A METHOD FOR ENABLING A BETTER COMPREHENSION OF USER BEHAVIOR

Gligor-Calin Loghin
SYSCOM - Université de Savoie*
Campus Scientifique 73376 Le Bourget du Lac – France**

Jean-Charles Marty* **
Thibault Carron* **

ABSTRACT
The work reported here takes place in the educational domain and deals especially with the observation of pedagogical activity. User behavior is observed through the raw traces (log files) of their activity. This method raises two problems: selecting only relevant traces and retrieving a high enough level of abstraction to make these traces clearly understandable by an analyst. A unifying approach is proposed to adapt and customize the raw traces in order to facilitate their interpretation. The objective is to obtain an adequate level of abstraction of traces that can better match a particular experience. In order to accomplish this objective the model described applies successive transformations that filter the traces of interest and change their level of abstraction.

KEYWORDS
Observation, raw traces, granularity, transformation of traces, educational experiment

1. INTRODUCTION
New developments in large families of applications as for instance in the education, e-commerce or enterprise organization domains are based on the study of users’ behavior. Such studies allow one to improve a site design (in the fields of e-commerce [Cooley, 99],[Burton, 01] or education [Corbiere, 04]) or to provide user on-line assistance (as for the educational domain [Zaiane, 02]).

User behavior can be observed from different sources, e.g. log files, eye-tracking data or even inspection by people outside the activity being studied. In this paper we mainly consider computer traces, i.e. numeric traces obtained from the explicit users’ actions. Examples of such traces can be server log files [Srivastava, 00].

In our team, we have used such an approach to evaluate educational workshops from several points of view. First of all, from a design point of view, as we obtain the relevant data needed to improve the "electronic schoolbag", a CSCW educational workspace developed and used in our university [Chabert, 05]; second, from an assistance point of view, as we support the teacher in a cooperative learning activity [France, 05]; finally, from a quality point of view, when we use the traces of a learning session to propose to the teacher an improvement to her/his learning scenario [Marty, 04] [Heraud, 05].

From all these experiments, we have learnt that traces contain rich behavioral data. But, the quantity of information is huge and the basic items extracted from the logs are very difficult to understand. Trying to explain this rough material through data mining methods [Srivastava, 00], scenario methods [Heraud, 05] or through change of granularity [France, 05] is an important topic in this area.

In this paper, we propose a unifyng approach to adapt and customize the raw traces in order to facilitate their interpretation. Our objective is to obtain an adequate level of abstraction of traces that can better match...
a particular experience. In order to achieve this objective our model applies successive transformations that filter the traces of interest and change their granularity (level of abstraction).

In the next section, we will present some existing methods that can be used to optimize and to abstract the volume of collected traces. Then, we will propose our model of transformation for the log file traces with an example. Finally we will conclude and give some perspectives for future work.

2. DIFFERENT TECHNIQUES FOR ADDRESSING TRACE UNDERSTANDING

Using log files as a data source raises two problems: the huge volume of data collected and the low level representation of this data. By removing useless data, we can reduce the volume of this data and minimize analysis time. But even a limited subset of data is difficult to interpret due to the use of raw parameters (e.g. ip, date, time, URL request) representing user traces.

Among the current research in this area, three methods emerge to optimize the log file dataset volume: data cleaning, data filtering and data summarization. Data cleaning removes from the log file the records that won’t be analyzed. Common methods use URL suffixes as in [Marquardt, 04] or [Cooley, 99] and they remove traces that are not explicitly requested by the user (images or robot requests). Although they significantly reduce the volume of traces (see results obtained in [Tanasa, 04]) these methods are too general to extract a restricted subset of data closely linked to the analysis objective. Their result generally produces an input dataset for filtering methods.

The goal of data filtering is to extract a subset of data that is relevant to the analysis objective. It allows optimization of the dataset size and contributes to the reduction of the search space for patterns [Zaiane, 01]. In their filtering operator implementation, Marquardt et al. [Marquardt, 04] use filtering conditions on data attributes. A filtering rule represents a logical expression involving attributes, logical connectives, relational operators and constants. In [Zaiane, 01] a similar idea is used to express constraints in order to define filters during the pre-processing phase.

From several points of view these existing solutions of data filtering are quite limited. First of all, it is difficult to handle raw attributes in order to express complex filtering constraints. Second, there is no support available to make the correspondence between the generated traces in the log file (in particular URL request) and their significance (which component has generated this request). This significance can help in the expression of filtering constraints and can facilitate the interpretation of traces by transforming raw attributes into more abstract description.

Data summarization transforms the set of URLs syntactically or semantically to consistently reduce their number. Syntactic generalization is used in [Rossi, 05] to simplify the URLs. In [Tanasa, 04] Tanasa et al. propose to group web pages with the same semantic topic under different generalized URLs. All these methods may cause loss of information (it will no longer be possible to differentiate elements of the same group). For that reason, these methods do not result in a very accurate analysis. The classification of files in different semantic topics is also difficult due to the large dimension of real sites.

The limited set of traces obtained must be prepared for interpretation. If we are able to get observed elements abstract enough to represent activities with clear semantics, the trace will be more easily interpretable by the teacher [Heraud, 05]. One solution to accomplish this task is to instrument the Web-based learning environments as in [Champin, 04]. They propose the creation of a use model that defines all observable elements of the system. By using this approach, traces can be collected at different levels of granularity. But it is difficult and time consuming to model (or re-model) systems by specifying all observable elements and their relations.

From the on-going work on these techniques, we can briefly conclude that existing filtering tools use raw attributes to express filtering requests. Moreover, existing approaches to abstract the collected traces are difficult to use. They suppose in general a particular development of the system.

Our proposition is based on a model of transformation for the log file traces. This model can filter the huge volume of traces by using a constraint-based approach. Users can thus express their needs at different levels of granularity: the level of raw attributes or higher levels of abstraction. A second transformation process can abstract the resultant traces at the level aimed at by the experiment, by applying patterns on lower level traces.
3. THE TRANSFORMATION MODEL

In this part, we first describe our general approach and we illustrate it via a concrete example.

In our approach, we consider that rough traces collected from an experiment constitute the input of the interpretation process. Most of the time, this material is not immediately interpretable by an analyst to draw conclusions on particular behavioral aspects. We thus need to make this data more appropriate for this interpretation. From our point of view, this must be done step by step, as in a cycle approach, each cycle aiming at solving a particular problem. For instance, we may focus on keeping only the relevant traces (filtering). Or, we can change the level of granularity of the traces, because they are too precise to allow good comprehension. Or, we may even need to represent the data graphically because this can be more appropriate for a given analysis case. These steps are not obligatory. They are strongly linked to the objective of the experiment. That is why we propose a transformation process of the rough traces leading to understandable material for the analyst (Figure 1).

Recently, we decided to start an experiment on our digital workspace in order to determine the impact of communication tools (email, instant messenger, forum, chat, etc.) on new usages in education. We gathered a large amount of numeric traces, and we faced two problems: 1) the extraction of a limited set of traces related to the use of these tools and 2) the abstraction of the raw attributes collected in order to facilitate their interpretation. We can notice that this example covers exactly the difficulties mentioned in section 2.

By identifying clearly these two problems, we can apply a two-step transformation process, one step per problem. Figure 1 presents an architecture supporting this process.

Figure 1. Transformation Model Architecture

First transformation: according to our experimental goal, we need to select the relevant traces, those related to the communication tools available in the workshop. A precise description of the domain (here the educational workshop) must be available at this point in order to perform an adequate filtering. The word "communication" must be recognized as a correct category of tools and the communication tools must be enumerated. Figure 2 gives an example of such a domain description. In addition, configuration parameters are also necessary. In Figure 3, one can find the path for accessing the communication tools. Similarly to MDE (Model Driven Engineering) approach, our aim is to obtain on the one hand a platform independent classification of the domain and on the other hand a platform specific configuration file. These descriptions make possible a higher level of abstraction for the analyst's requests (e.g. "consider the communication traces for people belonging to the observation group"). Applying this request to the rough traces provides us with a smaller set of traces, more relevant to this particular problem.

Second transformation: the second step changes the level of granularity of the traces obtained from the restriction step. In our example, we were interested in knowing how far the email tool and the forum tool were linked. We consider that the sequence of an email trace followed by a forum trace constitutes a more abstract action called 'incitation to join a forum'. This kind of abstraction can be represented by a rule. The left part of the rule is a sequence of traces and defines a pattern to find. The right part of the rule represents the more abstract trace. Very often, changing the granularity of the traces for better comprehension requires
the definition of a set of such rules. This change can be especially attractive when the experiment aim refers to another level of abstraction, for instance to discover how emails disturb educational activities or more generally to identify new usages in education due to the arrival of new technologies [Chabert, 05].

4. CONCLUSION

Collecting traces provides an interesting and rich way of better understanding users' behavior. Unfortunately, these rough traces are not easy to interpret. We therefore need to transform them into more abstract traces, understandable for an analyst. From different experiments, we have learnt that it is difficult to reproduce the transformation steps from one experiment to another. The reason is that the transformations are strongly dependent on the objective of the experiment. More generally, Observation has to be considered as a whole activity which must be prepared and eventually evaluated in order to be reused.

The cycle approach proposed in this paper is a general one, allowing at each step to have a sub-goal (filtering, abstraction, visualization) and to produce transformed traces closer to the analyst's expectations. This approach produces intermediate representations of the traces that can be exploited directly for other analysts' requests. This work opens several perspectives in relation with hot subjects in this domain. We can explore this method in order to evaluate students, experiments or even frameworks leading to a quality approach. Some results in collaboration with researchers in Information, Communication and Cognitive Psychology have already been obtained about emergent usages of new technologies in the education domain.

REFERENCES


