") [1][12], the states of the Markov chain of website users can be used to answer questions about the target distribution [14]. So the state of Markov chain represents the website users' mental representation.

In the website interface of human computer interaction, an excellent search system can direct users to his right web page, so he can get useful information contents. The location of search field in the website home page is the first website object that users interact with, and it's crucial to affect the efficiency of the utilization of search system whether location of search field meets the website users' mental model. The selection of location of search represents the transmission of website users' mental representations, so we select the location of the search field as exploratory experimental study of the MCMC application.

### **WEBSITE USERS' MENTAL REPRESENTATIONS MEASUREMENT ALGORITHM BASED ON MCMC**

MCMC algorithm for mental representations measurement involves two parts: one is the Monte Carlo

part, which approximates the average as the desired value; the other one is Markov simulation part, the main idea is to approximate the probability density of the target function by a Markov chain. So, as long as the parameters of the probability density function are given, we can approximate the probability density function, and get the desired value. Metroplis-Hastings sampling and Gibbs sampling are the most widely used sampling methods. And we choose the Metroplis-Hastings sampling method [8] in this paper. To measure the mental representations during the process of website users' information seeking, the basic principles of MCMC are as follows: Website users can be looked as elements of the accept function, the states of the Markov chain represent users' mental representations. The Markov chain can describe the users' mental states transmission process about the interface of the websites (such as the search field location), it involves two states: the current state and the proposed state, the website users need to make a choice between them. After that, the Markov chain can be used to analyze the website users' mental representations [5]. There are five main elements of MCMC algorithm: the initial state, the proposed state, the transmission kernel, the accept function and the steady state [4].

### Task Designing for the Accept Function

On website interface, there are many mental representations relevant to the users' choice making behaviors, in order to look them as an element in MCMC, we need to design a task, in which the website users need to make a choice between object  $x^*$  and  $x$ . In this task, the proposal distribution function  $p(x)$  is asymmetrically distributed, and the probability of selected object  $x^*$  obeys Barker accept function, which is named as Luce selection rule [9], it describes the task that website users need to make a choice between two stimuli. When the target function is a utility function or probability distribution, the rule can be used to convert the degree of website users' response into a probability [13]. The study of Anderson et al has presented the following result [2]: when a model of the website users' response probability created based on the Luce selection rule, a model of soft threshold by a logical function can be created. During the process of website users' information seeking, website users can be looked as an element in MCMC, a task can be designed based on the Luce selection rule to direct the participants to select, which a specific method to choose between a series of two options is utilized, one option is that website users can do selection based on previous selection of the experiment, the other option is that website users do selection based on a special arbitrary proposal distribution. It is a result of sampling from  $p(x) \propto f(x)$ . The algorithm can provide an efficient way of exploring the target distribution  $f(x)$ , and the characteristics of  $f(x)$  also can be explored through transmission of the selections.

### Transmission Kernel Construction

MCMC method means that establish a Markov chain with a stationary distribution  $\pi(x)$ , and then the samples of  $\pi$  can be obtained from it, and the certain statistical analysis can be done based on the samples. Overall, MCMC method involves following three steps [17]:

- (1) An "appropriate" Markov chain is selected in distribution, and its transmission kernel is  $p(\bullet, \bullet)$ . The "appropriate" means the corresponding stationary distribution is  $\pi$ .
- (2) Started with a state  $X(0)$  in  $D$ , and a sequence of states  $x(1), \dots, x(n)$  is generated by the Markov chain of step (1).
- (3) When  $m$  and  $n$  is large enough, the expectations of any function can be estimated as following:

$$\hat{E}_\pi f = \frac{1}{n-m} \sum_{t=m+1}^n f(X^{(t)}) \quad (1)$$

In this paper, we use Metroplis-Hastings method, the main idea of Metroplis-Hastings method is

following. If a Markov chain is established of which the corresponding stationary distribution is  $\pi(x)$ , an irreducible transmission probability function  $q(\bullet, \bullet)$  and a function  $a(\bullet, \bullet), (0 < a \leq 1)$  is selected, then any combination  $(x, y), (x \neq y)$  can be defined as following:

$$p(x, y) = p(x \rightarrow y) = a(x, y) \alpha \quad (2)$$

Based on above analysis, we can design experiment during the process of website users' selection the location of search field as follows. First, a method of sampling from a probability distribution can be defined, which starts with a set of parameterized states, and any state can selected as the initial state. Second, in each experiment, the proposed state is obtained by the proposal distribution around the initial state. The website users are asked to select a choice between the current state and proposed state, which is consistent with the question that which state is more consistent with the website interface. At last, assume that website users' selection behaviors obey the Luce selection rules, then the stationary distribution of the Markov chain is the probability distribution of website users' cognition about the location of search field, and samples obtained from the Markov chain can provide information about the mental representations on interface concepts.

## EMPIRICAL STUDY

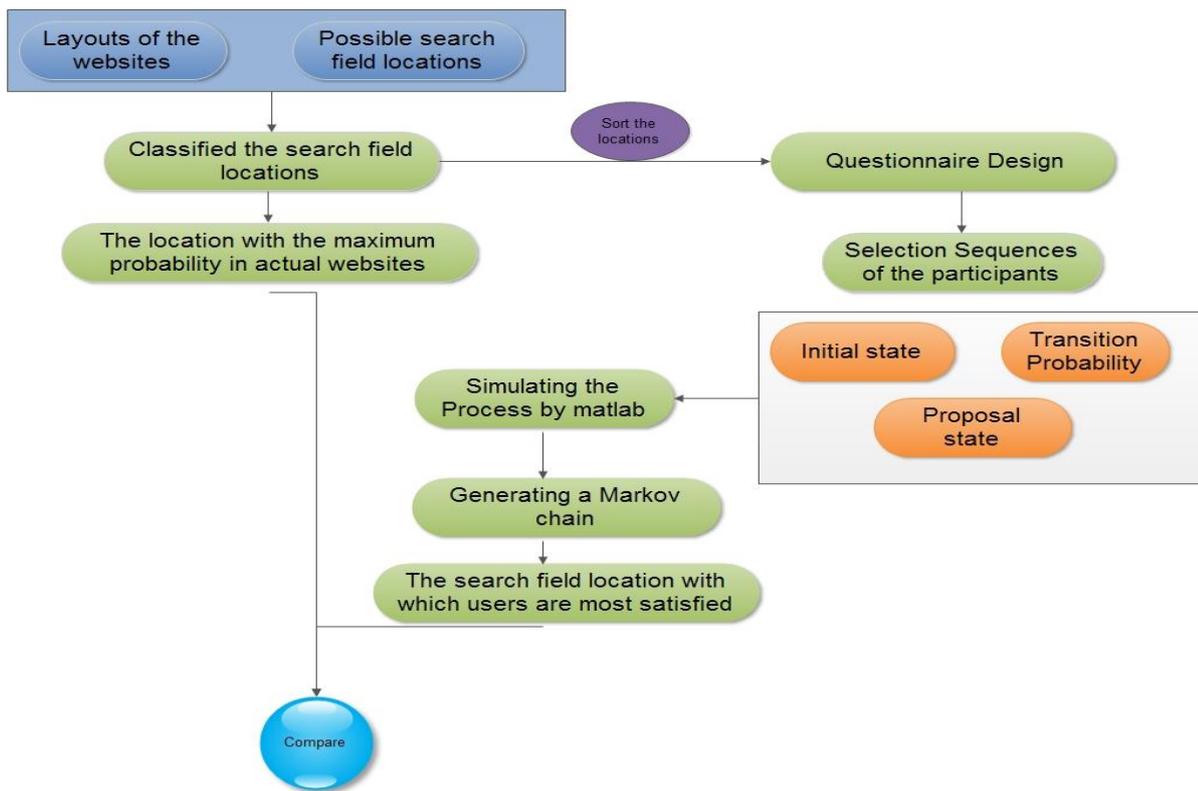
### Research Hypothesis

Website users' best satisfied search field location is consistent with the search field location that has the maximum probability in the actual sites.

### Experiment Design

Taking the website users as the elements of accept function, we can predict the mental representations transmission of website users by the Markov chain. In this chapter, we will study the process of website users' mental representations transmission on selection of best search field location of the government websites. Detailed experimental processes are following (Figure 1):

- (1) Investigation and analysis of main government portal website interface layouts.
- (2) Study of the possible locations of the search field. The task is to classify the locations of the websites layouts based on the investigation of the 100 main government portal websites of the whole world.
- (3) Calculating the probability of the possible locations of the search field. The task is to calculate the probability based on the investigation and survey of the 100 main government portal websites, identify the location with the maximum probability, generate the proposal probability distributions.
- (4) Design the selection sequence of the website users. The task is to get the sequence data of the website users' choices by a questionnaire. The sequence represents the users' mental states and transmissions. According to the interface layouts analysis result of step 3, the possible locations are arranged in a certain order, and the users are asked to make a choice between foremost two possible locations, remain the chosen one and delete the other one. At last we can get a choice sequence from each website users.
- (5) Calculating the transmission probability matrix. The transmission probability is the one of kernel elements of MCMC, the transmission probability of every possible couple of states can be calculated based on the website users' choice sequences, and all probability can construct a probability matrix.
- (6) Setting the initial state of MCMC and Getting the proposed state. The proposed state is the product result of the initial state of step 6 and the transmission probability matrix of step 5.
- (7) Generating a Markov chain. Steps (6) to (7) are implemented by Matlab software, and in order to prevent the Markov chain from pseudo convergence, we repeat the process three times and compare the three convergence results.



**Figure 1: The process of the empirical study**

## Participants

In the experiment, the subjects are the information management undergraduate students of first grade, second grade and third grade. They are 5 first grade undergraduates, 17 second grade undergraduates and 32 third grade undergraduates.

## Research About the Experience Distribution of the Search Field Location

The websites layouts can be broadly divided into 9 types: the "country" font type, corner type, title, body type, the left and right frame type, the upper and lower frame type, the integrated framework type, cover type, the flash type and the change type. We surveyed 100 of the world's major government websites (50 government websites of China and 50 foreign government web sites) about the search field locations in the homepage. According to the result of the investigation, the government websites layouts are divided into five main kinds: A, B, C, D, and E [7](Figure 2). According to the survey result of the search field locations of the existing 100 main government portal websites of the whole world, the locations can be summed up as 13 types. The location no is arranged by the order from small to large according to the corresponding numbers of websites(table 1),the first number represents the number of website which it's search field in this location, the number in parentheses represents the search field location serial number. The experiment homepage according 13 different locations of search field is designed for questionnaire, and in order to shield other factors, other elements of homepage interface such as global navigation, website log of each designed homepage are exactly the same except for the search field location.

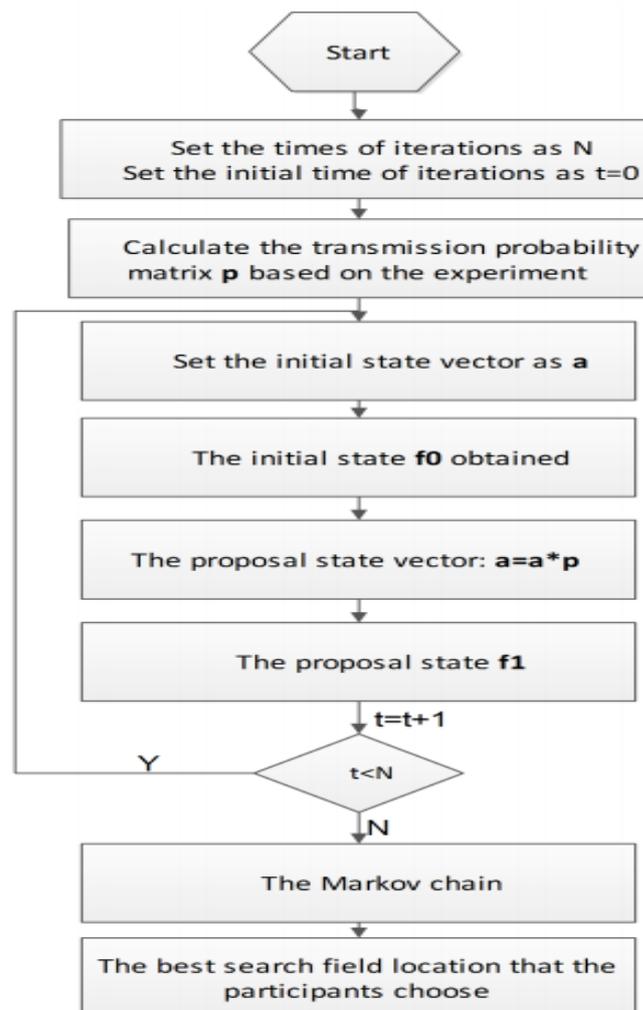




- (1) The locations of search field is represented by the sequence number of table 2, and the sequence number that website users selected are looked as the input data of the matrix.
- (2) The two close choices of the selection sequences are called a couple of data. And the former one represents the current state, and the last one represents the proposed state. The couple of data is analyzed for calculation of transition probability.
- (3) Among all couples of data, the variable  $m$  ( $0 < m < 14$ , and  $m$  is an integer) represents the current state, the variable  $SUM$  is the sum of variable  $m$ . The variable  $n$  ( $0 < n < 14$ , and  $n$  is an integer) represents proposed state, the variable  $SUM1$  is the sum of variable  $n$ . Then the transmission probability is  $SUM1/SUM$ , and it presents the column  $n$  row  $m$  element of the probability matrix;
- (4) All the probabilities is calculated based on step (3), and the transmission probability matrix is formed.

### Generating of Markov Chain

The flag that the Markov chain reaches the steady state is current state equal proposed state. The Markov chain is generated by Matlab software, and the logic process of Matlab program is shown in Figure 3. In the process, an initial state is set randomly within the allowable range, then program is executed, and a Markov chain is generated. To prevent it from pseudo-convergence, three initial states has been set on the same operations. The corresponding elements of MCMC in this experiment are shown in table 4.



**Figure 3: The logic process of generating the Markov chain**

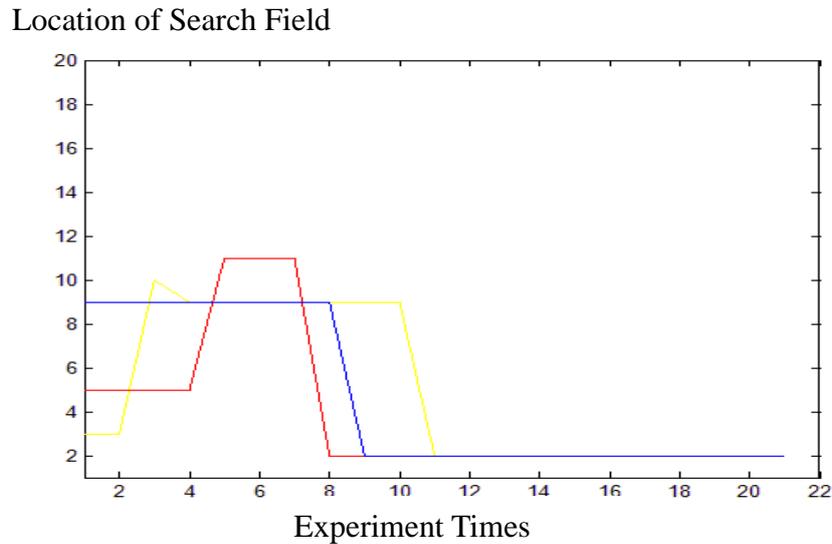
**Table 4: The elements of MCMC in this experiment**

Elements of MCMC	The corresponding element in this experiment
The Initial state	Randomly given in the experiment
The Proposed state	The next location needs to be compared
The transmission kernel	The transmission from one location to another location
The accept function	The participants
The steady state	The best search field location considered by the participants

In the experiment, an initial state vector is set. All the values of the vector represents the probability of the corresponding locations. The location of search field in questionnaire is equivalent to its column number in transmission probability matrix. The initial state is set as the column number with the maximum probability in the vector, the product result of  $a * p$  is the proposed state vector, and proposed state is represented by the column number with the maximum probability in vector  $a * p$ . The tree given initial state vectors are as follows:

$$\begin{aligned}
 a &= [0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0] \\
 a &= [0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0] \\
 a &= [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 0]
 \end{aligned} \tag{3}$$

For a example, the initial state is the column corresponding to  $\max(q)$ , namely location 3,5,9. The corresponding Markov chains generated are shown in Figure 5. In Figure 5, the abscissa is the number of experiments, which is equivalent to the times of proposed states, and also is equivalent to the times of iterations of  $a * p$ . The ordinate is the serial number of search field location selected by the website users.



**Figure 5: The Markov chain of the website users' mental representations during selection of location of search field**

### Experimental Results

Experimental results can be visualization based on the background of the design China Nanjing government website homepage. According to Figure 5, the steady state of the Markov chain is “location 2”, it means that location 2 is the search field location that website users are most satisfied with (Figure 6). But result of the survey of the actual websites shows the maximum probability of search field location is location 8 (Figure 7. There are 32 of the surveyed 100 websites which put the search field in

location 8). The finding can be concluded as following, the most satisfied search field location of the is not inconsistent with the user's mental model, and the hypothesis is not set up. Website designers can take advantage of this conclusion to design websites more satisfied with the website user's mental model.



Figure 6: The search field that website users are most satisfied with



Figure 7: The search field location of the maximum probability in actual websites

## DISCUSSION

The study explored the measurement method of website users' mental representations based on MCMC, which the location of search field is taken as the measure object, website users are taken as elements in MCMC algorithm, from which we can predict their behavior, and use this algorithms to learn more about their mental representations and establish connections between statistical inference and human behavior. The statistically finding of this study is that mental representations of users' information seeking behavior on Website Interface can be measured by the algorithm of Markov chain Monte Carlo (MCMC), and corresponding algorithm can be designed for the measurement of mental representations for website users during the process of information seeking.

This study consistent with and validate previous MCMC study [1][18-21], mental model study for the location of common web objects on web pages[15]. This study also provides practical insights for those who design and develop the website interface based on website users' cognition. The website designers need to think about ways of optimization design for meeting website users' expectations for the location of common web objects on web pages. A practical recommendation for the website designers is to explore the website users' mental representations that control their information seeking behavior, which metal representations measurement algorithm can be designer based on MCMC, to simulate and predict the website users' mental representations exactly in the decision process of information seeking, and get website users' transmissions rules of information seeking behavior.

## CONCLUSION AND FUTURE WORK

During the information seeking process, it's the first and most important step that website users browse the internet and build up certain expectations for the location of common web objects on website interface. Regardless of the study environment, laboratory setting or online survey, a range of common web objects seems to be expected at certain locations [15]. What is location of search field that website

users are most satisfied ,and how to measure the mental representations during the process of website users' information seeking, it's an import problem that need to be resolved. A couple of studies have already shown that applying mental models increase speed and orientation on web pages [15].However, a key question has not yet been resolved, it's the method how to investigate, simulate and predict of mental representations during process of website users' information seeking. Markov chain Monte Carlo is one of the basic tools in modern statistical computing, providing the basis for numerical simulations conducted in a wide range of disciplines. The study of Adam N. Sanborn presented MCMC can also be of use in psychology, not just for numerical simulation, but also for uncovering mental representations [1].The results of this paper validated the this idea, which MCMC algorithm could be used to measure the website users' mental representations during the process of selection of location of search field. The experiments in this paper have applied the MCMC method to location selection tasks. The MCMC method is designed to estimate these subjective functions. Using this technique, sample data could be collected by investigating the subjective response to a probe of memory or exploring the internal values of different response alternatives in decision making of search field location selection.

In conclusion, as an exploratory research at the intersection of information science and cognitive psychology, this research presented the algorithm based MCMC to simulate and predict the website users' mental representations that behaved the users' information seeking behavior. The research findings make a contribution to empirical knowledge on how to simulate and predict website users' information seeking behavior.

Regarding future research, additional populations of users need to be studied. In this study all the subjects were undergraduate students with information science background. As there may be other characteristics that influence website users' mental representations, studies of different populations need to be conducted to help understand the mental representations and information seeking behavior based on MCMC in best match circumstances among users in different fields of study or some other research parameters. A large-scale study with different task complexity and difficulty was also needed to determine the effects of tasks on mental representations, such as training. More research was needed to gain an understanding of website users' mental representations over time and to explore how website users' mental representations effect information seeking decision in various tasks and contexts. Particularly, we suggest that further research need to take a closer look based on MCMC at specific states of difficulty during each website user's search process. The research would help website designers to design best match systems that could better support users by appealing them the expectations for website interface and easily adapt their mental models from exact match to best match [21].

### ACKNOWLEDGEMENTS

This research was supported by the National Natural Science Foundation of China under contract No. 71003049, 71001052.

### REFERENCES

- [1]Adam N.Sanborn,ThomasL.Griffiths,RichardM.Shiffrin.*Uncovering mental representations with Markov chain Monte Carlo*. Cognitive Psychology ,2010,60(2):63–106
- [2]Anderson, J. R., & Milson, R. (1989). *Human memory: An adaptive perspective*. *Psychological Review*, 1989.96(4), 703–719.
- [3]Anderson, J. R. *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.1990
- [4]Billera, L. J., &Diaconis, P. A geometric interpretation of the Metropolis–Hastings algorithm.

- Statistical Science, 2001, 16(4):335–339.
- [5]Chater, N., Tenenbaum, J. B., & Yuille, A. *Special issue on Probabilistic models of cognition*. Trends in Cognitive Sciences, 2006.10(7).
- [6]Farrell, S., & Ludwig, C. *Bayesian and maximum likelihood estimation of hierarchical response time models*[J]. Psychonomic Bulletin & Review, 2008, 15(6):1209–1217.
- [7]GAN Liren et al. *On the Application of Information Architecture—Examination & Evaluation of the Government Websites in China*. Information Studies: Theory & Application, 2003, 26(6):487-491
- [8]Lei Hengheng Zhao Yongjun Huang Jie. *Website navigation systems based on user accessing sequences clustering*. Systems Engineering-Theory & Practice, 2010, 133(4):13-17
- [9]Luce, R. D. Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.). *Handbook of mathematical psychology*. New York, London: John Wiley and Sons, Inc. 1963:103–190
- [10]Morey, R. D., Rouder, J. N., & Speckman, P. L. *A statistical model for discriminating between subliminal and near-liminal performance*[J]. Journal of Mathematical Psychology, 2008, 52(1): 21–36.
- [11]Neal, R. M. *Probabilistic inference using Markov Chain Monte Carlo methods* (Tech. Rep. No. CRG-TR-93-1). Department of Computer Science, University of Toronto. 1993
- [12]Oulasvirta, A., Karkkainen, L., Laarni, J., 2005. *Expectations and memory in link search*. Computers in Human Behavior 21 (5), 773-789.
- [13]Rouder, J. N. *Modeling the effects of choice-set size on the processing of letters and words*. Psychological Review, 2004, 111(1):80–93.
- [14]Santa-Maria, L., Dyson, M. C. *The effect of violating visual conventions of a website on user performance and disorientation. how bad can it be?* In: SIGDOC'08. 2008. 47-54.
- [15]Sandra P. Roth, M. S. *Mental Models for Web Objects: Where do users expect the most frequent objects in Online Shops, News Portals, and Company Web Pages?* Interacting with Computers, 2010, 22, 140-152.
- [16]Shepard, R. N. *The analysis of proximities: Multidimensional scaling with an unknown distance function*[J]. Psychometrika, 1962, 27, 124–140.
- [17]Shepard, R. N., & Arabie, P. *Additive clustering: Representation of similarities as combinations of discrete overlapping properties*. Psychological Review, 1979, 86(2):87-123.
- [18]Shiffrin, R. M., & Steyvers, M. *A model for recognition memory: REM: Retrieving effectively from memory*[J]. Psychonomic Bulletin and Review, 1997, 4, 145–166.
- [19]Victor, J. *Analyzing receptive fields, classification images and functional images: challenges with opportunities for synergy*. Nature Neuroscience, 2005, 8(12), 1651–1656.
- [20]WANG You-wei, XUBo, WEIXue-qi, LING Hong. *Website navigation systems based on user accessing sequences clustering*. System Engineering-Theory & Practice, 2010, 30(7):1305-1311.
- [21]Wu He, Sanda Erdelez, Feng-Kwei Wang, Chi-Ren Shyu. *The effects of conceptual description and search practice on users' mental models and information seeking in a case-based reasoning retrieval system*. Information Processing and Management .2008. 44 .294–309
- [22]ZHANG Shao-gang. *Nonstation random signal analysis initial research based on MCMC method*. Automation & Instrumentation, 2010(3):115-116