Application of Clustering for Usability Evaluation in Web Systems

Alana R. B. S. Lisbôa, Luciano T. E. Pansanato, José A. Fabri, Carlos N. Silla Júnior

Universidade Tecnológica Federal do Paraná, Brazil
{alana, luciano, fabri, silla}@utfpr.edu.br

Abstract. This paper presents ongoing research for usability evaluation in web systems using clustering in log files. The method indicates the application of K-MEANS algorithm in a log file created by Web Application User Tracking Tool (WAUTT). K-MEANS is an unsupervised learning algorithm to grouping data (clustering). There are no standards or pre defined classes. The idea is to extract relevant information from unlabeled data. WAUTT is a tool that captures events related to user interaction with a web system. An event in WAUTT is any user interaction with a specific element of the interface, like mouse click, page scrolling, among others. An application that provides an interface to register curriculums has been used to simulate the method. Data analysis has been performed by checking the clustered instances in each cluster created by the algorithm. The proposal highlights the importance of data mining tools as a way to minimize human interference in the process of finding useful information in databases.

Keywords: Clustering, Data Mining, Usability Evaluation, Log analysis

1. INTRODUCTION

One of the most techniques used to remote usability evaluation is log analysis [Jansen 2006]. This technique analyzes the user interactions recorded in log files. Besides identifying usability problems, log analysis can help developers to understand users’ behavior in relation to the interface of the system [Rogers et al. 2011].

The aim of this article is to verify the effectiveness of log analysis to usability evaluation in web systems using clustering. Section 2 presents the concepts of log analysis. Section 3 presents the related works. Methodology will be detailed in Section 4. Section 5 presents the analysis of the results, followed by the conclusions in Section 6.

2. LOG ANALYSIS

In log analysis the user interactions recorded in files created by the use of the system are analyzed. This technique has no interference in the user tasks, thus it preserves the real work context. The data collection is low cost, does not require the participation of the evaluator during the usability test and stores all user actions, which is a difficult task in other techniques. [Rogers et al. 2011].

Server log files are usually used to support remote evaluation of web systems. A limitation of this technique is that server log files contain only information about the pages (URLs) accessed by the user and it does not offer details about the real user interaction with a web system. This limitation can be solved with the use of the Web Application User Tracking Tool (WAUTT) [Rivolli et al. 2008] which tracks the user interaction with each element contained in web page.

An event in WAUTT is any user interaction with a specific element of the interface, including: mouse click, considering the coordinate relative to the top left corner of the screen; page scrolling considering the position of the scroll bar from the top screen; focus changing whenever an element
3. RELATED WORK

The related works about usability evaluation using log analysis include: UseMonitor [Cybis 2009], WAUTER [Balbo et al. 2005], AWUSA [Tiedtke et al. 2002], WELFIT [Santana and Baranauskas 2010], WebRemUSINE [Paganelli and Paternò 2002] and Ergo-Monitor [Schwerz et al. 2007]. Table 1 presents a summary with the main features and limitations of related work.

<table>
<thead>
<tr>
<th>Tool</th>
<th>Features</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>UseMonitor</td>
<td>Transaction oriented analysis and runtime errors</td>
<td>Inability to know the user's goals</td>
</tr>
<tr>
<td>WAUTER</td>
<td>Log analyses are based on performed and expected tasks</td>
<td>Not transparent to the user; requires the installation of software on the client computer</td>
</tr>
<tr>
<td>AWUSA</td>
<td>Sequences of user actions are compared to find differences between user interaction and the task defined by the evaluator</td>
<td>Use pages/links from the analysis of the HTML (Hypertext Markup Language); this approach complicates the association of several alternative paths and a single task</td>
</tr>
<tr>
<td>WELFIT</td>
<td>Graph algorithms are used to identify behavior patterns on an individually page</td>
<td>It does not consider the paths taken by the user</td>
</tr>
<tr>
<td>WebRemUsine</td>
<td>The approach is based on comparing a task model and path made by users</td>
<td>Required plugin; user must select the tasks they are performing; reducing the usable space of the screen</td>
</tr>
<tr>
<td>Ergo-Monitor</td>
<td>The metrics are estimated by using models of expected behaviors for specific tasks</td>
<td>The model behavior is only based on URLs</td>
</tr>
</tbody>
</table>

The analysis of the existing approaches presented in Table 1 shows that the works do not have sufficient information about the usability in web systems. The difference of this work is related to track and record user interaction by WAUTT, which recorded user’s interaction with page elements, such as button and text boxes, beside the web pages, as server log files. The details captured by WAUTT can provide additional and important information about the user interaction with the system. Furthermore, these data are obtained without user participation and additional cost.

4. METHODS AND PROCEDURES

The WAUTT tool was prepared to analyze the interaction of users with a web system, which provides the service of curriculum registration. A hundred and twenty (120) logs created by the access to registration form were analyzed. Each log is a record of the task “curriculum registration” for a different user. Log analysis was performed after applying a clustering algorithm, which aimed to group records according to the similarity between them, i.e. data records with similar patterns were grouped in the same cluster. Clustering techniques use unsupervised learning: there are no standards or pre defined classes. The idea is to extract relevant information from unlabeled data. This paper is on going research, thus K-MEANS algorithm was chosen. K-Means is the most popular clustering algorithm due to its simple implementation, low run-time and space complexity and simple usage since no parameters (except the number of clusters) are involved [Bradley and Fayyad 1998] [Breaban and Luchian 2011]. Weblog data used as source for execution of the algorithm were captured by WAUTT and are structured as follows: date, time, IP, operating system, browser, URL, event. An example of log file entry created by WAUTT is:

```
```

The first step in data preparation is the exclusion of not relevant information to the process. This information refers to the following information in the WAUTT log file: date, browser, mouse position, keystroke and uploading of images. The next stage was data normalization. Clustering uses the concept of distance (similarity) between records and, generally, convert different data types (ordinal,
categorical, binary intervals) to a common scale, e.g. \([0.0, 1.0]\). The purpose of this normalization is to reduce discrepancies between the variables and standardize the information in a small range of values.

5. RESULTS

In this section we are interested in evaluating if the K-MEANS algorithm is able to distinguish between the different outcomes of tasks. The results were obtained by analyzing the log file instances grouped in clusters corresponding. The analysis of the results performed following criteria of representativeness of each cluster. For example, a cluster with the largest number of instances containing logs that indicate the page loading “error” and “help” may indicate a usability problem. Users may be having difficulty in finalizing the proposed task. Alternative paths used to achieve success in the task, e.g., constant access to “help” page can also show that users need further explanation about the steps to be taken. Another relevant factor is the time to finish the task or the amount of clicks on the forms. If the number of clicks are high comparing them with other records also show difficulty in filling/understanding the information requested. Likewise, access patterns with low completion time may indicate withdrawal of the task. In summary, data analysis must take into account: examination of the patterns presented in more and less representative clusters, analysis if the clusters presents the expected information, analysis of deviations in task execution, analysis of execution times, analysis of the amount of clicks to complete the task, analysis of the frequency of occurrence of the scroll event. Besides the representativeness criteria, it was analyzed the quality of the clusters by purity, which was calculated by the relationship between the principal class of each clusters and the size of this cluster. The results of purity indices were: Cluster 0 and Cluster 1 - 0.80, Cluster 2 - 0.91, Cluster 3 - 1.00. Results closer to 1 means a high quality cluster. We have employed the K-MEANS with \(k = 4\). Our expectation is that the task at hand could be divided into four classes. More precisely: 1) task executed successfully; 2) task executed successfully but with some alternative path - error or help; 3. dropouts; and 4) without success. The percentage of clustered instances in each cluster was distributed as shown in Figure 1.

![Fig. 1. Clustered Instances](image.png)

Cluster 2 (5 instances) refers to users who remained in the system for less time, dropped out of performing the task or closed the browser. There is no evidence that these same users returned to the system to submit your curriculum at another time, they just gave up at that point. Cluster 1 indicates that 49% of users were redirected to the “error” page, asked for help, but still spent a shorter time to complete the task than the instances of Cluster 0. All users were redirected to the “error” page, used scrolling, indicating that the interface may adapted to the screen space. Nielsen and Loranger [2006] claim that only 23% of the users of a system use the scroll bar on a webpage, the other 77% just visualize the contents of the page that are visible. This percentage of users who use the scroll decreases the second time you access the page. Cluster 0 and cluster 3 presents a very similar pattern. In both cases, users were successful in completing the task, but consulted the help page. The big difference is the completion time of each grouping. Instances in cluster 0 took on average 70% more time than the other users, changed the mouse focus and re-typed more fields in the forms. Overall it was possible to deduce that there is a lot of information needing review to minimize usability problems. In our experiments, almost half of the users have committed some kind of error, which may indicate a lack of
A. R. B. S. Lisbôa, L. T. E. Pansanato, J. A. Fabri and C. N. Silla Júnior

clear information and a poorly designed interface. Within the aspects that can be reviewed include: creating masks in some fields, suitability of the text contained in the “error” page to all information becomes visible on the first page, without the use of the scroll and create short texts with notes to explain how to complete each form field.

6. CONCLUSIONS

Clustering techniques facilitates the analysis of the data. The records are grouped according to the behavior pattern of users. The method automates part of the process of usability analysis, allowing evaluators to focus their efforts on finding the usability problems. The novelty of the proposed method is to use the log files generated by the WAUTT. The results indicate that the usability evaluation and the data mining techniques should be tested and combined to compare and to discover the best technique or combination of them can be used. This research should be continued by testing other clustering algorithms. Another extension is the application of rule extraction algorithms, after the use of clustering techniques, which minimize human intervention in the analysis phase of the clusters.

REFERENCES


Jansen, B. J. Search log analysis: What it is, what’s been done, how to do it, 2006.


