Extraction of User Interaction Patterns for Low-Usability Web Pages

Toshiya Yamada[1], Noboru Nakamichi[2], and Tomoko Matsui[3]

[1] Department of Statistical Science, The Graduate University for Advanced Studies,
10-3 Midiri-cho Tachikawa Tokyo Japan
[2] Faculty of Information Sciences and Engineering, Nanzan University,
27 Seirei-cho, Seto, Aichi Japan
[3] The Institute of Statistical Mathematics
10-3 Midiri-cho Tachikawa Tokyo Japan
tyamada@ism.ac.jp

Abstract. Our goal is to point out usability problems in web pages in order to improve the web usability. We investigate the relation between user interaction behaviors in web-viewing and evaluation results of web usability by subjects. And we extract discriminative patterns for user interaction behaviors in visited web pages with low usability by using the PrefixSpan based subsequence boosting (Pboost).

Keywords: Web Usability, PrefixSpan Boosting (Pboost), User Interaction, Machine learning.

1 Introduction

The usability of web sites is important because it reach sales and user interest. To improve usability, methods for evaluating web sites pages are required. A well known one is usability testing. In this method, multiple subjects actually visit the target web site and evaluators then analyze the interactions. The evaluators often discover unpredictable problems with the web site. To date, usability testing evaluators have used recording devices such as a video tape recorder and voice recorder to record the subjects' interactions. They have interviewed subjects asking them several questions about the web site usability in order to collect subjects' evaluations. Usability testing has a problem that evaluators must spend a huge amount of time on the interview and analysis replaying the recorded data.

Our goal is to identify problems on web pages that have low usability in order to reduce such prodigious labor for evaluators. We investigated the relation between user interaction behaviors in web-viewing and the subjects' evaluations. And we extracted discriminative patterns for user interaction behaviors in viewing low-usability web pages by using PrefixSpan-based subsequence boosting (Pboost) [S. Nowozin: 2007].

For our future tasks, we should do a lot of tests because we can evaluate the usability of web sites more easily. It is expected that the interaction that causes the new problem of usability can be discovered.

M. Kurosu (Ed.): Human Centered Design, HCII 2011, LNCS 6776, pp. 144–152, 2011.

2 Related Works

In this chapter, we describe methods of supporting evaluation objectively using quantitative data such as browsing time, mouse movement and eye movement.

Laila [Laila and Fabio 2002] analyzed the execution situation of a task from a user's operation event recorded by Java script. They supported the analysis based on quantitative data, such as page reference time and task execution time. They analyzed the usability of a Web page based on task execution time totaled for every Web page. Okada [Okada and Asahi 1999] developed the GUI-TESTER which extracts a common operation pattern from two or more users' operation history. If the tool is used, the operation pattern for mistaken operation can be extracted. And when the moving distance of a mouse cursor is long and an operation time interval is long, they suggest a possibility that a screen layout is bad.

WebTracer [Nakamichi et al. 2007] can collect the operation log of users on the Web pages. Collectable data include the information on users' sight line (the coordinates of the gazing point on the computer screen), operation log of a mouse, and the displayed screen images, together with their time information. The data collected by WebTracer characterize Web pages and have the possibility of being used for supporting the usability evaluation. However, the relation between such data and the problems in the Web usability was merely an example of the characteristics of the Web pages. Quantitative evaluation of the relation to the usability of Web pages was not done.

Heatmaps from user eyetracking studies based on fixations are used for observing in detail. Nielsen found that users' main reading behavior was fairly consistent across many different sites and tasks. [Nielsen 2006]. Eyetracking visualizations show that users often read Web pages in an F-shaped pattern: two horizontal stripes followed by a vertical stripe. Specialists may evaluate using eyetracking heatmaps. However it doesn't lead to the cost reduction of the web usability evaluation.

Eye-tracking methodologies are applying in the domain of Web search because gaze can be used as a proxy for a user's attention. Eye-tracking measures include pupil dilation, fixation information, and sequence information such as scan paths [Guan and Cutrell 2007]. They relied on measures related to gaze fixations with a minimum threshold of 100 ms in areas of interest. They found that as they increased the length of the query-dependent contextual snippet in search results, performance improved for informational queries but degraded for navigational queries. Analysis of eye movements showed that the decrease in search performance was partially due to the fact that users rarely looked at lower ranking results. Matsuda measured users' eye movements during web search tasks to analyze how long users spend on each result of the results pages [Matsuda et al. 2009]. They found the results displayed on the bottom of the page were viewed for a shorter time than the results displayed on the top of the next page.

In these conventional researches, the specialist had discriminated usability only using certain quantitative data. However, the effectiveness of combinational data was not verified statistically.

3 Experiment of Usability Testing

3.1 Quantitative Data of Users' Behavior

Browsing time, mouse movement, and eye movement are the quantitative data about users' behavior mainly used for web usability evaluation. This experiment recorded the interaction data for every Web page. Here, the interaction data is a record of the user's behavior during browsing such as "Where on the page is the user looking?" and "Where is the user's mouse cursor?". In this study, it adopts PageView (PV) as the count way of the Web page. We count the Web page which required of once with the browser and read from the web server as 1PV.

The interaction data is the vector which composed of the gazing point coordinates and the mouse cursor coordinates on the whole time of a subject browsing web 1PV. Gazing point is the point at the intersection of the users' look with the target screen.

3.2 Experimental Environment

The experiment environment used by this research is as follows.

- Display: 21 inches (Viewable screen size: H30 x W40cm)
- Device for measurement of sight line: NAC, EMR-NC (View angle: 0.28, resolution on the screen: approx. 2.4mm)
- Recording and playing of sight-line data: WebTracer (Sampling rate: 10 times per second)

WebTracer [Nakamichi et al. 2007] is an environment for recording and analyzing the users' operations in Web pages.

3.3 Experimental Procedure

We experimented with usability evaluation in the following procedures to five tasks. Subjects are 15 frequent users of the Internet. They have never visited the sites used in the experiment. We requested the subject to perform five tasks of looking for the starting salary of a master from the site of five companies, as a main experiment.

Procedure 1**:** The Web page for an experiment linked to the top page of each company is displayed by a subject. And the experiment is started from the time of a subject clicking the link.

Procedure 2**:** While subjects are doing the tasks, several types of quantitative data are recorded using WebTracer.

Procedure 3**:** The Web pages that subjects visited are displayed. We requested the subject to choose the ease of use for every visited Web page from the following five levels. We defines a low usability page as a page that a subject choose "hard to use" from four levels of the questionnaire.

1. Hard to use
2. Relatively hard to use
3. Relatively easy to use
4. Easy to use

***Procedure 4*:** We reproduce the operation history recorded by WebTracer, and a subject checks all the visited Web pages. At that time, we interviewed the subjects about the situation of their search.

We recorded the quantitative data for 275 pages which the subjects visited. We were not able to record correctly about 12 pages of them. The cause is a frequent blink and head movement. Moreover, there were 8 pages which the subjects answered "don't know" about the usability of the Web page. We measured the quantitative data in 263 pages except these pages: 20 pages of them were class 1 (hard-to-use group) and 243 pages were class 0 (other evaluation group).

4 Pattern Extraction Using Pboost

4.1 Pboost

We used Pboost to extract interaction patterns that were able to discriminate a web page evaluated as having low usability from other pages. Pboost was originally developed for sequential data classification which was proposed by Nowozin [9].

We assume that “S” is whole interaction pattern, and “s” is subspace of interaction pattern “S”: $s \in S$. The input data x_n is a vector of interaction data of a web page “n” ($n = 1 \cdots l$). And class label $y_n \in \{-1,1\}$ is input data too. Discriminant function $f(x)$ is follows:

$$f(x) = \sum \alpha_{s,\omega} h(x; s, \omega)$$

$$h(x; s, \omega) = \begin{cases} \omega & s \subseteq x \\ -\omega & otherwise \end{cases} \quad \omega \in \{-1,1\}$$

Where, $\alpha_{s,\omega}$ is called weight for pattern s . $h(x; s, \omega)$ are hypothesis function, an extra variable ω can decide for either class decision. To obtain an evaluable discriminant function, we need solve the optimization problem that formulated as linear problem. Using PrefixSpan algorithm as clever search strategy, we can solve this problem optimally.

4.2 Input Data

The interaction data with a class label is used as input data for Pboost. For the class label, we asked the subject to choose the ease of use for every visited web page from four levels: (1) hard to use, (2) relatively hard to use, (3) relatively easy to use, and (4) easy to use. We defined class 1 for low-usability pages for which a subject chose (1) and class 0 for the other pages as the other evaluation group.

Table 1. 10 criteria of dividing equally a screen

Criterion No.	Number of dividing equally a screen	Size of the division cell
1	2	512 pixel × 768 pixel
2	2	1024 pixel × 384 pixel
3	4	512 pixel × 384 pixel
4	12	256 pixel × 256 pixel
5	25	205 pixel × 154 pixel
6	48	128 pixel × 128 pixel
7	100	102 pixel × 77 pixel
8	192	64 pixel × 64 pixel
9	400	51 pixel × 38 pixel
10	768	32 pixel × 32 pixel

Meanwhile, it is possible to use only the data of the discrete-value as the input-data of Pboost but our recorded data is a continuation value. Due to this, we need to transform the interaction data from continuation value into discrete-value. First, we divide a screen into some the same size. Next, we give an integer value to each division cell. Lastly, it makes the coordinates of our recorded data correspond to the integer number of the contained cell. However, in this way, "the similarity between data" which could be expressed in the original coordinate-value cannot be expressed. Therefore, we used 10 criteria when dividing equally a screen. Then, we gave 10 values which were put by those 10 criteria to one interaction data. Table 1 shows these 10 criteria.

We show the example which was transformed into the integer value using these. In case of the interaction data such as $(x_{eye}, y_{eye}, x_{mouse}, y_{mouse}) = (250,1,250,1)$, the changed integer value vector becomes [1, 1003, 2001, 3001, 4002, 5002, 6003, 7004, 8005, 9008, 10001, 11003, 12001, 13001, 14002, 15002, 16003, 17004, 18005, 19008]. This vector consists of 20 elements about viewpoint and mouse cursor. When changing in this way, we can express " the similarity between data" as the number of the same elements of vectors.

We use the data which did such a change as the input data for Pboost. Our method using Pboost has two advantages: (i) the evaluator does not need to replay recorded data in order to analyze the subjects' interactions and (ii) the subjects' only need to answer simple questions and the burden on them is small.

5 Experiments

We extracted user interaction patterns for the Web page with low usability. All the data was used as input-data to train Pboost. We could completely classify the interaction data by using Pboost and extracted the interaction patterns. Total numbers of patterns extracted were 76, of which 40 were positively related to the identification of class 1 and the remaining 36 were related to class 0 identification. Figure 1 and the 2 show 10 visualized extraction patterns for each class.

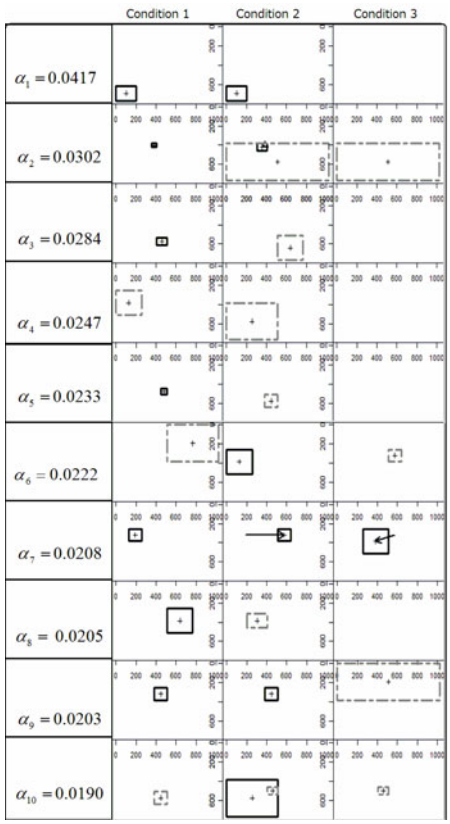

Fig. 1. Top 10 of the interaction pattern which has an influence on the direction which is identified by class 0

In the Figure 1 and 2, 10 patterns are showed at the longitudinal-direction. Also, "the condition" transition of each pattern is pictured in the crosswise direction. The outer frame with the scale shows the whole screen, the black square frame shows the area which a viewpoint is contained in and the gray chain line frame shows the area which the mouse cursor is contained in. Now, "the condition" is that a view point or the coordinates of the mouse are stored in the area in the screen. Moreover, it counts even if it doesn't transfer continuously between each condition and the condition. α_i shows the strength of the influence of each pattern. The bigger $|\alpha_i|$ is, the stronger the influence is.

These relations were found from the weight values for the interaction patterns in the discriminant function of Pboost. Many of the interaction patterns included mouse information.

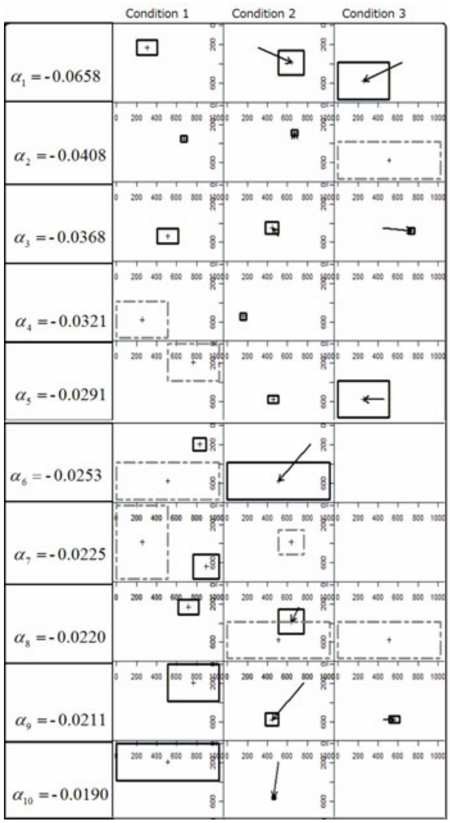

Fig. 2. Top 10 of the interaction pattern which has an influence on the direction which is identified by class 1

6 Conclusion Remarks

By using the proposal method, we were able to extract a total of 76 interaction patterns for the hard-to-use group and other evaluation group. Our method does not need the evaluator to reproduce of subjects' interactions and is effective at reducing the evaluator's load. Moreover, subjects only need to answer simple questions. Therefore, the method is effective at reducing the subject's load.

Figure 3 show two discriminant-function values $f(x)$ of Pboost when a user browses a "hard to use" page (class 1) and an "easy to use" page (class 0), where the horizontal axis shows browsing time and the vertical axis shows a discriminant-function value. If the discriminant-function value of a page i is $f(x_i) \leq 0$, it is classified into class1. We see from Figure 3 that this page was judged "hard to use" page 6 seconds later after beginning to browse it. Thus, we can understand when the web page was judge into "hard to use".

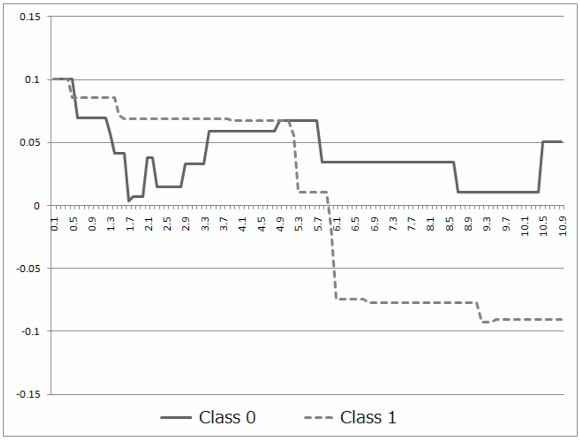

Fig. 3. Line graph of discriminant-function value

In addition to that it is possible to do extraction interaction patterns by using the proposed method, it is possible to understand when or why the user feels that the web page is "hard to use".

Acknowledgements. This work was supported in part by the Center for the Promotion of Integrated Sciences (CPIS) of Sokendai.

References

1. Guan, Z., Cutrell, E.: An eye tracking study of the effect of target rank on web search. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 417–420. ACM, New York (2007)
2. Joseph, S.D., Redish, A.P.: A Practical Guide to Usability Testing. Ablex Publishing, New Jersey (1993)
3. Kelly, G., Emily, G.: Web ReDesign. Pearson Education, London (2002)
4. Laila, P., Fabio, P.: Intelligent analysis of user interactions with web applications. In: Proceedings of the 7th International Conference on Intelligent User Interfaces, pp. 111–118. ACM, New York (2002)
5. Matsuda, Y., Uwano, H., Ohira, M., Matsumoto, K.: An Analysis of Eye Movements during Browsing Multiple Search Results Pages. In: Proceedings of the 13th International Conference on Human Computer Interaction, pp. 121–130. Springer, Heidelberg (2009)
6. Nakamichi, N., Sakai, M., Hu, J., Shima, K., Matsumoto, K.: Webtracer: A New Web Usability Evaluation Environment Using Gazing Point Information. In: Electronic Commerce Research and Applications, vol. 6(1), pp. 63–73. Elsevier Science, Amsterdam (2007)
7. Nakamichi, N., Sakai, M., Shima, K., Hu, J., Matsumoto, K.: Webtracer: A New Web Usability Evaluation Environment Using Gazing Point Information. Electronic Commerce Research and Applications 6(1), 63–73 (2001)
8. Nielsen, J.: Designing Web usability. New Riders, Indianapolis (2000)

9. Nielsen, J.: F-shaped pattern for reading web content. Jakob Nielsen's Alertbox (April 17, 2006), http://www.useit.com/alertbox/reading_pattern.html
10. Nowozin, S., Bakir, G., Tsuda, K.: Discriminative subsequence mining for action classification. In: Proceedings of the 11th IEEE International Conference on Computer Vision. IEEE Computer Society, Los Alamitos (2007)
11. Okada, H., Asahi, T.: GUITESTER: a log based usability testing tool for graphical user interfaces. IEICE-Trans. on Information and systems E82-D(6), IEICE, 1030–1041 (1999)
12. Oliveira, T.P.F., Aula, A., Russell, M.D.: Discriminating the relevance of web search results with measures of pupil size. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems, pp. 2209–2212. ACM, New York (2009)