47

UNDERSTANDING COLLABORATIVE AUTHORING IN SHARED WORKSPACES

Chaomei Chen and Roy Rada*

Department of Computer Science
University of Liverpool, Liverpool L69 3BX, England

Phone: +44 794 3706
Fax: +44 794 3715
*E-mail: chen1@csc.liv.ac.uk

KEY WORDS: Collaborative authoring, patterns of user behavior, shared workspaces, Markov chain analysis, collaborative hypertext systems.

ABSTRACT: This study investigates the intermediate role of shared workspaces in collaborative authoring. Collaborative authoring, as an interactive process, associates with the evolution of a shared workspace. This study develops a Markov chain model to describe the dynamics of these processes in an academic setting. Models from two groups of users, regular and occasional users, proved significantly different. The study revealed the structure of the interactive process corresponds to areas in the shared workspace that are activated by the interaction. These findings have both theoretical and empirical value for designing practicable computer systems for collaborative work.

1 INTRODUCTION

The intermediate role of computer systems in human activity is of growing interest in the fields of human-computer interaction and computer-supported cooperative work (CSCW). Designing a practicable CSCW system requires a good understanding of the dynamics.

A shared electronic workspace is a repository of knowledge for sharing. A shared workspace undergoes social or collective construction over a long period of time. An organized evolution is important so as to ensure the workspace retain its usability and maintainability.

Few empirical studies have described characteristics of users' behavior with shared electronic workspaces from a process-oriented view. Usage patterns with traditional information systems are usually not in a form that can be easily transformed for designing and evaluating collaborative support systems.

Our research aims to model the changes in user behavior of using a collaborative hypertext workspace, on the basis of longitudinal observations in an academic setting. The focus is on the role of a shared workspace in the process of collaborative writing.

The paper is organized into six sections. Section 2 reviews related theoretical and empirical work. Section 3 outlines the analytical and modeling techniques used in this study. Section 4 presents the most salient results of the analysis. Section 5 discusses these results and offers interpretations with respect to related findings in this area and section 6 summarizes the study.

2 RELATED WORK

In the 1970s and 1980s, a series of empirical studies explored the dynamic structure of an interactive process between users and online database systems. These studies conceptualized the interactive process as Markov chains. They employed different analytical techniques on the resultant Markov chains.

Recently, Markov chain models have been utilized to analyze the use of hypertext systems (Qiu 1993)

K. Nordby et al. (eds.), *Human-Computer Interaction*

and situated actions in broader human-computer interaction settings. In these studies, behavioral patterns were studied for information retrieval tasks with a hypertext system. Little is known about the extent to which these behavioral patterns can be generalized to the use of collaborative hypertext database systems. It is expected that utilization of Markov chain analysis will help us understand more about the use of a collaborative hypertext database.

Sociologist Robert Bales developed an analytical method to study interactions empirically in small groups, known as *interaction process analysis* (Bales 1951). The interaction process consists of several subordinate processes of problem solving. The interaction process alternates its focus dynamically between substantive and social-emotional problem solving processes.

2.1 Shared Workspaces

Supporting dynamic, smooth transitions from one mode of work to another has been a prevalent design paradigm for collaborative hypertext database systems. In this paradigm, working with a collaborative hypertext system involves three modes — independent, loosely-coupled, and tightly-coupled, as seen in systems such as gIBIS (Conklin 1988) and SEPIA (Haake 1992).

The key to this paradigm is to facilitate a smooth transition from one mode to another. For example, SEPIA facilitates a change from a loosely-coupled mode to a tightly-coupled mode. Transitions are triggered by the proximity of nodes activated by users in the workspace. On the other hand, empirical evidence is needed to verify whether these mechanisms fit well into users' workflow.

Awareness mechanisms in shared workspaces aim to help users organize their own actions (Dourish & Bellotti 1992). Mechanisms behind transitions are system-dependent and domain-specific. One needs to understand the characteristics of changing, the patterns associated with different users, and the triggers for the transitions.

2.2 Behavioral Patterns

Penniman's pioneer work significantly improved the understanding of user behavior of search in online databases. Penniman (1975) found that individuals may generate fairly complex and unique sequences of actions in the longer view.

Chapman (1981) compared online search patterns of several groups of users on Lockheed's DIALOG system. The group trained by computer systems had similar search patterns, whereas in the group trained by the professional searchers, individuals developed some unique search patterns. She concluded that experience in on-line searching has a great impact on user's information seeking behavior. A dominant pattern was found as PreparingQueryTerms → Querying → ViewingRetrievedTitles.

Qiu (1993) identified recurring patterns in searching a static hypertext document in Hyperties. The predominant sequences are Querying → DisplayTitle → DisplayArticle and DisplayTitle → DisplayArticle → Printing.

Task requirements in collaborative writing with computers were studied by Kraut, Galegher, Fish, and Chalfonte (1992). They characterized the process of collaborative writing as planning, drafting, and revising. Their study did not focus on the intermediate role of shared workspace in collaborative writing.

3 METHOD

Collaborative authoring in a shared workspace was analyzed as a stochastic process represented by a Markov chain model. Markov models of regular and occasional users were compared. The impact of the structure of a shared workspace was investigated by clustering the seven workspaces by the similarity between associated Markov models. The analysis particularly focused on user behavioral patterns in changing the modality of using the workspace

3.1 The MUCH System

The MUCH system, *M*ultiple *U*sing *C*ollaborative *H*ypermedia, supports collaborative authoring of scientific papers, allows knowledge sharing and reuse, and streamlines electronic publishing. A comprehensive account of the MUCH system can be found elsewhere (Zheng & Rada, 1994).

In this paper, *a collaborative hypertext database*, or *a shared workspace*, refers to the collaborative hypertext facilities in the MUCH system and the

database of documents viewed and edited via the MUCH system.

3.2 Users and Tasks

The MUCH system has been used to maintain several collaborative databases in a research group at the University of Liverpool. Seven databases were extensively used in supporting collaborative writing (the Groupware database), organizational management (the Management database), project management (the Oscar database), and as experimental systems in a course for Master of Science in Information Systems in 1993 and 1994 (the Class93 and Class94 databases).

3.3 Data Collection

Transactions of using the MUCH databases were logged for three months (April 20 - July 20, 1994). Each transaction was recorded with the name of a corresponding function in the MUCH system, the time of the call, and the title of the node which was activated by the function call. The database was identified at the beginning of a session. A total of 24 interactive functions of the MUCH system formed a state space with six states — constructing, browsing, reading, awareness, writing, and printing.

Users were classified into regular and occasional groups according to the number of transactions recorded. A regular user has transactions between 1,500 to 9,000. An occasional user has 1,000 or less. The two groups of users were expected to take part in tasks of different nature and therefore show different patterns of behavior.

3.4 Data Analysis and Modeling

The analysis used *Statistical Package for the Social Sciences* (SPSS). Markov models were specified in terms of hierarchical log-linear models in SPSS. Three major hypotheses were tested via log-linear models. Pearson's χ^2 goodness-of-fit statistics were used. A detailed description of using log-linear models in Markov analysis is found in Bishop et al. (1957) and was used in some studies on search patterns with static hypertext information space (Qiu 1993).

Hypothesis 1 The process of collaboratively using the MUCH databases is a second-order Markov chain. This hypothesis was tested by fitting hierarchical log-linear models.

Hypothesis 2 There is a significant difference between the Markov models of regular and occasional users. This hypothesis focuses on the effects of users' overall experience with the MUCH system.

Hypothesis 3 The structure of shared workspaces have significant effects on behavioral patterns of regular users. The differences between Markov models were measured by partial likelihood ratio chi-square statistics.

The analysis specified transitions that differ in Markov models compared. The question was investigated, whether regular users need more awareness support than occasional users.

4 RESULTS

The results of the study are presented in four parts: 1) interactive process, 2) patterns from different user groups, 3) process in different workspaces, and 4) aggregated behavioral patterns for groups of workspaces.

4.1 Interactive Process

Log-linear tests showed that an interactive process with the MUCH system is a second-order Markov chain. The process can be adequately modeled by a second-order Markov chain (Pearson's χ^2=660.86, df=780, p =0.9992), but not by first-order Markov chains (χ^2=6523, df=102, p < .0001). The tests were based on transactions from both regular and occasional users.

Most recurrently second-order transitions from regular users were dominated by Browsing (B), Reading (R), and Writing (W). Log-linear models suggest that, in essence, an interactive process (IP) has the following structure:

$B^p = B \rightarrow B \rightarrow ... \rightarrow B$, (p - *1 times)*

$IP = B^p \rightarrow (R^q \rightarrow W^r)^s \rightarrow IP$, $p, q, r, s = 1, 2, 3, ...$

A state of Structuring tends to be followed by a state of Browsing. Users adjust their views to a shared workspace as the organization of the workspace changes. The model indicates using a shared workspace not only involves searching and browsing, but also includes writing and organizing.

4.2 Process from Different User Groups

Log-linear tests, based on the use of seven hypertext databases, found the Markov chains of

regular and occasional users are significantly different (Pearson's χ^2= 5880.91, df = 180, p < .001). Differences between the two groups were specified by corresponding λ parameters in log-linear models.

Browsing acts on an organizational schema of the hypertext workspace, Writing acts on the associated content material in the workspace, and Reading transfers the focus of interaction from the organization to detailed information.

- Highly used R-R-B transitions by regular users (Z=4.17) suggested that they frequently adjust the focus and scope of their views. They browse the organizational structure more extensively than occasional users.
- The pattern of R-A-R (Z=2.57) indicates that, as users read the content of a shared workspace, they pay more attention to social-construction aspects of the workspace than occasional users.
- A-R-B (Z=2.48) and A-R-W (Z=2.23) implied regular users require the awareness of collaborative authoring at both the organizational and content level of the workspace.

4.3 Process in Different Workspaces

The organization and content of a workspace significantly affects the way that it is used. The influence was compared between the processes from two workspaces (Pearson's χ^2=19052.12, df = 1296, P < 0.001).

Databases were classified into three clusters by the similarity of user behavior (See Figure 1). Cluster A, for instance, includes the Management and Groupware databases which have similar patterns of user behavior.

4.4 Aggregated Patterns of Behavior

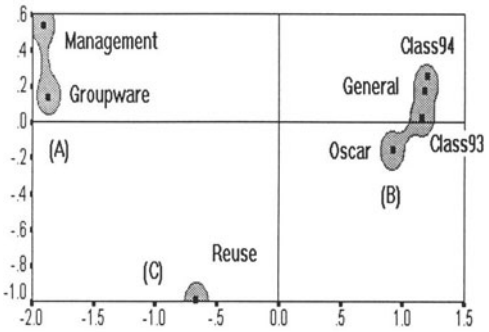
Log-linear models specified significant differences in patterns of user behavior associated with these clusters.

Cluster A Patterns such as R→A→B and W→A→B were predominant (Z=2.52 and Z=1.96, respectively). Users seek social awareness before they transit from the content level to the organizational level of the shared workspace. Awareness seeking appears as a local activity.

Cluster B The pattern R→B→B (Z=2.79) characterizes the behavior of an occasional user with general tasks and the pattern R→W→B (Z=2.34) suggests the focus of an interactive process alternates between the organizational and content level of a shared workspace. The R→X→A patterns in cluster B suggest that the stability alone may not explain the awareness needs which arise in the process of interaction with the workspace.

Cluster C The patterns of W→W→B (Z=3.17) and R→W→B (Z=2.06) were found. These patterns suggest an independent working mode in writing with a hypertext system. The lack of these patterns in cluster A indicates that those processes have more complicated structures because of the complexity of the organization and evolution of these workspaces. For an arbitrary state X, processes of cluster C have patterns of X→B→C, indicating an extensive evolution of the shared workspace.

Figure 1.Classification of hypertext databases according to the similarity of corresponding Markov chains. (Stress=0.0386, RSQ=0.9950).



Influential factors on browsing and awareness-seeking behavior are summarized as follows:

- Users expand and adjust the contextual views frequently to a rapidly changing shared workspace. Social awareness is required at both the organizational and content level. The heterogeneity of the information exchanged in the workspace and the commitment of the collaborating team significantly influence the patterns.
- Social awareness needs did not stand out in interactive processes associated with a stable workspace. These processes basically correspond to independent writing modes. Changes in the

workspace do not specifically increase awareness request.

5 Discussion

Analysis of the use of collaborative hypertext databases in MUCH has led to insights. Findings particularly related to collaborative use of a shared workspace are emphasized in the following discussion. The implications of these findings for designing collaborative hypertext database systems are outlined.

5.1 Markov Models

Research work in CSCW emphasizes the relationship between individual users and the context of their actions. Markov models describe the dynamic transitions and the relationship between a sequence of actions and situations. Collaborative using hypertext databases in the MUCH system is a second-order Markov chain. Although related studies suggested hypertext information retrieval is a second-order Markov model, the order of the Markov process of collaborative authoring has not been directly addressed.

5.2 Patterns of User Behavior

Qiu (1993) found the dominant search strategy used with a hypertext system is similar to traditional information retrieval strategies. The transition of B → B → B occurred 2.13% in her study, whereas in our study this pattern occurred in 17.68% of the second-order transitions.

There could be several explanations for the difference. First, the organization and content of these databases in the MUCH system have undergone an evolution. Browsing is an appropriate strategy in dealing with the dynamics. Second, the MUCH system provides a fisheye view browser. The fisheye view paradigm has been shown to be a promising approach to resolve the problem of accessing a large and evolving information structure from a dynamic view (Chen & Rada, 1994). In the category of Browsing, functions such as Unfolding and Folding are essential ways for users to move through the database.

There is a remarkable contrast between the predominant pattern Browsing → Reading → Writing in our study and the Querying → DisplayTitle → DisplayArticle in Qiu's study. Further analysis suggests that the difference distinguishes a process of collaborative authoring from a process of information retrieval.

Empirical studies in laboratory settings found that experienced users of hypertext systems tend to use browsing strategies more than inexperienced or novice users (Carmel et al. 1992). Experienced users tend to explore wider areas in a hypertext network, whereas inexperienced users tend to use backtrack to minimize the effects of getting off the main search paths. Our study found regular users access wider areas in the shared workspace to adjust their views of a changing the context for their actions. This finding has an implication for the design of hypertext database systems. A hypertext database system should provide sufficient facilities for experienced users to deal with the issues such as how to visualize a large hypertext database and how to balance the local details and the global structure.

Chapman(1981) found that users' experience has the greatest impact on their behavioral patterns with online information systems. The results of this study show differences in the awareness required between regular and occasional users. Regular users are more concerned with the social constructive nature of the shared workspaces.

These behavioral patterns reflect situations where contextual-social awareness might be called for. Collaboration is affected by various factors from strategic plans to the degree of one's involvement. One should bear in mind that the dynamics involved in the use of collaborative hypertext databases can only be partially captured in this approach. Coordination among users takes place with and without using a particular computer system. No single factor can reliably predict users' needs in a particular situation. These needs should be assessed dynamically and interactively.

6 Conclusion

The results of the analysis suggest that users read the content of the shared workspace with Reading and adjust the focal view to the shared workspace with Browsing as they navigate. In a more independent nature, users may switch from Reading to Writing of a node, or seeking Awareness information associated with the node.

Whether users need to be aware of the actions of their colleagues depends on the goal of the shared workspace and to what extent it is used as an intermediary for collaboration.

The stochastic approach used in this study can be extended to deal with situations involving a large number of users of a shared workspace. Further work is needed to understand fully the relationship between the process-oriented approach and the ethnographic approach.

The model for collaborative authoring extends basic writing models such as planning, drafting, revising to incorporate superordinate processes for browsing, balancing local details and the global structure of a large, shared workspace. The model also distinguishes the needs for awareness information from users with different experience of the system and expertise in the subject domain. Transitions with high probabilities require special consideration and support from the computer infrastructure underlying the collaboration.

This study, from a process-oriented view, connects existing theories and models on collaborative authoring with a shared workspace. Further work is needed to incorporate additional aspects of collaborative authoring into the descriptive and predictive model, including synchronous communication, the use of multimedia, and dynamically matching the allocations of roles and tasks.

References

Bales, R. (1951) *Interaction process analysis.* Cambridge, Massachusetts: Addison-Wesley Press.

Bishop, Y., Fienberg, S. & Holland, P. (1975) *Discrete multivariate analysis: theory and practice.* Cambridge, Massachusetts: MIT Press.

Carmel, E., Crawford, S., and Chen, H. (1992) Browsing in hypertext: a cognitive study. *IEEE Transactions on Systems, Man, and Cybernetics*, **22**(5), 865-884.

Chapman, J. (1981) A state transition analysis of online information seeking behavior. *Journal of the American Society for Information Science*, September 1981, 325-333.

Chen, C., Rada, R., & Zeb, A. (1994) An extended fisheye view browser for collaborative writing. *International Journal of Human-Computer Studies*, **40**(5), 859-878.

Conklin, J. & Begeman, M. (1988) gIBIS: a hypertext tool for exploratory policy discussion. *ACM transactions on office information systems*, **6**(4), 303-331.

Dourish, P. & Bellotti, V. (1992) Awareness and coordination in shared workspaces. *Proceedings of the ACM conference on computer-supported cooperative work* (CSCW'92). New York: ACM Press, 107-114.

Haake, J. & Wilson, B. (1992) Supporting collaborative writing of hyperdocuments in SEPIA. *Proceedings of the ACM conference on hypertext* (Hypertext'92). New York: ACM Press, 138-146.

Kraut, R., Galegher, J., Fish, R. & Chalfonte, B. (1992) Task requirements and media choice in collaborative writing. *Human-computer interaction*, **7**, 375-407.

Penniman, W. (1975) A stochastic process analysis of online user behavior. *Information revolution: Proceedings of the 38th American Society for Information Science annual meeting*. Washington, DC: ASIS, 1975 (12), 147-148.

Qiu, Liwen (1993) Markov models of search patterns in a hypertext information retrieval system. *Journal of the American Society for Information Science*, **44**(7), 413-427.

Zheng, M. & Rada, R. (1994) MUCH electronic publishing environment — principles and practices. *Journal of the American Society for Information Science*, **45**(5), 300-309.