Chapter XXI

Using Action-Object Pairs as a Conceptual Framework for Transaction Log Analysis

Mimi Zhang
Pennsylvania State University, USA

Bernard J. Jansen
Pennsylvania State University, USA

ABSTRACT

In this chapter, we present the action-object pair approach as a conceptual framework for conducting transaction log analysis. We argue that there are two basic components in the interaction between the user and the system recorded in a transaction log, which are action and object. An action is a specific expression of the user. An object is a self-contained information object, the recipient of the action. These two components form one interaction set or an action-object pair. A series of action-object pairs represents the interaction session. The action-object pair approach provides a conceptual framework for the collection, analysis, and understanding of data from transaction logs. We believe that this approach can benefit system design by providing the organizing principle for implicit feedback and other interactions concerning the user and delivering, for example, personalized service to the user based on this feedback. Action-object pairs also provide a worthwhile approach to advance our theoretical and conceptual understanding of transaction log analysis as a research method.

MOTIVATION

The ultimate purpose of search engine designers is to devise Web search engines that provide the most relevant information to each individual user. Since the user decides whether information is relevant or if the system is suitable, it is critical to understand the user’s system evaluation. Sun Tzu (n.d./1971), an ancient Chinese military strategist, said “know the enemy, know yourself;
Using Action-Object Pairs as a Conceptual Framework for Transaction Log Analysis

your victory will never be endangered” (p.129). This advice can be applied on the battlefield, but it can also apply to building information technology systems.

In a broad sense, one can understand Sun’s maxim as if you can know your own capability, and the characteristics and capabilities of people you deal with, it will be easier to devise processes appropriate to the situation. Therefore, in order to fulfill users’ information needs and serve them better, we should know the users, understand their goals, and recognize their information search tactics. If we can recognize users’ needs and their ways of approaching information, we can provide users with more suitable searching systems.

There are multiple ways to identify the individual user and provide tailored information systems. Search engines can learn about the users both explicitly and implicitly (Keenoy & Levene, 2005). In an explicit fashion, the users provide the necessary information to the system. The basis of this approach is that users would like to answer the questions, fill in a series of forms, or set up the profiles themselves. However, according to Keenoy and Levene (2005, p. 205), explicit feedback has low implementation rates due to the high cost of time and energy, unpredictable and unobvious benefits, and privacy concerns. This is in accordance with Zipf’s Law – an individual will only perform actions that cost “the least effort” (Case, 2002, p. 140). Zipf’s Law is a grounded and fundamental theoretical construct for information seeking studies. Zipf’s Law is used to guide user studies and understanding of human behaviors, as well as the development of information systems.

Rather than relying on explicit feedback by users, implicit feedback based on the analysis of interactions between the user and the system may be a better approach (Keenoy & Levene, 2005; Khopkar, Spink, Giles, Shah, & Debnath, 2003). Although it certainly depends on the design goals, the implicit approach is in many ways superior since the user does not need to perform more actions such as answering questions or setting up profiles. It is an unobtrusive method; therefore, the approach has less chance of altering users’ behavior.

The implicit approach is also highly dynamic. Since it analyzes and models current user interactions, it adapts well even if the users’ information needs change over time. White, Ruthven, and Jose (2001) compared the effectiveness of explicit and implicit feedback techniques and claimed no statistical difference between the two approaches. In addition, according to Zipf’s Law (1949), to users, the implicit feedback approach seems to be superior to the explicit feedback approach considering it costs them nothing but has the same effectiveness as the explicit feedback.

A search engine transaction log is “an electronic record of interactions that have occurred during a searching episode between a Web search engine and users searching for information on that Web search engine” (Jansen, 2006, p. 408). One can use the record of these interactions as a source of the implicit feedback. Dumais (2002) believes this is the only method for obtaining considerable amounts of data about users in a complex environment like the Web. Therefore, transaction log analysis seems a practical and convenient way to know the interactions of users with information systems. One can develop the user model by analyzing the data in transaction logs. Using this data, the system can make backward inferences to model the user and then make forward inferences to assist them with their information need.

However, there is a lack of theoretical frameworks for collecting, analyzing, and understanding data from transaction logs. Do we really need to analyze users’ every communication with the computer? If not, what kinds of user-system interactions do the transaction logs need to contain? Log files are usually huge and messy. How can we effectively and efficiently organize and analyze them? How can we get the data to make sense