Web User Modeling via Negotiating Information Foraging Agent

Xuehong Tao[1], Yin-Leng Theng[1], and Terence Ting[2]

[1] Wee Kim Wee School of Communication and Information
[2] School of Mechanical and Aerospace Engineering
Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798
{exhtao,tyltheng}@ntu.edu.sg, ekting@singnet.com.sg

Abstract. Information foraging theory lays a good foundation for web user modeling. However, the existing user modeling methods mainly focus on fixed information needs. In the real world, a user's information goal often evolves, and information foraging is a negotiation process between the user and the system. In this paper, we proposed an agent based approach that modeled the dynamic information seeking process of information foragers.

Keywords: Information foraging, information scent, information retrieval, user modeling, software agent.

1 Introduction

Usability evaluation is important in the software development lifecycle. However, it is usually difficult to afford a large number of real users to participate in the usability studies. Using software tools to simulate real users' behavior and therefore to conduct automated usability evaluation is a practical solution. There are some tools that based on information foraging theory for web usability evaluation, such as SNIF-ACT [1] and WUFIS [2]. They modeled web users' navigation from page to page searching for information with fixed information goals. However, a user's information goal may change based on what is learned during the seeking process. He/she may adapt to what the system has if the original information goal is not reached. The goal evolvement and the adaptation/negotiation process between searchers and systems have not been well addressed in the existing models.

This paper proposes an agent based web user model. The model distinguishes itself from the existing models in that it situated information goal in context and modeled the goal evolvement and the negotiation process of information seeking.

2 Information Foraging

Information foraging theory [3] is a foundation theory of web navigation. On the web, users typically forage for information by navigating from page to page along web links. The content of pages associated with these links is usually presented to the user by some snippets of text or graphic. Foragers use these proximal cues (snippets,

T. Gross et al. (Eds.): INTERACT 2009, Part I, LNCS 5726, pp. 436–439, 2009.

graphics) to assess the distal content (page pointed by the link). Information scent is a subjective perception of the value and cost of information sources obtained from proximal cues, such as web links, or icons representing the content sources [2]. The goal of the information forager is to use proximal information scent cues (eg. a web link) to predict the utility of distal sources of content (i.e. the web page associated with a web link), and to choose the links having the maximum expected utility [1]. Three important aspects of information foraging have not been well addressed in the existing web user modeling literature:

1) The goal of information foraging evolves. According to the berrypicking information seeking model [4], in real-life web search, users may begin with one feature of a broader topic, or a relevant reference, and move through a variety of sources. Each new piece of information encountered can offer new ideas and directions to follow, hence a new conception of query. The search goal evolves.

2) Information foraging is situated in context. Information foraging is usually a task embedded in the context of other tasks. Its value and cost structure is consequently defined in relation to the embedding task and often changes dynamically over time [5]. Such an embedding task can be choosing a good graduate school, developing a financial plan, or writing a scientific paper.

3) Information foraging is a negotiation process between the searcher and the system [6]. It is a trade-off between the value of information gained and the cost of performing a task to find the information. Therefore, foraging refers to the variety of strategies seekers exhibit in their quest for information and how humans adapt to their environments on a situational basis.

3 Negotiating Information Foraging Agent (NIF-Agent)

Software agents are autonomous entities that work towards their goals. They are suitable to work as representatives of human users. This section proposes a negotiating information foraging agent (NIF Agent) to simulate web information foragers. The NIF agent is able to model the goal evolvement, goal's context and the negotiation process of information foraging.

- **NIF Agent Knowledge Model**

Users may have different level of search goals. There can be four broad categories of goals: long term goal, leading search goal, current search goal and interactive intention [7]. For a specific search, higher level search goals may have different lower level options (sub goals). For example, to find out "why the baby is crying", the user may have a few search options in mind, such as "ear infection" or "infant colic". Lower level goals may further have next level of options. Different levels of goals form a hierarchical structure. The goals and their relationships form the knowledge base of a NIF agent, and it can be defined as a directed graph KB= <V, E>, where $V = \{ v_i \mid i = 1, 2, \dots n. \}$, it is a set of all search goals $E = \{(v_i, v_j) \mid v_i$ is the higher level goal of $v_j \}$, it is a set of goal relationships, v_i is termed as father of v_j, v_j is termed as son of v_i.

- **Proximal Scent Matrix**

The scent of information comes from the linguistic relationships between words expressing an information need and words contained in links to web pages. People are more likely to select the link on a page that appears to have the highest probability of leading them to the page best matching their information need.

The proximal scent of a link is calculated as a degree of similarity between the proximal cues and the information need. Proximal Scent Matrix PS' and PS can be computed based on WUFIS (Web User Flow by Information Scent) algorithm [2, 8] which calculates the similarity based on the weight of word importance. Each PS'(i,j) describes the similarity degree between the proximal cues of a link (link from page j to page i) and the information need. Matrix PS is obtained by normalizing the PS' matrix so that each of the columns sums to 1. Each entry in PS(i,j) specifies the probability of a user flowing down the link from page j to page i.

- **Automation of Information Foraging**

Information goal evolves and is embedded in certain context. The knowledge base of NIF agent models the context. For an information goal (a node in KB), its higher level goals is the search purpose and sibling goals are other search options. When a user is not able to achieve a search goal, he/she may change to other options (sibling goals) or a more general goal (father goal). For example, for the purpose of finding out "why my baby is crying, and always draw up his knees against his abdomens", the user may search for the possible options, from "ear infection" to "rashes". He/she may also shift to a more general search about "crying baby". The shifting of goals models the dynamic evolvement of the information need.

A user might bring some initial information needs to start the foraging. However, the system may not be able to satisfy the user's needs exactly. Other information is still acceptable to the user if it has considerably strong information scent (satisfying certain threshold) to the user. This is the negotiation nature of information foraging. As in the "crying baby" example, if the user hasn't found the perfect explanation of the baby's situation, a link "infant colic" with high information scent may also catch his/her attention. In this case, "infant colic" is actually the most likely answer.

Algorithm *Forage* and *Find* is provided below. Algorithm *Forage* automates the information foraging processes with evolving goals, and the algorithm *Find* is a sub function that forages for a fixed information goal. In the algorithm, v.InfoNeed, v.father/v.son and v.visited are used to refer to the information need vector of goal v, the super/sub goal of v and whether v is processed respectively.

Algorithm. Forage ()
For all $v \in$ V in KB, v.visited = false
CurrentPage = n_1, add n_1 to the end of *PageLog //PageLog records the history*
v = StartNode //Suppose the search starts from a goal StartNode $\in$ *V*
While ($v \neq \Phi$)
{ If Find(*v.InfoNeed*)= true Return true
Else
v.visited = true, v = *v.father*
if ($v \neq \Phi$) *// change goal to the most detailed level of other relevant goals*
while (exist *v.son* $\neq \Phi$ and *v.son.visited* = false) v = v.son
} Return false
End Forage.

Algorithm. Find (InformationNeed) *// forage for a fixed information need*
Step = 0, Build matrix PS' and PS based on *InformationNeed*
While (*Step* < *MaxSteps*) *// give up after a certain number of steps*
{ Randomly generate *NextPage* based on the probability, i.e. the value of
PS[1, *CurrentPage*], PS[2, *CurrentPage*],… ,PS[*n*, *CurrentPage*]
Put *NextPage* at the end of *PageLog*, *Step*=*Step*+1
If (PS'(*NextPage, CurrentPage*) $\geq \theta$) *//θ is the user satisfaction threshold*
CurrentPage = *NextPage*, Return true
// suppose the link cues are meaningful, i.e. if the link cue is similar with the
// user need, the page pointed by the link is also similar to the user need
Else *CurrentPage* = *NextPage // flow to the next page and continue foraging*
} Return false
End Find.

4 Conclusion and Future Works

The NIF agent we proposed models the exploratory search activity [9]. The model distinguishes from the existing models in that it models the evolvement of information need and the negotiation process of information seeking. Future works include the implementation of a NIF agent tool and evaluation of its effectiveness.

References

1. Pirolli, P.L., Fu, W., Chi, E.H., Farahat, A.: Information scent and Web navigation: theory, models and automated usability evaluation. The Next Wave: NSA's Review of Emerging Technologies 15(2), 5–12 (2006)
2. Chi, E.H., Pirolli, P., Chen, K., Pitkow, J.: Using information scent to model user information needs and actions on the Web. In: Proceedings of the SIGCHI conference on Human factors in computing systems, Seattle, USA, pp. 490–497 (2001)
3. Pirolli, P., Card, S.K.: Information foraging. Psychological review 106(4), 643–675 (1999)
4. Bates, M.J.: The design of browsing and berrypicking techniques for the online search interface. Online Review 13, 407–424 (1989)
5. Pirolli, P., Card, S.: Information foraging in information access environments. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 51–58 (1995)
6. Kalbach, J.: Designing for Information Foragers: A Behavioral Model for Information Seeking on the World Wide Web. Internetworking ITG Newsletter (3.3) (2000), http://www.internettg.org/newsletter/dec00/article_information_foragers.html
7. Xie, H.: Shifts of interactive intentions and information seeking strategies in interactive information retrieval. Journal of the Smerican Society for Information Science 51, 841–857 (2000)
8. Lawrance, J., Bellamy, R., Burnett, M., Rector, K.: Using information scent to model the dynamic foraging behavior of programmers in maintenance tasks. In: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, Florence, Italy, pp. 1323–1332 (2008)
9. White, R.W., Drucker, S.M., Marchionini, G., Hearst, M., Schraefel, M.C.: Exploratory search and HCI: designing and evaluating interfaces to support exploratory search interaction. In: CHI 2007 extended abstracts on Human factors in computing systems, April 2007, San Jose, USA (2007)