CHI 99 SIG: Automated Data Collection for Evaluating Collaborative Systems

Jill Drury, Tari Fanderclai, Frank Linton

Motivation

The purpose of this CHI 99 Special Interest Group (SIG) session was to share lessons learned about using automated logging techniques to collect data for evaluating collaborative (multi-user) systems. Automated logging techniques are frequently used in evaluating the human-computer interaction of single-user systems. There has been much less experience in using logging techniques for evaluating collaborative systems, thus prompting the SIG proposal. We discussed issues surrounding using logging systems, methods, and metrics to collect data that are useful for evaluating collaborative systems.

Procedure

We asked the SIG participants to introduce themselves and to state the main collaborative systems evaluation logging issue they would like to discuss. From the issues put forth, we selected six at random to discuss in more detail. Because it is difficult to discuss automated logging issues in the abstract, we approached each issue by asking ourselves, "What advice would we give a colleague facing this issue?" While many of the suggestions given were specific to the particular circumstances, we felt that these types of issues are pervasive enough so that the discussion will provide insight to others even in significantly different situations.

The SIG participants divided into two groups. To promote a brainstorming approach, we allocated only 10 minutes of discussion time to each topic. At the end of the session, we reconvened in one group and each group chose one topic for full-group discussion.

Issues Discussed:

Automated measurements of events with emotional significance

A SIG participant is in the process of designing an experiment using automated logging to capture whether the quality of internet users' interpersonal interactions degrades with increasingly greater use of the internet. He wishes to collect data while respecting a minimally acceptable level of privacy, and while not inducing artificial behaviors. The logging will take place in subjects' homes, and would ideally include information such as the number, duration, content of, and rationale for interactions with family members; the frequency and duration of internet sessions, and the subjects' emotional state at various times.

Other SIG participants suggested logging/data collection methods involving clothing tags, beepers, and palm pilots. Clothing tags could measure proximity to other family members. Beepers could be used to request reports of feelings at the time of the beep. Palm pilots could be used for self-reporting of activities and emotions. Sampling of Internet usage could be used at irregular intervals so that the subject would not know when data was being taken (thus hopefully avoiding artificial behavior).

Metrics and data collection techniques for a large-scale collaborative system evaluation

A SIG participant must evaluate a large-scale collaborative system that is installed in many locations in a university. Besides automating collection of quantitative data (e.g., straightforward collection of time on task, numbers of keystrokes to complete task, etc.), he needs to efficiently collect

qualitative data (e.g., the more difficult issue of determining whether users appear to be confused or unhappy when using the system). The logging will occur as students use the system to complete class projects.

A SIG participant pointed out that there is a significant body of literature surrounding unobtrusive observation techniques, so that some qualitative data collection could, in fact, take place using trained (human) observers. Another participant noted that professors' grades of students' projects could serve as a source of independent evaluations of the results of using the collaborative system. Videotapes of people working on several "A" projects could be contrasted with videotapes of those working on the lowest-graded projects to determine if there is a particularly successful "style" of collaboration supported by the system. Once characteristics of this style are identified, it may be possible to search for it by fast-forwarding the videotapes. The system can then be at least partially evaluated by how well it supports a successful collaboration style.

Logging large amounts of telephone interaction

Another SIG participant has the task of evaluating how well a computerized system supports operators as they handle queries, requests, and complaints received from customers via telephone. The logging to support this evaluation needs to take place in the field, as operators handle calls from customers. The logging needs to be non-intrusive and must not hinder the operators in performing their jobs.

SIG participants discussed potential measures, such as the number of times a person called before their problem was resolved, the time it took the operator to provide an answer, the correctness of the answer, and the length and number of pauses in the conversation necessitated by the operator concentrating on the computer. To avoid an overwhelming amount of data, a sampling of complete conversations could be used when ascertain the correctness of the operator's answers. Participants agreed that it is difficult to get operators to log conversation information manually, so automated logging can be particularly useful in this situation.

Long-term logging of multi-media, multi-modal events

One SIG member needs to log multi-media, multi-modal events over a long time period. This task will generate a great quantity of diverse kinds of data, and the group discussed the problems of summarizing and using such data. One suggestion was that the researchers consider what questions they would want to ask of the data before beginning to collect data, and avoid collecting what wasn't needed. For example, if they record many hours of video, they will need to sample and code all of that video; they should, therefore, make sure that the video will tell them something that simpler logging methods, such as recording local computer system events, will not. Another suggestion was that the researchers look for ways to make the data as coherent as possible, for example by recording the various logs in similar formats.

Large scale field logs

Another SIG member who also has to collect large scale field logs brought up the problems of handling, storing and transporting all of the data. Many group members had experience with this problem, and added related problems to the list. For instance, one member mentioned the problem of having to keep vast quantities of data even though it is in an obsolete format. One helpful practice, which the group discussed, is to collect the data by logging on the client side, sending the logs to the server, and storing them in a database for analysis. As with the problem of long-term logging of multi-media and multi-modal events, the group stressed the importance of collecting only needed data.

Software for ill-defined data-collection tasks

The SIG member who brought up this problem explained that researchers frequently have to collect data before they know for sure what they will do with that data. It isn't possible in those cases to know exactly what to collect or the optimum format in which to store it for later visualization and analysis. The group brainstormed features of tools that would structure the task in some way. For example, the software would define a format in which actions would be logged. It would provide the user with ways of clustering and classifying the data. Accompanying visualization tools would allow visualization of the data at varying levels of detail.

Software for Logging or Analysis

The group identified the following tools for logging or for the analysis of logged data. The list is not meant to be exhaustive. The name of each tool is followed by a descriptive phrase and contact information.

Wosit: logs Unix GUI actions. The URL for wosit is: www.mitre.org/centers/cafc3/wosit/. Contact: M. Geier, megak@mitre.org.

OWL: logs Word commands; may be modified to log any application with a VisualBasic API. The URL for OWL is www.mitre.org/technology/tech_tats/modeling/owl/owl_hp.html. Contact F. Linton, linton@mitre.org.

McShapa: used for analysis of team activities. The URL for McShapa is: http://www.aviation.uiuc.edu:80/institute/acadprog/epjp/macshapa.html. Contact P. Sanderson, psanderson@swin.edu.au

Multi-modal logger: logs, stores, annotates, and displays multimodal events. The URL for MML is: http://www.mitre.org/technology/logger/. Contact: S. Bayer, sam@mitre.org.

An event recorder has been developed by E. Hyder. Contact: E. Hyder, Ehyder@cs.cmu.edu.

Custom logging code has been developed by J. Campbell. Contact: J. Campbell, Jeffc@sis.pitt.edu.

Server log analysis. a popular server logfile analyser may be found at: http://www.statslab.cam.ac.uk/~sret1/analog/.

Habanero: enables logging of Java applications. The URL for wosit is: http://havefun.ncsa.uiuc.edu/habanero/.

About the Organizers

Jill Drury is a Principal Engineer at The MITRE Corporation and a doctoral candidate at University of Massachusetts Lowell researching evaluation methods for collaborative systems.

Tari Fanderclai is a Human Factors Engineer at the MITRE Corporation with extensive experience developing, evaluating, and immersing herself in collaborative systems.

Frank Linton is a Lead Scientist at the MITRE Corporation. His research interests include computer-supported collaborative learning (CSCL) and recommender-based learning.

Authors' Address

Jill Drury, Tari Fanderclai, Frank Linton
The MITRE Corporation,
Bedford MA 01730.
Email: {jldrury, tari, linton}@mitre.org