Proceedings of the 2007 IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2007)

A Quantitative Method for Analyzing Scan Path Data Obtained by Eye Tracker

Haruhiko Takeuchi, and Yoshiko Habuchi
National Institute of Advanced Industrial Science and Technology (AIST)
AIST Tsukuba Central 6, Tsukuba, Ibaraki, 305-8566, JAPAN
takeuchi.h@aist.go.jp, habuchi.y@aist.go.jp

Abstract— **Scan path is one of the most important metrics measured by eye tracking systems. This paper describes a new method for analyzing scan-path data based on the string-edit method that is popular for correcting human errors made at the input stage. We defined several cost functions for the substitution costs in the string-edit method, and applied the method to the scan-path data we had collected in a series of experiments for studying Web browsing behavior. We demonstrate the usefulness of our method and discuss the appropriate cost functions for the eye-tracking data.**

I. INTRODUCTION

Eye-tracking systems have been used in various fields for measuring people's viewing behavior. There is a variety of metrics derived from the eye-tracking systems [1]. Those metrics are classified into two groups: static metrics and dynamic metrics. Static metrics are indices calculated by accumulating or averaging observed data, such as fixation durations, cumulative fixation times, number of fixations, and gaze rate. Dynamic metrics are obtained from time series data that contain information on the time axis such as the scan path, which is a spatial arrangement of a sequence of fixations.

Static metrics are widely used in the community because there are standard algorithms to derive these metrics. In contrast, utilization of dynamic metrics is limited. Qualitative representations of the scan path are available, such as a gaze plot that connects fixations superimposed on the visual stimuli according to the order of their occurrences. New methods for visualization have also been proposed [2] [3]. However, quantitative metrics that compare with those for static metrics have been little utilized due to the lack of standard tools for quantitatively analyzing time series data.

Qualitative representations for dynamic metrics are useful for understanding overall tendencies of viewing behavior along the time axis, but there is a serious limitation: viewing patterns from two users, for example, can not be quantitatively compared. Nonetheless, there are cases where we need to quantitatively compare a number of viewing patterns, and this paper focuses on this quantitative analysis of scan-path data.

The quantitative approach treats the data numerically. The only method proposed so far is the application of the so-called string-edit method to scan-path data. The string-edit method was originally developed in the field of coding theory. The string-edit method is also referred as the Levenshtein distance method or optimal matching analysis. The string-edit method has been used in various fields such as pattern analysis, speech research, and human behavior analysis. The method is particularly well known for analyzing human errors in typing.

The string-edit method calculates the distance between two strings. This method was formalized by Levenshtein. There are three basic operations for calculating the distance in this method. Those operations are called deletion, addition, and substitution. Using these operations, one string is transformed into another string. Every time an operation is applied, a pre-assigned cost is accumulated. The distance between the two strings is defined as the smallest cost for matching those two strings [4].

Josephson and Holmes [5] first applied the string-edit method to the scan-path data. Although they noted that substitution costs should be assigned so that the cost between the closer targets was small, they focused on "adjacency" instead of distance. They counted the number of regions between the two target areas based on the concept of adjacency. Therefore their substitution costs were not always monotonic with physical distance. In our method, we define the cost functions that directly reflect physical distance between the two target areas. We thus apply the string-edit method to scan-path data more appropriately.

II. METHODS

A. Data

The scan-path data are time series data of fixations, which are the main gaze points of measured eye-tracking data. Originally, the scan-path data are expressed as values of x and y axis of a graphic display with duration time calculated from raw eye-tracking data. We divide the graphic area into several parts, and assign a unique letter for each portion. We then derive a string of letters from the original scan-path data. Figure 1 shows an example of the graphical partition. In this case, the whole area is divided into nine parts, and letters "A" to "I" are assigned. An example of a string is "ABBBCBDGAB." The way the area is divided depends on the contents of Web pages and the purpose of the analysis. The division should be in precise in order to deal with detailed contents. However, the division should be larger to facilitate understanding the general tendency of the scan path.

1-4244-0705-2/07/$20.00 ©2007 IEEE

A	B	C
D	E	F
G	H	I

Fig. 1. Example of graphic partition

B. Cost functions

As described in the previous section, the string-edit method consists of three basic operations: insertion, deletion, and substitution. As a default, the same cost is assigned to each operation. As Josephson and Holmes [5] noted, it would be more reasonable to assign appropriate values for substitution cost.

We will introduce cost functions for analyzing scan-path data. In this paper, we compare three types of cost functions: (1) uniform cost function, (2) city block cost function, and (3) Euclidian cost function. The uniform cost function is used as a reference. All substitution costs are set to equal for all substituting pairs. The second and third cost functions are our original methods for analyzing scan-path data. Substitution of spatially close gaze points costs little, while substitution between long distance points costs much. The city block cost function uses city block distance to calculate substitution cost in the string-edit method. In the Euclidian cost function, the substitution cost is defined by the Euclidian distance between the original geographical positions.

Let us define the geographical positions for each character of the strings as follows. The specified position (u_1, u_2) for the character u corresponds to the center of the assigned geographical area of the character.

$$u = (u_1, u_2) \qquad (u \in S)$$

Here, S denotes the set of characters whose elements stand for the respective area names. Let us define the cost function based on city block distance and Euclidian distance. The substitution cost f from u to v, and vice versa, is defined below. In the following equations, the parameter α denotes a kind of normalization parameter that differs according to the effective screen size and number of partitions.

The substitution cost for city block distance is defined by the following formula:

$$f(u,v) = \alpha \sum_{i=1}^{2} | u_i - v_i | \qquad (u, v \in S)$$

The substitution cost for the Euclidian distance is defined by the following formula:

$$f(u,v) = \alpha \sum_{i=1}^{2} \sqrt{\left(u_i - v_i\right)^2} \qquad (u, v \in S)$$

Table 1 lists the substitution costs based on the city block distance and Euclidian distance for the data in Fig. 1. Here, we assume that the effective screen size is 750 x 750 pixels, and α is 0.001.

TABLE 1

Substitution costs for the data in Fig.1

(a) City block distance model

	A	B	C	D	E	F	G	H	I
A	0	0.25	0.5	0.25	0.5	0.75	0.5	0.75	1
B	0.25	0	0.25	0.5	0.25	0.5	0.75	0.5	0.75
C	0.5	0.25	0	0.25	0.5	0.25	0.5	0.75	0.5
D	0.25	0.5	0.25	0	0.25	0.5	0.25	0.5	0.75
E	0.5	0.25	0.5	0.25	0	0.25	0.5	0.25	0.5
F	0.75	0.5	0.25	0.5	0.25	0	0.25	0.5	0.25
G	0.5	0.75	0.5	0.25	0.5	0.25	0	0.25	0.5
H	0.75	0.5	0.75	0.5	0.25	0.5	0.25	0	0.25
I	1	0.75	0.5	0.75	0.5	0.25	0.5	0.25	0

(b) Euclidian distance model

	A	B	C	D	E	F	G	H	I
A	0	0.25	0.5	0.25	0.35	0.56	0.5	0.56	0.71
B	0.25	0	0.25	0.5	0.25	0.35	0.56	0.5	0.56
C	0.5	0.25	0	0.25	0.5	0.25	0.35	0.56	0.5
D	0.25	0.5	0.25	0	0.25	0.5	0.25	0.35	0.56
E	0.35	0.25	0.5	0.25	0	0.25	0.5	0.25	0.35
F	0.56	0.35	0.25	0.5	0.25	0	0.25	0.5	0.25
G	0.5	0.56	0.35	0.25	0.5	0.25	0	0.25	0.5
H	0.56	0.5	0.56	0.35	0.25	0.5	0.25	0	0.25
I	0.71	0.56	0.5	0.56	0.35	0.25	0.5	0.25	0

III. APPLICATION

A. Eye-tracking experiments

In this section, we will first show the details of our eye-tracking data. These data were measured as a series of Web usability tests.

Subjects Twenty subjects were recruited for this study. They were divided into two Web user groups, ten heavy Internet users and ten light Internet users. The heavy Internet users browse various sites for more than 10 hours per week; the light Internet users use the Internet less than 6 hours per week.

Stimuli The stimuli presented to the subjects were three Web pages in three different Web categories, portal, news, and advertisement. These three Web pages are shown in Fig. 2. The three Web categories were chosen, because each Web page in these three Web categories has a distinct visual imagery as stated in Josephson & Holmes [5].

Apparatus A Tobii x50 eye-tracking system with a TFT 21” display and maximum resolution of 1600 X 1200 was used in this study. The eye tracker has a tracking rate of 50Hz.

Procedure Subjects were instructed to browse each Web page for 20 seconds and to evaluate it. The eye tracking data for the 20 seconds were recorded.

(a) Portal page

(b) News page

(c) Advertisement page

Fig. 2. Three Web pages used in the experiment

B. Data preparation

This section describes the data preparation. First, fixations were extracted from the measured eye-tracking data. Here, the minimum time of fixation was set to 100msec, and the greatest radius was set to 30 pixels. Second, we transformed the fixation data into string data. We divided each Web page into small areas, dividing the effective screen size of 1400 x 1200 pixels horizontally into seven partitions and vertically into six partitions. Thus each area was a 200-pixel square. We then assigned a letter to each small area, using "A" to "Z" and then "a" to "p." We next translated each fixation data point into a letter that represents the corresponding area. We now had the strings of characters for each subject's fixation data. Two examples of the strings for the advertisement page are given below.

CBCDCCDDCCCCCDDDDEECCDEJIJRRRSSL
TTLEECCRRQPPWIICICCKPPPQQSRa

CBBKQIICDDKCLRRRLLTKTTTTTSSTTMML
aPPPPIRBBCCBBCCDKDDKKRKIIJ

In the subsequent analysis, we used only the data of heavy Internet users. We selected five subjects for each of the three Web categories, so we had to analyze a total of 15 scan paths. The data were selected to simplify the analysis and comparison. This data set was analyzed by the string-edit method with three different cost functions.

C. Results and discussion

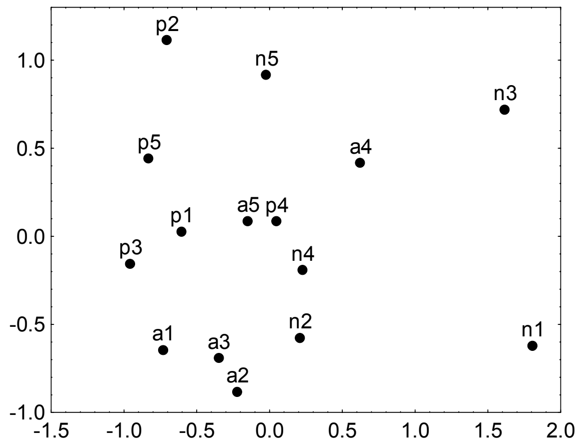
We calculated the dissimilarity between each pair of the strings based on the string-edit method, using three different cost functions for the calculation. The cost of deletion and addition was set to one, and the parameter α was set to 0.001. As a result, we obtained final string-edit costs for each pair in the 15 selected scan paths. We will call the final cost the dissimilarity between the two strings.

We can calculate the average distance among identical Web pages (A) and the average distance between the different Web pages (B). The ratio of B to A expresses the goodness of separation; the greater the value, the better the method. We repeated the random selection of 15 scan paths and calculated the average dissimilarities. Table 2 lists the result of one hundred trials and indicates that the separation ratio of the uniform distance model is smallest. The separation ratio of the city block distance model and Euclidian distance model is bigger than that of uniform distance model. This result proves that the string-edit method with appropriate substitution costs performs better.

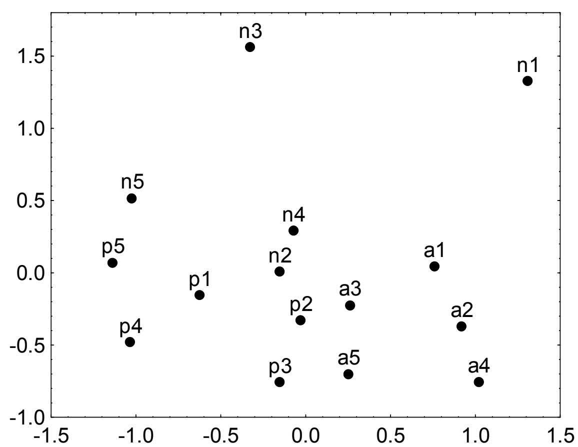
TABLE 2

Separation ratio calculated by three distance models

Distance model	Average distance in the same page Mean (A), SD	Average distance between the pages Mean (B), SD	ratio B/A
Uniform	0.869 (0.015)	0.897 (0.010)	1.032
City block	0.526 (0.016)	0.558 (0.017)	1.061
Euclidian	0.459 (0.015)	0.479 (0.015)	1.044



(a) Example of uniform distance model



(b) Example of city block distance model

Fig. 3. Results of two-dimensional configurations derived from multidimensional scaling

We analyzed the dissimilarity matrix by multidimensional scaling analysis (MDS) in order to understand the effect of cost functions. Figure 3 presents the two- dimensional configurations of the uniform distance model and city block distance model. Here, the label "p" denotes portal page, "n" denotes news page, and "a" denotes advertisement page. And the number following those letters corresponds to the subject number.

As demonstrated in Fig. 3 (a), the three Web pages were not properly separated when we used the uniform cost function. However, in Fig. 3 (b), we observed that advertisement page scan sites concentrated at the bottom right of the page, portal page scan sites concentrated at the bottom left of the page, and news page scan sites were distributed around the center and top areas. As illustrated in Fig. 3, the scan paths belonging to different Web categories were classified better by the string-edit method based on the city block distance model than that based on the uniform distance model.

IV. CONCLUSION

We analyzed scan-path data with string-edit method. This is the first paper to adopt adequate cost functions for applying the string-edit method to eye-tracking data. We compared three types of cost functions and concluded that the city block distance and Euclidian distance model gave better results. In this paper, we focused on the geographical distance for the substitution cost. For future work, we would like to consider other types of substitution costs, such as the functionality of the objects or visually emphasized objects.

REFFERENCES

[1] R.J.K. Jacob and K.S. Karn, "Eye tracking in Human-computer interaction and usability research: Ready to deliver the promises," in the *Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, ed. by J. Hyona, R. Radach, and H. Deubel, pp. 573-605, Amsterdam, Elsevier Science, 2003.

[2] Ramloll, R., Trepagnier, C., Sebrechts, M. and J. Beedasy, "Gaze data visualization tools: Opportunities and challenges," in Proc. *International Conference on Information Visualization*, 2004.

[3] Granka, L.A., Hembrooke, H., and Gay, G., "Location location location: Viewing patterns on WWW pages," in *Proc. Symposium on Eye Tracking Research & Applications (ETRA)*, pp.43, 2006.

[4] Sankoff, D. & Kruskal, J.B. (eds.), *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*, CSLI Publications, 1999.

[5] Josephson, S., & Holmes, M.E., "Attention to repeated images on the World-Wide Web: Another look at scanpath theory," *Behavior Research Methods, Instruments, & Computers*, vol. 34, pp. 539-548, 2002.

[6] Josephson, S., & Holmes, M.E., "Clutter or content? How on-screen enhancements affect how TV viewers scan and what they learn," in *Proc. Symposium on Eye Tracking Research & Applications (ETRA)*, pp. 155-162, 2006.

[7] Hembrooke, H., Feusner,M, & Gay, G., "Averaging scan patterns and what they can tell us," in *Proc. Symposium on Eye Tracking Research & Applications (ETRA)*, pp. 41, 2006.