Tracing Temporal Changes of Selection Criteria from Gaze Information

Kei SHIMONISHI[†] **Hiroaki KAWASHIMA**[†] **Erina Schaffer**[†] **Takashi MATSUYAMA**[†]

†Kyoto University, Yoshida-Honmachi, Kyoto 606-8501, Japan

This is a author version of "Tracing Temporal Changes of Selection Criteria from Gaze Information" (the 21st International Conference on Intelligent User Interfaces (IUI '16 Companion), pp. 9–12, Sonoma, California, USA, 2016.3.7.)

abstract

To design interactive systems that proactively assist users' decision making, the users' gaze information is an important cue for the system to estimate users' selection criteria. Users sometimes change selection criteria while browsing content. Therefore, temporal changes of those criteria need to be traced from gaze data in short time scales. In this paper, we propose an approach to detecting users' distinctive browsing periods with its appropriate time-scale by leveraging multiscale exact tests so that the system can trace temporal changes of selection criteria. We demonstrate the applicability of the proposed method through a toy example and experiments.

keywords

Gaze information; decision making; multiscale exact test

category

H.1.2. User/Machine Systems Human factors

1. Introduction

Making decisions among alternatives is a fundamental part of people's daily lives. However, people sometimes only have a fuzzy understanding of their internal selection criteria (a set of criteria for that decision). Designing recommender systems that assist users by interactively converging the users' selection criteria has thus been an important research topic [2, 4]. Understanding decision states of users such as which criteria are currently in focus is essential to recommender system design.

In this work, we propose a method to detect users' currently active selection criteria (active criteria) from gaze information. Gaze information is considered to be an important clue to understand various aspects of decision making [3]. To trace the temporal changes of selection criteria, we need to consider (1) how to estimate active criteria during short periods, and (2) how to decide the window size for analysis.

To deal with the above problems, we first model neutral browsing behavior of the users when they are not focusing on any specific criteria. Then, a multiscale exact test is applied to detect the users' distinctive browsing behavior by its significance level. As a result, temporal changes of selection criteria can be successfully traced.

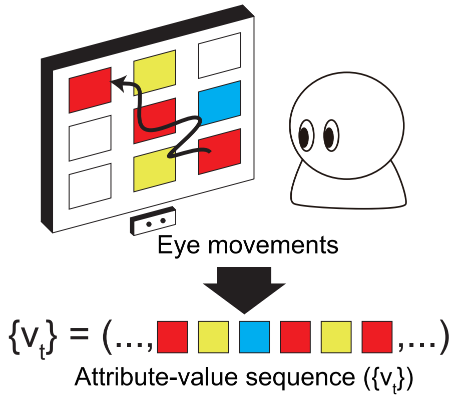

Figure 1: An obtained attribute-value sequence from eye movements. Each color corresponds to a different attribute value.

2. Method

We assume the situation where a user is browsing a digital catalog on a screen. Each catalog has information about a set of items $\mathcal{I}$ described by A types of attributes, e.g., *price*. A set of possible values of the a-th attribute type is denoted as $\mathcal{V}^{(a)} = \{V_1^{(a)}, V_2^{(a)}, \ldots, V_{K_a}^{(a)}\}$.

2.1 Gaze information

Recorded gaze information is represented as a sequence of gaze items $(i_1, \ldots, i_T)$ $(i_t \in \mathcal{I})$. Note that the time, t, is defined by the transition of gaze targets, that is, $i_{t-1} \neq i_t$. By referring to the attribute values of each item i_t, we obtain A attribute-value sequences (sequences of attribute value) $\{v_t^{(a)}\}_{t=1}^T = (v_1^{(a)}, \ldots, v_T^{(a)})$ $(v_t^{(a)} \in \mathcal{V}^{(a)}, a = 1, \ldots, A)$ as shown in Fig. 1. In this paper, we treat each sequence of attribute values independently for simplicity.

2.2 Modeling neutral browsing behavior

Neutral browsing behavior is defined as browsing behavior of users when they are not focusing on any specific criteria. We simply assume that the users look at items randomly when they are in the neutral browsing. When the catalog contains $N_k^{(a)}$ items that have the k-th value of the a-th attribute, $V_k^{(a)}$, how likely

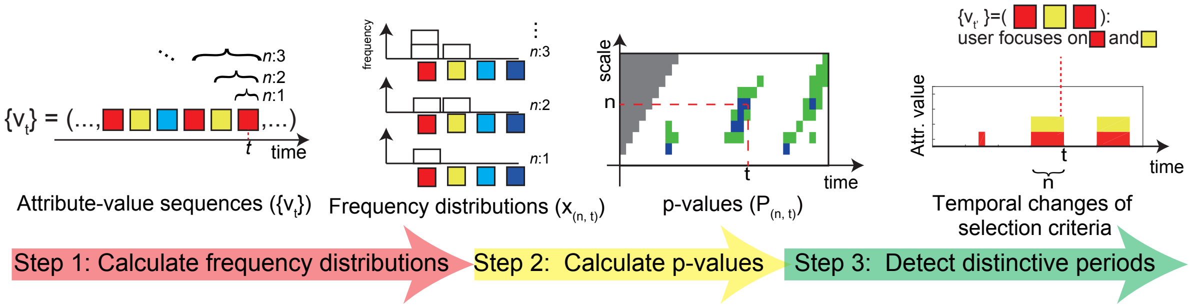

Figure 2: Proposed method to trace temporal changes of selection criteria by multiscale exact test.

the attribute value is looked at can be represented as $p_k^{(a)} = N_k^{(a)}/N$, where N is the total number of items on the catalog.

2.3 Multiscale exact multinomial test

We utilize the exact multinomial test to detect users' distinctive browsing behavior by its significance level to the neutral browsing behavior. This method assumes all attributes are categorical variables. Since it is difficult to decide the window size for analysis, we extend the exact multinomial test to multiple time scales (Fig. 2).

Assume we have an attribute-value sequence of size n and the k-th attribute value is observed x_k times in the sequence ($k = 1, \ldots, K$). The probability of $\boldsymbol{x} = (x_1, \ldots, x_K)$ in the neutral browsing follows a multinomial distribution, $f(\boldsymbol{x}; n, \boldsymbol{p}) = n! \prod_{k=1}^{K} p_k^{x_k}/x_k!$, where $\boldsymbol{p} = (p_1, \ldots, p_K)$ is multinomial parameters defined in the previous section. Note that the attribute indicator, a, is omitted for simplicity.

Let $\{v_{t'}\}_{t'=t-n+1}^{t}$ be a substring of an attribute-value sequence $\{v_t\}_{t=1}^{T}$, where $1 \le t \le T$ and n ($1 \le n \le 10$) is the window size. The frequency distribution of attribute values $\boldsymbol{x}_{(n,t)}$ is calculated for each substring (Fig. 2 Step 1). Then, the distinctive periods (as substrings with time t and scale n) can be detected by its significance level of the p-value (Fig. 2 Step 2). The p-value can be calculated as

$$P_{(n,t)} = \sum_{\hat{\boldsymbol{x}}: f(\hat{\boldsymbol{x}}; n, \boldsymbol{p}) \le f(\boldsymbol{x}_{(n,t)}; n, \boldsymbol{p})} f(\hat{\boldsymbol{x}}; n, \boldsymbol{p}), \qquad (1)$$

where $\hat{\boldsymbol{x}}$ is a potentially observed frequency distribution of attribute values ($\sum_{k=1}^{K} \hat{x}_k = n$). In each detected distinctive period, if the relative frequency of the k-th attribute value x_k/n is higher than the multinomial parameter p_k, the k-th attribute value, V_k, is considered the active criteria (Fig. 2 Step 3). As a result, temporal changes of users' selection criteria can be traced (e.g., the number of active criteria gradually increases).

3. Evaluation

To evaluate the proposed method of tracing temporal changes of selection criteria, it is applied to two types of data: a toy example and actual gaze data collected from eye tracking. The toy example was used to explain and confirm how the proposed method works.

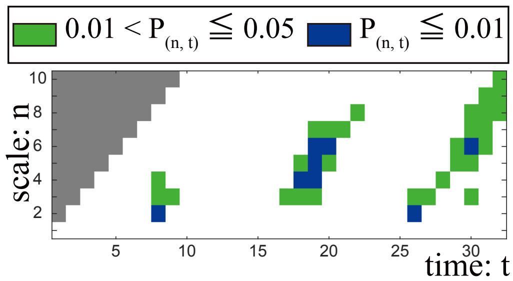

Figure 3: Calculated p-values ($P_{(n,t)}$) of toy example.

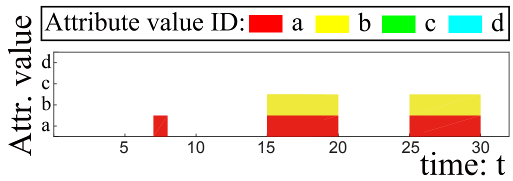

Figure 4: Traced temporal change of selection criteria on toy example. Each color shows each attribute value.

3.1 Toy example

Let us consider the following attribute-value sequence,
$\{v_t\} = (c, d, d, d, b, d, a, a, d, d, d, d, c, d, b, b,$
$a, b, b, d, c, c, d, d, a, a, d, b, b, a, d, d)$,
where possible attribute values are $\mathcal{V} = \{a, b, c, d\}$. Here, the multinomial parameters are given as $\boldsymbol{p} = (1/12, 1/6, 1/4, 1/2)$.

3.2 Results on toy example

The calculated p-values and the traced temporal change of selection criteria from the toy example are shown in Fig. 3 and Fig. 4, respectively. In Fig. 3, the short distinctive periods are detected at the lower scale at $(t, n) = (8, 2)$, and the long distinctive periods are detected at both lower and higher scale at $(t, n) = (26, 2)$ and at $(t, n) = (30, 6)$. These results show that the proposed method can detect both short and long periods of distinctive behavior. In Fig. 4, the attribute value a and b are detected as selection criteria. Note that even though the attribute value d was often looked at in the sequence, this was simply be-

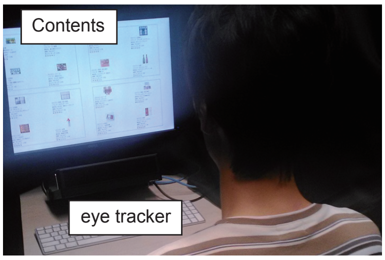

Figure 5: Experimental environment.

cause many items had the value d, and not because it was part of the hypothetical user's selection criterion. These results show that the proposed method can trace the active criteria even when the numbers of items having each attribute value are not uniformly distributed in the catalog.

3.3 Gaze data collection

We conducted a preliminary experiment to evaluate the proposed method with gaze data. In the experiments, participants were asked to select one item out of 16 items displayed on a screen (see Fig. 5). Each item had four attribute types: *category, price, ranking*, and *review*. We gave participants instructions that specified the requirements for items to select for each trial, for example: "Select an item in Alcohol category and with more than 4-star review.".

3.4 Results on gaze data

An example result of calculated p-values and traced temporal change of selection criteria from the actual gaze data is shown in Fig. 6 and Fig.7, respectively. The task was to select an item with the fourth category $V_4^{(cat)}$ and more than 4-star reviews, $V_4^{(rev)}$ or $V_5^{(rev)}$. The multinomial parameters were $\boldsymbol{p}^{(cat)} = \boldsymbol{p}^{(pri)} = \boldsymbol{p}^{(ran)} = (1/4, 1/4, 1/4, 1/4)$, and $\boldsymbol{p}^{(rev)} = (3/16, 1/4, 1/8, 3/16, 1/4)$.

The top row in Fig. 6 shows that the participant kept the focus on specific attribute value after $t = 20$. In this period, the participant focused on the attribute value $V_4^{(cat)}$, which is specified by the task (see the top row in Fig. 7).

Moreover, the other rows in Fig. 6 show that the participant focused on specific attribute values intermittently. Although the attribute values specified by the task, $V_4^{(rev)}, V_5^{(rev)}$, are included in the detected selection criteria, attribute values that are not related to the task are also included (Fig. 7). This is because the participant compared few items that have $V_4^{(cat)}$ and the variety of attribute values decreased.

4. Conclusion and Future work

We presented a method to detect the active criteria of users from gaze information so that we can trace processes of decision making. An adapted multinomial test to detect distinctive periods with multi time scale is introduced to accommodate varying comparison patterns of users. We demonstrate the applicability of the

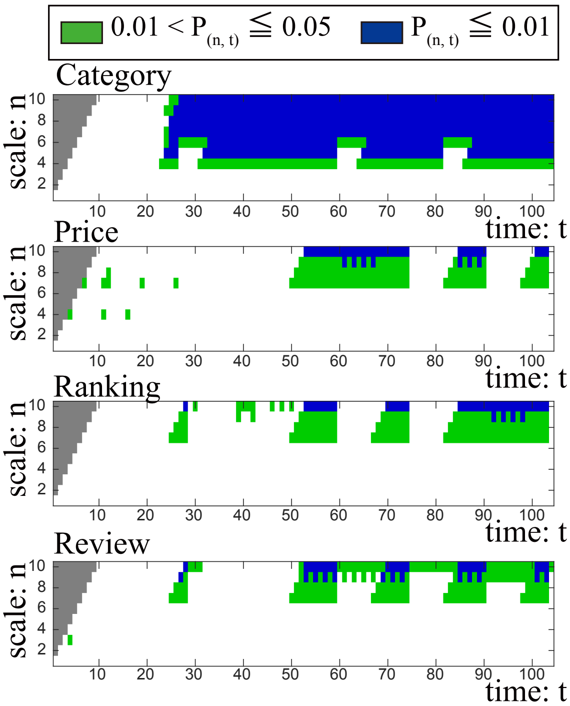

Figure 6: Calculated p-values ($P_{(n,t)}$) of an example gaze sequence.

proposed method with a toy example and with gaze data collected from eye tracking.

In the future, we will apply the method to an interactive system that can assist users in decision making. For instance, the system can probe users' decision state by suggesting alternatives based on detected selection criteria and observing the users' reaction.

Moreover, we consider extending the proposed method to overcome the following current limitations. First, we assume each attribute takes a categorical value. However, some attributes (e.g., price) can have order relations and that might be important to understand users' decision making behavior. To deal with this, we consider exact test for ordinal variables. Second, users' gaze behavior is affected not only by their selection criteria but also by the layout of the content (e.g., center bias [1]), thus, users may not browse content uniformly. We therefore consider adjusting the multinomial parameter $\boldsymbol{p}$ by considering the bias of the content layout.

5. Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers 15J06965, 13J05396, 26280075 and JST, PRESTO.

References

[1] Ali Borji. 2012. Boosting Bottom-up and Top-down Visual Features for Saliency Estimation. In *CVPR*. 438–445.

[2] Robin Burke. 2002. Interactive Critiquing for Catalog Navigation in E-Commerce. *Artif. Intell. Rev.* 18, 3-4, 245–267.

[3] Jacob L. Orquin and Simone Mueller Loose. 2013. Attention and choice: A review on eye movements in decision making. *Acta Psychologica* 144, 1, 190 – 206.

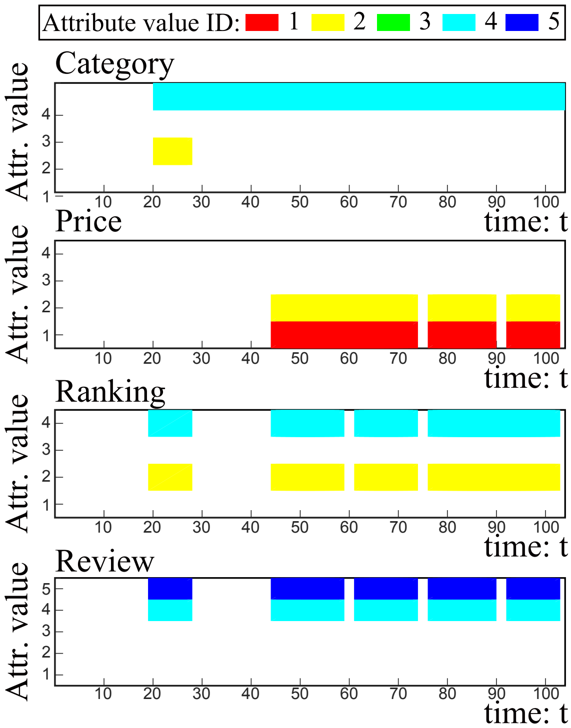

Figure 7: Traced temporal change of selection criteria on gaze data. Each color shows each attribute value.

[4] Joseph Polifroni and Marilyn Anne Walker. 2008. Intensional Summaries as Cooperative Responses in Dialogue: Automation and Evaluation. In *ACL*.